Online Appendix

The Local Root of Wage Inequality

Hugo Lhuillier

C Additional tables

Table C.3: City cluster-level summary statistics

Cluster	# CZs	Size	Avg. wage	St. dev.	P10	P90	Rent	Smallest CZ	Largest CZ
1	100	8,518	16.61	0.31	11.33	23.48	8.41	Le Blanc	La Roche-sur-Yon
2	53	15,924	17.34	0.34	11.35	25.38	9.58	Châteaudun	Vannes
3	41	21,214	18.04	0.35	11.54	26.73	9.62	Commercy	Metz
4	29	28,976	18.54	0.37	11.55	28.15	10.38	Tergnier	Tours
5	24	36,931	19.44	0.38	11.79	29.88	10.48	Houdan	Rouen
6	13	62,581	19.37	0.39	11.64	30.16	11.34	Wissembourg	Bordeaux
7	19	50,164	20.33	0.41	11.75	32.48	11.62	Ambert	Toulouse
8	8	101,150	21.43	0.43	11.92	34.61	13.96	Chinon	Roissy
9	8	91,755	22.39	0.45	12.00	36.28	14.54	Étampes	Lyon
10	2	924,781	27.63	0.52	12.53	46.09	22.25	Saclay	Paris

The columns are: the cluster ID, the number of CZs, the average number of employed worker, the average wage, the standard deviation of log wages, the 10^{th} and 90^{th} percentiles of local wages, the rent per meter square, and the smallest and largest CZs. All statistics computed at the CZ level and averaged at the cluster.

Table C.4: Switching rates within and across locations

	All switch	Switch up
Switching rate (%)	7.56	4.46
Within city (%)	84.1	86.8
Switchers' wage growth (%)	1.20	13.0
City stayers (%)	2.00	12.4
City movers (%)	-3.35	17.1
Days between jobs	75	60
City stayers	61	49
City movers	136	121

First row: fraction of job switches per quarter, and fraction of switches taking place within a location. Second row: average wage growth conditional on a switch for all switches, switches within location, and switches across locations. Third row: days between employment spells conditional on a switch for all switches, switches within location, and switches across locations. A worker switches job if they switch establishments or occupations. Rows 1 and 2 restrict switches to take place within 90 days of previous employment spell. Column 1 presents the statistics for all switches, and column 2 for switches associated with non-negative wage growth.

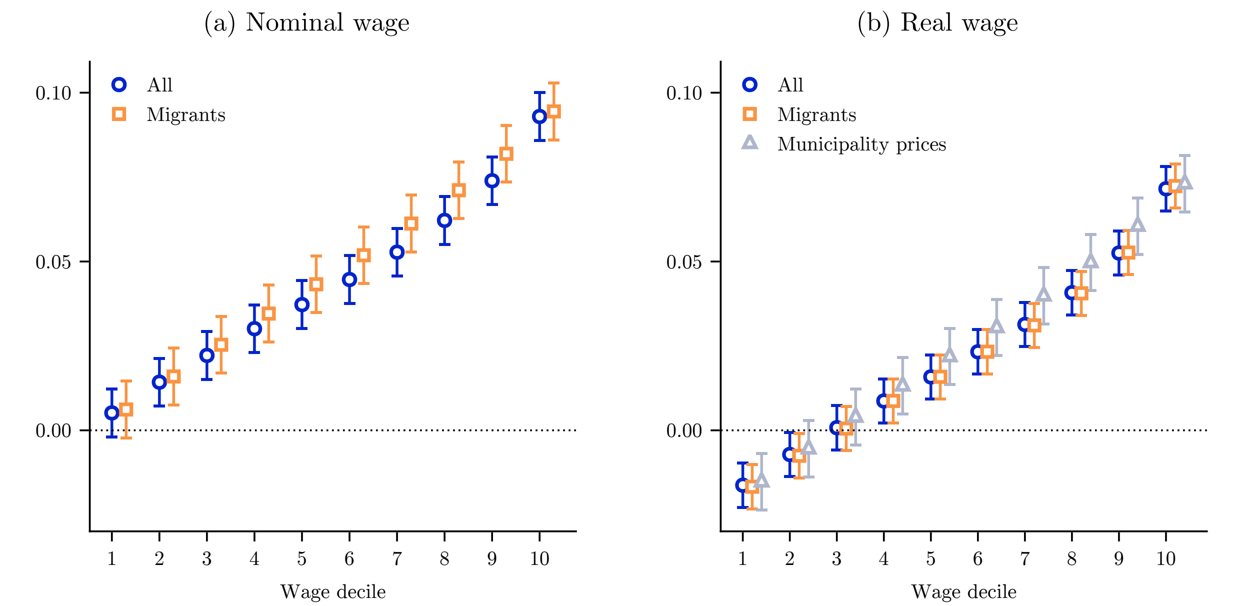
Table C.5: The local returns to job switching estimated on all switches

	(1)	(2)	(3)	(4)	(5)	(6)
Aggregate	0.51	0.22	0.13	0.22		
	(0.013)	(0.011)	(0.011)	(0.011)		
By occupation						
Blue-collar					0.29	
					(0.013)	
Low-skill service					0.59	
					(0.048)	
Intermediary					0.34	
					(0.033)	
White-collar					0.29	
					(0.012)	
By rank						
1st quartile						0.59
						(0.072)
2nd quartile						0.30
						(0.024)
3rd quartile						0.17
						(0.030)
4th quartile						-0.19
						(0.014)
Dep. variable	Wage	Job F.E.	Job F.E.	Job F.E.	Job F.E.	Job F.E.
W/in occupation	·	·	✓	·	·	·
Worker slope	·	·	·	✓	·	·
Occupation F.E.	·	·	·	·	✓	·
Rank F.E.	·	·	·	·	·	✓
R^2	0.01	0.04	0.16	0.04	0.05	0.08
Observations	26,246,229	26,246,229	26,246,229	26,246,229	26,246,229	26,246,229

Table presents the coefficient of interest ζ estimated in (6). $J2J_{it} = 1$ if worker switches establishments or occupations within 90 days of previous employment spell and within the CZ. Dependent variable is quarterly growth in wages (column 1) or job fixed effects (column 2 — 6). Job fixed effects estimated in (1). Column 3 includes origin-occupation × destination-occupation fixed effects at the 3-digit level. Column 4 uses job fixed effects estimated with a worker-specific experience slope. Column 5 estimates (6) fully interacted with origin-occupation dummy. Column 6 estimates (6) fully interacted with quartile dummy, where the quartile refers to the rank in the local job fixed effect distribution. Standard errors clustered at the city level.

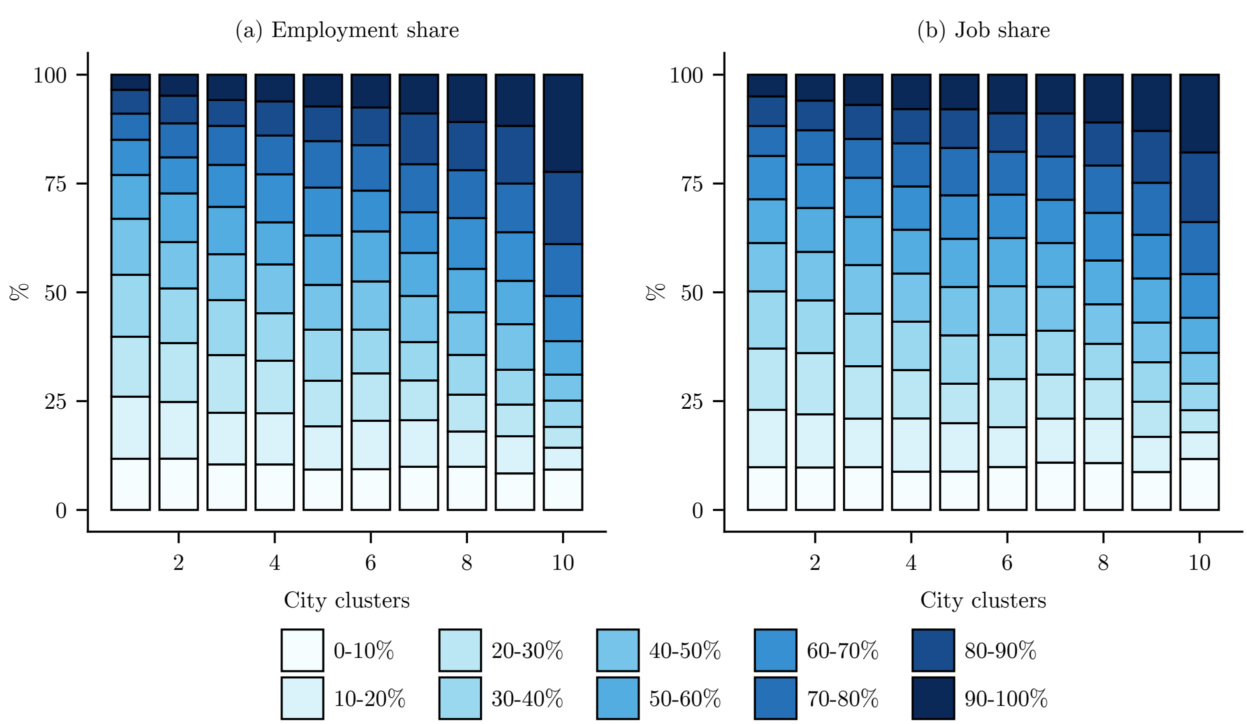
C.1 Additional figures

Figure C.2: Wage distribution by commuting zone



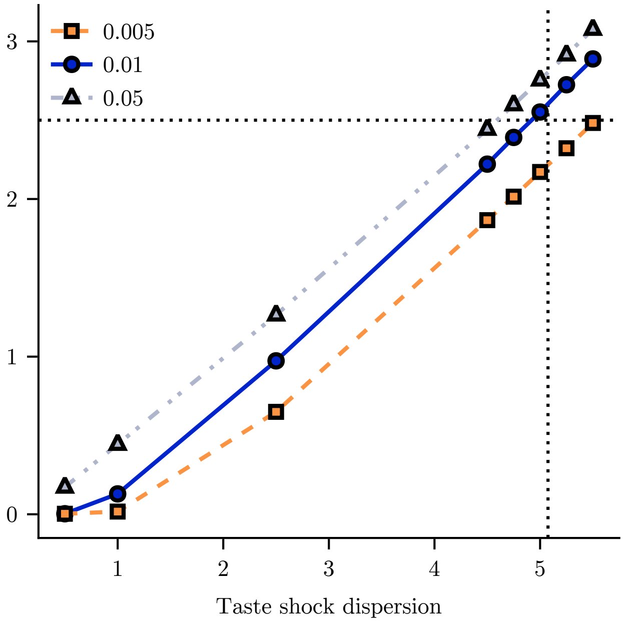
Let $w_{q\ell}$ denote the average wage in the qth decile of location ℓ's wage distribution, and m_ℓ be ℓ's size. Both panels display the size elasticities estimated by $\log w_{q\ell} = \alpha_q + \beta_q \log m_\ell + u_{q\ell}$. Panel (a) plots $\hat{\beta}_q$ for nominal wages. Panel (b) plots $\hat{\beta}_q$ for real wages, defined as nominal wages deflated by a Cobb-Douglas price index with a housing expenditure share of 0.2. The blue circles depict $\hat{\beta}_q$ for the entire sample. The orange rectangles for workers living in different CZs than their birthplace. The grey triangles for wages deflated by a municipality housing price index, $P^m_{\ell q} = (\sum_{m \in \ell} \omega^q_{\ell m} p_m)^{0.2}$, where $\omega^q_{\ell m}$ is the fraction of workers in (q, ℓ) that lives in municipality m, and p_m is the average housing price in m. The vertical bars are 95th confidence intervals.

