Dynamic Monopsony with Large Firms and Noncompetes*

Axel Gottfries† Gregor Jarosch‡

April 2, 2024

Abstract

How do noncompete agreements between workers and firms affect wages and employment in equilibrium? We build a tractable framework of wage posting with on-the-job search and large employers that provides a natural laboratory to assess anti-competitive practices in the labor market. We characterize the impact of market structure and show that noncompetes can sharply suppress wages. We validate the quantitative model with empirical evidence on the impact of mergers and noncompetes on employment and wages. Banning noncompetes in the US would raise wages by 4%. Wage gains are large when demand is inelastic, training costs are high, and when noncompetes are widespread.

*Thanks to our discussants Brigitte Hochmuth, Simon Mongey, and Liyan Shi. Thanks to David Berger, Mike Elsby, Kyle Herkenhoff, Leqing Huang, Matt Johnson, Laura Pilossoph, Elena Prager, Isaac Sorkin, and Daniel Xu for thoughts and comments. Thanks to Zhaosheng Li for research assistance.

†University of Edinburgh, axel.gottfries@gmail.com

‡Duke University and NBER, gregorjarosch@gmail.com

1 Introduction

There is now a large and thriving literature on imperfect competition in the labor market. At the same time, policymakers in the US are increasingly focusing on anticompetitive practices in the labor market. Most notably, the Federal Trade Commission (FTC) has proposed an outright ban on noncompete agreements between workers and firms in the US labor market.[1] Noncompetes, which cover approximately 18% of the US workforce, including many workers in low-skilled, low-pay jobs, often occur in highly concentrated labor markets and restrict worker mobility to a few close competitors (Starr et al., 2021).

In this paper, we assess the consequences of a ban such as that proposed by the FTC from both a theoretical and a quantitative perspective. To do so, we develop a new model of wage posting and on-the-job search in the tradition of the canonical Burdett–Mortensen (1998, BM) model. This framework is widely considered a workhorse model in the literature on monopsony in labor markets (Manning, 2003). We extend it along four key dimensions.

First and most importantly, the model has a finite number of employers, all large with respect to the labor market. This allows us to capture the granular nature of many local labor markets, where anticompetitive practices might have particularly adverse impacts on workers. Second, it has decreasing returns and hence can endogenize firm size and market structure more flexibly than the standard model, which is restrictive since firms can adjust employment only through wages. Third, we introduce a downward-sloping market-level product demand curve. This allows to also connect with settings where adjustment to shocks operates primarily through prices rather than employment and output. Finally, we work with a hiring rather than a vacancy cost. This captures that the most substantive costs associated with turnover are due to hiring and training, rather than merely locating workers (Manning, 2011; Blatter et al., 2012).

Jointly, these innovations substantially generalize the textbook model and make it a natural laboratory for the analysis of anticompetitive practices in the labor market.

We demonstrate key properties of the environment by studying how a granular market structure affects wages. When there are fewer firms, there is less outside competition along the job ladder, which means that any firm's effective labor supply curve becomes more in-

[1] www.ftc.gov/noncompetes

elastic, leading to wage compression. This extends the arguments in Manning (2003) to a granular setting. In addition, we also endogenize the reservation wage and firms' employment decisions so as to analyze the full general-equilibrium response of employment and wages to shocks and policies. We relate equilibrium wages, markdowns, and the quit elasticity to the concentration of the labor market.

We then use the model as a laboratory to formally study the economic effects of noncompetes. We model these as wage offers that come with a stipulation that the worker must not transition—job-to-job—to another employer. To understand their impact, it is useful to begin with the Diamond (1971) Paradox. It states that, in an equilibrium model of wage posting, no firm should post any wage above the reservation wage which will equal the flow value of unemployment. BM overcome this by incorporating competition along the job ladder; a poacher offering higher pay is rewarded with lower turnover. This form of competition shifts rents to workers and erodes the Diamond equilibrium.

We show that firms with access to noncompetes offer a mass of identical jobs, all of which deliver the lowest possible value to workers. Why is the logic of BM no longer operative? Because workers at the mass point cannot be poached, they are under a noncompete.

When some firms introduce noncompetes this reduces competition along the job ladder and spills over to all other firms in the market. In the limit where all firms can offer noncompetes, the Diamond (1971) equilibrium reappears, and wages collapse. This shows that the widespread adoption of noncompetes can sharply hurt workers by undermining labor market competition.

