

---

# An Anatomy of Monopsony: Search Frictions, Amenities, and Bargaining in Concentrated Markets

**David Berger**, *Duke University and NBER, United States of America*

**Kyle Herkenhoff**, *University of Minnesota, Federal Reserve Bank of Minneapolis, and NBER, United States of America*

**Andreas R. Kostøl**, *Arizona State University W. P. Carey School of Business, United States of America, Norges Bank, Norway, and NBER, United States of America*

**Simon Mongey**, *Federal Reserve Bank of Minneapolis and NBER, United States of America*

## I. Introduction

There is a growing consensus that imperfect competition in the labor market is pervasive.<sup>1</sup> Many local labor markets are dominated by a few firms, which gives them the ability to set wages and pay workers less than their marginal product. In his 2022 AEA presidential address, David Card argued that developing a tractable framework combining preference heterogeneity (in the Daniel McFadden-industrial organization tradition) and search and matching frictions with job ladders (in the Postel-Vinay and Robin [2002] tradition) were key to understanding the importance of market power in the labor market. The goal of this paper is to take a first step toward answering this charge. We develop a theory of monopsony that incorporates the three paradigms of the modern monopsony literature: worker-firm-specific preference heterogeneity (Robinson 1933), search frictions (Burdett and Mortensen 1998), and firm granularity (Berger, Herkenhoff, and Mongey 2022a; Jarosch, Nimcsik, and Sorkin 2024). We then quantify our framework using Norwegian worker-firm data and use it to answer several pertinent questions: How do the three sources of monopsony interact to shape wages, job flows, and welfare? Which sources of monopsony account for the wedge between a worker's pay and marginal product (henceforth, the wage mark-down)? How does monopsony power affect wage inequality?

*NBER Macroeconomics Annual*, volume 38, 2024.

© 2024 National Bureau of Economic Research. All rights reserved. Published by The University of Chicago Press for the National Bureau of Economic Research. <https://doi.org/10.1086/729194>

We make three contributions. Empirically, we use Norwegian administrative data to document that, within an occupation, wage levels, wage inequality, and job flows correlate systematically with local labor market concentration. Theoretically, we develop a model of frictional labor markets with a finite number of firms, as well as on-the-job search, worker-firm-specific nonwage amenities, and vacancy posting. Although some recent papers include the first two, vacancy posting closes the model in equilibrium and provides the first theory of search in granular markets that admits counterfactuals. Quantitatively, we use the Norwegian data and the structure of our model to discipline and then quantify the wage, welfare, and job flow implications of each source of monopsony power. Our empirics motivate a number of assumptions in our theory and allow us to conduct overidentification tests on the role of concentration in the quantitative framework.

Our framework implies that granularity in the wage bargaining process, amenities, and search frictions account for one-half, one-third, and two-thirds of wage markdowns, respectively, with the nonadditivity arising from the nonlinearity of our model. Although amenities and search frictions are studied extensively in the literature, our results suggest that more exploration of granularity and strategic interactions in the bargaining process are an important next step for the monopsony literature.

### *A. Empirics*

We begin by using detailed data about workplace locations and workers' line of work to document the relationship between concentration levels, wages, wage inequality, and job flows. Using two separate fixed-effect specifications that isolate within occupation-year, across-region variation and within occupation-region, across-time variation, we document a set of covariances between market concentration and labor market characteristics. More concentrated markets are associated with lower wages, less wage dispersion, lower employer-to-employer job flow rates, and lower job-finding rates. It is well known that job-to-job transitions are a key source of wage growth (see, e.g., Postel-Vinay and Robin 2002). These strong links between job flows and concentration suggest that on-the-job search may be an important mechanism through which market structure affects the level and dispersion of wages.

### *B. Theory*

Motivated by these findings, we develop a theory that incorporates neoclassical sources of monopsony as well as frictional job flows and

concentration. Our model features a finite number of firms, on-the-job search, worker-firm-specific nonwage amenities, and strategic wage setting. The first two are necessary to replicate our empirical finding that both employer-to-employer flows and wages are lower in more concentrated markets. Firm contact rates are determined in general equilibrium by optimal firm vacancy posting, given the endogenous distribution of workers across firms.

Our framework substantially extends Postel-Vinay and Robin (2002) to accommodate a finite number of firms in each labor market and worker-firm-specific nonwage amenities. The number of firms in each market is given by  $M < \infty$ , and firms differ in their idiosyncratic but fixed productivity levels. Unemployed workers randomly meet vacancies of the  $M$  firms within their labor market, whereas employed workers randomly meet vacancies of the remaining  $M - 1$  firms, excluding their current employer. When an unemployed worker meets a firm, the worker draws a worker-firm-specific nonwage amenity, and the parties Nash bargain over surplus. When an employed worker meets a firm, the worker draws a new worker-firm-specific amenity, and the incumbent and poaching firms compete via alternating offers (e.g., Cahuc, Postel-Vinay, and Robin 2006), yielding equilibrium values in which outside options are determined by Bertrand competition and the remaining surplus is split by Nash bargaining. Last, we assume that firms optimally choose vacancies, taking match surplus and market contract rates as given. Modeling of a firm's vacancy-posting decisions explicitly generates endogenous contact rates, which both closes the model and delivers a meaningful firm size distribution.

The surplus-sharing protocol yields strategic complementarity between wage offers of incumbent and poaching firms. In the Cahuc et al. (2006) class of models, regardless of  $M$ , only two firms bargain. This "hardwired" duopsony is an important source of markdowns that operates through strategic wage setting. Allowing more firms at the bargaining table will reduce markdowns and increase efficiency. We contribute a stylized counterfactual that raises this point, and we leave more careful extensions of the theory to future research. Future researchers may try to (i) understand how changes in  $M$  affect the number of firms that are bargaining over any particular match and (ii) provide a theory for bargaining with more than two firms.

There are several caveats to our approach. Neoclassical models of monopsony focus on worker-firm-specific amenities as generating a mechanism by which firms must increase wages to attract more workers. We take those same worker-firm-specific amenities and study them in a

search and matching model where we find they are necessary to match features of the data and provide market power for firms. However, we assume that they operate differently than in the neoclassical case. We allow firms to observe amenity draws and perfectly price discriminate. This differs with Robinson (1933), who treats amenities as unobserved and thus only allows for third-degree discrimination. An important implication of perfect price discrimination is that amenities themselves do not necessarily yield inefficient allocations.

### C. *Quantification*

We use linked employer-employee data from Norway to discipline the quantitative model. The administrative data offer several advantages over other data sources. The Norwegian data include information about the type of work that each employee is hired to do (i.e., their occupation) and the workplace location of every employment contract from 2006 to 2016. These complete records allow us to accurately measure job flows and thus better classify local labor markets.

We contribute a simple clustering algorithm to define local labor markets, which we apply to these data. Rather than working with connected sets of firms (which is computationally demanding), our approach uses the much lower-dimensional occupation-to-occupation flow transition matrix. We first isolate single-occupation markets with high self-flow rates. Among the remaining occupations, we K-means cluster the rows of the occupation-to-occupation flow transition matrix. The resulting groups are occupations with similar job flow patterns. We determine the optimal number of clusters using an objective function that is increasing in the self-flow rate but decreasing in the concentration of occupations in each cluster. This rewards the lowering of self-flow rates but penalizes the classification of all occupations in one large cluster and thus “overfitting” the data.

We do not innovate on the dimension of geography; we simply define the boundary of a market to be the boundary of the commuting zone as computed in Bhuller (2009). This yields approximately 5,000 markets with a self-flow rate of 51% (among job transitioners, 51% transition back into the same market). Using three-digit occupation by commuting zone to define the market yields a similar number of markets but a lower self-flow rate of 45%. In an approach that intuitively groups connected sets of firms (stochastic blocks), Jarosch et al. (2024) find 376 markets with a self-flow rate of 40% in Austria. Worker occupation data allow

us to define markets that both have high self-flow rates and are computationally easy to compute.

With market definitions in hand, we estimate the model to match key moments between 2006 and 2016 in Norway. First, we directly import market structures observed in the Norwegian data, including the number of firms in a market and the labor force in the market. Second, we discipline the role of amenities using the fraction of employer-to-employer (E-to-E) job moves down the ladder, where rungs are defined by the poaching index following Bagger and Lentz (2019). Third, we discipline the role of search frictions based on unemployment and E-to-E rates. Fourth, we discipline the bargaining power of workers using wage growth. We estimate the model on an overidentified set of moments to ensure that our model is consistent with observed covariances of market Herfindahl values and the level and standard deviation of wages. Despite its parsimony, the quantitative model fits nontargeted cross-market moments from our earlier empirics, generating lower E-to-E rates and unemployment-to-employment (U-to-E) rates in more concentrated markets.

#### *D. Results*

In our main results, we use the model to generate five counterfactuals that highlight the importance of labor market competition for markdowns, welfare, and wage inequality.

##### Concentration

We explore the role of concentration in depth because it is the newest element of our analysis. Our goal is to vary the number of firms ( $M$ ) while holding the distribution of productivity ( $z$ ) and bargaining protocol fixed. To implement this “idealized” heuristic experiment, we draw a vector of 10 productivities from the ergodic distribution and duplicate this productivity vector 10 times to construct 10 markets. We then organize the productivity vectors into three counterfactual economies (from most to least concentrated): 10 identical 10-firm markets, two identical 50-firm markets, and one 100-firm market. We hold the ratio of firms per worker fixed to remove mechanical effects from adding firms to a market. We find that workers’ share of surplus monotonically increases as markets become less concentrated. A side effect of a higher surplus share is a reduction in compensating differentials; that is, the amenity wage penalty shrinks. With less concentration, wages  $w$  rise, markdowns

narrow (i.e.,  $w/z$  increases), and welfare increases. Inequality increases: across firms, more productive firms pay higher wages, and within firms, some workers are paid more due to better bargaining opportunities. Consolidating all 100 firms into a single market allows more workers to reach the highest-productivity firms, increasing output. Accordingly, unemployment falls, and E-to-E rates increase as more meetings result in job transitions.

We next explore the role of concentration in the actual Norwegian economy. We double the number of firms in the Norwegian economy by duplicating the existing productivity vector in every market (i.e., every firm's *doppelgänger* enters the market, leaving the job ladder rungs untouched) and double the number of workers in the market such that the number of firms per worker remains constant. Although this experiment yields results that are qualitatively consistent with our idealized heuristic experiment, the quantitative effects of concentration on wages and welfare are limited. Markdowns narrow by 1 percentage point, average wages rise by 0.68%, and the standard deviation of log wages increases by 0.81%. Why are the effects of changing  $M$  so small? There are two reasons: (1) approximately 70% of the labor force resides in markets with more than 150 firms, and hence, doubling the number of firms in these markets is irrelevant, and (2) the duopsony wage-setting assumption—that is, that only two firms at the bargaining table strategically set wages—remains unchanged regardless of  $M$ .

### Exclusion

To explore the effect of firms removing themselves from future E-to-E contacts with the worker, we allow a worker at a given firm to meet that firm again, rebargain, and thus extract all the surplus. This mechanism is related to the approaches in Zhu (2012) and Jarosch et al. (2024).<sup>2</sup> We find that this leads to an economically small change in observed markdowns, narrowing them by approximately 1 percentage point compared with the 21-percentage-point markdown in the baseline economy. Again, the reason is that the bulk of the labor force resides in markets with many firms, limiting the impact of self-exclusion.

### Preference Heterogeneity

To quantify the effects of preference heterogeneity (i.e., nonwage, worker-firm-specific amenities), we eliminate all variations in amenities and set

them to a single value, resulting in a uniform level of amenities across the economy, while we maintain the same level of aggregate amenities as in our baseline economy. In this counterfactual, we find that the Herfindahl index more than triples. Workers now flow to the highest-productivity firm and stay there, unlike in our baseline economy, where differences in amenities can cause workers to leave the highest-productivity firms. High-productivity firms also post more vacancies, understanding that workers are less likely to leave due to idiosyncratic tastes. As a result, output, productivity, and welfare increase substantially. Without amenity dispersion, wage inequality falls, but wage levels rise as workers flow to—and stay at—more productive firms. Despite the level of amenities being the same, markdowns narrow by 7 percentage points.<sup>3</sup> This represents a 30% reduction in markdowns from the level in the baseline economy.

### Search Frictions

To quantify the effects of search frictions, we increase worker contact rates to 100% per period, leading workers to always meet a firm in every period regardless of their employment status. Markdowns narrow by 14 percentage points, which represents a 60% reduction from the level in the baseline economy. Greater contact with lower-ranked firms allows workers at higher-productivity firms to bid up their share of surplus, and they are also more likely to meet the highest-ranked firms. The Herfindahl index rises to 0.75 (an eightfold increase), as workers rapidly climb the job ladder. Wage inequality falls dramatically, as a majority of workers work at the highest-productivity firm and quickly negotiate the highest-possible surplus share.

### Bargaining

As in Cahuc et al. (2006), in our model, there are only ever two firms at the bargaining table. Duopsony is a feature of the economic environment regardless of the number of firms in a given market. We remove this “hardwired” role for duopsony by assuming that whenever a worker meets a firm whose surplus rank is  $K$ , the worker also meets the next-best firm (i.e., the rank  $K - 1$  firm). Now three firms are always at the bargaining table. Holding vacancies fixed, this protocol does not alter allocations—the worker either stays put or goes to firm  $K$ —however,

surplus is redistributed from firms to workers, increasing wages. In general equilibrium, however, vacancies adjust (because the firm's share of surplus is now lower), shifting the allocation of workers to firms. Our main result is that markdowns narrow to approximately half of the level in the baseline economy. This experiment suggests that future work on the precise structure of strategic wage setting is valuable, in particular how variation in  $M$  endogenously leads to differences in the number of firms at the bargaining table.<sup>4</sup>

Our counterfactuals have implications for policy makers who may seek to address inefficiencies arising from labor market power. Our wage decomposition results point to a significant role for policies that alleviate markdowns due to amenities and strategic wage setting. Merger policy—primarily focused on the number of firms in a market,  $M$ —may have more moderate effects on wage markdowns. However, we anticipate that in a model with multiparty bargaining, this conclusion may be different. The extent to which bargaining is between two, three, or more parties is an empirical question that deserves more attention in light of our findings.

Further research can use our framework to study the distributional consequences of policies in granular labor markets. It is tractable enough to incorporate realistic policies (e.g., minimum wages, taxation, and antitrust), richer theories of the household and firm (e.g., costly human capital accumulation), and alternative contractual environments (e.g., non-competes, as in Gottfries and Jarosch 2023; Shi 2023).