Figure C.3: Spatial distribution of job wage premia



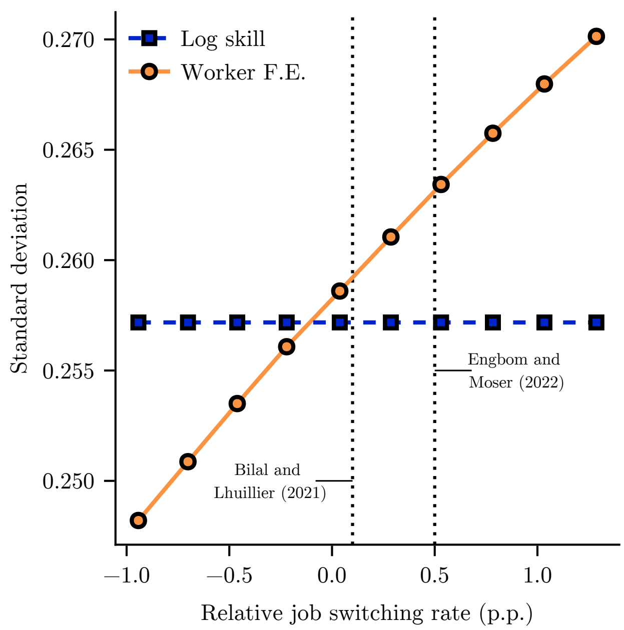
Panel (a) displays the local employment shares of jobs in the x^{th} decile of the aggregate employment-weighted job fixed effect distribution. Panel (b) displays the share of jobs in the x^{th} decile of the aggregate job fixed effect distribution. The job fixed effects are obtained from (1).

Figure C.4: Reduced-form local labor supply elasticity



Each line shows the reduced-form local labor supply elasticity computed in the model following local tax reforms against the taste shock dispersion χ. Appendix G.2 presents in details how the elasticities are computed. The blue circles, orange rectangles, and grey triangles depict the elasticities when the standard deviation of the tax reforms are 0.01, 0.005, and 0.05 respectively. The horizontal dotted line is the empirical reduced form elasticity targeted in the calibration, and the vertical dotted line is shows the estimated taste shock dispersion.

Figure C.5: Mis-specification bias in AKM

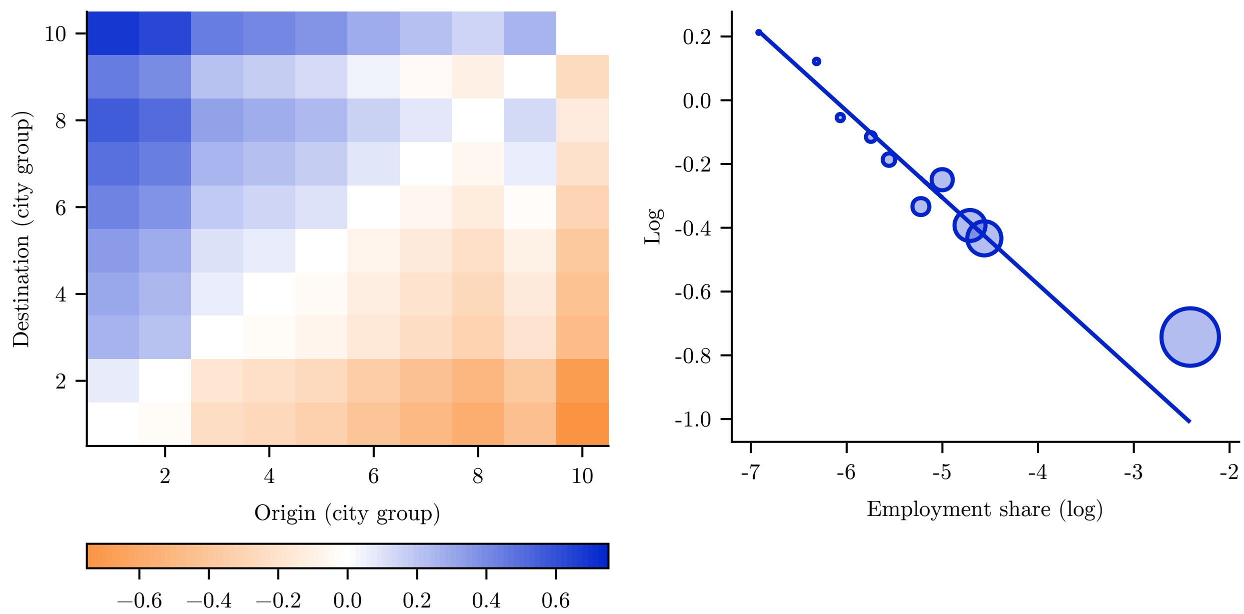


Blue rectangles show the variance of log skill in the heterogeneous-skill model presented in Section F.3. Orange circles plot the variance of the AKM worker fixed effects estimated on simulated data. Both statistics plotted against the difference in job switching rates between the fifth and first skill quintile. Dotted lines depict estimates of the relative job switching rates in Bilal and Lhuillier (2021) and Engbom and Moser (2022).

Figure C.6: The causal impact of cities on firm size

(a) Size change conditional on origin and destination

(b) Labor market tightness

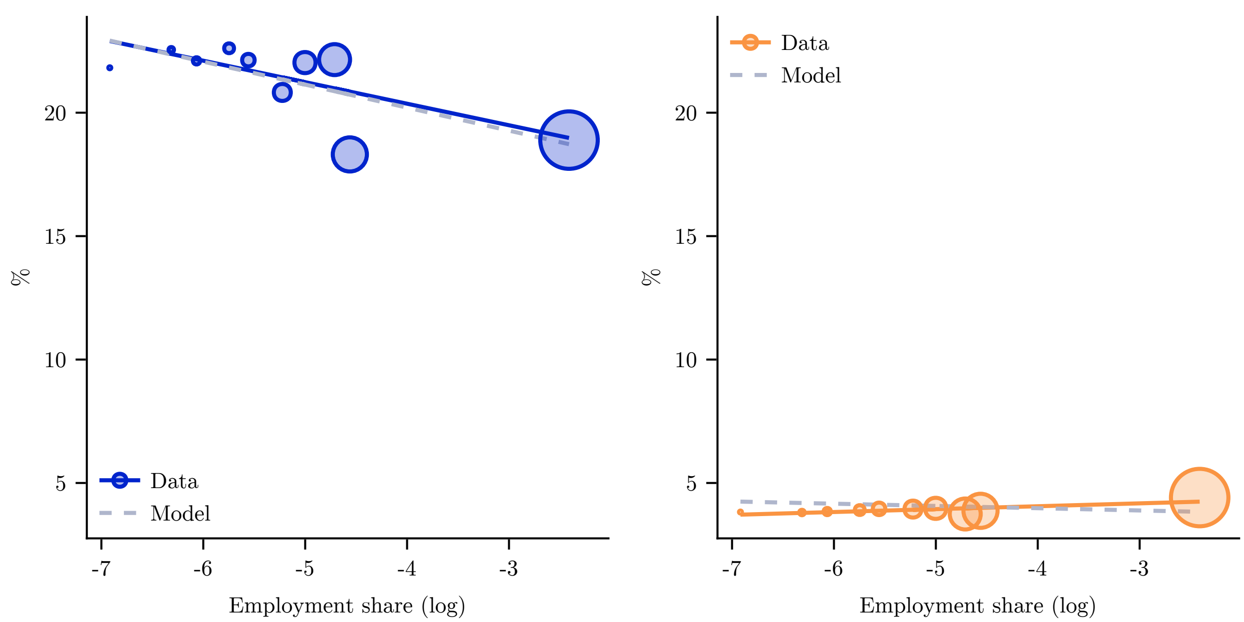


Left panel shows the average change in log firm size conditional on an origin (x-axis) and destination (y-axis) location as computed in (23). Right panel shows the inferred log labor market tightness (V_ℓ/e_ℓ) against the average employment share of CZs in each group. The labor market tightness

Figure C.7: Search frictions in the model and the data

(a) Unemployment to employment rate

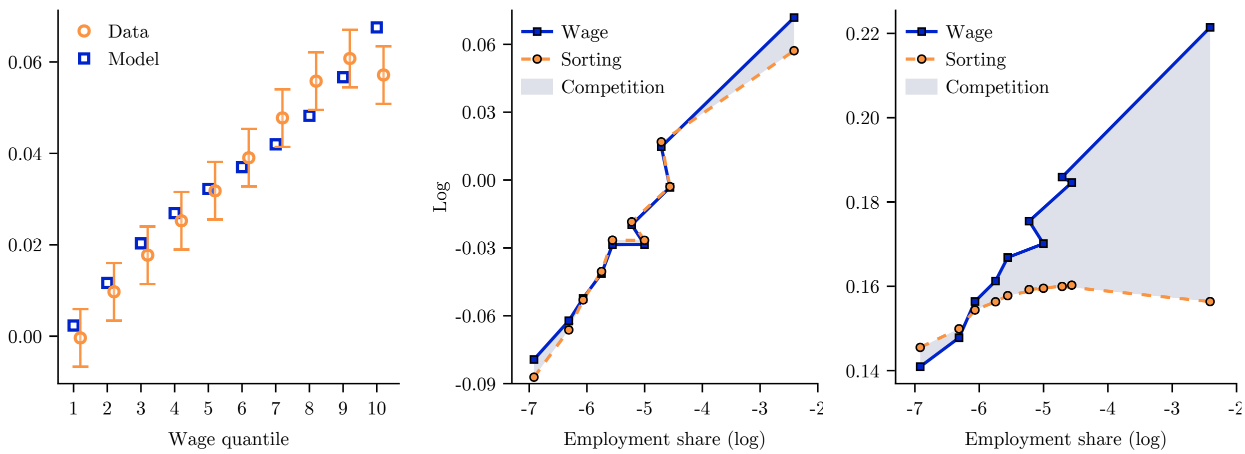
(b) Job switching rate



Left panel displays the UE rate against city size in the data. Right panel displays the job switching rate against city size in the data. The job switching rates are net of worker fixed effects and age controls as in (5). In both panels, dashed grey lines show the linear fit in the model.