We also discuss the efficiency of banning noncompetes. We show that firms with noncompetes are inefficiently large relative to the firms without noncompetes which introduces misallocation. Banning the practice hence improves the allocation of workers to firms. But noncompetes also reduce turnover costs because the reduce worker churn. Banning noncompetes increases these costs. In addition, production in our setting is already inefficiently low because of product market power. This is exacerbated by the rise in costs.

We conduct the theoretical analysis in a setting with symmetric firms but then show how the model can be extended to allow for a richer, asymmetric market structure where firms differ in terms of productivity and hiring cost. We show that this richer setting remains tractable and can easily be solved numerically. We calibrate the model and then turn to our quantitative applications.

We set out by confronting the model with existing empirical evidence on firm mergers and noncompete agreements. We view both exercises as model validation. To begin with, we use information on the wage and employment effect of hospital mergers in Prager and Schmitt (2021). We show that our model replicates their headline results on the response of wages and employment to mergers that occur in highly concentrated labor markets. In many other cases, good ex post data might not yet be available, and so our model can straightforwardly be used to predict the impact of a merger. We also discuss how our approach contrasts with a neoclassical, static approach to merger analysis (Berger et al., 2023b).

Our second validation exercise considers the impact of Oregon's 2008 ban on noncompetes for low-skilled workers, studied in Lipsitz and Starr (2022). These authors show that the average low-skilled wage increased by 2.2–3.1% while job-to-job mobility increased by 12–18%. Our model closely captures this. The model also predicts strong spillover effects that are quantitatively in line with the results in Starr et al. (2019) and Johnson et al. (2020).

In light of these validation exercises, we view the model as a natural laboratory to assess the potential impact of the FTC's proposed ban on noncompetes in the US. Our baseline analysis suggests that a ban on noncompetes would increase wages by 4% as a consequence of the rise in competition. The model predicts a large increase in worker turnover, and strong wage spillovers to firms that did not initially use noncompetes.

We conduct substantive heterogeneity analysis with regard to (local) labor market features. The wage increase is larger in markets with high hiring costs, widespread initial use of noncompetes, inelastic demand, and when highly productive firms use noncompetes. Wage gains are typically in the range of 2–6% but can rise to 15% when noncompetes are common and hiring costs are high, a case that plausibly describes many labor markets.

While reducing misallocation, a ban on noncompetes generates a sharp rise in worker churn that is wasteful from an aggregate perspective. This rise in cost additionally further exacerbates product market distortions. Whether a ban benefits, in utility terms, the workers directly affected therefore depends on the extent to which they consume the goods they produce.

Overall, our analysis suggests that it is difficult to make the case for a ban purely on efficiency grounds, in particular if output is already distorted. At the same time, we show that noncompetes, if widespread, are a powerful tool of wage suppression in equilibrium and a ban is a relatively cheap way of restoring wage competition.

1.1 Relation to Literature

It has long been recognized that labor markets are imperfectly competitive. Robinson (1933) was the first to formulate a notion of monopsony in labor markets. Similarly, work in the search tradition has long emphasized frictions that lead to rents and market power. However, only in recent years has a literature specifically focused on the origins and consequences of employer labor market power taken off. This literature is particularly interested in the impact of (changes in) market structure on wages and employment.

An early, important contribution to this literature by Manning (2003) conceptualized labor market power through the lens of the Burdett–Mortensen model of wage posting and on-the-job search. In this framework, now sometimes called the "modern" or "dynamic monopsony" model, labor market power is rooted in search frictions and can be assessed by empirically measuring the elasticity of the recruiting and quit functions (see Manning (2021) for an overview). Much of the follow-up literature has taken a reduced-form approach and focused on settings where a firm-level elasticity can cleanly be measured; see, e.g., Dube et al. (2019) and Dube et al. (2020).

In comparison, our microfounded equilibrium approach allows us to assess policies and quantify the consequences of shifts in market structure in settings where clean variation that allows a reduced-form approach is not (yet) available. In addition, our approach can be used to gauge the cost and benefits of a change in policy or market structure and quantify the equilibrium employment and output response. It goes further than existing model-based work since it allows for both product and labor market power and endogenizes reservation wages, the structure of the labor market, and markdowns.