We review the literature and then proceed as follows. Section II describes the Norwegian administrative data and offers motivating empirics. Section III describes the model and defines the equilibrium. Section IV provides details on model calibration and fit. Section V decomposes wages in the steady state to analyze the mechanisms through which concentration shapes wages. Section VI conducts the main counterfactual exercises and discusses potential policy implications of our findings.

### *E. Related Literature*

We contribute to a growing theoretical and quantitative literature by integrating the three existing monopsony paradigms into one framework: search frictions, nonwage amenities, and granularity. There are two main classes of monopsony models, each with two subgroups: (i) models in which frictional markets generate monopsony power with a continuum of firms (e.g., Burdett and Mortensen 1998; Manning 2003; Engbom and



Moser 2022; Hurst et al. 2022) and a finite number of firms (Burdett, Shi, and Wright 2001; Zhu 2012; Jarosch et al. 2024; Bagga 2023; Bloesch and Larsen 2023; Gottfries and Jarosch 2023) and (ii) models in which neoclassical markets in the presence of amenities generate monopsony power with a continuum of firms (e.g., Robinson 1933; Card et al. 2018; Kroft et al. 2020; Taber and Vejlín 2020; Lamadon, Mogstad, and Setzler 2022) and a finite number of firms (e.g., Bhaskar and To 1999; Bhaskar, Manning, and To 2002; Azkarate-Askasua and Zerecero 2024; Berger, Herkenhoff, and Mongey 2022a, 2022b). Unlike these existing frameworks, our model simultaneously features (1) search and matching frictions, (2) neoclassical nonwage amenities, (3) price discrimination within the firm, and (4) a finite number of firms. In addition, we model vacancy posting, and thus our general equilibrium model links employer concentration to both prices and quantities. This allows us to discuss welfare and conduct normative counterfactual exercises.

We contribute to a growing empirical literature that explores the relationship between worker and firm outcomes and market granularity. Recent work has documented cross-sectional relationships between standard measures of concentration (Herfindahl index values) and wages or employment (Benmelech, Bergman, and Kim 2022; Rinz 2022; Yeh, Macaluso, and Hershbein 2022) and vacancies (Azar et al. 2018; Azar, Berry, and Marinescu 2022; Azar, Marinescu, and Steinbaum 2022). To our knowledge, we are the first to (i) document reduced-form, cross-market relationships between Herfindahl values, job flows, and various measures of wage inequality, including within- and between-firm wage inequality, and (ii) combine occupational flow-data and clustering techniques to define markets.

## II. Empirical Analysis

This section presents new evidence on market structure, job mobility, and wage-setting behavior in Norwegian labor markets.

### A. *Data and Measurement*

Although the use of linked employer-employee data covering the universe of firms, establishments, and employees is now common among researchers, the Norwegian data have the key advantage that employers must record the type of work that each employee is hired to do and the workplace location of every employment contract from 2006 to 2018.

These bibliographic records allow us to define labor markets using geography and occupation rather than industry (as in Berger et al. 2022a).<sup>5</sup> We can then count the employers within each occupation-location market and track changes to individuals' wages when they move between employers.

**Data.** Data collection consists of two steps. In the first step, we use the information about the work contract that the employer submits to the employment agency (NAV). To comply with labor laws, the employer must enter a specific position from a list of more than 6,000 possible job titles and the workplace location, wage income, and work hours.<sup>6</sup> Job titles are then grouped into 354 four-digit occupations by Statistics Norway based on similarity of work.<sup>7</sup> We cluster these occupations using novel techniques to compute the occupational scope of labor markets. In the second step, we combine linked employer-employee data with socioeconomic variables from longitudinal population registers. These include demographic information (e.g., sex, age, residential municipality, education). We can therefore determine commuting distances between residence and workplace, which facilitates computation of the geographical scope of labor markets. We can therefore allow labor markets to cross administrative borders of municipalities and counties.<sup>8</sup>

**Institutional detail.** Norway has a population of 5 million, and Oslo, the capital, accounts for approximately one-fifth. The labor force aged 25–66 is some 2 million, and the labor share of income is approximately 70%.<sup>9</sup> In 2016, unemployment was approximately 4.5%. There were 176,019 firms and 234,941 establishments with workers on payroll.

Firms can hire employees on either fixed-term or permanent contracts and can dismiss workers if they underperform relative to their peers or if the firms are operating at a loss. Employment protection in Norway ranks near the median among Organization for Economic Cooperation and Development (OECD) countries and is comparable with that in France and Sweden.<sup>10</sup> Wages and typical working hours, in turn, tend to be set by collective bargaining at the industry level, after which wages are supplemented by local adjustments or wage drift, bargained over at the worker-firm and collective agreement level, which may vary by occupation within a firm (see, e.g., Bhuller et al. 2022). This two-tier framework gives rise to a relatively compressed wage structure. The Norwegian safety net covers lost income from unemployment. The primary insurance source is unemployment insurance (UI) benefits, which begin after a 3-day waiting period, replacing approximately two-thirds of workers' past earnings net of tax (see, e.g., Røed and Zhang 2003).<sup>11</sup>

### B. Defining Labor Markets

In our empirical analysis, we build markets from the ground up using individual data. To avoid issues of entry and exit from the labor force, we focus on residents aged 25–60. We define a local labor market as a group of four-digit occupations within a commuting zone region indexed by  $r$ , where the commuting zones are taken from Bhuller (2009). As shown later, the self-flow rate is the fraction of job-to-job transitions from one group of occupations back into the same group.<sup>12</sup> We then group occupations as follows:

1. First, we isolate single-occupation markets with high self-flow rates (e.g., those with rates of more than 50%, such as the market for dentists).
2. Among the remaining occupations, we  $K$ -means cluster the rows of the occupation-to-occupation flow transition matrix.
3. For each commuting zone region indexed by  $r$ , we compute the Herfindahl-Hirschman index of employment across clusters ( $\text{HHI}_r^K$ ).<sup>13</sup> We then determine the optimal number of  $K$  clusters by maximizing an objective function that is (i) increasing in the self-flow rate but (ii) decreasing in  $\text{HHI}_r^K$ , such that we penalize the classification of all occupations in one large cluster:

$$\underbrace{\frac{\text{Average self-flow rate of all } K \text{ clusters}}{\text{Standard deviation self-flow rate of all } K \text{ clusters}}}_{\text{(i) Reward fit}} \times \underbrace{[1 - \text{HHI}_r^K]}_{\text{(ii) Penalize overfitting}}.$$

Table 1 summarizes our market definitions. Our procedure yields a self-flow rate of 51%. We obtain approximately 103 clusters per commuting zone, and the average market has 404 workers. The unweighted number

**Table 1**  
Market Definition Summary Statistics

Moment	Value
Fraction of E-to-E flows within market 2006–16 (%)	51.40
Number of markets per region	102.8
Average firm employment per market	6.20
Average labor force per market	404.7
Average markets per firm	2.30
Total number of markets	4,783

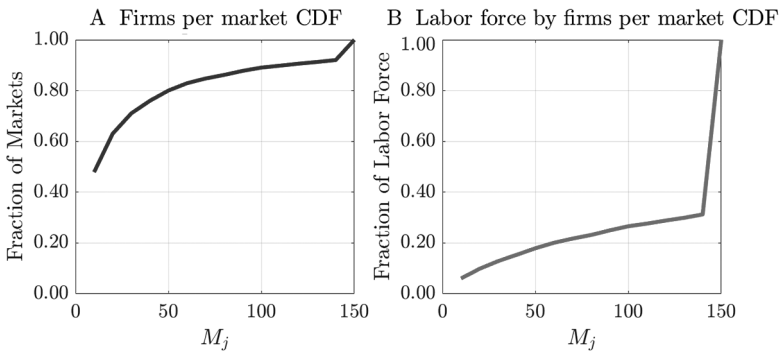
Note: E-to-E = employer-to-employer. Summary statistics are unweighted. All rows except top row are calculated from December 2016.

of employees per firm market is 6, and the average firm operates in 2.3 markets.

To ground our definition of markets, consider the fictitious example of a dental care firm in Oslo named ABC Dental. ABC Dental is an eight-person firm that hires workers in three occupations corresponding to its six dentists, one janitor, and one groundskeeper. The commuting zone region of ABC Dental is Oslo. Suppose dentists are in a single dentist cluster, and janitors and groundskeepers are both in a manual low-skill service cluster.<sup>14</sup> Markets are cluster–commuting zone pairs and indexed by  $j$  (e.g., “manual low-skill services-Oslo” is one market and “dentists-Oslo” is another). Firms are firm–market pairs (e.g., “manual low-skill services-ABC Dental-Oslo” is treated as a different firm than “dentists-ABC Dental-Oslo”). Occupations are the four-digit raw occupation (e.g., dentist, janitor, groundskeeper).

Figure 1 plots the distribution of firms and employment across markets, ordered by the number of firms in each market  $M_j$ . As we do in the calibration, we truncate the graph at  $M_j = 150$  firms in a market. Only 10% of markets have more than 150 firms (panel A), but these markets employ more than 70% of the Norwegian labor force (panel B).

Table 2 provides summary statistics on key labor market outcomes for 2006–16. The economy-wide unemployment rate averaged 4%. The monthly job-to-job transition rate was 0.7%, the job-finding rate was 8% per month, and the layoff rate was 0.4%. Relative to those in US data, worker flows in the Norwegian labor market are noticeably lower. We note that all flow rates in the main body of the text are computed within



**Fig. 1.** Cumulative distribution functions (CDFs) of market size and labor force. Panel A is the CDF of the number of firms per market. We winsorize the data at 10 and 150 firms per market. Panel B is the CDF of the labor force by number of firms per market. A color version of this figure is available online.

**Table 2**  
Summary Statistics: Worker Flows, Concentration, Wage Inequality

Moment	Value	Moment	Value
Unemployment rate (%)	4.01	Average firms per market	75.2
E-to-E rate (monthly %)	.65	HHI wage bill (wage bill weighted)	.09
U-to-E rate (monthly %)	8.08	HHI employment (employment weighted)	.08
E-to-U rate (monthly %)	.35	Standard deviation of log wages	.63
Fraction E-to-E moves down poach index	.15	Fraction of log wage variance within firms	.61

Note: E-to-E = employer-to-employer, E-to-U = employment-to-unemployment, U-to-E = unemployment-to-employment, and HHI = Herfindahl-Hirschman index. Flow rates are computed within markets (flows within markets divided by workers remaining within markets) to be consistent with the model definition of a market. Similar statistics are obtained when we include job flows outside of the market.

markets (flows within markets divided by workers remaining within markets) to be consistent with the model definition of a market. Including job flows outside of the market yields similar statistics and covariances (see app. D, <http://www.nber.org/data-appendix/c14848/appendix.pdf>). We then rank firms by the fraction of their hires coming from other firms—that is, the poach rank index (see Bagger and Lentz 2019)—and find that 15% of job-to-job transitions are down this ladder. We use this moment to discipline the role of nonwage amenities.

The average number of firms in a market is large: 75.2. However, markets are concentrated. The average wage bill Herfindahl index (weighted by the wage bill) is 0.09.<sup>15</sup> This is the same concentration as in a market with only 11 equally sized firms.<sup>16</sup> Similarly, the average employment Herfindahl index (employment weighted) is 0.08. The wage bill HHI is higher due to larger firms paying higher wages. The standard deviation of log wages is 63%, with the bulk of this (61%) accounted for by dispersion in wages within firms.

C. *Regression Framework*

Our goal is to study the relationship between employer concentration wage levels, the wage distribution, and job-to-job transitions. We provide a set of covariances between concentration and labor market outcomes that any theory of concentration and labor market dynamics should replicate. We do not attempt to attribute causality. We, along with the existing literature, lack credible instruments for measuring concentration. For example, take the change in local concentration due to a national firm exit used by Azkarate-Askasua and Zerecero (2024). Through

the lens of our theory in Section III, this shock alters the rungs of the job ladder. Thus, it is not a pure “concentration” shock, *ceteris paribus*.<sup>17</sup> The same is true of instrumenting changes in local exposure with changes in national concentration (e.g., Azar, Marinescu, and Steinbaum 2022) because the second stage still implies changes in the number of rungs on the local job ladder.

Instead, we consider two different dimensions of variation in the Herfindahl values by using across-region variation within occupation-years and across-time variation within occupation-regions. Importantly, however, the unit of observation remains four-digit occupations in a commuting zone, allowing us to use occupation fixed effects. Each approach differs in its interpretation. Comparisons across regions may reflect sorting across space on unobservables. Comparisons across time may reflect changing demand patterns. We do not take a stance on what drives these covariances. However, both approaches provide similar negative correlations between Herfindahl values, wages, job flows, and wage inequality.

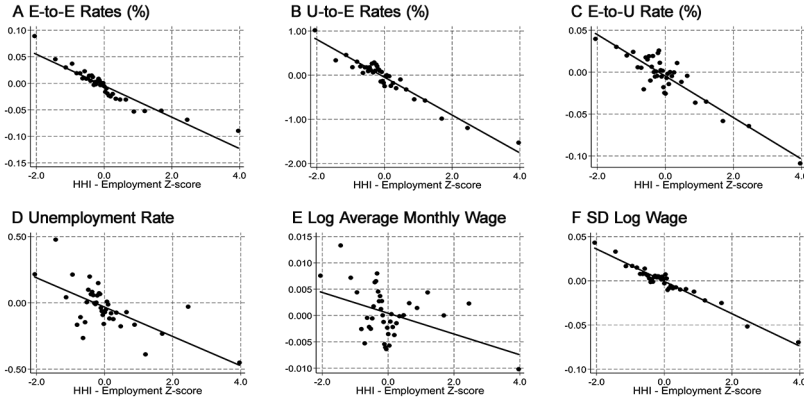
Let  $o$  denote the occupation,  $r$  the region (commuting zone),  $t$  the time (the data are monthly, and we denote the corresponding year as  $\tau(t)$ ), and  $m(o, r)$  the market to which the occupation-region was assigned by our algorithm.<sup>18</sup> Let  $\gamma_{FE}$  denote either (1) occupation-year fixed effects ( $\gamma_{o\tau(t)}$ ), thus isolating across-region variation, or (2) occupation-region fixed effects  $\gamma_{ort}$ , thus isolating across-time variation. Given our focus on market-level outcomes, we do not weight our regressions.<sup>19</sup> We estimate the following equation using ordinary least squares:

$$y_{ort} = \gamma_{FE} + \beta \text{HHI}_{m(o)r,t} + X_{ort} + \varepsilon_{ort}. \quad (1)$$

We include a vector of controls,  $X_{ort}$ , that vary at the occupation-region-time level. As we discuss in Section III, our model removes mechanical variation in the number of firms per worker. We therefore control for lagged quintiles of firms per worker measured at the market-time level. We also control for month of the year to hold seasonal fluctuations fixed, lagged labor force growth and age, gender, and education composition at the  $ort$  level.