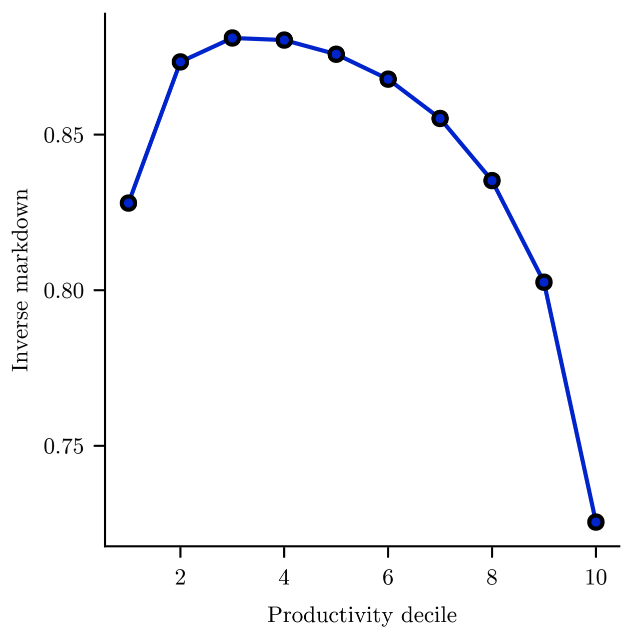
Figure C.8: Spatial wage inequality in the model and the data

(a) City-size wage decile elasticity (b) Average wage (c) Wage standard deviation



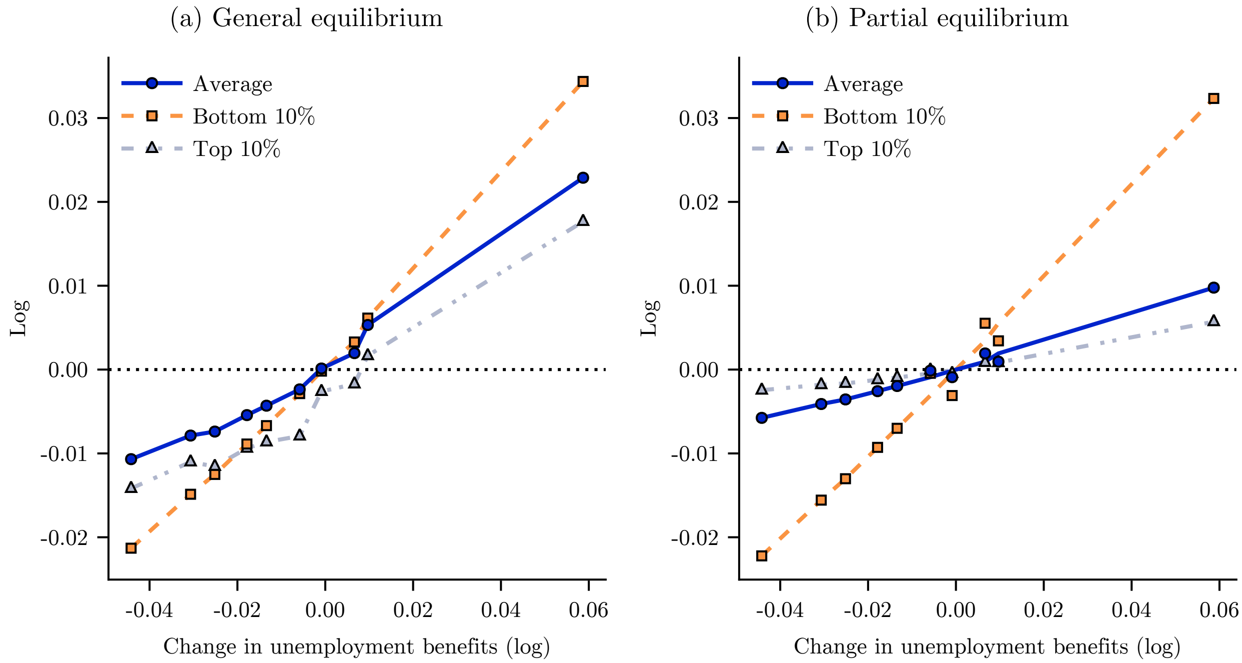
Let $w_{q\ell}$ denote the average wage premia in the qth decile of location ℓ's wage premia distribution, and m_ℓ be ℓ's size. Panel (a) displays the size elasticities estimated by $\log w_{q\ell} = \alpha_q + \beta_q \log m_\ell + u_{q\ell}$, in the data (orange circles) and in the model (blue rectangles). The vertical bars are 95th confidence intervals. Panel (b) applies the decomposition (24) to the average wage ($\mathcal{W} = \mathbb{R}$) in the model. The blue rectangles show the average wage, the orange circles the composition effect (first two terms), and the grey area is the competition effect (third term). Panel (c) computes the counterfactual within-city inequality if markdowns were constant across space. The blue rectangles show the wage standard deviation, the orange circles are the counterfactual statistics under $\mu_\ell(z) = \bar{\mu}(z)$, and the grey area is the difference between the two.

Figure C.9: Markdowns distribution



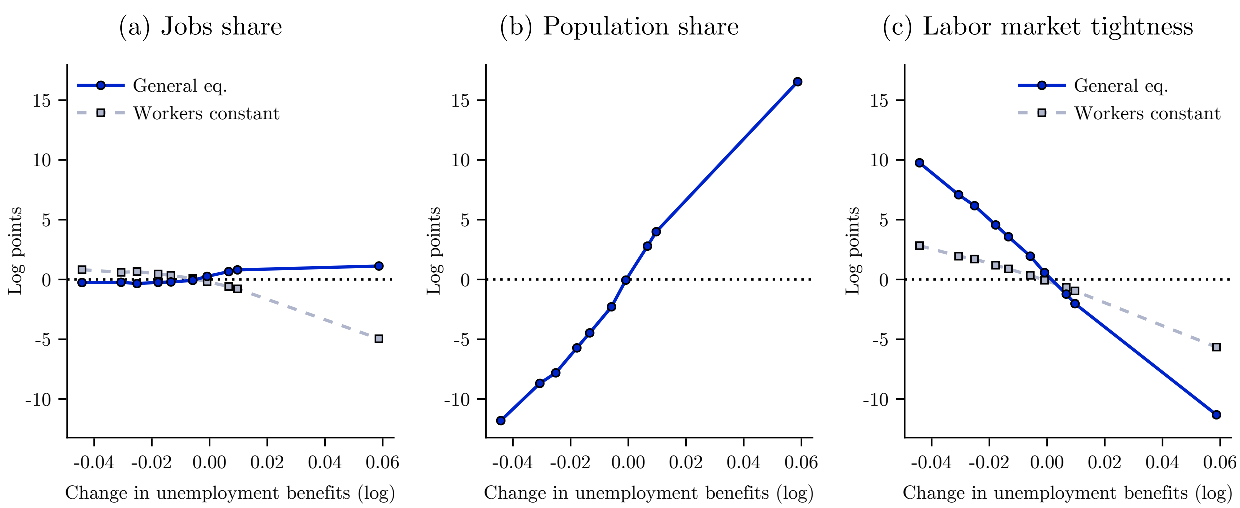
Plot displays the inverse average markdown by productivity decile in the model. Markdowns are defined as $\mu_\ell(z) = zT_\ell / w_\ell(z)$, so that $1/\mu_\ell(z) \in (0, 1)$. Productivity deciles are employment weighted.

Figure C.10: The consequences of housing-adjusted unemployment on the wage distributions



In both panels, the blue circles show the change in the average wage, and orange rectangles and grey triangles the change in the average wage within the bottom and top 10% of the local wage distributions. Left panel is in general equilibrium, right panel is holding fixed the spatial allocation of workers and employers at their laissez-faire equilibrium. All statistics plotted against the change in unemployment benefits.

Figure C.11: The consequences of housing-adjusted unemployment benefits on location choices



Panel (a) and (b) show the change in the measure of employers and workers per location. Panel (c) depicts he change in labor market tightness (V_ℓ/e_ℓ). Blue circles show the response in general equilibrium and grey rectangles show the changes holding constant the spatial allocation of workers. The three variables are expressed in log points and projected against the change in unemployment benefits.

D Data

D.1 France

D.1.1 Matched employer-employee dataset

Sample restrictions. I apply the following restrictions on the Déclaration Annuelle de Données Sociales:

1. Exclude workers younger than 25 and older than 55 year old;
2. Exclude workers employed in the public sector;
3. Exclude the agriculture, education and health industries;
4. Exclude workers outside of metropolitan France;
5. Keep only workers employed full-time;
6. Exclude workers who are non-employed for more than three years;
7. Exclude employment spells that last less than 30 days;
8. Exclude employment spells with no labor income or hours worked.

Construction of the panel. Once a worker enters the long-panel, each of their employment spells are recorded. An employment spell is defined as a pair establishment $\times$ occupation. The dataset provides the start and end days of each employment spell. Workers can be observed multiple times within a given period if they work for multiple employers or if they switch employers. By contrast, a worker is observed only once per year if they work for a unique employer during that entire year.

I aggregate the data at the quarterly level. If workers hold multiple jobs within a quarter, I keep the job that provides the highest total labor income. I only keep the information (e.g., employer's ID, occupation, etc.) associated to that employment spell.

Clustering. I group cities into ten clusters according to their average wage. The clusters are population-weighted, and thus contains approximately 10% of the workforce. Table C.3 provides summary statistics on each of the city cluster. I cluster jobs in 100 employment-weighted groups according to their average wage. I implement this clustering within city group to respect the geographic nature of an establishment. Lastly, I cluster workers in 100 employment-weighted groups according to their average wage.

Structural estimation. I measure wages through the job fixed effects estimated in (1). I define the size of a job as the average number of workers per job within each job cluster in the panel.

D.1.2 Other data sources

Carte des Loyers. I measure residential housing prices via rents. I obtain data on rents from the "Carte des Loyers" dataset (Rental Map). The dataset covers the 2011 — 2019 period. It provides the average rent per square meter at the municipal level after controlling for housing characteristics. I compute commuting zone-level residential housing prices as the population-weighted average of the municipal prices.

Demande de Valeurs Foncières. I obtain commercial housing prices and housing surfaces using the Demande de Valeurs Foncières dataset. The dataset covers the period 2015 — 2019. It provides the value (euros), size (square meters), and usage (residential or commercial) for the universe of housing transactions in France. I compute the price per square meter as the value of the transaction over the size of the unit. The dataset is noisy, and I winsorize the size and price distribution at the 10%.

I use the average commercial housing price and housing surface to infer local commercial prices from residential prices. Specifically, I suppose that the relative housing prices are the same in the commercial and residential markets. I then adjust the scale of the commercial prices to reflect the national commercial housing prices in the Demande de Valeurs Foncières dataset. Furthermore, I adjust commercial housing prices by the average housing size in the commercial market since employers do not face an intensive margin.

Enquête emploi en continu. I apply as much as possible the same restrictions on the Enquête emploi en continu as for the Déclaration Annuelle de Données Sociales:

1. Exclude workers younger than 25 and older than 55 year old;
2. Exclude workers employed in the public sector;
3. Exclude the agriculture, education and health industries;
4. Exclude workers outside of metropolitan France;
5. Exclude workers not reporting an employment status or hours worked when employed;
6. Exclude workers who are non-employed for more than three years.

D.2 United States

I obtain wage data from the American Community Survey. I use sample restrictions similar to the French data:

1. Exclude workers younger than 25 and older than 55 year old;
2. Exclude workers employed in the public sector;
3. Exclude the agriculture industry;
4. Keep only workers employed full-time ($\geq$ 30 hours per week and 48 weeks per year);
5. Exclude workers with no labor income.

I define hourly wages as yearly labor income over the number weeks worked last year and the usual amount of hours worked per week. I truncate the left-tail of the hourly wage distribution by US State at the 2.5% to reduce measurement error in hours worked.[30]

I also obtain rent data from the American Community Survey. To ensure consistency, I compute local rents on the same sample as local wages for those who report positive rents.