A more recent wave of papers takes a "neoclassical" approach, starting with Card et al. (2018), who build on a static model of monopsonistic competition. This is a frictionless approach in which market power derives from employers being differentiated from the perspective of the workforce, yielding an upward-sloping labor supply curve at the employer level. Berger et al. (2022) extend this approach so that it can connect with (locally) granular labor markets, with a finite number of large employers that strategically compete for workers locally. Jungerman (2023) extends this framework to incorporate human capital dynamics.

In this paper, we revisit the "dynamic" perspective in the search tradition but extend it to connect with granular markets. Doing so connects this perspective with a new range of questions centered around worker mobility and labor market structure. It is not obvious how to approach these in a frictionless setting. A further advantage of this approach is that it grounds labor market competition in readily observable frictions.

Jarosch et al. (2023, JNS) first modeled a granular market structure in a frictional labor market. They do so in the context of the canonical random search model with bargaining, the Diamond–Mortensen–Pissarides (DMP) model. This paper is complementary in that it introduces similar considerations into the canonical random search model with wage posting, the BM model. The economic forces at play are also different. In JNS, competition for workers operates through outside options. Here, outside competition works along the job ladder and more competition leads to more quits, which drives up wages. Bagga (2023) studies the same setting with offer matching along the lines of Cahuc et al. (2006) instead of Nash bargaining. Berger et al. (2023a) add dispersed amenities.

Gouin-Bonenfant (2022) uses a BM setting to explore the fanning-out of the firm productivity distribution, which might lead to less local competition for workers, depressing wages. Our setting shares the emphasis on competition in a BM setting but focuses on a granular market structure.[2] Potter et al. (2022) model noncompetes as a decrease in the efficiency of on-the-job search in a wage posting model. Finally, Shi (2023) is closely related and studies noncompetes in a frictional labor market with bargaining. She argues that worker-firm pairs can use noncompetes to extract rents from outside employers that poach workers. Noncompetes in her setting are thus used to extract rents from a third party, outside employers, while they are used by employers to reduce competition and extract rents from workers in our setting, representing a different channel.

2 Model

The framework shares many features with the canonical Burdett–Mortensen model. The labor market features search frictions. Search is random, and workers search both on and

[2] The paper also builds on the recent literature on models with random on-the-job search and decreasing returns to scale by incorporating a granular market structure (Lentz and Mortensen, 2012; Bilal et al., 2022; Elsby and Gottfries, 2022).

off the job. Firms post wages and commit to paying the posted wage. They may post a mix of wages. There is a measure 1 of workers, a fraction e of whom are employed, with the remainder $u = 1 - e$ being unemployed. Workers lose their jobs at rate δ and become unemployed. Unemployed workers receive flow income b. We focus on a stationary equilibrium.

The following features are nonstandard. First, firms operate a decreasing-returns-to-scale production function, $F(N) = xN^{\alpha}$, with N denoting firm-level employment and x denoting firm-level productivity.

Second, firms choose a contact rate at which workers receive an offer from them but, instead of a vacancy cost, pay a hiring cost, denoted c. One interpretation is that this includes a firm's investment into its workers' firm-specific human capital. It follows that firms face turnover cost because they lose workers and need to replace them. However, firms always reach their desired employment level.

Third, workers have quasi-linear utility, which results in a downward-sloping market-level demand curve for output. Since—as we discuss next—firms are large, they internalize the price impact of their employment decisions.

Finally, there is a finite number M of firms, each accounting for a strictly positive fraction of employment in the labor market and output in the goods market. Joint with random on-the-job search, this implies that currently employed workers sometimes encounter jobs posted by their own employers. To the extent that an employer posts a mix of wages, this gives rise to the possibility of an internal transition to a higher wage. We assume that, in such an event, the firm does not need to pay the hiring cost again. Since firms are, in equilibrium, indifferent across any wage they post, they are thus indifferent about such a raise. They are compensated for the higher wage by the associated reduction in expected turnover costs.

Demand in the Product Market

We assume that all firms in a market produce an identical composite good. The boundaries of the product market are, by assumption, the same as those of the labor market. Total output is the combined output of the set of firms that compete for the same workers.

Workers have linear preferences over an outside good ("money"), so their utility is quasi-linear. Employers care only about the outside good.[3] Thus, there is transferable utility between workers and firms. Workers choose consumption C to maximize the instantaneous utility function

$$v = \frac{\eta}{\eta - 1} \bar{Q}^{\frac{1}{\eta}} C^{\frac{\eta-1}{\eta}} + I - pC.$$

They do so taking income I, which is either equal to the wage or the flow income of unemployment, as given.