#### D. Empirical Results

**Across regions.** Figure 2 provides a graphical representation of our regression evidence using across-region, within-occupation-year variation. The  $x$ -axis (Herfindahl index) and  $y$ -axis (labor market outcome) variables are residualized on occupation-year fixed effects and the controls  $X_{ort}$ . This leaves across-region variation (e.g., Oslo vs. Bergen, for dentists



**Fig. 2.** The figures show scatter plots of employer-to-employer (E-to-E) rates, unemployment-to-employment (U-to-E) rates, employment-to-unemployment (E-to-U) rates, unemployment rates, log average monthly wage and the standard deviation (SD) of the log wage against labor market concentration. Concentration and labor market outcomes residualized on occupation-year fixed effects (FEs), leaving across-region variation. For each market (where a market is defined, as in Sec. II, as a cluster of occupations within a commuting zone), we compute the employment Herfindahl-Hirschman index (HHI). For each four-digit occupation-commuting zone-year, we compute the average of the dependent variable within 40 centiles of the market HHI, unweighted. We then residualize all  $x$  and  $y$  variables on occupation-year FEs, age composition, gender composition, education composition, lagged firms-per-worker ventiles, lagged labor force growth, and month-of-year dummies. The average NOK/USD in 2021 was 9.

in 2008). We normalize the residualized Herfindahl value by its standard deviation and subtract its mean (i.e., convert it to a z-score) to ease interpretation. We also perform inference by clustering standard errors at the market level.

We find a statistically significant negative relationship between employment-to-employment transition rates and the market Herfindahl index (panel A). To interpret the relationships, we note that the unweighted employment Herfindahl index has a mean of 0.28 and a standard deviation of 0.27. The slope of the bin-scatter implies that a one-standard-deviation increase in the market employment Herfindahl index is associated with a 0.06-percentage-point reduction in the E-to-E rate ( $= -0.00214 \times 0.27$ ), which corresponds to approximately 10% of the sample average E-to-E rate (see table 2). Panels B and C show similar negative relationships for U-to-E rates and E-to-U rates. On net, these yield a negative relationship with the unemployment rate (panel D). Later, we find that this negative unemployment-HHI relationship is not robust to the choice of fixed effects.

There is also a negative relationship between wages and the Herfindahl index (panel *E*). This relationship is significant with occupation-year fixed effects alone but insignificant with occupation-year fixed effects and controls. Nonetheless, with occupation-year fixed effects and controls, a one-standard-deviation increase in the market employment Herfindahl index is associated with a 0.27-percentage-point reduction in the wage ( $= -0.0104 \times 0.27 \times 100$ ). Last, panel *F* illustrates a strong negative relationship between concentration and wage inequality (the standard deviation of log wages).

One concern is that these covariances reflect the sorting of better workers into less concentrated markets. We think that this is unlikely for two reasons. First, we control for education level and show that the main conclusions remain unaltered if we omit these controls in table 3. Second, high-skill workers are more likely to have high U-to-E rates but lower E-to-U rates. Nevertheless, we consider an alternative specification that isolates within-occupation-region across-time variation, thereby mitigating concerns regarding spatial sorting on unobservables.

**Across time.** Figure 3 repeats the exercise with both the Herfindahl index and labor market outcome residualized on occupation-region fixed effects and the controls  $X_{ort}$ . This leaves across-time variation (e.g., 2006 vs. 2007, for dentists in Oslo). Again we observe the negative correlation between concentration and labor market flows (panels *A*, *B*, and *C*), wages (panel *E*), and wage inequality (panel *F*). Here, however, the relationship with unemployment is flipped (panel *D*). However, with occupation-region fixed effects, the relationship between wages and concentration is robustly negative, regardless of controls. Panel *E* implies that a one-standard-deviation increase in the market employment Herfindahl index is associated with a 4.5-percentage-point reduction in the wage ( $= -0.166 \times 0.27 \times 100$ ).

**Regressions.** The significant, negative relationship between concentration and all three worker flows and wage inequality is robust across all specifications (table 3, panels *A*, *B*, *C*, and *F*). Table 3 provides the regression tables corresponding to figures 2 and 3. We estimate equation (1) with and without controls and for both sets of fixed effects (occupation-year, denoted O-Y, and occupation-region, denoted O-R). The dependent variable in column 1 is the monthly E-to-E transition rate (not expressed as a percentage). The coefficient can be interpreted as follows: a one-standard-deviation increase in the HHI is associated with 0.05% ( $= -0.00194 \times 0.27 \times 100$ ) reduction in the employment-to-employment transition rate. The relationship between concentration and unemployment rates is sometimes insignificant, negative, or positive depending on the fixed



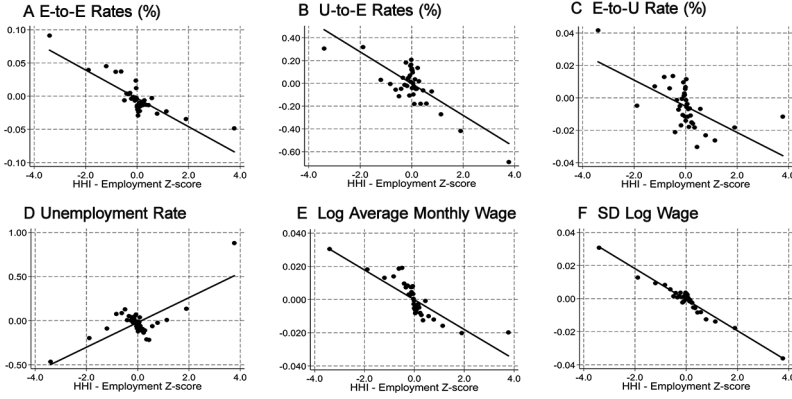
**Table 3**  
Regression Analysis: Concentration and Labor Market Outcomes

	A. E-to-E rate				B. U-to-E rate			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
HHI	-.00194*** (7.58e-05)	-.00214*** (9.11e-05)	-.00361*** (.000232)	-.00475*** (.000312)	-.0356*** (.00208)	-.0330*** (.00234)	-.0231*** (.00443)	-.0284*** (.00520)
FE		O-Y	O-R	O-R	O-Y	O-Y	O-R	O-R
Controls		Y	N	Y	N	Y	N	Y
Obs.	1,035,450	892,774	1,035,382	892,731	628,097	553,705	627,933	553,544
R <sup>2</sup>	.066	.181	.047	.164	.065	.075	.094	.101
	C. E-to-U rate				D. Unemployment rate			
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
HHI	-.00191*** (.000154)	-.00177*** (.000177)	-.000825** (.000325)	-.00172*** (.000415)	.00251 (.00302)	-.00781*** (.00298)	.0631*** (.00650)	.0420*** (.00668)
FE		O-Y	O-R	O-R	O-Y	O-Y	O-R	O-R
Controls		Y	N	Y	N	Y	N	Y
Obs.	1,035,450	892,774	1,035,382	892,731	1,043,861	892,927	1,043,790	892,885
R <sup>2</sup>	.069	.082	.095	.110	.408	.425	.509	.528
	E. Log wage				F. Standard deviation of log wage			
	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
HHI	-.0328*** (.00956)	-.0104 (.00970)	-.167*** (.0197)	-.166*** (.0211)	-.126*** (.00892)	-.134*** (.00960)	-.226*** (.0166)	-.216*** (.0198)
FE		O-Y	O-R	O-R	O-Y	O-Y	O-R	O-R
Controls		Y	N	Y	N	Y	N	Y
Obs.	1,043,861	892,927	1,043,790	892,885	1,042,160	892,176	1,042,089	892,134
R <sup>2</sup>	.793	.817	.784	.825	.382	.402	.518	.537

Note: E-to-E = employer-to-employer, U-to-E = unemployment-to-employment, E-to-U = employment-to-unemployment, HHI = Herfindahl-Hirschman index. Standard errors are clustered at the market level. For the fixed effects, (i) O-Y refers to occupation-year fixed effects, and (ii) O-R refers to occupation-region fixed effects. For the controls, Y indicates controls for quintiles of firms per worker measured at the market-month level, month of the year, lagged labor force growth, age, gender, and education composition at the ort level. In this table (unlike the figures), E-to-E, U-to-E, E-to-U, and unemployment rates take values between zero and one and are thus not expressed in percentage points.

\*\* $p < .01$ .

\*\*\* $p < .001$ .



**Fig. 3.** The figures show scatter plots of employer-to-employer (E-to-E) rates, unemployment-to-employment (U-to-E) rates, employment-to-unemployment (E-to-U) rates, unemployment rates, log average monthly wage and the standard deviation (SD) of the log wage against labor market concentration. Concentration and labor market outcomes residualized on occupation-region fixed effects (FEs), leaving across-time variation. For each market (where a market is defined, as in Sec. II, as a cluster of occupations within a commuting zone), we compute the employment Herfindahl-Hirschman index (HHI). For each four-digit occupation-commuting zone-year, we compute the average of the dependent variable within 40 centiles of the market HHI, unweighted. We then residualize all  $x$  and  $y$  variables on occupation-region FEs, age composition, gender composition, education composition, lagged firms-per-worker ventiles, lagged labor force growth, and month-of-year dummies. The average NOK/USD in 2021 was 9.

effects and inclusion of controls (panel D). Log wages are significantly negatively related to concentration in all specifications except for that with occupation-year fixed effects with controls (panel E).

**Robustness to alternative labor market definitions.** We explore the robustness of our empirical results to defining labor markets in alternative ways in appendix C, <http://www.nber.org/data-appendix/c14848/appendix.pdf>. Rather than clustering occupations by K-means using the occupational flow matrix within each commuting zone, we follow a recent literature (Lindenlaub and Postel-Vinay 2021) and extract the relevant clusters of occupations within a commuting zone using modularity maximization following the work of Schmutte (2014). Reassuringly, the patterns that we find are similar to those obtained when we use our baseline clustering algorithm. Last, an alternative approach to defining markets is to simply use raw three-digit occupations and commuting zones. We used this market definition in an earlier draft of this paper and found quantitatively similar results.

**Summary.** In the rest of the paper, we ask whether a benchmark theory of frictional labor markets with on-the-job search and bargaining (Cahuc

et al. 2006) can quantitatively replicate these empirical relationships when extended to accommodate (i) concentrated markets, (ii) a vacancy-posting equilibrium, and (iii) worker-firm-specific tastes. We then use the model to decompose the role of different economic forces in determining wage markdowns, employment, and inequality.

### III. Model

In our model, time is discrete and runs forever. We assume that there are  $J$  markets indexed by  $j \in \{1, \dots, J\}$ . Within each market, there are  $\bar{N}_j$  workers and  $M_j$  firms. Firms are indexed by  $i \in \{1, \dots, M_j\}$ . The measure of firms is exogenous, and workers are assumed to be immobile across markets. For the remainder of the section, we suppress the market index  $j$ .

**Workers.** Workers have linear utility and maximize the net present value of discounted utility. Both workers and firms discount the future at a rate  $\beta$ . Workers are either employed or unemployed. Among the employed, a measure  $n_i$  is employed at firm  $i$ . Total employment in a labor market is therefore  $n = \sum_{i=1}^M n_i$ , and thus, the measure of unemployed individuals in a labor market is  $u = \bar{N} - n$ .

A worker's per period utility is the summation of income and a worker-firm-specific taste shock,  $\varepsilon$ . This taste shock is a stand-in for commuting times, how well a worker gets along with their boss and colleagues, and any other worker-firm-specific nonwage characteristics. We assume that the taste shock is independently drawn from the distribution  $F(\varepsilon)$  when a worker contacts a firm, and we assume that  $\varepsilon$  is constant throughout a worker-firm match. Both the firm and worker observe and contract are on the amenity draw. Thus, we allow firms to first-degree price discriminate on amenities.<sup>20</sup>

Search is random. Each period, a random fraction  $\phi$  of unemployed individuals searches for job openings. Unemployed individuals receive utility from home production,  $b$ . Employed workers search on the job with probability  $\xi\phi$  and do not apply to jobs at their existing firm. We refer to this search protocol as partially directed search.

**Firms.** Firm  $i$ 's productivity is fixed and is denoted  $z_i$ . Posting  $v_i$  vacancies costs  $c(v_i, M, \bar{N})$ , where  $c$  is convex in vacancies. Empirically, unemployment rates vary relatively little across markets that vary widely in terms of the number of firms per worker (see app. table 9, <http://www.nber.org/data-appendix/c14848/appendix.pdf>). For the model to scale and achieve this stylized fact, we remove variation in firms per worker by scaling the vacancy costs accordingly. Anticipating the

calibration, we assume that vacancy costs are given by  $c(v_i, M, \bar{N}) = (M/\bar{N})^\gamma 1/(1 + \gamma) v_i^{1+\gamma}$ . The scaling factor achieves neutrality of the unemployment rate with respect to firms per worker ( $M/\bar{N}$ ). For convenience, we suppress the market employment and firm arguments of  $c(\cdot)$  and write  $c(v_i)$  for vacancy costs.

We assume that firms post vacancies nonstrategically. In an earlier version of this paper, we solved the strategic vacancy-posting decision for markets with few firms and found that strategic vacancy-posting motives yield no discernible effects on aggregates while simultaneously making the model less tractable. Strategic vacancy posting, when combined with amenities, also gives rise to complex hiring rules.<sup>21</sup> We assume away this behavior and leave it to future researchers to find tractable solutions to this problem.

**Meeting rates.** When workers apply for jobs, only a fraction of those applications actually result in a meeting with a prospective employer via a vacancy. After meeting, the worker and prospective firm observe the nonwage amenity draw  $\varepsilon$  upon which they base their matching decision. A match occurs if the worker moves to the prospective employer. Matches end through on-the-job search as well as at an exogenous rate  $\delta$ .