E Additional proofs and derivations

E.1 Lifetime utility

Equation (11) follows from a guess and verify. Guess that

$$U_\ell = \frac{A_\ell \mathcal{W}_\ell}{P_\ell} \quad \text{and} \quad V_\ell(w) = \frac{A_\ell \mathcal{V}_\ell(w)}{P_\ell}.$$

Plugging the guesses into (8) and (9) verifies the guesses and delivers

$$\begin{aligned} \rho \mathcal{W}_\ell &= b + \zeta\lambda \int \max\{\mathcal{V}_\ell(w) - \mathcal{W}_\ell, 0\} dF_\ell(w), \\ \rho \mathcal{V}_\ell(w) &= w + \lambda \int \max\{\mathcal{V}_\ell(w') - \mathcal{V}_\ell(w), 0\} dF_\ell(w') + \delta(\mathcal{W}_\ell - \mathcal{V}_\ell(w)). \end{aligned} \tag{28}$$

[30]Labor incomes are already right-truncated in the ACS.

E.2 Proof of Proposition 1

Reservation wage. Suppose for simplicity that F_ℓ admits a density. Differentiate the HJB of employed workers (9) with respect to w, integrate it back to w, and use $V_\ell(\underline{w}_\ell) = U$ to obtain

$$V_\ell(w) = U + \frac{A_\ell}{P_\ell} \int_{\underline{w}_\ell}^{w} \frac{1}{\rho + \delta + \lambda \bar{F}_\ell(w')} dw'.$$

Combine this expression into (8) and (9) to get

$$\rho U = \frac{A_\ell}{P_\ell} \left(b + \zeta\lambda \int_{\underline{w}_\ell} \frac{\bar{F}_\ell(w)}{\rho + \delta + \lambda \bar{F}_\ell(w)} dw \right), \tag{29}$$

$$\rho V_\ell(\underline{w}_\ell) = \frac{A_\ell}{P_\ell} \left(\underline{w}_\ell + \lambda \int_{\underline{w}_\ell} \frac{\bar{F}_\ell(w)}{\rho + \delta + \lambda \bar{F}_\ell(w)} dw \right) = \rho U. \tag{30}$$

Equating both equations and solving for $\underline{w}_\ell$ yields (12).

Nominal. Let $H_F(x)$ be the operator defined by

$$H_F(x) = \int_x \frac{1 - F(w)}{\rho + \delta + \lambda(1 - F(w))} dw,$$

and $J_F(x)$ be defined by $J_F(x) = x - b - \lambda(\zeta - 1) H_F(x)$. The reservation wage solves $J_{F_\ell}(\underline{w}_\ell) = 0$. H satisfies two properties. First, $x \to H_F(x)$ is decreasing. Second, $F_1 \succ F_2$ implies $H_{F_1}(x) > H_{F_2}(x)$ for any x. These properties imply that J_F: $x \to J_F(x)$ is increasing and $F \to J_F(x)$ is weakly decreasing, strictly decreasing if $\zeta > 1$. Therefore, if $F_1 \succ F_2$, $J_{F_1}(\underline{w}_2) \leq J_{F_2}(\underline{w}_2) = 0$, and it must be that $\underline{w}_1 \geq \underline{w}_2$, strictly if $\zeta > 1$.

Real. From (30), the lifetime utility of a worker employed at the reservation wage is

$$\rho V_\ell(\underline{w}_\ell) = \frac{\underline{w}_\ell A_\ell}{P_\ell} \left(1 + \frac{\lambda H_F(\underline{w}_\ell)}{b + \lambda\,(\zeta - 1)\, H_F(\underline{w}_\ell)} \right),$$

where I have also used (12). H_F is (strictly) increasing in F in the FOSD sense, and therefore so must $\lambda H_F/(b + \lambda[\zeta - 1] H_F)$. However, indifference (10) requires $V_\ell(\underline{w}_\ell) = \bar{U} = V_{\ell'}(\underline{w}_{\ell'})$. Therefore, $\underline{w}_\ell A_\ell / P_\ell$ is decreasing in F_ℓ in the FOSD sense.

E.3 Expected profits

KFE. I first solve the two KFEs that dictate local employment and the employment wage distribution. For notational simplicity, let $\lambda^u = \zeta\lambda$. In ℓ, the measure of unemployed and employed workers are such that the flow out of unemployment equates the flow into unemployment: $\lambda^u u_\ell = \delta e_\ell$. Consistency requires $u_\ell + e_\ell = m_\ell$, and therefore $u_\ell = \delta/(\lambda^u + \delta) m_\ell$ and $e_\ell = \lambda^u/(\lambda^u + \delta) m_\ell$.

Meanwhile, in steady state, the flow of workers into the wage interval $[\underline{w}_\ell, w)$ equals the flow of workers out of the interval. This pins down the wage distribution among employed workers, G_ℓ:

$$\lambda^u F_\ell(w) u_\ell = [\delta_\ell + \lambda \bar{F}_\ell(w)] e_\ell G_\ell(w) \quad \Rightarrow \quad G_\ell(w) = \frac{F_\ell(w)}{1 + k(1 - F_\ell(w))}, \tag{31}$$

where the second expressions uses $u_\ell/e_\ell = \delta/\lambda^u$.

Labor supply curves. I now derive expected profits when $r > 0$, and show how this converges to (13) when $r \to 0$. Take a filled job with productivity z in location ℓ paying wage w. The discounted value of this job is

$$rJ_\ell(z, w) = z - w - s_\ell(w)J_\ell(z, w),$$

where $s_\ell(w) = \delta + \lambda(1 - F_\ell(w))$ is the separation rate. Inverting: $J_\ell(z, w) = \frac{z-w}{r+s_\ell(w)}$. Expected profits are given by the value of a filled job multiplied by the hiring rate $h_\ell(w)$:

$$\pi_\ell(z) = h_\ell(w)J_\ell(z, w) - p^c_\ell.$$

The hiring rate depends on the vacancy contact rate and how many workers accept the offer conditional on being contacted:

$$h_\ell(w) = q_\ell \left(\frac{\zeta u_\ell}{\zeta u_\ell + e_\ell} + \frac{e_\ell}{\zeta u_\ell + e_\ell} G_\ell(w) \right) = \frac{\lambda}{\theta_\ell} \frac{1}{1 + k(1 - F_\ell(w))},$$

where the second expression uses (31) and $q_\ell = \lambda/\theta_\ell$. Combined with the previous expressions, expected profits converge to (13) when $r \to 0$ and the local labor supply curves are defined by $n_\ell(w) = h_\ell(w)/s_\ell(w)$.

E.4 Local wage offer distributions

I derive here the wage equation (16). This section follows Burdett and Mortensen (1998). I first show that there cannot be holes or mass points in the wage offer distributions.

To start, suppose there is a hole in F_ℓ between $\underline{x} \geq \underline{w}_\ell$ and $\bar{x} \leq \bar{w}_\ell$. Hence, $F_\ell(\underline{x}) \leq F_\ell(\bar{x})$, where the inequality is strict if there is a mass point at either $\underline{x}$ or $\bar{x}$. Therefore $n_\ell(\underline{x}) \leq n_\ell(\bar{w})$. However, by offering any wage in $(\underline{x}, \bar{x})$, an employer that used to post wage $\bar{x}$ would keep the same size while lowering its wage bill. This constitutes a profitable deviation.

Suppose now that there is a mass point at $w \in [\underline{w}_\ell, \bar{w}_\ell]$. Take an employer z that offers wage w, and consider the deviation $w + \varepsilon$ for $\varepsilon > 0$ small. For $\varepsilon \to 0^+$, it must be $n_\ell(w + \varepsilon) > n_\ell(w)$ since there is a mass point at w, but $w + \varepsilon \to w$. Hence, the profits under wage offer w and $w + \varepsilon$ are respectively $(z - w)n_\ell(w) < (z - w - \varepsilon)n_\ell(w + \varepsilon)$, and offering wage $w + \varepsilon$ is a profitable deviation.

I now show that $F_\ell[w_\ell(z)] = \Gamma_\ell(z)$. Since n_ℓ is strictly increasing in w, π_ℓ is continuously differentiable and strictly supermodular in (z, w). Directly applying Theorem 2.8.5. in Topkis (1998), it follows that w is strictly increasing in z, and therefore $F[w_\ell(z)] = \Gamma_\ell(z)$.

Finally, I derive the wage equation (16). Since there are no mass point in the wage offer distribution, I take the first-order conditions of (17) with respect to w for any $z \in \text{supp}\, \Gamma_\ell$:

$$(z - w)\, \frac{n'_\ell(w)}{n_\ell(w)} = 1.$$

Evaluating this equation at $w_\ell(z)$ and using the change of variable $n_\ell(z) = n_\ell[w_\ell(z)]$ yields

$$w'_\ell(z) = \frac{n'_\ell(z)}{n_\ell(z)} \left(z - w_\ell(z)\right). \tag{32}$$

Integrating this ODE with respect to w and using the boundary condition $w_\ell(\underline{z}_\ell) = \underline{w}_\ell$ returns

$$w_\ell(z) = \underline{w}_\ell \left(\frac{n_\ell(\underline{z}_\ell)}{n_\ell(z)}\right) + \int_{\underline{z}_\ell}^{z} \zeta \left(\frac{n'_\ell(\zeta)}{n_\ell(z)}\right) \mathrm{d}\zeta, \quad \forall z \in \operatorname{supp} \Gamma_\ell. \tag{33}$$

Since $\int_{\underline{z}_\ell}^{z} n'_\ell(t)dt = n_\ell(z) - n_\ell(\underline{z}_\ell)$, this implies

$$w_\ell(z) = \underline{w}_\ell \left(\frac{n_\ell(\underline{z}_\ell)}{n_\ell(z)}\right) + \left(1 - \frac{n_\ell(\underline{z}_\ell)}{n_\ell(z)}\right) \int_{\underline{z}_\ell}^{z} t \left(\frac{n'_\ell(t)}{\int_{\underline{z}_\ell}^{z} n'_\ell(x)dx}\right) dt,$$

which corresponds to (16).

E.5 Proof of Proposition 2

Proposition 2.1. I first show that $\Gamma_\ell \succ \Gamma_{\ell'}$ implies $F_\ell \succ F_{\ell'}$. Let w_ℓ^q denote the q-th quantile of the wage offer distribution in ℓ, $q = F_\ell(w_\ell^q)$. Likewise, let z_ℓ^q denote the firm that offers the q-th quantile of the wage offer distribution in ℓ, $w_\ell(z_\ell^q) = w_\ell^q$. Combining the two and using the rank preserving property of F_ℓ returns $q = F_\ell[w_\ell(z_\ell^q)] = \Gamma_\ell(z_\ell^q)$, or $z_\ell^q = \Gamma_\ell^{-1}(q)$. Using the wage expression (33), the q-th quantile of the wage offer distribution is

$$w_\ell^q = \underline{w}_\ell \left(\frac{1 + k(1-q)}{1+k}\right)^2 + \int_0^q \Gamma_\ell^{-1}(u) \frac{2k[1 + k(1-q)]^2}{[1 + k(1-u)]^3} du, \tag{34}$$

where the above expression uses the fact that employment size (net of market tightness) is constant across space given a rank on the ladder. The integral is increasing in Γ_ℓ (in the FOSD sense). Therefore, it only remains to show that $\underline{w}_\ell$ is increasing in Γ_ℓ. Starting from (12), using the change of variable $w \to w_\ell(z)$, substituting in the wage FOC (32) and the wage equation (33), I obtain:

$$\underline{w}_\ell = \alpha b + (1-\alpha) \int_0^1 \Gamma_\ell^{-1}(u) \phi(u) du,$$

where $\alpha \equiv \frac{1}{1 + \frac{k(k^u - k)}{(1+k)^2}} \in (0,1)$,

$$\phi(u) \equiv \frac{2(1+k)^2(1-u)}{[1 + k(1-u)]^3}$$

is a density (i.e., $\phi \geq 0$ and $\int \phi(u)du = 1$), and I have assumed $\rho/\delta \to 0^+$ for simplicity. Therefore, $\underline{w}_\ell$ is also increasing in Γ_ℓ. Combined, $w_\ell^q \geq w_{\ell'}^q$ for all q, and strictly for some; i.e., $F_\ell \succ F_{\ell'}$. Finally, since G_ℓ is a monotonically increasing function on F_ℓ, $G_\ell \succ G_{\ell'}$.