Quasi-linear preferences imply that consumption is equalized across all workers and, since there is a measure 1 of them, equilibrium consumption of each worker is simply $C = \sum_j x_j N_j^{\alpha}$. Optimal consumption requires $p = \bar{Q}^{\frac{1}{\eta}} C^{\frac{-1}{\eta}}$, which results in an iso-elastic market-level inverse demand function where the output price satisfies

$$p = \bar{Q}^{\frac{1}{\eta}} \quad \sum_{j=1}^{M} x_j N_j^{\alpha} \Bigg)^{-\frac{1}{\eta}}. \tag{1}$$

Setting the model up in this fashion delivers two convenient features. First, the utility of workers and employers is linear in income and profits, respectively, as in the standard model. Second, the distribution of income does not matter for consumption. Finally, it allows straightforward calculation of the welfare consequences of a shock or policy change, as we show below.

The literature on imperfect competition in the labor market frequently assumes that the output price is fixed, as in, e.g., Berger et al. (2022). Introducing a generic downward-sloping demand curve allows us to generalize and cover cases where adjustment operates primarily through prices rather than quantities. It also allows the markup to depend on market structure and hence permits us to trace out the impact of a shock to market structure on both the input and output markets.

Workers

Unemployed and employed workers make contact with employer j at endogenous rate ψ_j and $s\psi_j$, respectively. If that happens, they draw from the firm's distribution of posted

[3] This assumption allows us to sidestep an interesting and natural issue that otherwise arises with large firms: if their output accounts for a large share of their owners' consumption, firms must take into account the impact of their output decision on the prices faced by their owners.

wages $F_j(w)$. Workers are assumed to move with small, strictly positive probability when indifferent.[4]

Workers' preferences are linear in income, following from the quasi-linear utility assumption discussed above. A worker's value of unemployment and employment at wage w hence satisfies, respectively,

$$\begin{aligned} rU &= b + \sum_{j=1}^{M} \psi_j \int_{w_r}^{\infty} \left(W(\tilde{w}) - U\right) dF_j(\tilde{w}), && (2) \\ rW(w) &= w + \delta\left(U - W(w)\right) + \sum_{j=1}^{M} s\psi_j \int_{w}^{\infty} \left(W(\tilde{w}) - W(w)\right) dF_j(\tilde{w}). && (3) \end{aligned}$$

Since search opportunities are the same in any firm, employed workers move whenever they receive a higher wage offer, including from their own employer. Unemployed workers decide on a reservation wage w_r that satisfies $W(w_r) = U$. Standard arguments yield

$$w_r = b + (1-s) \int_{w_r}^{\infty} \frac{\sum_j \psi_j \left(1 - F_j(\tilde{w})\right)}{r + \delta + \sum_j s\psi_j \left(1 - F_j(\tilde{w})\right)} d\tilde{w}. \tag{4}$$

Firms

There are M firms in the market.[5] In this granular setting, several subtle, nonstandard issues arise when setting up the firm problem. We discuss these jointly at the end of this section.

We cast the firm problem as one of directly choosing employment and the distribution of wages across employed workers.[6] We will restrict attention to stationary equilibria. This is less restrictive than it might seem. It is straightforward to verify that the firm problem is such that if firms other than i opt for stationary strategies, a stationary policy is optimal for

[4]This assumption rules out equilibria in which all firms pay the same wage and workers never move. Such equilibria do not arise in the standard BM environment where firms care about both the pace of hiring and the retention rate. In our setting, firms only care about the retention rate and so such equilibria are possible if indifferent workers never move. With the assumption that workers do move with positive probability when indifferent the usual BM deviation argument rules out such equilibria. This assumption is common in the literature, e.g. Shimer (2006).

[5]While the analysis applies to the full monopsonist case with $M = 1$, this case is not particularly interesting since that firm simply posts b. We focus on $M > 1$.

[6]We restrict attention to cases where total equilibrium employment demand is below 1, the normalized size of the workforce. This is always the case if firms are homogeneous and alternatively satisfied via a parameter restriction such that the marginal revenue product at full employment is below $b + (r + \delta)c$.

i. The stationary equilibrium studied below is therefore an equilibrium even if we allow for unrestricted, time-dependent choices.