From the worker's perspective, let  $\lambda_{ik}$  denote the rate at which a firm  $i$  worker meets a firm  $k$  vacancy. From the firm's perspective, let  $\lambda_{ik}^f$  denote the rate at which firm  $i$ 's workers meet a firm  $k$  vacancy. Our convention with subscripts is origin first ( $i$ ), then destination ( $k$ ).

To describe the meeting process, we must keep track of the origin and destination of job applicants. Let  $x_{ik}$  ( $x_{uk}$ ) denote the measure of firm  $i$  workers (unemployed workers) who randomly apply for jobs at firm  $k$ . Then,  $x_i = \sum_{k \neq i} x_{ki} + x_{ui}$  is the total measure of workers who randomly apply for jobs at  $i$ .

Throughout, we assume that meetings at each firm are governed by a constant-returns-to-scale meeting function  $\bar{m}(v_i, x_i) \leq \min\{v_i, x_i\}$ . In the absence of on-the-job search, firm-specific constant-returns-to-scale matching functions and the usual pooled random search model of Diamond-Mortensen-Pissarides are equivalent. Let  $A$  denote match efficiency. In practice, we use a Cobb-Douglas matching function:

$$\bar{m}(v, x) = \min\{A v^{\alpha-1} x^\alpha, 1\}.$$

We let  $f(\theta_i)$  denote the job-finding rate at firm  $i$ , where the tightness at firm  $i$  is defined to be  $\theta_i = v_i/x_i$ . Constant returns to scale imply  $\bar{m}(v_i, x_i) = x_i f(\theta_i)$ .

From each firm  $i$ , there is a measure  $\xi\phi n_i$  of workers who engage in partially directed on-the-job search. Because employed workers randomly apply only for vacancies posted by firms other than the one where they are currently employed, the measure of workers at firm  $i$  who apply to firm  $k$  is given by

$$x_{ik} = \frac{v_k}{\sum_{j \neq i} v_j} \xi\phi n_i.$$

A fraction  $\phi$  of unemployed individuals apply to all  $M$  firms randomly. Therefore, the measure of workers who are unemployed and apply to firm  $k$  is given by

$$x_{uk} = \frac{v_k}{\sum_j v_j} \phi u.$$

We can now derive the probability that a worker at firm  $i$  meets firm  $k$  into three terms. The first term is the probability that a worker searches,  $\xi\phi$ . The second term is the conditional probability that the worker applies to a vacancy at firm  $k$ ,  $v_k/\sum_{j \neq i} v_j$ . The third term is the conditional probability that a meeting occurs,  $\bar{m}(v_k, x_k)/x_k$ . These yield the worker meeting rate:

$$\lambda_{ik} = \xi\phi \times \left( \frac{v_k}{\sum_{j \neq i} v_j} \right) \times \left( \frac{\bar{m}(v_k, x_k)}{x_k} \right) = \xi\phi \times \left( \frac{v_k}{\sum_{j \neq i} v_j} \right) \times f(\theta_k). \quad (2)$$

The contact rate of unemployed workers is defined similarly, except that unemployed workers may meet any firm and, hence,  $\lambda_{uk} = \phi(v_k/\sum_j v_j)f(\theta_k)$ .

We similarly divide the probability of a meeting a firm  $k$ 's vacancy with a worker at firm  $i$  into two terms: (i) the probability of a meeting between a worker and a vacancy,  $\bar{m}(v_k, x_k)/v_k$ , and (ii) the probability that the worker originated from firm  $i$ ,  $x_{ik}/x_k$ . These yield the firm meeting rate:

$$\lambda_{ik}^f = \frac{\bar{m}(v_k, x_k)}{v_k} \times \left( \frac{x_{ik}}{x_k} \right). \quad (3)$$

The probability that firm  $k$  meets an unemployed worker is given by  $\lambda_{uk}^f = \bar{m}(v_k, x_k)/v_k(x_{uk}/x_k)$ .

Flow balance holds. Using the definitions of  $x_{ik}$  and  $x_k$ , one can check that  $n_i\lambda_{ik} = v_k\lambda_{ik}^f$ : firm  $i$  workers' rate of meeting firm  $k$  vacancies equals firm  $k$ 's rate of meeting vacancies with firm  $i$  workers.

**Bargaining over promised values.** We follow Cahuc et al. (2006). When a worker meets a new firm, the incumbent and poaching firm propose sequential offers. We assume that firms offer workers promised values that they are committed to. There are three possible cases: the worker meets a firm that can deliver a maximum promised value that is (i) less than the value promised to the worker by their current employer, (ii) greater than the value promised to the worker by their current employer but less than the maximum promised value of the worker's current employer, or (iii) greater than the maximum promised value that the worker's current employer can offer.

In case (i), we assume no change to the worker's wage, and the worker remains with the incumbent firm. In case (ii), we assume that the incumbent firm offers a new promised value that delivers the full joint value of the match with the poaching firm. In case (iii), the worker moves to the poaching firm, and Nash bargaining determines the split of the remaining surplus between the poaching firm and worker, where the full joint value of the match with the incumbent firm constitutes the worker's outside option. Let  $\theta \in [0, 1]$  denote the worker's Nash bargaining parameter. Let  $\sigma \in [\theta, 1]$  denote the worker's share of the match surplus. Because of cases (ii) and (iii),  $\sigma$  may increase above the worker's Nash bargaining weight  $\theta$ .

**Wage determination.** The bargaining protocol pins down the promised values, but the wage that delivers the promised values is indeterminate. We assume that firms deliver the promised values to workers as a constant wage unless the worker receives an credible outside option. This is a common assumption and delivers wage dynamics consistent with the empirical evidence in Cahuc et al. (2006), Jarosch (2023), and Lise and Robin (2017).

Let the continuation value of a worker at firm  $i$  with bargained surplus share  $\sigma$  and taste shock  $\varepsilon$  be given by  $W_i(\sigma, \varepsilon)$ . Likewise, let firm  $i$ 's continuation value of a match with bargained surplus share  $\sigma$  and taste shock  $\varepsilon$  be given by  $J_i(\sigma, \varepsilon)$ . The continuation value of an unemployed individual is given by  $U$ . We will frequently work with both the joint value of a match,  $P_i(\varepsilon) = W_i(\sigma, \varepsilon) + J_i(\sigma, \varepsilon)$ , and the match surplus,  $S_i(\varepsilon) = W_i(\sigma, \varepsilon) - U + J_i(\sigma, \varepsilon) \equiv P_i(\varepsilon) - U$ . Because firms commit to the promised values and workers and firms have linear utility, it can be shown that the match surplus and joint value are independent of the division of surplus,  $\sigma$ .

We assume that once workers and firms separate, the job position is destroyed. To hire again, then, a firm needs to post new vacancies. Hence,

implicit in these definitions is that the value of an unfilled vacancy is zero, similar to the setup in Lise and Robin (2017).

Before we exposit the continuation values of the worker, we note that appendix E (<http://www.nber.org/data-appendix/c14848/appendix.pdf>) provides a full derivation of the main equations in the text, including the joint value of a match, the surplus of a match, and the wage equation. In addition, appendix E shows how one may solve for surplus using a simple matrix inversion.

**Unemployed workers.** Unemployed workers enjoy home production  $b$  and meet with firm  $k$  with probability  $\lambda_{uk}$ . When they meet with firm  $k$ , they draw a taste  $\varepsilon' \sim F$  for working at firm  $k$ . They receive a share  $\theta$  of surplus. The continuation value of an unemployed worker is therefore

$$U = b + \beta \left[ U + \theta \int \sum_k \lambda_{uk} \max\{S_k(\varepsilon'), 0\} dF(\varepsilon') \right].$$

**Employed workers.** The worker value is the value of unemployment plus some share  $\sigma$  of the match surplus (following Lise and Postel-Vinay 2020). In other words,  $\sigma$  is defined to be the number that satisfies the following equation:

$$W_i(\sigma, \varepsilon) = U + \sigma[P_i(\varepsilon) - U].$$

The Cahuc et al. (2006) bargaining protocol implies that when a worker at firm  $i$  with amenity  $\varepsilon$  meets a vacancy at firm  $k$  and draws amenity  $\varepsilon'$ , there are three possible outcomes:

- i) If  $P_i(\varepsilon) > W_i(\sigma, \varepsilon) > P_k(\varepsilon')$ , the worker stays at firm  $i$  with promised value  $W_i(\sigma, \varepsilon)$ .
- ii) If  $P_i(\varepsilon) > P_k(\varepsilon') > W_i(\sigma, \varepsilon)$ , the worker stays at firm  $i$  but is now delivered a promised value  $W_i(\sigma', \varepsilon) = P_k(\varepsilon')$ , where  $\sigma' = S_k(\varepsilon')/S_i(\varepsilon)$ .
- iii) If  $P_k(\varepsilon') > P_i(\varepsilon)$ , the worker moves to firm  $k$  and Nash bargains over the gains from trade  $[P_k(\varepsilon') - P_i(\varepsilon)]$ , with the full joint value at firm  $i$ ,  $P_i(\varepsilon)$ , as their outside option.

Under this protocol, the worker policy function is to move to the firm with the greatest surplus. Note that workers may move down the productivity ladder if the amenity draw increases surplus above that associated with the incumbent firm.

In case (iii), the Nash bargaining solution delivers a worker continuation value that maximizes:

$$\max_{\hat{W}} (P_k(\varepsilon') - \hat{W})^{1-\theta} (\hat{W} - P_i(\varepsilon))^\theta.$$

The resulting promised value is equal to the entire joint value between the worker and firm  $i$  plus a fraction  $\theta$  of the gains from trade:

$$\hat{W} = P_i(\varepsilon) + \theta[P_k(\varepsilon') - P_i(\varepsilon)],$$

which is convenient to express as a fraction of the match surplus:

$$W_k(\sigma', \varepsilon') = \hat{W} = U + \sigma'[P_k(\varepsilon') - U], \text{ where } \sigma' = \theta + (1 - \theta) \frac{S_i(\varepsilon)}{S_k(\varepsilon')}.$$

Given the above, it can be verified that the worker's share of surplus evolves according to:

$$\sigma' = \begin{cases} \left( \frac{\theta S_k(\varepsilon') + (1 - \theta) S_i(\varepsilon)}{S_k(\varepsilon')} \right) & \text{if } S_k(\varepsilon') > S_i(\varepsilon) \\ \max \left\{ \sigma, \frac{S_k(\varepsilon')}{S_i(\varepsilon)} \right\} & \text{if } S_k(\varepsilon') \leq S_i(\varepsilon) \end{cases}.$$

As discussed earlier, we assume that the promised values are delivered as a constant wage  $w_i(\sigma, \varepsilon)$  unless the worker has a meeting with an employer that triggers renegotiation. Employed workers at firm  $i$  meet with firm  $k$  with probability  $\lambda_{ik}$ . Under the bargaining protocol of Cahuc et al. (2006), the continuation value of the worker can be written as (see app. E):

$$\begin{aligned} W_i(\sigma, \varepsilon) = & w_i(\sigma, \varepsilon) + \varepsilon + \beta \left[ W_i(\sigma, \varepsilon) - \delta \sigma S_i(\varepsilon) \right. \\ & + \int \sum_{k \neq i} \lambda_{ik} \max\{0, \min\{\theta[S_k(\varepsilon') - S_i(\varepsilon)], S_k(\varepsilon') - S_i(\varepsilon)\} \\ & \left. + (1 - \sigma) S_i(\varepsilon)\} dF(\varepsilon') \right]. \end{aligned} \quad (4)$$

**Joint value.** Rather than exposit the value of a firm directly, we focus on the joint value of a match between worker and firm  $i$  with taste shock  $\varepsilon$ . They jointly produce  $z_i$  unless (i) the worker receives an outside offer at a firm that generates greater surplus or (ii) the match exogenously dissolves. Thus, the joint value takes into account the worker's future value of a new match or unemployment:



$$P_i(\varepsilon) = z_i + \varepsilon + \beta \left[ P_i(\varepsilon) + \theta \int \sum_{k \neq i} \lambda_{ik} \max\{S_k(\varepsilon') - S_i(\varepsilon), 0\} dF(\varepsilon') - \delta S_i(\varepsilon) \right].$$

**Surplus.** The surplus of a match relative to unemployment can be expressed similarly. The costs include the flow value of unemployment,  $b$ , and the option value of unemployment forfeited by being employed at  $i$ :  $\theta \int \sum_k \lambda_{ik} \max\{S_k(\varepsilon''), 0\} dF(\varepsilon'')$ . The benefits are production, amenities, and potential gains from on-the-job search,  $\theta \int \sum_{k \neq i} \lambda_{ik} \max\{S_k(\varepsilon') - S_i(\varepsilon), 0\} dF(\varepsilon')$ :

$$S_i(\varepsilon) = (z_i + \varepsilon) - b + \beta \left[ (1 - \delta) S_i(\varepsilon) + \theta \int \sum_{k \neq i} \lambda_{ik} \max\{S_k(\varepsilon') - S_i(\varepsilon), 0\} dF(\varepsilon') - \theta \int \sum_k \lambda_{ik} \max\{S_k(\varepsilon''), 0\} dF(\varepsilon'') \right]. \quad (5)$$

**Wage equation.** Combining the worker's value (eq. [4]) and surplus (eq. [5]), we can compute the wage based on surplus values alone. The wage function  $w_i(\sigma, \varepsilon)$  delivers a surplus share  $\sigma$  at firm  $i$ :

$$w_i(\sigma, \varepsilon) = \sigma z_i - (1 - \sigma)\varepsilon + (1 - \sigma) \left[ b + \beta \theta \int \sum_k \lambda_{ik} \max\{S_k(\varepsilon'_u), 0\} dF(\varepsilon'_u) \right] - \beta \int \sum_{k \neq i} \lambda_{ik} \max\{0, \min\{(1 - \sigma)\theta(S_k(\varepsilon') - S_i(\varepsilon)), (S_k(\varepsilon') - S_i(\varepsilon))\} + (1 - \sigma)S_i(\varepsilon)\} dF(\varepsilon'). \quad (6)$$

The wage equation includes four terms: (i) workers obtain  $\sigma$  of production, (ii) the firm can offer a lower wage to workers with higher taste shocks to deliver any given promised value, (iii) workers obtain  $(1 - \sigma)$  of their outside option, and (iv) there is backloading because firms that offer greater future pay prospects can initially pay less.