Proposition 2.2. Using the wage expression (33), the average wage in ℓ reads

$$\mathbb{E}_\ell[W] = \frac{\int w_\ell(z) n_\ell(z) d\Gamma_\ell(z)}{\int n_\ell(z) d\Gamma_\ell(z)} = \beta \underline{w}_\ell + (1-\beta) \int \Gamma_\ell^{-1}(u) \phi(u) du,$$

where $\beta = \frac{1}{1+k} \in (0,1)$. Combined with the previous derivation, the mean-to-min ratio reads

$$\frac{\mathbb{E}_\ell[W]}{\underline{w}_\ell} = \beta + \frac{(1-\beta)\int \Gamma_\ell^{-1}(u)\phi(u)du}{\alpha b + (1-\alpha)\int_0^1 \Gamma_\ell^{-1}(u)\phi(u)du}.$$

The integral is increasing in $\Gamma_\ell \underline{w}_\ell$, and therefore so is $\frac{\mathbb{E}_\ell[W]}{\underline{w}_\ell}$.

E.6 A competitive spatial matching model

Consider a general spatial matching model with competitive labor markets. There is a measure M of heterogeneous firms indexed by their productivity z distributed according to Γ. Firms use a single input, labor. They face a decreasing returns to scale technology: $R(z,n) = zn^\rho$. Decreasing returns could arise due to span-of-control costs or love-for-variety across differentiated goods.

There are L cities. Cities differ in their size, $\{m_\ell\}_{\ell=1}^L$, and their revenue TFP, $\{T_\ell\}_{\ell=1}^L$. City size can either be exogenous or endogenous (e.g., free entry, preference shocks, migration costs, etc.). Likewise, local TFPs can either be exogenous or endogenous (e.g., productivity spillovers, market access, etc.). Both local characteristics are treated as given by firms.

Labor markets are competitive and segmented by locations. The law of one price holds within locations, and the local wage is denoted w_ℓ. Employers pay housing cost r_ℓ to produce in location ℓ. The housing supply is $L_\ell = \bar{L} r_\ell^\chi$.

Employers solve

$$\pi(z) = \max_{\ell,n} \; T_\ell z n^\rho - w_\ell n - r_\ell.$$

Their labor demand and optimal profits are

$$n_\ell(z) = \left(\frac{T_\ell z \rho}{w_\ell}\right)^{\frac{1}{1-\rho}} \quad \text{and} \quad \pi(z) = \max_\ell \; \kappa_1 \psi_\ell z^{\frac{1}{1-\rho}} - r_\ell,$$

where $\kappa_1 \equiv (1-\rho)\rho^{\frac{\rho}{1-\rho}}$ is a constant, and ψ_ℓ summarizes the profitability of ℓ: $\psi_\ell \equiv \left(\frac{T_\ell}{w_\ell^\rho}\right)^{\frac{1}{1-\rho}}$.

The spatial allocation of firms is described by $\{M_\ell, \Gamma_\ell\}_{\ell=1}^L$ for M_ℓ the measure of employers and Γ_ℓ the local productivity distribution. As in the model of Section 2, the spatial allocation of firms is determined by the profit condition (17) and the feasibility condition (7). Finally, labor market clearing in every location demands

$$m_\ell = M_\ell \int n_\ell(z) d\Gamma_\ell(z) = M_\ell \left(\frac{T_\ell \rho}{w_\ell}\right)^{\frac{1}{1-\rho}} \mathbb{E}\left[z^{\frac{1}{1-\rho}}\right]. \tag{35}$$

The allocation of firms across space is determined by the complementarity between firms' productivity and local profitability, ψ_ℓ. More productive firms have a higher willingness to pay to access high local profitability. The profitability of a location depends in turn on its local revenue TFP, and on the price of the inputs. Absent revenue TFP, productive firms thus sort across space to access cheap labor. As a result, there is a negative correlation between the average firm productivity and local wages. In equilibrium, large cities attract more employers, which are relatively more productive, because they offer cheap labor.

Proposition 6 (Spatial wage inequality in competitive matching models)**.**
Suppose that revenue TFPs are homogeneous, $T_\ell = 1$. In any equilibrium,

- *Employers are more productive in larger cities, and there are more of them: $m_\ell > m_{\ell'}$ implies $M_\ell > M_{\ell'}$ and $\mathbb{E}_\ell[z] > \mathbb{E}_{\ell'}[z]$;*
- *Wages are lower in larger cities: $m_\ell > m_{\ell'}$ implies $w_\ell < w_{\ell'}$.*

Proof. Since $T_\ell = 1$, I order cities by city size wlog, $m_1 < m_2 < \cdots < M_L$. Using the equilibrium condition for wages (35), local profitability rewrites

$$\psi_\ell \propto \left(\frac{m_\ell}{M_\ell \mathbb{E}_\ell \left[z^{\frac{1}{1-\rho}} \right]} \right)^\rho = \left(\frac{\rho}{w_\ell} \right)^{\frac{\rho}{1-\rho}},$$

where the constant of proportionally is a function of ρ.

The constant ψ_ℓ is thus an equilibrium object that depends on the allocation of employers across space. However, given a distribution of $\{\psi_\ell\}_{\ell=1}^L$, the complementarity between (ψ_ℓ, z) implies that there is pure positive assortative matching between $\boldsymbol{\psi}$ and $\boldsymbol{z}$ — or pure negative assortative matching between $\boldsymbol{w}$ and $\boldsymbol{z}$ (Topkis, 1998).

With these initial remarks, I start by proving the second part of the proposition. For the sake of contradiction, suppose that wages are increasing in city size, $w_1 < w_2 < \cdots < w_L$. From the expression for wages, it must then be that

$$\frac{M_{\ell+1}}{M_\ell} \frac{\mathbb{E}_{\ell+1} \left[z^{\frac{1}{1-\rho}} \right]}{\mathbb{E}_\ell \left[z^{\frac{1}{1-\rho}} \right]} > \frac{m_{\ell+1}}{m_\ell} > 1. \tag{36}$$

That is, $M_\ell \mathbb{E}_\ell[z]$ must be increasing in m_ℓ. I also know firms sort in the opposite direction of w_ℓ. Hence, $\mathbb{E}_\ell[z^{\frac{1}{1-\rho}}]$ is decreasing in m_ℓ. A necessary condition for (36) is therefore that M_ℓ is increasing in m_ℓ.

In any equilibrium with pure sorting, marginal firms must be indifferent between two locations. Since w_ℓ is increasing in m_ℓ and firms sort in the opposite direction of w_ℓ, this indifference is $\pi_\ell(\underline{z}_\ell) = \pi_{\ell+1}(\bar{z}_{\ell+1})$. Using the expression for profits and housing prices, the indifference condition reads

$$\kappa_2 \left[\psi_\ell - \psi_{\ell+1} \right] \underline{z}_\ell^{\frac{1}{1-\rho}} = \left(\frac{M_\ell}{\bar{L}} \right)^{1/\chi} - \left(\frac{M_{\ell+1}}{\bar{L}} \right)^{1/\chi} > 0,$$

for $\kappa_2 > 0$ a parametric constant. Therefore, M_ℓ is decreasing in ℓ, and (36) does not hold –a contradiction. The first part of the proposition automatically follows from Topkis (1998). □

F Quantitative model

F.1 Setup

Workers. The HJB dictating the lifetime utility of an unemployed workers born in l living in ℓ with preferences is

$$(\rho + \xi)U_{l\ell}(\boldsymbol{\omega}) = \frac{\kappa_{l\ell}\omega_\ell A_\ell b}{P_\ell} + \zeta\lambda_\ell \int [V_{l\ell}(w, \boldsymbol{\omega}) - U_{l\ell}(\boldsymbol{\omega})]^+ dF_\ell(w), \tag{37}$$

where $\kappa_{l\ell} = 1 - \kappa 1\{l \neq \ell\}$ and $[\cdot]^+ = \max\{\cdot, 0\}$. That of an employed workers with the same characteristics working at wage w is

$$(\rho + \xi)V_{l\ell}(\boldsymbol{\omega}, w) = \frac{\kappa_{l\ell}\omega_\ell A_\ell w}{P_\ell} + \lambda_\ell \int [V_{l\ell}(w', \boldsymbol{\omega}) - V_{l\ell}(w, \boldsymbol{\omega})]^+ dF_\ell(w') + \delta(U_{l\ell}(\boldsymbol{\omega}) - V_{l\ell}(\boldsymbol{\omega}, w)). \tag{38}$$

As in Section E.1, I guess and verify that the lifetime utility of unemployed workers rewrite

$$U_{l\ell}(\boldsymbol{\omega}) = \frac{\kappa_{l\ell}\omega_\ell A_\ell \mathcal{W}_\ell}{P_\ell},$$

where $\mathcal{W}_\ell$ solves a similar equation as (28) adjusted for the exit rate ξ.

The solution to the worker problem can therefore still be solved recursively. Given a spatial allocation of workers $\{m_\ell\}_{\ell=1}^L$, the search block continues to satisfy the expressions in Section E.2 and E.3, adjusting the separation rate δ by the exit rate ξ.

Turning to the spatial allocation of workers, two preliminary remarks are warranted. First, since taste shocks are constant over time and the economy is in steady state, workers move only once — when they enter the labor force. As a result, the measure of people who decide to live in ℓ upon entry is the same as the measure of people who actually live there. Second, given the birth process, the measure of workers born in ℓ is the same as the measure of people living in ℓ.

Given these remarks, let $m_{l\ell}$ denote the measure of workers born in l who choose to start their career in ℓ. Given the Frechet shocks, this is given by

$$\frac{m_{l\ell}}{m_l} = \frac{[\kappa_{l\ell} A_\ell \mathcal{W}_\ell / P_\ell]^\chi}{\sum_{\ell'} [\kappa_{l\ell'} A_{\ell'} \mathcal{W}_{\ell'} / P_{\ell'}]^\chi}. \tag{39}$$

The total number of people starting their career in ℓ — which equates the total number of people living in ℓ — is $m_\ell = \sum_l m_{l\ell}$.