Of course, in this physical environment firms cannot actually directly choose these two objects. To implement the optimal time-invariant solution, firms i) hire, at time zero, to reach their desired workforce and distribution of wages; and ii) choose a time-invariant contact rate ψ_i and a distribution of posted wages $F_i(w)$ to sustain these.[7] That is, the transition is immediate as the economy jumps directly to steady state which is feasible and optimal due to the linear hiring technology. The standard timeless equilibria studied in the BM literature correspond to $r \to 0$, a special case that makes the transition irrelevant.

What are the contact rate and wage offer distribution that implement a firm's optimal employment level and pay distribution? Given all other firms' choice of contact rate and posted wages, we can construct the firm's implied contact rate and posted wages as follows. Let $g_i(w)$ denote the steady state density of workers employed at wage w at firm i. Since no firm posts a wage below the reservation wage w_r, total employment solves

$$\sum_j N_j = \frac{\sum_j \psi_j}{\delta + \sum_j \psi_j}. \tag{5}$$

We can then use the usual flow balance relating posted and paid wages to solve for the distribution of wages in the economy. The outflow rate of workers from employment at wages lower than w is $\delta + s\sum_j \psi_j \left(1 - F_j(w)\right)$ whereas the inflow rate is $\sum_j \psi_j F_j(w)$ from unemployment $u = \frac{\delta}{\delta + \sum_j \psi_j}$. This implies that the fraction of workers employed at a wage below w satisfies

$$\sum_j N_j G_j(w) \quad = \quad \frac{\sum_j \psi_j F_j(w)}{\delta + s\sum_j \psi_j \left(1 - F_j(w)\right)} u = \frac{1}{s} \frac{\delta + s\sum_j \psi_j}{\delta + s\sum_j \psi_j \left(1 - F_j(w)\right)} u - \frac{1}{s} u.$$

The density of workers at a wage w in firm i has the total inflow from unemployment and lower wages given by $\psi_i f_i(w) \left(u + s\sum_j N_j G_j(w)\right)$ and outflow rate to higher wages or unemployment is $\delta + s\sum_j \psi_j \left(1 - F_j(w)\right)$. The density of workers at a wage w therefore

[7]This requires that the initial distribution of workers is such that all firms want to weakly increase employment at all wages. The firm's value function below posits that all workers are initially unemployed which satisfies this requirement. This is innocuous and for notational simplicity. If the firm had some initial workers this would just add a constant in the value function.

satisfies

$$N_i g_i(w) \quad = \quad \psi_i f_i(w) \frac{\delta + s\sum_j \psi_j}{\left(\delta + s\sum_j \psi_j \left(1 - F_j(w)\right)\right)^2} \frac{\delta}{\delta + \sum_j \psi_j}. \tag{6}$$

Total employment at firm i is then

$$N_i \quad = \quad \int_{w_r}^{\infty} \psi_i f_i(w) \frac{\delta + s\sum_j \psi_j}{\left(\delta + s\sum_j \psi_j \left(1 - F_j(w)\right)\right)^2} \frac{\delta}{\delta + \sum_j \psi_j} dw, \tag{7}$$

which we can use to recover the density from (6). Using these equations one can construct the contact rate ψ_i and posted wages $F_i(w)$ that implement a firms optimal employment N_i and paid wages $G_i(w)$.

Taken together, firms choose total employment and a distribution of pay to maximize the present value of profits,

$$\begin{aligned}\Pi_i \quad = \quad & \max_{N_i, G_i(w)} -cN_i + \int_0^{\infty} e^{-rt} \times \\ & \left(p\left(N_i, \mathbf{N}_{-i}\right) x_i N_i^{\alpha} - N_i \int_{w_r}^{\infty} \left(w + c \left(\delta + \sum_{j \neq i} s\psi_j \left(1 - F_j\left(w\right)\right)\right)\right) dG_i(w)\right) dt. \end{aligned} \tag{8}$$

The firm hires its workforce upfront and then sustains it by replacing those it loses to unemployment and higher paying jobs at competitors at cost c. Flow profits are given by gross revenue net of the wage bill and these effective turnover cost.

We assume that, when making these decisions, firms take the reservation wage w_r as given. They also take their competitors' actions—the contact rates ψ_j and offer distributions $F_j(w)$—as given. In that sense, we consider a Nash equilibrium where the agents, despite being large, take each others actions as given.

We emphasize that the output price is endogenous and given by (1) where $\mathbf{N}_{-i} \equiv \left\{N_j\right\}_{j \neq i}$ denotes the employment at firms other than i. Even taking their competitors' actions ψ_j and $F_j(w)$ as given, firms might in principle recognize that their actions can affect $\mathbf{N}_{-i}$ because they can affect employment at their competitors. We assume that is not the case, an assumption we discuss in more detail below.