**Optimal vacancy posting.** The firm vacancy-posting problem requires knowledge of the distribution of workers across amenity values and employers. The probability that a worker at firm  $k$  has amenity draw  $\varepsilon$  is given by the endogenous ratio  $n_k(\varepsilon)/n_k$ . As discussed earlier, we assume that the vacancy-posting decision is nonstrategic. Thus, the firm chooses  $v_i$  to maximize the following objective, taking all contact rates, worker stocks, and surplus values as given:<sup>22</sup>

$$\begin{aligned} & \max_{v_i} -c(v_i) + \underbrace{(1 - \theta)v_i \int \lambda_{ii}^f \max\{S_i(\varepsilon'), 0\} dF(\varepsilon')}_{\text{Hire from unemployment}} \\ & + \underbrace{(1 - \theta)v_i \iint \sum_{k \neq i} \lambda_{ki}^f \left( \frac{n_k(\varepsilon)}{n_k} \right) \max\{S_i(\varepsilon') - S_k(\varepsilon), 0\} d\varepsilon dF(\varepsilon')}_{\text{Hire from employment}}. \end{aligned} \quad (7)$$

This yields the following optimality condition for firms:

$$v_i = c'^{-1} \left( (1 - \theta) \int \lambda_{ui}^f \max\{S_i(\varepsilon'), 0\} dF(\varepsilon') \right. \\ \left. + (1 - \theta) \iint \sum_{k \neq i} \lambda_{ki}^f \left( \frac{n_k(\varepsilon)}{n_k} \right) \max\{S_i(\varepsilon') - S_k(\varepsilon), 0\} d\varepsilon dF(\varepsilon') \right).$$

**Laws of motion for employment.** The laws of motion for employment and unemployment are given by the following equations, where primes denote next-period values.<sup>23</sup>

$$n'_i(\varepsilon) = \left( 1 - \delta - \sum_{k \neq i} \lambda_{ik} \int 1_{[S_i(\varepsilon') \geq S_k(\varepsilon)]} dF(\varepsilon') \right) n_i(\varepsilon) + \lambda_{ui} 1_{[S_i(\varepsilon) > U]} f(\varepsilon) u \\ + \sum_{k \neq i} \lambda_{ki} \int 1_{[S_i(\varepsilon) \geq S_k(\varepsilon')]} f(\varepsilon) n_k(\varepsilon') d\varepsilon' \quad (8) \\ u' = \left( 1 - \sum_i \lambda_{ui} \int 1_{[S_i(\varepsilon') \geq U]} dF(\varepsilon') \right) u + \delta(\bar{N} - u), n_i = \int n_i(\varepsilon) d\varepsilon, u = \bar{N} - \sum_i n_i.$$

**Equilibrium.** Because markets do not interact, it suffices to define the equilibrium for a single market. We continue to suppress the market  $j$  index, and we note that, in the quantitative model, all  $J$  markets satisfy the following equilibrium definition.

In a given market with a mass of firms  $M$  and labor force  $\bar{N}$ , a *stationary equilibrium* is a stock of vacancies and employment  $\{v_i, n_i(\varepsilon)\}_{i=1}^M$ , an unemployed value  $U$ , surplus values  $\{S_i(\varepsilon)\}_{i=1}^M$  (which implicitly define the worker's mobility policy function), meeting rates  $\{\lambda_{ui}^f, \lambda_{ui}\}_{i=1}^M$  and  $\{\{\lambda_{ki}^f, \lambda_{ik}\}_{k \neq i}\}_{i=1}^M$  such that

1. Worker optimality: Given surpluses and contact rates, worker mobility decisions are optimal—that is, mobility decisions are consistent with surpluses (eq. [5]) and deliver the worker value (eq. [4]).
2. Firm optimality: Given surpluses, contact rates, and worker stocks,  $v_i$  solves equation (7).
3. Market clearing: Worker mobility decisions (implicitly defined by  $\{S_i(\varepsilon)\}_{i=1}^M$  and  $U$ ) and optimal firm vacancy postings deliver a stationary distribution of workers given by equation (8) consistent with  $\{n_i(\varepsilon)\}_{i=1}^M$ .

#### IV. Calibration and Model Fit

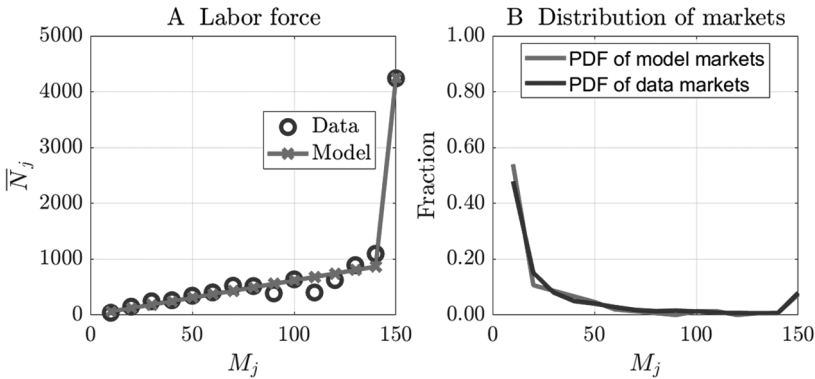
This section describes our calibration approach. We then explore the model fit relative to the data moments in Section II.

### A. Calibration

**Markets.** We adopt the clustering method described in Section II to define markets. With these market definitions in hand, we can compute the labor force per market,  $\bar{N}_j$ , and the distribution of the number of firms per market,  $G(M_j)$ . Rather than parameterize the joint distribution of  $M_j$  and  $\bar{N}_j$ , we note that a linear relationship between labor force and firms per market fits the data quite well (see fig. 4 and the discussion in the following paragraph). Thus, we impose  $\bar{N}_j = a\bar{M}_j$  so that once we draw the market size  $M_j$ , we know  $\bar{N}_j$ . We truncate the data at  $M_j = 150$ , above which we find very little difference with respect to equilibrium as we raise  $M_j$ . At the final truncated market size, for the labor force to add up to that of Norway, we deviate from the linear relationship and simply compute the necessary  $\bar{N}_j$  to match the size of the Norwegian labor force (in particular, the characteristic that 70% of workers are in  $M_j = 150$  markets).

Figure 4A illustrates the relationship between  $\bar{N}_j$  and  $M_j$ , as well as our fitted values. We assume there are  $J = 200$  markets (despite there being 3,700 in our data), and for each market, we draw the number of firms  $M_j \sim G(M_j)$  where  $G(\cdot)$  is the empirical cumulative distribution function (CDF) of firms per market. Figure 4B plots the distribution  $G(M_j)$  in the data and in our  $J = 200$  markets, illustrating the limited Monte Carlo simulation error in our quantitative experiments.

**Preferences and technology.** A period is 1 month. On an annual basis, the discount rate is 4% ( $\beta = 0.96^{1/12}$ ). We assume that the matching function elasticity is given by  $\alpha = 0.50$  and that unemployed workers



**Fig. 4.** Calibration of market size and labor force. Panel A plots the relationship between labor force size  $\bar{N}_j$  and market size  $M_j$ . Panel B plots the probability density function of market sizes in the model versus the data. A color version of this figure is available online.

search every period,  $\phi = 1$ . Employed workers' intensity of on-the-job search  $\xi = 0.32$  delivers the job-to-job transition rate shown in table 2. The match efficiency parameter  $A = 0.18$  then delivers the unemployment rate. The home production parameter  $b = 0.85$  is estimated to yield a 66% replacement rate as in Kostøl (2017).

We calibrate the amenity distribution to match the fraction of employment-to-employment moves that occur down the poaching index ladder. We assume that amenities are distributed uniformly  $\varepsilon \sim U[0, \bar{\varepsilon}]$  and calibrate  $\bar{\varepsilon} = 0.76$ . As discussed in Section II, the poaching index is simply a firm's share of hires who are poached from competitors rather than hired from unemployment.<sup>24</sup> We construct the same index in our model to map this moment to the data. If  $\bar{\varepsilon} = 0$ , so that there were no amenities, then all job-to-job transitions would be to firms with a higher poaching index (we demonstrate this and provide deeper discussion of identification in app. F, <http://www.nber.org/data-appendix/c14848/appendix.pdf>). Through the lens of our model, the fact that only 85% of moves are up this ladder provides evidence for idiosyncratic tastes and disciplines  $\bar{\varepsilon}$ . The bargaining power  $\theta = 0.18$  is calibrated to match the average wage growth rate in Norway.<sup>25</sup>

The remaining parameters control dispersion in productivity  $\sigma_z$  and vacancy cost convexity  $\gamma$ . We assume that firm productivity is log normal:  $\log z \sim N(-\frac{1}{2}\sigma_z^2, \sigma_z)$ . Conditional on other parameters,  $\sigma_z = 0.14$  is determined by matching the standard deviation of log wages. Given the amount of productivity dispersion and number of firms in a market,  $\gamma$  determines the share of employment at large firms. A higher  $\gamma$  compresses vacancy posting at the most productive firms, impeding their growth. The vacancy cost convexity parameter  $\gamma = 1.16$  is therefore pinned down by the employment Herfindahl index of 0.09.

To further discipline the parameters that govern concentration ( $\gamma, \sigma_z, b, \theta$ ) and its relationship with wages, we also target the regression coefficients of wages and the standard deviation of wages on the Herfindahl values in table 3 (cols. 18, 20, 22, and 24). Thus, our estimation is overidentified. We take a simple average of the regression coefficients based on occupation-year and occupation-region fixed effects. The results of the estimation are shown in table 4 along with the corresponding data moment that identifies each parameter.

### B. Model Fit

Table 5 compares the model's fit with the remaining reduced-form estimates from Section II. The first two rows are the wage moments that we explicitly target. These coefficients are negative, but this is not by

**Table 4**  
Calibration

Parameter		Value	Moment	Model	Data
Match efficiency	$A$	.18	Unemployment rate	.03	.04
OJS intensity	$\xi_e$	.32	Aggregate E-to-E rate	.01	.01
Home production	$b$	.85	Replacement rate	.86	.66
Vacancy cost elasticity	$\gamma$	1.16	Employment HHI	.09	.08
Bargaining power	$\theta$	.18	Average log wage growth	.01	.01
Upper-bound amenity	$\bar{\varepsilon}$	.76	Fraction of $EE$ moves down poach index ladder	.21	.15
Standard deviation of productivity	$\sigma_z$	14	Standard deviation of log wages	.69	.63
			Regression $\beta$ : Log market wage on $HHI$	-.09	-.09
			Regression $\beta$ : Standard deviation of log wages on $HHI$	-.18	-.18

Note: E-to-E = employer-to-employer, HHI = Herfindahl-Hirschman index, OJS = on-the-job search (OJS).

construction. The model can generate a counterfactually positive relationship between concentration and wages. How? With a sufficiently high productivity dispersion, the most concentrated markets are the markets in which one firm has drawn an outlier draw of productivity,  $z_i$ . Workers bargain over a share of surplus that is proportional to  $z_i$ , and thus concentrated markets can offer higher wages.

In terms of nontargeted moments, the model naturally generates the negative relationship between employment-to-employment transition rates and concentration. In the extreme, a market with a solo monopsonist  $N = 1$  has zero employer-to-employer job transitions. The model also generates the negative observed relationship between job-finding rates and concentration. Note that the model can generate a counterfactually positive relationship between concentration and the job-finding rate. If concentrated markets are also the most productive (i.e., a firm in such a market drew an extremely high  $z_i$ ), then surplus and vacancy postings reflect the high surplus value, and hence, high job-finding rates and high concentration occur simultaneously. That the model correctly generates the right negative relationship between the UE rates and Herfindahl values is thus a positive contribution of the model.

In summary, we have a quantitative model that matches the cross-sectional empirical relationship between concentration and (i) wages, (ii) worker flows, and (iii) wage inequality.

## V. Wage Decomposition and Mechanisms

Before we discuss our model counterfactuals, we first provide a decomposition of wages. Rewriting equation (6), we can express the wage as

**Table 5**  
Model Fit Relative to Regression Estimates in Section II

Dependent Variable ( $Y_{ort}$ )	Targeted in Estimation	Data		
		Model	Occ-Year FE	Occ-Region FE
		(1)	(2)	(3)
Log wage	Yes—Target average of (2) and (3)	-.0915	-.0104	-.166***
Standard deviation of log wage	Yes—Target average of (2) and (3)	-.1799	-.125***	-.230***
E.-to-E rate	No	-.0027	-.00214***	-.00475***
U-to-E rate	No	-.0103	-.0330***	-.0284***

Note: E-to-E = employer-to-employer, U-to-E = unemployment-to-employment. Regression estimates are taken from table 3. Log wages correspond to columns 18 and 20. The standard deviation of wages corresponds to columns 22 and 24. E-to-E rate estimates correspond to columns 2 and 4. U-to-E rate estimates correspond to columns 6 and 8.

\*\*\* $p < .001$ .

the sum of four components: the output share, the opportunity cost, the amenity discount, and the quit/promotion discount. In our calibrated steady state, the sum of the output share components over the sum of total wages is 91.4%. The opportunity cost is approximately half as important, and the amenity discount and quit/promotion discounts lower wages by approximately 20%. We summarize these results as follows:

$$w_k(\sigma, \varepsilon) = \quad (9)$$

1. Output share (91.4%)  $\sigma z_k$
2. Opportunity cost (42.0%)  $+(1-\sigma)(b + \beta \theta \sum_k^M \lambda_{kk'} \int \max\{S_{k'}(\varepsilon'), 0\} dF(\varepsilon'))$
3. Amenity discount (-18.8%)  $-(1-\sigma)\varepsilon$
4. Quit/promotion discount (-14.6%)  $-(1-\sigma)\beta \int \sum_{k' \neq k}^M \lambda_{kk'} \mathbf{1}_{[S' > S]} [S_k(\varepsilon) + \theta(S_{k'}(\varepsilon') - S_k(\varepsilon))] dF(\varepsilon') - \beta \int \sum_{k' \neq k}^M \lambda_{kk'} \mathbf{1}_{[\sigma S \leq S' < S]} [S_{k'}(\varepsilon') - \sigma S_k(\varepsilon)] dF(\varepsilon')$

The wage equation provides insight into how wage concentration, duopsonistic bargaining, and search frictions affect wages. First, concentration directly limits the surplus share  $\sigma$ . The finiteness of the number of firms

implies that  $\sigma$  is bounded below one at the highest productivity-amenity match (i.e., the highest-ranked/highest-surplus firms). Consider the economy without amenities, and denote the highest attainable surplus share at the most productive firm  $\bar{\sigma}_1$ . In a single-firm market, workers coming out of unemployment can meet only the most productive firm, and thus,  $\bar{\sigma}_1 = \theta < 1$ . With two firms in a market where the  $k = 1$  firm is the most productive, the highest-possible share of surplus is  $\bar{\sigma}_1 = (\theta S_1 + (1 - \theta)S_2)/S_1$ , which occurs when a worker at the lower-ranked firm meets the top-ranked firm via on-the-job search. A similar argument holds with amenities.