Employers. Employers solve (20). Consider first the optimal wage and vacancy postings given a spatial allocation of employers. The wage and vacancy optimality condition demands

$$\frac{n'_\ell[w_\ell(z)]}{n_\ell[w_\ell(z)]}(zT_\ell - w_\ell(z)) = 1, \tag{40}$$

$$(zT_\ell - w_\ell(z))\, n_\ell[w_\ell(z)] = v_\ell(z)^\gamma. \tag{41}$$

The wage offer distribution satisfies

$$F_\ell(w) = \frac{M_\ell}{V_\ell} \int 1\{w_\ell(z) \leq w\} v_\ell(z) d\Gamma_\ell(z). \tag{42}$$

The solution to (40):(42) can be recasted as a system of two differential equations. Let $\Upsilon_\ell(z) \equiv F_\ell[w_\ell(z)]$ denote the rank of an employer in the local wage offer distribution under their optimal wage strategy. This implies $d\Upsilon_\ell(z) = dF_\ell[w_\ell(z)]w'_\ell(z)$. Used in the wage optimality condition (40):

$$w'_\ell(z) = \frac{2k_\ell \mathrm{d}\Upsilon_\ell(z)}{1 + k_\ell \bar{\Upsilon}_\ell(z)} [zT_\ell - w_\ell(z)]. \tag{43}$$

Meanwhile, (42) implies $d\Upsilon_\ell(z) = M_\ell v_\ell(z) d\Gamma_\ell(z)/V_\ell$. Used in the vacancy optimality condition (41), I obtain a second differential equation:

$$d\Upsilon_\ell(z) = \frac{M_\ell}{V_\ell} \left[(zT_\ell - w_\ell(z))\, n_\ell(z)\right]^{\frac{1}{\gamma}} d\Gamma_\ell(z). \tag{44}$$

Equations (43) and (44) are subject to two boundary conditions: $w_\ell(\underline{z}_\ell) = \underline{w}_\ell$ and $\Upsilon_\ell(\underline{z}_\ell) = 0$.

The local productivity distributions are given by the local profit opportunities, π_ℓ, together with the employers' idiosyncratic costs. The probability that a firm with productivity z produces in city ℓ is

$$\Omega_\ell(z) = \frac{\mathrm{e}^{\vartheta \pi_\ell(z)}}{\sum_{\ell'} \mathrm{e}^{\vartheta \pi_{\ell'}(z)}}. \tag{45}$$

The local productivity distribution in city ℓ is then

$$\Gamma_\ell(z) = \frac{M}{M_\ell} \int_{\underline{z}_\ell}^{z} \Omega_\ell(x) \mathrm{d}\Gamma(x), \tag{46}$$

and the mass of firm in city ℓ is

$$M_\ell = M \int \Omega_\ell(x) \mathrm{d}\Gamma(x). \tag{47}$$

The idiosyncratic entry costs ensure full support and $\underline{z}_\ell = \underline{z}$ in every location. Finally, M adjusts such that the free entry condition $\mathbb{E}[\pi_\ell(z, \boldsymbol{\varepsilon})] = f$ holds.

Frictions. The contact rates are given by the local matching function:

$$\lambda_\ell = \mu \left(\frac{V_\ell}{\zeta u_\ell + e_\ell} \right)^{\psi}. \tag{48}$$

Local prices. There are three sets of aggregate quantities. The reservation wages $\{\underline{w}_\ell\}_{\ell=1}^{L}$ are given by (12). Commercial housing prices clear the commercial housing market. The housing demand is M_ℓ. The housing supply reads $H^c_\ell = L^c \left(p^c_\ell \beta^c\right)^{\frac{\beta^c}{1-\beta^c}}$. Combined:

$$M_\ell = L^c \left(p^c_\ell \beta^c\right)^{\frac{\beta^c}{1-\beta^c}}. \tag{49}$$

Likewise, the residential housing prices clear the residential housing market. The residential housing demand is $\alpha/p^r_\ell(u_\ell b + e_\ell \mathbb{E}[w_\ell])$. The housing supply is $H^r_\ell = L^r \left(p^r_\ell \beta^r\right)^{\frac{\beta^r}{1-\beta^r}}$. Combined, this implies

$$\frac{\alpha}{p^r_\ell} \left(u_\ell b + e_\ell \mathbb{E}_\ell[w]\right) = L^r \left(p^r_\ell \beta^r\right)^{\frac{\beta^r}{1-\beta^r}}. \tag{50}$$

F.2 Consumer surplus

Accounting. I measure consumer surplus via the ex-post lifetime utility of workers. The ex-post discounted lifetime utility of new entrants (and unemployed) born in l and working in ℓ is straightforward. From the Frechet algebra, it is equal to their ex-ante lifetime utility

$$\bar{U}_{l\ell}^{\text{post}} \equiv \mathbb{E}[U_{l\ell}(\boldsymbol{\omega}) \mid l, \ell] = \bar{U}_l^{\text{ante}} \quad \Rightarrow \quad \mathbb{E}[\kappa_{l\ell}\omega_\ell \mid l, \ell] = \left(\frac{P_\ell}{A_\ell \mathcal{W}_\ell}\right) \bar{U}_l^{\text{ante}},$$

where the second expression follows from $\bar{U}_{l\ell}^{\text{post}} = A_\ell \mathcal{W}_\ell \mathbb{E}[\kappa_{l\ell}\omega_\ell \mid l, \ell]/P_\ell$, and $\bar{U}_l^{\text{ante}} = \mathbb{E}[U_{l\ell}(\boldsymbol{\omega}) \mid l] \propto (\sum_\ell [\kappa_{l\ell} A_\ell \mathcal{W}_\ell / P_\ell])^{\frac{1}{\chi}}$. For employed workers, their preferences are time-invariant and they cannot move with their job. Hence, the average preferences of employed workers equate those of new entrants. It follows that the average discounted lifetime utility of those born in l and working in ℓ at a wage w is[31]

$$\bar{V}_{l\ell}^{\text{post}}(w) = \frac{A_\ell \mathcal{V}_\ell(w) \mathbb{E}[\kappa_{l\ell}\omega_\ell \mid l, \ell]}{P_\ell}$$

Given these, the average ex-post lifetime utility of workers born in ℓ and working in ℓ is

$$\overline{UV}_{l\ell}^{\text{post}} = \frac{u_\ell}{m_\ell} \bar{U}_{l\ell}^{\text{post}} + \frac{e_\ell}{m_\ell} \int \bar{V}_{l\ell}^{\text{post}}(w) dG_\ell(w).$$

Then, the average ex-post lifetime utility of workers born in ℓ is

$$\overline{UV}_l^{\text{birth}} = \sum_\ell \text{Prob}(\text{work} = \ell \mid \text{birth} = l) \overline{UV}_{l\ell}^{\text{post}},$$

where $\text{Prob}(\text{work} = \ell \mid \text{birth} = l) = (A_\ell \mathcal{W}_\ell \kappa_{l\ell}/P_\ell)^\chi / \sum_{\ell'} (A_{\ell'} \mathcal{W}_{\ell'} \kappa_{l\ell'}/P_{\ell'})^\chi$. Finally, aggregate ex-post lifetime utility is given by

$$\overline{UV}^{\text{post}} = \sum_l m_l \overline{UV}_l^{\text{birth}}.$$

Lifetime earnings equivalence. I express welfare changes in lifetime earnings equivalence: the increase in lifetime earnings that renders workers indifferent between two equilibria holding other prices constant. Specifically, let $\overline{UV}_l^{\text{post}}(\zeta, \boldsymbol{\tau})$ be the ex-post average lifetime utility of workers born in l when the policy is $\boldsymbol{\tau}$ and their lifetime earnings are multiplied by ζ. The lifetime earnings equivalence for workers born in l is defined by $\overline{UV}_l^{\text{birth}}(\zeta_l, 1) = \overline{UV}_l^{\text{birth}}(1, \boldsymbol{\tau})$, and in the aggregate it is $\overline{UV}^{\text{birth}}(\zeta, 1) = \overline{UV}^{\text{birth}}(1, \boldsymbol{\tau})$.

I now show that $\zeta_l = \overline{UV}_l^{\text{birth}}(1, \boldsymbol{\tau}) / \overline{UV}_l^{\text{birth}}(1, 1)$. Using consistent notation throughout, it is straightforward to guess and verify from (28) that $\mathcal{W}_\ell(\zeta, \boldsymbol{\tau}) = \zeta \mathcal{W}_\ell(1, \boldsymbol{\tau}) = \zeta \mathcal{W}_\ell(\boldsymbol{\tau})$ and $\mathcal{V}_\ell(\zeta, \boldsymbol{\tau}) = \zeta \mathcal{V}_\ell(\boldsymbol{\tau})$. Accordingly, holding housing prices constant, $\bar{U}_l^{\text{ante}}(\zeta, \boldsymbol{\tau}) = \zeta \bar{U}_l^{\text{ante}}(1, \boldsymbol{\tau})$. It follows that $\overline{UV}_{l\ell}^{\text{post}}(\zeta, \boldsymbol{\tau}) = \zeta \overline{UV}_{l\ell}^{\text{post}}(1, \boldsymbol{\tau})$. Lastly, since ζ affects all workers born in l equally, it does not affect $\text{Prob}(\text{work} = \ell \mid \text{birth} = l)$, and $\overline{UV}^{\text{birth}}(\zeta, \boldsymbol{\tau}) = \zeta \overline{UV}^{\text{birth}}(1, \boldsymbol{\tau})$, as claimed.

[31] For employed workers, $\bar{\mathcal{V}}_{l\ell}^{\text{post}}(w) \neq \bar{\mathcal{V}}_{l\ell'}^{\text{post}}(w)$ since free mobility does not hold.

F.3 Worker heterogeneity

This appendix extends the quantitative model to allow for worker heterogeneity (Engbom and Moser, 2022; Bilal and Lhuillier, 2021). I use it to provide conditions under which the AKM model (1) is well-specified. I quantify the strength of the mis-specification bias when the conditions are not met. The mis-specification bias is only a function of firms' wage setting strategy, and I thus take as given the spatial allocation of skills and firms throughout the section.

Workers are endowed with a time-invariant skill $s \in \mathcal{S}$. The home production technology scales up with workers' productivity: unemployed workers earn $b_s = bs$. The density of skill s in location ℓ is denoted $\phi_{\ell s} ds$. Labor markets are segmented by skills, and workers only search in their skill-specific market. Search frictions may vary across skills, and $k_{\ell s}$ is a sufficient statistics. The yield of a vacancy in market (ℓ, s) is

$$n_{\ell s}(w) = \frac{e_{\ell s}}{V_{\ell s}} \frac{1 + k_{\ell s}}{[1 + k_{\ell s}(1 - F_{\ell s}(w))]^2}.$$

The productivity distribution in location ℓ is Γ_ℓ. The revenue generated by a match (z, s) is zs. Skills are perfect substitutes in production, and I maintain the constant returns to scale assumption. Accordingly, the total revenues generated by a firm z when they hire $\{n_s\}_s$ is

$$R(z, \{n_s\}_s) = \int zsn_s ds.$$

Firms decide how many vacancies to post for each skill, and the vacancy cost elasticity is identical across skills. Firms solve

$$\pi_\ell(z) = \max_{\{l_s, w_s, v_s\}_s} \int zsl_s ds - \int w_s l_s ds - \int \frac{v_s^{1+\gamma}}{1+\gamma} ds \tag{51}$$

subject to $l_s \leq v_{\ell s} n_{\ell s}(w)$.