We can then use (8) to express the firm problem as maximizing steady-state flow profits,

$$r\Pi_i = \max_{N_i, G_i(w)} p\left(N_i, \mathbf{N}_{-i}\right) x_i N_i^{\alpha} - N_i \int_{w_r}^{\infty} \left(w + c \left(r + \delta + \sum_{j \neq i} s\psi_j \left(1 - F_j\left(w\right)\right) \right) \right) dG_i(w). \quad (9)$$

We henceforth refer to the term under the integral as the *user cost of labor*. It consists of the wage paid, along with the turnover cost, which is wage-specific because higher wages might come with a lower quit rate. The turnover cost include, in annuitized form, the cost of hiring the initial workforce.

Discussion of assumptions in firm problem We discuss several key aspects of this firm problem in turn. First, we set up the firm problem in a time-consistent fashion since it takes the instantaneous transition to steady state into account. An alternative formulation has firms only maximize over static long-run profits, akin to the "timeless" equilibria usually studied in the BM literature. This would lead to higher employment choices but would not take the cost of the initial hires into account, analogous to the "golden rule" savings case. Of course, as $r \to 0$, the two formulations converge.

Second, and relatedly, we use a linear hiring technology. Locating workers is effectively costless to firms, and the cost of adding workers does not depend on how many workers are available. This allows for the "jump" to steady state just discussed. It also reflects that hiring and training, rather than locating workers is the most important cost associated with turnover.

Third, we consider a Nash equilibrium where firms, when choosing their actions, take their competitors' and workers' choices as given. We can microfound this by assuming that firms commit to their actions at time zero and that workers do not observe firms' choices but instead act on their rational expectations about them.

A final issue is that a firm might recognize that, by hiring additional workers, it reduces the workers available to its competitors given their choice of contact rate, which would have a price impact. Our formulation shuts down this consideration. To microfound this myopia, assume that firms commit to their employment, hence output, choices. Should employment fall below target, they hire workers at high cost from an outsourcing firm outside the model.

Welfare

We can now define a notion of welfare. Summing over all workers, unemployed and employed, along with employers, utilitarian flow welfare can be measured as

$$rV = \frac{\eta}{\eta-1}\bar{Q}^{\frac{1}{\eta}} \left(\sum_{j=1}^{M} x_j N_j^{\alpha} \right)^{\frac{\eta-1}{\eta}} + ub - \sum_{i=1}^{M} N_i \int_{w_r}^{\infty} c \left(r + \delta + \sum_{j \neq i} s\psi_j \left(1 - F_j\left(w\right)\right) \right) dG_i(w).$$

This consists of the utility of the output produced and home production of the outside good by the unemployed ub, net of the realized turnover costs. We note that this is independent of worker pay because of quasi-linear preferences and the assumption that firm owners care only about the outside good.

Equilibrium

Definition 1 A stationary equilibrium is, for all firms i, employment N_i, a distribution of wages $G_i(w)$, a contact rate ψ_i, and an offer distribution $F_i(w)$ along with a reservation wage w_r such that

(i) all firms i choose the level employment N_i and a distribution of wages $G_i(w)$ to maximize (9) taking ψ_j, $F_j(w)$, and N_j for $j \neq i$ as given;

(ii) for all firms i, a contact rate ψ_i and an offer distribution $F_i(w)$ that implement the optimal N_i and $G_i(w)$, implied by (6) and (7);

(iii) workers choose a reservation wage w_r that solves (4).

Given the objects in the equilibrium definition, additional objects such as firm and worker values are straightforward to compute.

2.1 Optimal Hiring

To characterize the optimal distribution of wages $G_i(w)$, consider the following deviation at fixed employment. The firm can always deviate to a distribution of wages $(1-\epsilon)G_i(w) + \epsilon 1(w \geq \tilde{w})$ that places some small mass ϵ at some acceptable wage $\tilde{w} \geq w_r$. For this not to be a profitable deviation for any offered wages w and wage $\tilde{w} \geq w_r$ requires that

$$w + c\left(r + \delta + \sum_{j \neq i} s\psi_j \left(1 - F_j(w)\right)\right) \leq \tilde{w} + c\left(r + \delta + \sum_{j \neq i} s\psi_j \left(1 - F_j(\tilde{w})\right)\right)$$

Of course, any offered wage is also an acceptable wage weakly larger than the reservation wage. It follows that this weak inequality has to hold in both directions for any pair of offered wages and thus that the user cost of labor is equated across all wages offered by a firm.