Concentration also manifests itself through future contact rates. Similar to Jarosch et al. (2024), we assume that workers cannot meet their current employer.<sup>26</sup> Hence, for a worker at firm  $k$ , the quit/promotion discount sums only over  $k' \neq k$ . Without nonwage amenities, this does not affect the surplus value but does affect the split of the surplus. With nonwage amenities, this affects the surplus value because the worker is restricted from drawing a new amenity value at their current employer (i.e., working for a new boss, changing departments), potentially creating surplus. Therefore, firms' ability to exclude themselves from future job-to-job transitions generates lower total surplus and thus lowers wages.

Neoclassical monopsony forces working through nonwage amenities contribute negatively to wages and thus drive a gap between a worker's wage and marginal product due to compensating differentials. Note that when  $\sigma = 1$ , workers are paid all the way up to their output  $z_i$ , and the amenity discount disappears. That is, a worker could obtain a high amenity value from working at firm  $i$ , but in a dynamic labor market with a high rate of outside offers,  $\sigma$  would quickly increase, and this idiosyncratic taste would no longer reduce pay. Hence, search frictions and amenities interact; furthermore, the mere finiteness of firms in the market bounds  $\bar{\sigma}_1 < 1$ , generating positive amenity discounts.

Given the wage-setting protocol in Cahuc et al. (2006), the worker's bargaining power  $\theta$  limits wage payments. Perhaps more subtly, the duopsonistic bargaining protocol—that is, only two firms ever compete simultaneously for a worker's services—also lowers the share of worker surplus  $\sigma$ . We explain this more in detail later when we allow more than two firms to bargain for a worker's services.

## VI. Counterfactuals

We now use the model to investigate a number of counterfactuals that isolate the roles of (1) concentration, (2) amenities, and (3) search frictions

for wage markdowns, welfare, and inequality. As discussed, there are many facets of concentration. Concentration affects the set of possible meetings, attainable surplus shares, and outside options. In what follows, we attempt to thoroughly explore the various dimensions of concentration, as how we model the granularity of firms is arguably the most novel aspect of our framework. We then proceed to isolate the effects of amenity dispersion and search frictions on wages and welfare.

#### *A. Isolating the Effect of the Number of Firms, $M$*

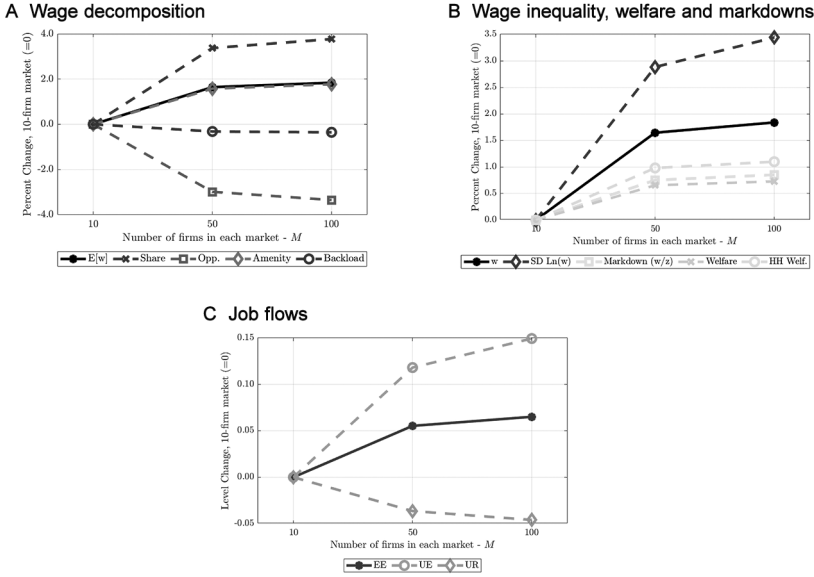
##### An Ideal Experiment

Our first counterfactual exercise aims to isolate the role of firms per market,  $M$ . We first consider an idealized experiment in which we solve an economy with 10 identical 10-firm markets—that is, with the same vector of productivities in each market—and then combine these into two identical 50-firm markets and, finally, one 100-firm market. When we combine markets, we combine the labor force as well, ensuring that  $M/\bar{N}$  is constant across exercises, thus removing any mechanical changes to firms per worker. The vector of productivities  $\mathbf{z} = (z_1, \dots, z_{10})$  that we consider is evenly spaced between the 10th and 90th percentiles of the ergodic distribution of  $z_i$ . When we combine five of the 10-firm markets to produce a single 50-firm market, our exercise keeps the rungs of the productivity ladder fixed but reduces concentration.

Figure 5A uses equation (9) to plot the effect of  $M$  on the average wage and its four components. We plot percentage changes relative to the 10-firm benchmark, reducing concentration as we move from left to right. We find that average wages increase by 2% (solid). As  $M$  increases from 10 to 100, the output share component increases by 4%, being responsible for more than all of the gains ( $x$ 's). The output share component consists of  $\sigma \times z$ , and both increase. The share parameter  $\sigma$  increases by 3% alone due to a higher inflow of outside offers. The average productivity level  $z$  increases as workers flow to higher  $z$  firms.

When shares of surplus are higher, the opportunity cost of employment is lower, which reduces the wage (squares). Notably, this nearly offsets all gains due to reallocation to higher-productivity firms ( $x$ 's). Once these offsetting effects are accounted for, the dominant force in increasing the wage is the change in the amenity discount. As the worker's surplus share increases due to more competitive outside offers in a denser labor market, the reduction in wages due to worker-firm-specific tastes is reduced.





**Fig. 5.** Altering  $M$  in isolation: ideal experiment. EE = employer-to-employer, UE = unemployment-to-employment, EU = employment-to-unemployment. A color version of this figure is available online.

Wage discrimination on the basis of idiosyncratic factors is impossible when competition is tight, as workers receive more and better outside offers. The wage penalty due to amenities declines, pushing up wages by 2% (diamonds).

With more competition, workers are now more likely to meet with higher  $z$  firms in the future, which leads to more backloading of wages (circles). In the 50-firm and 100-firm markets, workers at  $z_{10}$  can meet with another  $z_{10}$  firm that they might like more due to personal preferences. As the worker will gain some surplus from that later transition, a lower wage is required today to deliver the promised values. Quantitatively, we find that this effect is small, contributing only a small negative effect on wages as we increase the number of firms per market by a factor of 10.

Figure 5B plots four other moments of interest, including wage inequality, welfare, and markdowns. First, wage inequality rises as we move from the 10-firm market to the 100-firm market. Workers at the top of the ladder now have many more possibilities in terms of competitive outside options, leading to higher wages to retain workers. Lucky workers receive many outside offers from good firms, fanning out the wage distribution and increasing wage inequality.

Total welfare—the sum of household utility and firm profits net of vacancy-posting costs—rises by 0.75%. As worker surplus shares rise, firm profits net of vacancy costs fall. As a result, household welfare, ignoring firm profits, rises by more than 1%.

A central focus of the literature on monopsony is the static wage markdown (e.g., Robinson [1933] and Berger et al. [2022a], among many others). In static economies, the definition of the markdown is universally agreed upon in the literature—it is the ratio of the spot wage to the marginal revenue product of the worker. In the dynamic context, markdowns can be defined in a number of ways. For example, the markdown could be logically defined as  $\sigma$  or perhaps via a comparison of the net present value of wages with the net present value of productivity. To stay as close as possible to the literature, we define markdowns following the static definition, here defined at the worker-firm level:

$$\text{Markdown definition: } \mu = w/z.$$

Thus, in our dynamic setting,  $\mu$  reflects a variety of forces, including imperfect competition and its interaction with amenities, bargaining, and backloading. As a conceptual point, we are able to decompose what researchers would estimate in static models into its various components including backloading, which may not necessarily reflect noncompetitive behavior. We aggregate by taking the average across matches.

Figure 5B shows that markdowns narrow, with workers' spot wages relative to their marginal product increasing by 1% in the 100-firm market over the wages in the 10-firm markets. Importantly, the majority of the reductions in markdowns occur between the ten 10-firm markets and the two 50-firm markets. There is very little impact on markdowns between the two 50-firm markets and the single 100-firm market. This foreshadows our counterfactuals in Norway. The bulk of employment in Norway is in markets with 150 or more firms. Doubling the number of firms in an  $M_j = 150$  firm market does little to markdowns, with the rungs of the productivity ladder held fixed. Conversely, markdowns are extremely sensitive to the number of firms in highly concentrated markets.

Last, figure 5C shows the increase in mobility that underpins the improvement in workers' shares of output. The effect of the ability to meet with a same-rank firm of higher personal preference is that job-to-job transition rates rise (open circles). A knock-on effect is that, in equilibrium, poaching workers becomes easier, incentivizing vacancy posting. Higher vacancies per worker support an increase in the U-to-E rate (filled circles) and hence a fall in the unemployment rate (diamonds). Importantly, with

endogenous recruiting effort—here via vacancies—it is not only the split of surplus that is affected by the density of firms in the market but also quantities.

Application to Norway

We now extend the idealized experiment to the full Norwegian economy. Limited by computational resources, we conduct the following experiment: we duplicate the productivity vector in every simulated Norwegian market while simultaneously doubling the number of workers in each market. As in the ideal experiment, this preserves the rungs of the job ladder (i.e., the vector of  $z$  is duplicated but not altered otherwise), and the number of firms per worker remains the same. Table 6 column A contains statistics for our baseline economy, whereas column B considers the counterfactual economy. With the doubling of  $M$  in each market via duplication, the Herfindahl value is approximately halved from 0.09 to 0.05, which is mostly mechanical. As tighter competition leads workers to obtain a greater surplus share, the wage markdown narrows by 1 percentage point. The employer-to-employer job transition rate increases marginally, and labor productivity improves by 0.42% relative to that in

**Table 6**  
Counterfactuals in Norway

Variable	A.	B.	C.	D.	E.
	Baseline $\{M_{jt}, \bar{N}_j\}$	Duplicate $\mathbf{z}_j$ $2 \times \{M_{jt}, \bar{N}_j\}$	Meet yourself	One $\varepsilon$ $\frac{1}{N} \int \varepsilon_i di$ fixed	No search $\xi_\varepsilon = 1,$ $A = \infty$
<b>I. Outcomes</b>					
HHI <sub><math>n</math></sub>	.09	.05	.12	.28	.75
Markdown ( $w/z$ ), employ- ment weighted	.79	.80	.80	.86	.93
E-to-E worker flow rate (%)	.55	.56	.62	.63	1.17
<b>II. Comparison with baseline economy—percent difference</b>					
Labor productivity	–	.42	.92	7.59	12.74
Welfare	–	.25	.57	6.60	14.75
Average wage	–	.68	1.43	16.82	30.88
Standard deviation of log wage	–	.81	2.05	–8.98	–91.66

Note: E-to-E = employer-to-employer, HHI = Herfindahl-Hirschman index.

our baseline economy. Similarly to the ideal experiment, welfare increases, the average wage increases, and wages become more dispersed. The welfare and labor productivity effects are nonnegligible: welfare increases by 0.25 percent. Throughout, the mechanisms are identical to those in the ideal experiment.

These results suggest a limited role of concentration in shaping Norwegian markdowns, but importantly, the exercise captures only one facet of concentration. There are two reasons that the doubling of  $M$  yields small effects on markdowns: (1) the duopsonistic wage setting remains unchanged: when we double  $M$  there are still only ever two firms at the bargaining table competing on wages, and (2) approximately 70% of the labor force resides in markets with more than 150 firms, and hence, doubling the number of firms in those markets is irrelevant (fig. 5B). To dig deeper, we next conduct two exercises designed to measure the effects of concentration and granularity operating through outside options and duopsonistic strategic wage setting.

#### *B. Additional Sources of Monopsony Power*

##### Granularity and Outside Options

In our model, similar to the setups in Zhu (2012) and Jarosch et al. (2024), firms are able to partially exclude themselves from future meetings when bargaining with workers. Such an assumption makes sense in a granular market. In our model, this is operationalized by a restriction that workers cannot meet with their current employer through job-to-job transitions in the next period. Jarosch et al. (2024) go one step further and also exclude the current employer from the next period's possible U-to-E transitions during the bargaining process, which we do not consider.

Table 6, column C reports what happens when we remove this assumption. Now workers redraw their amenity value at the firm. One can interpret this as life events that change the utility of a firm. Even if they draw the same amenity value, the firm now bargains against itself. One can interpret this as a worker changing departments or bosses. We find that wage dispersion increases substantially. Because the firms may have to compete internally for workers, wage dispersion increases at the top of the ladder because surplus shares can now go all the way to one. Two offsetting forces limit the narrowing of markdowns, which narrow by only 1 percentage point. On the one hand, workers' ability to meet the

same firm leads to higher surplus shares, especially at the top, which increases wages. On the other hand, their ability to redraw amenity values leads to large amenity wage penalties, which lower markdowns. The latter improves welfare, however, which increases by 0.57%.

### Amenities

Similarly to Robinson (1933), we include a role for nonwage amenities in driving the employment and wage-setting decisions of firms. Unlike Robinson (1933), however, we allow firms to first-degree price discriminate over the amenity value. This puts greater downward pressure on wages but, in many settings, does not necessarily result in inefficient allocations.

We find that our removing amenity dispersion yields a large reduction in the wage markdown. In table 6, column D, we restrict the amenity distribution to one value, calibrated to deliver the same aggregate amenity per capita as in our baseline economy. In other words, we study a mean-preserving removal of amenity dispersion. Workers' spot wages as a ratio to their productivity are now 7 percentage points greater than in the baseline economy. This represents a 33% narrowing of markdowns compared with the initial 21-percentage-point markdown observed in our baseline economy. With no heterogeneity in the amenity penalty, wage dispersion declines by a tenth relative to its level in the baseline economy.