This maximization problem can be solved skill-by-skill. For each skill, firms are on their labor supply curves, and wages are increasing in firms' productivity. The optimality conditions thus resemble (41) and (43) with s-subscripts. In particular, the wage equation is given by

$$w_{\ell s}(z) = \underline{w}_{\ell s} \left(\frac{n_{\ell s}(\underline{z}_\ell)}{n_{\ell s}(z)} \right) + \int_{\underline{z}_\ell}^{z} ts \frac{n'_{\ell s}(t)}{n_{\ell s}(z)} dt, \tag{52}$$

where

$$n_{\ell s}(z) \equiv \frac{e_{\ell s}}{V_{\ell s}} \frac{1 + k_{\ell s}}{[1 + k_{\ell s}(1 - \Upsilon_{\ell s}(w))]^2},$$
$$\Upsilon_{\ell s}(z) \equiv F_{\ell s}[w_{\ell s}(z)] = \frac{1}{V_{\ell s}} \int_{\underline{z}_\ell}^{z} v_{\ell s}(t) d\Gamma_\ell(t),$$

and the reservation wage is given by (12) adjusted for the s-subscript.

Using (52), Proposition 7 shows that the AKM model (1) is well-specified if search frictions are identical across skills (within a location).

Proposition 7 (AKM in BM)**.**
Suppose $k_{\ell s} = k_\ell$ for all s. Then, the AKM model (1) *is well-specified. In particular, $\alpha_i = s_i$,*

$\beta_\ell = 0$, *and*

$$\gamma_j = \underline{w}_{\ell_j}\left(\frac{n_{\ell_j}(\underline{z}_{\ell_j})}{n_{\ell_j}(z_j)}\right) + \int_{\underline{z}_{\ell_j}}^{z_j} t\frac{n'_{\ell_j}(t)}{n_{\ell_j}(z_j)}dt.$$

Proof. The proof is a guess and verify. Suppose $k_{\ell s} = k_\ell$. Guess that $w_{\ell s}(z) = sw_\ell(z)$ and $\Upsilon_{\ell s}(z) = \Upsilon_\ell(z)$ for some function $w_\ell(z)$. Since search frictions are constant across skills, the guesses imply $n_{\ell s}(z) = a_{\ell s}n_\ell(z)$ for $a_{\ell s}$ some constants. The vacancy optimality condition (41) then implies $v_{\ell s}(z) = \nu_{\ell s}v_\ell(z)$ for $\nu_{\ell s}$ another constant. The constant $\nu_{\ell s}$ scales aggregate vacancies $V_{\ell s}$, and it follows that $v_{\ell s}(z)/V_{\ell s} = v_\ell(z) \perp s$. This verifies $\Upsilon_{\ell s}(z) = \Upsilon_\ell(z)$. Moving onto wages, (52) reveals that $w_{\ell s}(z) = sw_\ell(z)$ if $\underline{w}_{\ell s} = s\underline{w}_\ell$. Using the guess $w_{\ell s}(z) = sw_\ell(z)$ into the expression for the reservation wage (12) yields $\underline{w}_{\ell s} = s\underline{w}_\ell$, and thus verifies the guess. □

Whether search frictions vary across skills is an empirical question. I assess through simulations the strength of the mis-specification bias under several calibrations of $k_{\ell s}$.

I use the parameters reported in Table B.2 whenever possible. I take the local productivity distributions Γ_ℓ from the estimated quantitative model. I assume $\mathcal{S} \in \{1, \dots, \bar{s}\}$ is discrete with five skill types for numerical tractability. I set a uniform skill distribution in the aggregate without loss of generality, and calibrate $\bar{s}$ to match the aggregate variance of worker fixed effects in the data (0.066, which yields $\bar{s} = 1.9$). I calibrate the spatial allocation of skills to match the between-city worker fixed effects variance in the data. Specifically, I suppose the likelihood of skill s in location ℓ is $(m_\ell + A_\ell s^{\nu^m})/\sum_{\ell'}(m_{\ell'} + A_{\ell'}s^{\nu^m})$, for ν^m a parameter. The higher ν^m, the more high skill workers sort to high amenities cities. I set ν^m so that $\mathbb{V}\text{ar}[\mathbb{E}_\ell(\log s)] = 0.011$, and obtain $\nu^m = 7.7$.

I compute the mis-specification bias under various calibrations of $k_{\ell s}$. I set $k_{\ell s} = s^{\nu^k}k_\ell$, where $\nu^k \in \mathbb{R}$ is another parameter. For a given ν^k, I compute the wage distribution for each location and skill using (52). I then simulate a panel dataset at the worker level with 10 quarters and 1,000,000 workers. I estimate (1) within the simulated panel restricting $\beta_\ell = 0$ for all ℓ. I compute $\mathbb{V}\text{ar}[\hat{\alpha}_i]$, and report the difference between $\sqrt{\mathbb{V}\text{ar}[\hat{\alpha}_i]}$ and $\sqrt{\mathbb{V}\text{ar}[\log s_i]}$. Whenever $\sqrt{\mathbb{V}\text{ar}[\hat{\alpha}_i]} > \sqrt{\mathbb{V}\text{ar}[\log s_i]}$, the AKM model under-estimates the importance of employers for wage inequality.

The blue rectangles and orange circles in Figure C.5 report $\sqrt{\mathbb{V}\text{ar}[\log s_i]}$ and $\sqrt{\mathbb{V}\text{ar}[\hat{\alpha}_i]}$ for various ν^k. Since the magnitude of ν^k is hardly interpretable, I report these statistics against the difference in average job switching rate between the most skilled (top quintile) and least skilled (bottom quintile) workers. The dotted black lines present empirical estimates of this relative job switching rate in Bilal and Lhuillier (2021) and Engbom and Moser (2022).

Aligned with Proposition 7, the AKM model is well-specified when there is no dispersion in job switching rates across skills. AKM underestimates the skill variance when low-skill workers switch jobs more often, and overestimates it when high-skill workers move more often across jobs. For empirically plausible estimates of $k_{\ell s}$, the mis-specification bias is small, and if anything, AKM under-estimates the importance of employers for wage inequality.

G Estimation

G.1 Proof of Proposition 5

Frictions. The average job switching rate, average EU rate, and location-specific average UE rate identifies the search frictions. First, the EU rate equates δ. Second, the average career lifespan is $1/\xi$. Hence, $\bar{\delta} = \delta + \xi$. Third, the location-specific average UE rate equates the contact rate of unemployed workers, $\zeta\lambda_\ell = \mathrm{UE}_\ell$. Fourth, the job switching rate in ℓ is

$$\mathrm{J2J}_\ell = \lambda_\ell \int \bar{F}_\ell(w) dG_\ell(w) = \bar{\delta}\left[\left(1 + \frac{\bar{\delta}}{\lambda_\ell}\right)\log\left(1 + \frac{\lambda_\ell}{\bar{\delta}}\right) - 1\right].$$

The right-hand side is monotonically increasing in ζ, and therefore so is the aggregate switching rate $\sum_\ell (e_\ell/E)\mathrm{J2J}_\ell$, where e_ℓ/E is ℓ's employment share. This identifies ζ and λ_ℓ separately.

Worker allocation. The total mass of workers in ℓ reads $m_\ell = (\lambda_\ell^u + \bar{\delta})e_\ell/\lambda_\ell^u$, where $\lambda_\ell^u = \zeta\lambda_\ell$ for notational simplicity. Consistency imposes $\sum_\ell m_\ell = 1$. Together, $\frac{1}{E} = \sum_\ell \left(\frac{\lambda_\ell^u + \delta}{\lambda_\ell^u}\right)\frac{e_\ell}{E}$, which identifies the aggregate mass of employed workers. I then recover the location-specific measure of employed and unemployed workers via $e_\ell = \left(\frac{e_\ell}{E}\right)E$, $u_\ell = \left(\frac{\delta}{\lambda_\ell^u}\right)e_\ell$, and $m_\ell = u_\ell + e_\ell$. I treat the spatial allocation of workers as given for now, and later show that any allocation can be rationalized by local amenities.

Unemployment insurance. The home production technology is identified from the aggregate replacement rate. The location-specific replacement rate is $\mathrm{RE}_\ell = b/\mathbb{E}[w_\ell]$. The aggregate replacement rate is therefore

$$\mathrm{RE} = \frac{\sum_\ell \left(\frac{b}{\mathbb{E}[w_\ell]}\right) e_\ell}{\sum_\ell e_\ell},$$

which identifies b.

Employers. Let j denote an employer. For each employer, I observe the wage it offers, w_j, its size, n_j, and its location, ℓ_j. Consistency requires that the number of workers hired in each location equates the number of employed workers: $M_\ell \mathbb{E}_\ell[n_j] = e_\ell$. This allows to back out the measure of employers per location $\{M_\ell\}_{\ell=1}^L$. The total measure of employers follows $M = \sum_\ell m_\ell$.

I then recover the productivity and vacancy of each employer from its size and wage. Given employers' wage and size, I know its rank in the wage distribution, G_j. Inverting (31) returns its rank in the wage offer distribution F_j, and therefore the number of workers per vacancy that employer can hire: $\eta_j = \frac{k_\ell(\zeta u_\ell + e_\ell)}{[1+k_\ell(1-F_j)]^2}$. Employers are on their labor supply curves (18), which allows to recover their vacancy share $v_j/V_{\ell_j} = n_j/\eta_j$. Finally, employers are on their wage optimality condition (40), which implies[32]

$$z_j T_{\ell_j} \equiv \zeta_j = w_j + \left(\frac{1 + k_\ell(1 - F_j)}{2k_\ell}\right) \left.\frac{\partial w}{\partial F}\right|_{w=w_j}.$$

Given data on vacancy shares, the vacancy cost elasticity is identified from the vacancy optimality condition. Equation (41) indeed implies

$$\gamma = \sqrt{\frac{\sum_\ell e_\ell \mathbb{V}\mathrm{ar}_\ell[\log((\zeta_j - w_j)\eta_j)]}{\sum_\ell e_\ell \mathbb{V}\mathrm{ar}_\ell[\log \frac{v_j}{V_\ell}]}}. \tag{53}$$

[32]In practice, I approximate $w \to F_\ell$ with a spline and use the approximation to compute $\partial w/\partial F$.

Given an estimate for γ, the measure of vacancy in each location is then recovered as a location-specific residual from the optimality condition:

$$(1+\gamma)\log V_\ell = \mathbb{E}_\ell[\log(\zeta_j - w_j)\eta_j] - \gamma \mathbb{E}_\ell\left[\log \frac{v_j}{V_\ell}\right].$$

Together, I can compute vacancy costs $c(v_j) = v_j^{1+\gamma}/(1+\gamma)$.

From (45), the log probability that an employer with total productivity ζ produces in ℓ is

$$\log \Omega_\ell(z) = -\vartheta B_\ell + \vartheta \tilde{\pi}_\ell(z) - \log \sum_{\ell'} \mathrm{e}^{\vartheta \pi_{\ell'}(z)}, \tag{54}$$

where $\tilde{\pi}_\ell(z)$ is profits gross of the entry costs B_ℓ. According to (54), projecting employers' location probabilities onto the local profit opportunities conditional on location and productivity fixed effects identifies ϑ and $\{B_\ell\}$. Given the parameters already identified, realized profits can be computed

$$\tilde{\pi}_j = (\zeta_j - w_j)n_j - c(v_j) - r_{\ell_j}.$$

When TFP gaps are small, $\zeta_j \approx z_j$. Then, realized profits can be used to infer potential profits in other locations. Likewise, $\Omega_\ell(z)$ can be computed from the location choice of employers conditional on productivity.