Differentiating (9) with respect to total employment N_i, we obtain the following condition for optimal employment,

$$\begin{aligned} m &\equiv \left(\frac{\sum_j x_j N_j^{\alpha}}{\bar{Q}}\right)^{-1/\eta} \alpha x_i N_i^{\alpha-1}\left(1 - \frac{1}{\eta}\frac{x_i N_i^{\alpha}}{\sum_j x_j N_j^{\alpha}}\right) \\ &= \int_{w_r}^{\infty}\left(w + c\left(r + \delta + \sum_{j \neq i} s\psi_j \left(1 - F_j(w)\right)\right) dG_i(w)\right) \\ &= w + c\left(r + \delta + \sum_{j \neq i} s\psi_j \left(1 - F_j(w)\right)\right) \end{aligned} \tag{10}$$

The first line simply defines the marginal revenue product of labor m. The assumption that firms, when hiring, act myopically with respect to the output produced by their competitors retains the usual form for m. It captures that firms internalize their own impact on the output price. The resulting markup between marginal product and marginal cost increases in the sales share of firm i, $\frac{x_i N_i^{\alpha}}{\sum_j x_j N_j^{\alpha}}$.

The second line gives the optimality condition, naturally stating that the marginal revenue product of labor equals the user cost. The third line applies to any wage actively posted by firm i and uses the fact that the user cost is equated across all wages paid by a firm, as we just established. Alternatively, we can write

$$\frac{m - w}{r + \delta + \sum_{j \neq i} s\psi_j \left(1 - F_j(w)\right)} = c, \tag{11}$$

which simply states that the present value of profits generated by the marginal hire must equal its hiring cost for all wages posted by a firm.

It is instructive to note that the markdown between the marginal revenue product of labor and the wage, $\frac{m}{w}$, is larger than one and endogenous. Furthermore, equation (11) suggests that, as competition for workers from outside competitors as encoded by the last term in the denominator rises, markdowns must rise, too. The flow profits from the match must cover the cost of creating it. If workers churn to competitors faster markdowns must hence rise. This is the opposite of the relation between competition and markdown one might expect from a neoclassical, frictionless perspective, an issue we revisit below.

2.2 Equilibrium with Homogeneous Firms

We now consider the case where there are M homogeneous firms with equal productivity. Appendix A.1 shows that the unique equilibrium is symmetric, so we impose symmetry. Solve (11) for $F(w)$ under symmetry, and use that the lowest wage must equal the reservation wage to obtain that[8]

$$F(w) = \frac{w - w_r}{(M-1)\, s\psi c}. \tag{12}$$

It follows that the equilibrium wage distribution is uniform. A higher wage must pay off in terms of a higher retention rate. Intuitively, a uniform (outside) offer distribution guarantees this in our setting where firms care only about the expected duration of the match. A key difference to the BM model is that there firms additionally care about the pace of hiring, which gives rise to the well-known convex density in that framework.

It is instructive to use (12) in a partial equilibrium fashion to think about the impact of M on posted wages. To do so, use that $F(w_u) = 1$, where w_u denotes the highest posted wage. Then, the rate at which workers receive wage offers above w is given by $Ms\psi\,(1 - F(w)) = \frac{M}{M-1}\frac{w_u - w}{c}$, while, likewise following from (12) evaluated at the highest wage w_u,

$$w_u = w_r + c\left(1 - \frac{1}{M}\right) s\psi M. \tag{13}$$

These equations can be used to show that, for a given reservation wage and given offer arrival rate $M\psi$, fewer firms means lower wages, with a full monopsonist naturally offering only the reservation wage.

The intuition is as follows. In this model, an additional dollar of pay is compensated for by a higher retention probability. Now, suppose that the number of distinct employers

[8]We also use that, with symmetry, there can be no mass point in the wage offer distribution, as a firm would then prefer to offer a wage just above the mass point.

falls and that firms still post the same density of wages as before. In this case, lowering the wage would no longer come with the same increase in the quit rate because there are fewer outside jobs on any given wage interval. Hence, all firms reduce wages, the distribution of posted wages steepens, and the highest wage falls. An increase in concentration thus lowers outside competition and leads to falling wages.