Workers' sorting due to the combination of firm productivity and idiosyncratic tastes reduces the level of output in the economy relative to that in an economy where only firm productivity determines mobility. Recall that heterogeneity in amenities was necessary to ensure consistency with the empirical frequency of down-the-ladder job moves in the Norwegian data. In the economy in column D, there are zero down-the-ladder job moves. Hence, when we remove amenities, workers are sorted across firms only on firm productivity. This causes aggregate productivity to increase substantially, by more than 7%.

### Search Frictions

We next explore the role of search frictions by steeply increasing match efficiency  $A$ .<sup>27</sup> In practice, our counterfactual implies that all employed and unemployed workers meet with a firm every period with certainty. Because our model is in discrete time, it remains the case that workers

can at most meet with one firm per period. Hence, search frictions still exist, but they are strongly mitigated.

In an economy with on-the-job search, mitigating search frictions leads to a large increase in productivity (12.74%), as workers rapidly ascend the job ladder (E-to-E rates triple). Table 6 column E shows that this is accompanied by a large increase in concentration as workers agglomerate at the most productive firms. The Herfindahl index increases by a factor of 8, from 0.09 to 0.75. When at the highest-ranked firm, workers now quickly bump into the next-highest-ranked firm, which bids up their surplus share. As a result, reducing search frictions leads to a substantial narrowing of the markdown, with wages now only 7% below marginal product. Markdowns remain positive for two reasons: (i) the recuperation of vacancy costs for firms that still operate and (ii) the fact that, with granular firms, some workers at the top of the productivity ladder are employed at an amenity discount and run out of potential outside offers to increase their wage. Both the sorting to higher  $z$  firms and greater surplus share contribute to a 30% increase in wages; meanwhile, inequality compresses dramatically, as workers end up in approximately the same boat relatively quickly.

### Bargaining Protocol and Bargaining Power

Our final set of counterfactuals addresses a source of market power that is implicit in the assumptions of this framework: duopsony in the bargaining protocol. What if a worker could instead have three firms at the negotiating table? Or four? It is beyond the scope of this paper to solve a model with these features; however, we think that future progress in this area is important. The aim of our final set of counterfactuals is to show that work in this area may be of future interest.

As a first step, we consider a counterfactual bargaining protocol that involves three firms at the table. First, we abstract from amenities, which gives an unambiguous ranking of firms. Second, we assume that whenever a worker meets with a firm of rank  $K$ , it simultaneously meets with the firm at rank  $K - 1$  (i.e., the next-best firm). We assume that the arrival of firm  $K - 1$  at the table is frictionless and does not come at the cost of any vacancies. We refer to this as the next-highest-ranked bargaining protocol. Because firm  $K$  presents the highest surplus value, the protocol affects only how surplus is split. In partial equilibrium, with contact rates held fixed, there would be no change to allocations. In general equilibrium, making the bargaining table more competitive reduces firms' values of vacancy posting, which is taken into account.

**Table 7**  
Bargaining Counterfactuals in Norway

Variable	A.	B.	C.	D.
	Single $\varepsilon$ $\frac{1}{N} \int \varepsilon_i di$ fixed	... and meet next highest rank	... and $\theta = .5$	... and $\theta = .9$
<b>I. Outcomes</b>				
HHI <sub>n</sub>	.28	.84	.54	.72
Markdown ( $w/z$ ), employment weighted	.86	.97	.97	1.00
E-to-E worker flow rate (%)	.63	.48	.64	.67
<b>II. Comparison with single <math>\varepsilon</math> economy—percent difference</b>				
Labor productivity	—	4.90	3.75	5.08
Welfare	—	7.20	5.48	7.34
Average wage	—	18.3	15.4	20.3
Standard deviation of log wage	—	−95.2	−87.2	−93.8

Note: EE = employer-to-employer, HHI = Herfindahl-Hirschman index.

Table 7 reports the results, where the comparison case is the model with a single amenity value (col. A). The markdown narrows by 11 percentage points, which is around half of the baseline markdown. A lower surplus share due to more direct competition reduces the return to vacancy posting, and especially so for the least productive firms. The shift in the distribution of vacancies toward more productive firms increases labor productivity by an additional 5%. As workers tend toward more productive firms, concentration increases and the Herfindahl index rises from 0.28 to 0.84. Despite this, the reallocation of production and more competitive labor markets increases welfare by 7.2%. Wage inequality falls, driven by compression in wages within firms as all workers are hired against a highly competitive outside offer.

Expanding the bargaining table has implications similar to those of massive increases in the bargaining power of workers.<sup>28</sup> Counterfactual levels of  $\theta$  provide context for the large effects of the next-highest-ranked protocol. A more straightforward way to understand the effect of bargaining power could be to study comparative statics with respect to the worker bargaining weight  $\theta \in [0, 1]$ . However, this is difficult to interpret by itself. Column C shows that the next-highest-ranked protocol reduces markdowns as much as an increase in  $\theta$  from 0.18 (baseline) to 0.50. Column D shows that the next-highest-ranked protocol increases welfare as much as an increase in  $\theta$  to 0.90. These are substantial increases in bargaining power, and similarly increase wages by 15%–20%.

### C. *Summary and Policy Implications*

Our experiments have explored how the three paradigms of monopsony—concentration, amenities, and search frictions—interact. Concentration is multifaceted, and we find a significant role for the hardwired duopsonistic wage setting built into models descended from Cahuc et al. (2006). Giving a seat at the bargaining table to the next-highest-ranked firm lowers markdowns by 50%. On the other hand, we find little effect of firms excluding themselves from future contacts. We also find that leaving the bargaining protocol untouched and doubling the number of firms in Norway does very little to markdowns. These results do not necessarily imply futility of stricter antitrust enforcement. Welfare significantly increases when the number of firms in the market doubles (see table 6, row 5, col. B) and even more so when firms are prevented from removing themselves from future contracts.

Although outside the scope of the current formulation of our model, antitrust law may be an effective way to alter the bargaining protocol of firms. For instance, theoretical work by Shi (2023) argues that a near-ban on noncompetes is an efficient policy. Moreover, such a ban may be within the scope of the Federal Trade Commission’s mandate. As this paper is written, the Federal Trade Commission has proposed a rule to ban noncompetes.<sup>29</sup> Our framework can be easily modified to include noncompete restrictions and thus be used to assess the effect of such policies in a setting with firm concentration.<sup>30</sup>

We find that amenity dispersion explains one-third of observed markdowns in Norway. Because our model features first-degree price discrimination, however, the presence of amenity-driven markdowns does not imply inefficiency per se. Nonetheless, any policy that tilts the bargaining protocol in favor of the worker (raising  $\sigma$  and  $\theta$ ) increases pay by an erosion of the compensating differential. Put simply, a firm obtains output from the worker, and when pushed—via outside offers—they will pay up to that level of output. This raises consumer surplus but may lower total surplus.

Finally, search frictions in our discrete-time setting account for two-thirds of observed markdowns. We view search frictions as immutable, with no practical policy to alleviate them (see also Naidu and Posner 2022). Despite search frictions alone accounting for the majority of observed markdowns, the nonlinear interaction of the sources of monopsony in our model implies that addressing concentration and neoclassical sources of monopsony can still yield sizable improvements in worker welfare and wages.



One direction that we hope that future work will explore further is the distributional consequences of monopsony and market power. As Schmitz (2016) and Herkenhoff and Raveendranathan (2019) emphasize, the deadweight costs of monopoly are sizable and borne primarily by low-income households. This might also be true for any deadweight costs due to monopsony. Our framework is well suited to addressing these questions, as it allows for wage dispersion both within and across firms.

Two additional factors that can be accommodated relatively well in the model are human capital and risk aversion. Adding human capital to the model would enable researchers to explore the impact of monopsony and market power on investment in education and skill development. This would help us understand how these market structures affect workers' long-term career prospects and their ability to adapt to changes in the labor market. Work by Jungerman (2023) makes progress along these lines by incorporating human capital accumulation into a framework with dynamic oligopsony. Incorporating risk aversion would allow a more comprehensive analysis of the welfare costs of market power. This would help researchers better understand the costs associated with job loss and the ways in which individuals' decisions and well-being are affected by market power dynamics. Last, similarly to Berger et al. (2022a), we observe that worker mobility across markets may be important for quantifying the aggregate consequences of policies. By exploring these additional factors, future research could provide a more complete picture of the distributional consequences of monopsony and market power and guide policy interventions aimed at mitigating their negative effects on workers and society at large.

## VII. Conclusion

In this paper, we develop a general equilibrium theory of monopsony that features (1) search frictions, (2) amenities, and (3) firm granularity. We estimate the strength of each source of monopsony using administrative data from Norway. Our approach introduces a novel method for defining markets and offers an extensive empirical overview of the conditional covariances of concentration, job flows, wages, and wage dispersion. Our model successfully replicates these relationships.

We use our model as a testing ground to investigate the sources of labor market power, focusing on wage markdowns—a measure defined as the ratio of a worker's current wage to the worker's productivity at

the firm. Our findings indicate that more than 50% of observed wage markdowns can be attributed to firm granularity and strategic wage setting during bargaining. Amenities account for 33% of observed markdowns, and eliminating search frictions (by setting contact rates to one in a discrete-time setting) reveals that they account for 66% of markdowns. Due to strong nonlinear interactions between the three monopsony channels, the sum of this decomposition does not equal 100%.

These results can inform policy discussions. Markdowns related to concentration could potentially be influenced by antitrust laws or restrictions on noncompete agreements, which may help reduce amenity wage penalties. In contrast, markdowns driven by search and matching processes are likely immutable and more influenced by technology than policy.

Our paper points to a number of fruitful avenues for future research. Allowing worker mobility across markets, incorporating human capital, and allowing for risk aversion are necessary to account for these first-order factors that may have an important bearing on policy recommendations. Likewise, structurally modeling natural experiments that affect the composition and productivity of firms within a market may yield novel insights into how labor markets respond to market structure and provide strong tests of our own and other existing theories of oligopsony.

## Endnotes

Authors email addresses: David Berger (david.berger@duke.edu), Kyle Herkenhoff (kyle.herkenhoff@gmail.com), Andreas R. Kostøl (andreas.r.kostøl@gmail.com), Simon Mongey (simonmongey@gmail.com). This paper is for the *NBER Macroeconomics Annual 2023*, and we thank the editors, Martin Eichenbaum, Erik Hurst, and Valerie Ramey. We also thank our discussants, Giuseppe Moscarini and Katarína Borovičková. Any opinions and conclusions expressed herein are those of the author(s) and do not necessarily represent the views of the Federal Reserve Bank of Minneapolis, the Federal Reserve System, or Norges Bank. Thanks to Alexander Weinberg for excellent research assistance. This work was supported by the Norwegian Research Council grant numbers 227115 and 315437 and National Science Foundation award #SES-2214431. For acknowledgments, sources of research support, and disclosure of the authors' material financial relationships, if any, please see <https://www.nber.org/books-and-chapters/nber-macroeconomics-annual-2023-volume-38/anatomy-monopsony-search-frictions-amenities-and-bargaining-concentrated-markets>.

1. See Manning (2003) for a summary of the literature and recent papers by Card et al. (2018), Dube et al. (2020), Azar, Marinescu, and Steinbaum (2022), Benmelech et al. (2022), Berger et al. (2022a), Lamadon et al. (2022), Schubert, Stansbury, and Taska (2022), and Yeh et al. (2022), as well as Brooks et al. (2019), Dodini et al. (2020), Dodini, Løken, and Willén (2022), Felix (2022), and Rubens (2023) outside of the United States, among others.

2. In our baseline model, firms do not exclude themselves from future U-to-E transitions, only contemporaneous E-to-E meetings. Thus, this particular experiment holds fixed the set of possible U-to-E transitions, differentiating what we do from the approach in Jarosch et al. (2024).

3. Roussille and Scuderi (2022) find a similarly large role for amenities in empirical analysis of online job-board wage postings.

4. Early progress on this question is being made by Flinn and Mullins (2021), among others.

5. For the United States, economists have used Burning Glass data for occupation and wage information (Schubert et al. 2022). These data pose serious issues for analyses such as the one here. First, the data do not contain information on the universe of employees, employers, and jobs or wages paid to employees. Second, data on advertisements lack information on the quantity of positions and hence cannot be used to compute market shares. Third, only 6% of the advertisements scraped and collated by Burning Glass have wage, employer, and occupation information. Table A4 of Hazell et al. (2022) shows that although the 2010–19 data contain 239 million ads, dropping those without wages or a range of wages posted and without firm, county, sector, or occupation data leaves only 15 million ads, which is 6.27% of the initial sample. Further screening reduces their analysis sample to less than 1.6% of all ads.

6. The four-digit occupational classification is based on the International Standard Classification of Occupations adapted to Norwegian labor markets by Statistics Norway and the employment agency. There are strong incentives for correct reporting. First, the Norwegian labor law stipulates that firms undergoing a mass layoff, defined as laying off more than 10 workers over 30 days, must follow the last-in, first-out principle. The ordering is typically defined within position and establishment. Second, the employment agency uses information about occupation and the workplace location for targeted job search assistance. In practice, employers report positions by a seven-digit system (see <https://www.ssb.no/klass/klassifikasjoner/145/>, in Norwegian), with new job titles added at regular intervals.

7. For example, the occupational code for “economics and business” includes consultants, controllers, junior and senior credit analysts, and research and chief economists, to name a few. The codes also cluster unskilled positions, such as maintenance workers and janitors, and different levels of management positions into distinct groups.

8. See data appendix A, <http://www.nber.org/data-appendix/c14848/appendix.pdf>, for a more detailed description of the sources and variables and Bhuller (2009) for commuting patterns.

9. The petroleum sector accounts for a large fraction of income but is excluded from the calculation of the labor share.

10. Union membership in Norway is high relative to that in other countries in the OECD and the United States but fell from 58% to 53% from 1992 to 2013 (OECD Statistics Trade Union Statistics, <https://stats.oecd.org/Index.aspx?DataSetCode=TUD>). Unions play an important role in ensuring that firms comply with labor law, stating, e.g., that downsizing requires a 1-month notification to employees, with the dismissal time varying from 1 to 6 months, depending on age and tenure. Wrongful discharge can end with a lawsuit, where firms must compensate the dismissed employees for lost income.

11. Payroll taxes finance the UI system, and there is no experience rating on the firm. The potential benefit period is 52 weeks for workers who have earned less than twice the National Insurance basic amount for the last 3 years. The “basic amount” of benefits is currently approximately USD 1,000 per month. UI benefits are capped at a maximum level of previous earnings, currently six times the basic amount, which creates a kink in the benefit formula. To remain eligible for the cash benefits, work hours must have fallen by at least 50%, and recipients must be actively looking for work and willing to take any employment.