I operationalize (54) with three modifications. First, I instrument local profit opportunities with employers' wage to reduce measurement errors.[33] Second, I cluster employers in groups of MPL within cities, I aggregate all the variables in (54) at the cluster level, and I replace $\log \sum_{\ell'} \mathrm{e}^{\vartheta \pi_{\ell'}(z)}$ with a group fixed effect. Third, I implement the same aggregation and instrument variable strategy in the model, and set ϑ to match the conditional correlation in the data.

Matching function. The matching function (48) implies

$$\log \lambda_\ell = \log \mu + \psi \log\left(\frac{V_\ell}{\zeta u_\ell + e_\ell}\right). \tag{55}$$

Estimating (55) by OLS given our estimates of $\{\lambda_\ell, V_\ell, u_\ell, e_\ell\}_{\ell=1}^{L}$ and ζ identifies the matching function parameters (μ, ψ).

Housing supply. The residential housing market clearing condition implies

$$\log p_\ell^r = (1-\beta^r)\log\frac{\alpha}{L^r} - \beta^r \log \beta^r + (1-\beta^r)\log I_\ell, \tag{56}$$

where $I_\ell = u_\ell b + e_\ell \mathbb{E}[w_\ell]$ are total expenditures in ℓ. Accordingly, given data on residential housing prices and an estimate for α, estimating the above relationship by OLS identifies $\{L^r, \beta^r\}$. In practice, residential housing prices are a function of the total local housing expenditures, not only the spending reflected in the employer fixed effects. To account for this, I estimate (56) controlling for worker heterogeneity:

$$\log p_\ell^r = (1-\beta^r)\log\frac{\alpha}{L^r} - \beta^r \log \beta^r + (1-\beta^r)\log I_\ell + (1-\beta^r)\mathbb{E}_\ell[s], \tag{57}$$

[33] The instrument is relevant since, within a city, wages, productivity, and profits are all monotonic functions of each other. The instrument satisfies the exclusion restriction because wages are raw data.

where $\mathbb{E}_\ell[s]$ is the average worker effect estimated in (1). This specification is consistent with the extended model that allows for skill heterogeneity (Section F.3).

The commercial market clearing condition is

$$\log p_\ell^c = -\frac{1-\beta^c}{\beta^c}\log L^c - \log\beta^c + \frac{1-\beta^c}{\beta^c}\log M_\ell^c, \tag{58}$$

which identifies (L^c, β^c).

Worker preferences. I conclude with the identification of worker preferences, $(\alpha, \boldsymbol{\kappa}, \boldsymbol{A})$. The Cobb-Douglas parameter α equates the aggregate housing expenditure share. The bilateral flows of workers born in l and working in ℓ relative to the number of stayers identify the migration costs given the dispersion in preferences:

$$\frac{m_{l\ell}}{m_{ll}}\frac{m_{\ell l}}{m_{\ell\ell}} = \kappa^{2\chi},$$

which holds for all $\ell \neq l$. Finally, amenities are identified from the location choice of workers. Equation (39), together with $m_\ell = \sum_l m_{l\ell}$, imply

$$\nu_\ell = m_\ell \left(\sum_l \frac{m_l \kappa_{l\ell}^\chi}{\sum_{\ell'}\kappa_{l\ell'}^\chi \nu_{\ell'}}\right)^{-1},$$

where $\nu_\ell \equiv (A_\ell \mathcal{W}_\ell/P_\ell)^\chi$. Given $(\kappa, \chi, \boldsymbol{m})$, the above equation identifies $\boldsymbol{\nu}$ up to a normalization. I normalize mean amenities to unity. Given $(\chi, \boldsymbol{\nu})$, housing prices $\boldsymbol{P}$, and the net present value of unemployed's lifetime earnings, $\boldsymbol{\mathcal{W}}$, I invert $\boldsymbol{\nu}$ to obtain amenities. I compute unemployed's lifetime earnings from

$$(\rho+\xi)\mathcal{W}_\ell = b + \lambda_\ell^u \int_{\underline{w}_\ell} \frac{1-F_\ell(w)}{\rho+\xi+\delta+\lambda_\ell(1-F_\ell(w))} dw,$$

where all the elements on the right-hand side have already been identified. This conclude the proof.

G.2 Estimating the taste shock dispersion

I first show that standard reduced-form local labor supply elasticity does not map into the taste shock dispersion χ in this model. I then describe how I use indirect inference to calibrate χ.

Static framework. To start, I lay down a standard static spatial framework. The spatial allocation of workers is given by

$$m_\ell = \frac{(A_\ell \bar{w}_\ell/P_\ell)^\chi}{\Phi},$$

where $\Phi = \sum_l (A_l \bar{w}_l/P_l)^\chi$ is a general equilibrium constant, and $\bar{w}_\ell$ refers to post-tax average wages. Taking logs and letting Δ denote time differences:

$$\Delta\log m_\ell = \chi\Delta\log A_\ell + \chi\Delta\log\frac{\bar{w}_\ell}{P_\ell} - \Delta\log\Phi. \tag{59}$$

Equation (59) justifies estimating the reduced form equation

$$\Delta \log m_\ell = \alpha + \beta \Delta \log \frac{\bar{w}_\ell}{P_\ell} + u_\ell, \tag{60}$$

to estimate χ. In (60), α is the GE constant, u_ℓ is the unobserved change in amenities, and $\beta = \chi$ is the coefficient of interest. Since amenities are not observed, an instrument for $\Delta \log \bar{w}_\ell / P_\ell$ orthogonal to $\Delta \log A_\ell$ is required. Changes in local taxes $\Delta \tau_\ell$ are often used (e.g., Fajgelbaum et al., 2019). This yields the 2SLS estimate

$$\gamma = \beta = \frac{\mathbb{C}\text{ov}(\Delta \tau_\ell, \Delta \log m_\ell)}{\mathbb{C}\text{ov}(\Delta \tau_\ell, \Delta \log \bar{w}_\ell / P_\ell)}. \tag{61}$$

Dynamic framework. Our framework departs in two ways from (59). First, the framework is dynamic, and individuals care about lifetime earnings. Second, individuals face migration costs. I start with the first component, and integrate migration costs afterwards. The allocation is now

$$\Delta \log m_\ell = \chi \Delta \log A_\ell + \chi \Delta \log \left(\frac{\bar{u}_\ell b + (1 - \bar{u}_\ell) \bar{w}_\ell}{P_\ell} \right) - \Delta \log \Phi, \tag{62}$$

where $\bar{u}_\ell$ is the unemployment rate, and I have assumed $\rho / \delta \to 0$. Accordingly,

$$\beta = \chi \frac{\mathbb{C}\text{ov}\left[\Delta \tau_\ell, \Delta \log \left(\frac{\bar{u}_\ell b + (1 - \bar{u}_\ell) \bar{w}_\ell}{P_\ell} \right) \right]}{\mathbb{C}\text{ov}(\Delta \tau_\ell, \Delta \log \frac{\bar{w}_\ell}{P_\ell})}$$

Suppose for now that $u_\ell = \bar{u}$.[34] When $\bar{u} = 0$, this returns to $\beta = \chi$. However, for $u > 0$, it implies

$$\beta \approx \chi \left(1 + b\bar{u} \frac{\mathbb{C}\text{ov}\left[\Delta \log \tau_\ell, \Delta \frac{1}{\bar{w}_\ell} \right]}{\mathbb{C}\text{ov}(\Delta \log \tau_\ell, \Delta \log \frac{\bar{w}_\ell}{P_\ell})} \right),$$

where the approximation is to first order around $u = 0$. In practice, we expect $\mathbb{C}\text{ov}\left[\Delta \tau_\ell, \Delta \frac{1}{\bar{w}_\ell} \right] > 0$. As a result, $\beta < \chi$.

Migration costs. Suppose now I also introduce migration costs of the form $\kappa_{l\ell} = 1 - \kappa 1\{\ell \neq l\}$. The allocation is then

$$\Delta \log m_\ell = \Delta \chi \log A_\ell + \Delta \chi \log \left(\frac{\bar{u}_\ell b + (1 - \bar{u}_\ell) \bar{w}_\ell}{P_\ell} \right) - \Delta \log \left(\theta X_\ell + (1 - \theta) \bar{X} \right),$$

where $\theta \equiv 1 - (1 - \kappa)^\chi$, $X_\ell \equiv \frac{1}{\theta (A_\ell \mathcal{W}_\ell / P_\ell)^\chi + (1 - \theta) \sum_{\ell'} (A_{\ell'} \mathcal{W}_{\ell'} / P_{\ell'})^\chi}$, and $\bar{X} \equiv \sum_l X_l$. The last term is now non-constant. In particular, it depends on how good ℓ is vis-à-vis the rest of the locations.

To see the bias this introduces, suppose I shut down the unemployment channel by setting $\bar{u}_\ell = 0$ for all ℓ. Then:

$$\beta = \chi - \frac{\mathbb{C}\text{ov}(\Delta \tau_\ell, \Delta \log \left(\theta X_\ell + (1 - \theta) \bar{X} \right))}{\mathbb{C}\text{ov}(\Delta \tau_\ell, \Delta \log \bar{w}_\ell / P_\ell)}.$$

[34] The argument extends to $u_\ell = \bar{u} + \varepsilon_\ell$ for ε_ℓ i.i.d. and small variance.

To first order around $\theta = 0$ (or $\kappa = 0$):

$$\beta \approx \chi - \frac{\theta}{\bar{X}} \frac{\mathbb{C}\text{ov}(\Delta\tau_\ell, \Delta X_\ell)}{\mathbb{C}\text{ov}(\Delta\tau_\ell, \Delta \log \bar{w}_\ell / P_\ell)}.$$

From the expression for X_ℓ: higher taxes lowers $\mathcal{U}_\ell$, and thus $\mathbb{C}\text{ov}(\Delta\tau_\ell, \Delta X_\ell) > 0$. As a result, $\beta > \chi$.

Indirect inference. The reduced-form local labor supply elasticity thus does not map into the taste shock dispersion. Instead, I use indirect inference to calibrate χ. For a given χ, I compute the amenities and migration costs that rationalize the data. I then simulate the new equilibrium under a vector of local tax reforms $\{\tau_\ell\}_{\ell=1}^{L}$, and compute (61) within the model. I set χ to minimize the distance

$$F(\chi) = \hat{\beta} - \frac{\mathbb{C}\text{ov}(\Delta\tau_\ell, \Delta \log m_\ell)}{\mathbb{C}\text{ov}(\Delta\tau_\ell, \Delta \log \bar{w}_\ell / P_\ell)},$$

where $\hat{\beta}$ are empirical estimates of the reduced-form local labor supply elasticity, and the ratio of covariances is computed in the model. I target $\hat{\beta} = 2.5$, which is in the middle of the range of the empirical estimates (e.g., Table A.17 in Fajgelbaum et al., 2019). I simulate i.i.d. local tax reforms drawn from a normal distribution with standard deviation σ_τ. For each χ, I simulate ten vectors of tax reforms. In the baseline, I set $\sigma_\tau = 0.01$. The blue line in Figure C.4 shows the reduced-form local labor supply elasticity in the model as a function of τ. The horizontal dotted line depicts $\hat{\beta} = 2.5$, and the vertical dotted line represents the estimated taste shock dispersion. The orange and grey lines show the reduced-form local labor supply elasticity in the model when the standard deviation of the tax reforms is 0.005 and 0.05 respectively.