12. In appendix A.3, <http://www.nber.org/data-appendix/c14848/appendix.pdf>, we describe the construction of self-flow rates and plot the employment distribution of occupations by their self-flow. Approximately 50% of the workforce has a rate above 50%.

13.  $HHI_r^k = \sum_{i=1}^K (s_r^k)^2$ , where  $s_r^k$  is the employment share of occupation cluster  $k \in \{1, \dots, K\}$  in commuting zone region  $r$ .

14. Note the clustering algorithm does not produce labels. We only use the label “manual low-skill services” for heuristic purposes.

15. See Berger et al. (2022a) for why weighting by the wage bill is appropriate.

16. A market with  $M$  identically sized firms has an HHI of  $\text{HHI}_M = \sum_i (1/M)^2 = 1/M$ . Hence, an HHI of 0.09 is what one would obtain from a market with  $1/0.09 \approx 11$  equally sized firms.

17. These natural experiments can only be interpreted through a structural model, as Azkarate-Askasua and Zerecero (2024) adeptly do.

18. Recall that our clustering approach has potentially clustered four-digit occupations into different groupings in different regions due to heterogeneity across regions in the occupation-to-occupation flow matrix. Hence, our unit of analysis is the occupation, but statistics such as concentration or wage inequality are measured at the level of the market (cluster-commuting zone) to which the occupation-commuting zone is assigned by our algorithm.

19. Small concentrated markets vs. large less concentrated markets are precisely the comparison that we want to study and thus should not be downweighted by labor force size.

20. Note that Robinson's (1933) class of models does not allow for wage discrimination based on amenity draws, thus yielding inefficient allocations. First-degree price discrimination, on the other hand, yields the efficient allocation, with zero consumer surplus and wages set below marginal products.

21. For example, a firm may turn down all hires with the lowest amenity draw so as to wait until it meets a worker with a better amenity draw, to whom it can pay lower wages via a compensating differential. In a market with 150 firms and 3 potential amenity draws,  $3^{150}$ , such possible complex hiring rules exist.

22. Specifically, for example, firm  $i$  does not internalize that its vacancies affect the contact rates of workers at firm  $k$  and hence affect the surplus  $S_k(\epsilon')$ , which then affects the cost of hiring a worker from firm  $k$ .

23. Note that search efficiency is accounted for by  $\lambda_{ik}$  and  $\lambda_{ui}$  and that the density of  $F(\epsilon)$  is denoted  $f(\epsilon)$ .

24. See Bagger and Lentz (2019) table 2 for more detail.

25. In this class of models, low values of  $\theta$  can deliver negative wages, or near-zero wages, generating what look like fat-tailed wage growth distributions (e.g., in reasonable calibrations, workers can move from wages of 0.001 to a wage of 1, for instance, yielding 1,000% growth rates). Higher values of  $\theta$  remove negative and near-zero wages, thus bringing the wage growth rates in line with the data.

26. Note that Jarosch et al. (2024) further exclude the unemployed from meeting their former employer.

27. Numerically, we implement this by setting  $A$  to 100, which is more than 500 times the level in our baseline (table 4). We also set  $\xi_u$  to 1, so that unemployed and employed workers have the same search efficiency. Because  $A$  is so large, this is almost irrelevant, but we do it nonetheless for completeness.

28. Anecdotally, this is well understood in the post-PhD economics job market and provides incentives for firms to make exploding offers, limiting workers' length of search and hence the number of competing firms at the negotiation stage.

29. See <https://www.ftc.gov/legal-library/browse/federal-register-notice/non-compete-clause-rulemaking>.

30. See Berger, Hasenzagl, et al. (2022) for an extensive discussion of the role antitrust authorities could play in addressing monopsony power.

## References

- Azar, José, Ioana Marinescu, and Marshall Steinbaum. 2022. "Labor Market Concentration." *Journal of Human Resources* 57 (S): S167–S199.
- Azar, José A., Steven T. Berry, and Ioana Marinescu. 2022. "Estimating Labor Market Power." Discussion paper. <https://www.nber.org/papers/w30365>.
- Azar, José A., Ioana Marinescu, Marshall I. Steinbaum, and Bledi Taska. 2018. "Concentration in US Labor Markets: Evidence from Online Vacancy Data." Working Paper no. 24395, NBER, Cambridge, MA.

- Azkarate-Askasua, Miren, and Miguel Zerecero. 2024. "Union and Firm Labor Market Power." [https://mzerecero.netlify.app/working\\_papers/labor\\_conc\\_occ.pdf](https://mzerecero.netlify.app/working_papers/labor_conc_occ.pdf).
- Bagga, Sadhika. 2023. "Firm Market Power, Worker Mobility, and Wages in the US Labor Market." *Journal of Labor Economics* 41 (S1): S205–S256.
- Bagger, Jesper, and Rasmus Lentz. 2019. "An Empirical Model of Wage Dispersion with Sorting." *Review of Economic Studies* 86 (1): 153–90.
- Benmelech, Efraim, Nittai K. Bergman, and Hyunseob Kim. 2022. "Strong Employers and Weak Employees: How Does Employer Concentration Affect Wages?" *Journal of Human Resources* 57 (S): S200–S250.
- Berger, David, Thomas Hasenzagl, Kyle Herkenhoff, Simon Mongey, and Eric Posner. 2022. "Merger Guidelines for Labor Markets." Working paper. <https://www.nber.org/papers/w31147>.
- Berger, David, Kyle Herkenhoff, and Simon Mongey. 2022a. "Labor Market Power." *American Economic Review* 112 (4): 1147–93.
- . 2022b. "Minimum Wages, Efficiency and Welfare." Working Paper no. 29662, NBER, Cambridge, MA.
- Bhaskar, V., A. Manning, and T. To. 2002. "Oligopsony and Monopsonistic Competition in Labor Markets." *Journal of Economic Perspectives* 16 (2): 155–74.
- Bhaskar, V., and T. To. 1999. "Minimum Wages for Ronald McDonald Monopsonies: A Theory of Monopsonistic Competition." *Economic Journal* 109 (455): 190–203.
- Bhuller, M. 2009. "Classification of Norwegian Labor Market Regions." *SSB Notater*, 24. [https://www.ssb.no/a/publikasjoner/pdf/notat\\_200924/notat\\_200924.pdf](https://www.ssb.no/a/publikasjoner/pdf/notat_200924/notat_200924.pdf).
- Bhuller, M., K. O. Moene, M. Mogstad, and O. L. Vestad. 2022. "Fact and Fantasies about Wage Setting and Collective Bargaining." *Journal of Economic Perspectives* 36 (4): 29–52.
- Bloesch, Justin, and Birthe Larsen. 2023. "When Do Firms Profit from Wage Setting Power? New vs. Classical Monopsony." Manuscript. [https://drive.google.com/file/d/1KIO2elh7fbcKovtxqxS26prfkxOt\\_00A/view](https://drive.google.com/file/d/1KIO2elh7fbcKovtxqxS26prfkxOt_00A/view).
- Brooks, Wyatt J., Joseph P. Kaboski, Yao Amber Li, and Wei Qian. 2019. "Exploitation of Labor? Classical Monopsony Power and Labor's Share." Working Paper no. 25660, NBER, Cambridge, MA.
- Burdett, K., and D. T. Mortensen. 1998. "Wage Differentials, Employer Size, and Unemployment." *International Economic Review* 39 (2): 257–73.
- Burdett, Kenneth, Shouyong Shi, and Randall Wright. 2001. "Pricing and Matching with Frictions." *Journal of Political Economy* 109 (5): 1060–85.
- Cahuc, Pierre, Fabien Postel-Vinay, and Jean-Marc Robin. 2006. "Wage Bargaining with On-the-Job Search: Theory and Evidence." *Econometrica* 74 (2): 323–64.
- Card, David, Ana Rute Cardoso, Jörg Heining, and Patrick Kline. 2018. "Firms and Labor Market Inequality: Evidence and Some Theory." *Journal of Labor Economics* 36 (S1): S13–S70.
- Dodini, Samuel, Katrine Vellesen Løken, and Alexander Willén. 2022. "The Effect of Labor Market Competition on Firms, Workers, and Communities." Discussion Paper no. 17, NHH Department of Economics, Oslo. <https://openaccess.nhh.no/nhh-xmlui/handle/11250/3032278>.
- Dodini, Samuel, Michael Lovenheim, Kjell G. Salvanes, and Alexander Willén. 2020. "Monopsony, Skills, and Labor Market Concentration." <https://ideas.repec.org/p/cpr/ceprdp/15412.html>.
- Dube, Arindrajit, J. Jacobs, Suresh Naidu, and Siddharth Suri. 2020. "Monopsony in Online Labor Markets." *American Economic Review: Insights* 2 (1): 33–46.

- Engbom, N., and C. Moser. 2022. "Earning Inequality and the Minimum Wage: Evidence from Brazil." *American Economic Review* 112 (12): 3803–47.
- Felix, Mayara. 2022. "Trade, Labor Market Concentration, and Wages." Job Market Paper. [https://www.mayarafelix.com/papers/Felix\\_JMP.pdf](https://www.mayarafelix.com/papers/Felix_JMP.pdf).
- Flinn, Christopher, and Joseph Mullins. 2021. "Firms' Choices of Wage-Setting Protocols." Discussion paper, New York University.
- Gottfries, Axel, and Gregor Jarosch. 2023. "Dynamic Monopsony with Large Firms and an Application to Non-competes." Manuscript. <https://www.nber.org/papers/w31965>.
- Hazell, Joe, Christina Patterson, Heather Sarsons, and Bledi Taska. 2022. "National Wage Setting." [https://www.nber.org/system/files/working\\_papers/w30623/w30623.pdf](https://www.nber.org/system/files/working_papers/w30623/w30623.pdf).
- Herkenhoff, Kyle F., and Gajendran Raveendranathan. 2019. "Who Bears the Welfare Costs of Monopoly? The Case of the Credit Card Industry." <https://www.nber.org/papers/w26604>.
- Hurst, E., P. Kehoe, E. Pastorino, and T. Winberry. 2022. "The Distributional Impact of the Minimum Wage in the Short and Long Run." Discussion paper. <https://www.nber.org/papers/w30294>.
- Jarosch, Gregor. 2023. "Searching for Job Security and the Consequences of Job Loss." *Econometrica* 91 (3): 903–42.
- Jarosch, Gregor, Jan Sebastian Nimcsik, and Isaac Sorkin. 2024. "Granular Search, Market Structure, and Wages." *Review of Economic Studies*. <https://doi.org/10.1093/restud/rdae004>.
- Jungerman, William. 2023. "Monopsony and Human Capital." Manuscript, University of Minnesota, Minneapolis.
- Kostøl, Andreas R. 2017. "Mismatch and the Consequence of Job Loss." Discussion paper, Preliminary Working Paper, Arizona State University, Phoenix.
- Kroft, K., Y. Luo, M. Mogstad, and B. Setzler. 2020. "Imperfect Competition and Rents in Labor and Product Markets: The Case of the Construction Industry." Working Paper no. 27325, NBER, Cambridge, MA.
- Lamadon, Thibaut, Magne Mogstad, and Bradley Setzler. 2022. "Imperfect Competition, Compensating Differentials, and Rent Sharing in the US Labor Market." *American Economic Review* 112 (1): 169–212.
- Lindenlaub, Ilse, and Fabien Postel-Vinay. 2021. "The Worker-Job Surplus." Working Paper no. 28402, NBER, Cambridge, MA.
- Lise, Jeremy, and Fabien Postel-Vinay. 2020. "Multidimensional Skills, Sorting, and Human Capital Accumulation." *American Economic Review* 110 (8): 2328–76.
- Lise, Jeremy, and Jean-Marc Robin. 2017. "The Macrodynamics of Sorting between Workers and Firms." *American Economic Review* 107 (4): 1104–35.
- Manning, Alan. 2003. *Monopsony in Motion: Imperfect Competition in Labor Markets*. Princeton, NJ: Princeton University Press.
- Naidu, S., and E. A. Posner. 2022. "Labor Monopsony and the Limits of the Law." *Journal of Human Resources* 57 (2): S284–S323.
- Postel-Vinay, Fabien, and Jean-Marc Robin. 2002. "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity." *Econometrica* 70 (6): 2295–350.
- Rinz, Kevin. 2022. "Labor Market Concentration, Earnings, and Inequality." *Journal of Human Resources* 57 (S): S251–S283.
- Robinson, Joan. 1933. *The Economics of Imperfect Competition*. London: Palgrave Macmillan.
- Røed, K., and T. Zhang. 2003. "Does Unemployment Compensation Affect Unemployment Duration?" *Economic Journal* 113 (484): 190–206.

- Roussille, Nina, and Benjamin Scuderi. 2022. "Bidding for Talent: Equilibrium Wage Dispersion on a High-Wage Online Job Board." Working paper, MIT, Cambridge, MA.
- Rubens, Michael. 2023. "Market Structure, Oligopsony Power, and Productivity." *American Economic Review* 113 (9): 2382–410.
- Schmitz, James A. 2016. "The Cost of Monopoly: A New View." Federal Reserve Bank of Minneapolis Region.
- Schmutte, Ian. 2014. "Free to Move? A Network Analytic Approach for Learning the Limits to Job Mobility." *Labour Economics* 29 (C): 49–61.
- Schubert, Gregory, Anna Stansbury, and Bledi Taska. 2022. "Employer Concentration and Outside Options." Mimeo. <https://equitablegrowth.org/wp-content/uploads/2021/01/030822-WP-Employer-Concentration-and-Outside-Options-Schubert-Stansbury-and-Taska.pdf>.
- Shi, Liyan. 2023. "Optimal Regulation of Noncompete Contracts." *Econometrica* 91 (2): 425–63.
- Taber, C., and R. Vejlin. 2020. "Estimation of a Roy/Search/Compensating Differential Model of the Labor Market." *Econometrica* 88 (3): 1031–69.
- Yeh, Chen, Claudia Macaluso, and Brad Hershbein. 2022. "Monopsony in the US Labor Market." *American Economic Review* 112 (7): 2099–138.
- Zhu, Haoxiang. 2012. "Finding a Good Price in Opaque Over-the-Counter Markets." *Review of Financial Studies* 25 (4): 1255–85.