Of course, this partial-equilibrium argument ignores that both the reservation wage and total employment are endogenous and respond to a change in the number of firms M. To endogenize these, we need to solve for the equilibrium offer rate ψ. Total employment has to satisfy the usual flow balance; hence, $\sum_i N_i = \frac{M\psi}{M\psi+\delta}$. This has to equal aggregate labor demand, which can be derived by combining the pricing equation (1) with optimal labor demand at the firm level (10) evaluated at w_r and using the fact that each firm's sales share is just $\frac{1}{M}$. This gives

$$\frac{M\psi}{M\psi+\delta} = \left(\left(\frac{w_r + (r+\delta+(M-1)s\psi)c}{\alpha\left(1-\frac{1}{\eta}\frac{1}{M}\right)} \right)^{-\eta} \bar{Q} M^{(\eta-1)(1-\alpha)} x^{\eta-1} \right)^{\frac{1}{(1-\alpha)(\eta-1)+1}}. \tag{14}$$

To interpret labor demand on the right-hand side, note that increasing the number of firms increases competition in both the labor market and the product market. Granularity in the product market is captured by the term $\frac{1}{\eta}\frac{1}{M}$ in the denominator. An increase in M results in lower sales shares, which imply, all else equal, more elastic demand at the firm level and therefore lower markups. Large firms suppress output and employment because they have the product market in mind. The strength of this effect naturally depends on the demand elasticity η.

Finally, the derivations in Appendix A.2 show that the reservation wage simplifies to

$$w_r = b + (1-s)c\left(1-\frac{1}{M}\right)\left(M\psi - \frac{r+\delta}{s}\log\left(\frac{r+\delta+sM\psi}{r+\delta}\right)\right) \tag{15}$$

Jointly, the reservation wage (15), the labor market clearing condition (14), and the wage offer distribution (12) fully characterize the equilibrium and can be solved for the unknowns $F(w)$, w_r, and ψ.

2.3 Impact of Market Structure

We can then formally study the impact of market structure on firms and workers. Specifically, we consider what happens when the number of firms M rises.[9] We focus on two limiting cases. The first one fixes total employment and output; that is, it shuts down any demand response to a rise in cost. The second one makes demand perfectly elastic $\eta \to \infty$.[10] We think of the first case as a good description of settings such as the one in Prager and Schmitt (2021), who study hospital mergers.

PROPOSITION 1. Suppose that the number of firms rises and aggregate employment is constant. Then, the reservation wage, mean wage, and highest wage all rise, while profits fall.

Proof. See Appendix A.2. □

Thus, with inelastic product demand such that employment remains unchanged, workers are worse off as the number of distinct employers falls. This is directly consistent with Prager and Schmitt (2021), who argue that hospital mergers that result in a large increase in employment concentration lead to sizable wage losses without any significant employment losses.

The intuition from the previous subsection carries over. The density of wages on any given interval rises with a decrease in M because there is less competitive pressure from outside employers. This direct effect leads workers to lower their reservation wage, which further decreases mean wages and reduces profits. By construction of the case covered in the previous proposition there is no aggregate employment response to the change in market structure. We next turn to the setting where product demand is perfectly elastic.

When $\eta \to \infty$, the reservation wage and highest wage are still given by (15) and (13), and wage offers are uniform in between as before. The only difference is that the contact rate $M\psi$ is now endogenous according to equation (14). We can then again ask what happens as M increases.

[9] For all our results we shut down any purely mechanical productivity gains due to decreasing returns by keeping $x^{\frac{1}{1-\alpha}} M$ constant.

[10] These are the natural limits in our setting where firms are strategic about the product market.

PROPOSITION 2. Suppose the number of firms rises. When industry demand is elastic, $\eta \to \infty$, the highest wage increases, while profits fall. Unemployment increases. The response of mean wages and the reservation wage is ambiguous.

Proof. See Appendix A.3. $\square$

To understand this result, note that the same forces we have previously discussed are at play. Now, however, firms respond to the falling profitability due to higher turnover cost by seeking a lower level of employment, hence employment falls. The highest wage still increases, but once we let quantities adjust endogenously, the response of mean wages is ambiguous. The important role of the demand elasticity for the transmission of shocks and policies to wages and employment will also appear in our quantitative exercises.

Discussion: Quit Elasticity and Relation to "Classical" Monopsony

While we apply this framework to the study of noncompetes, it carries some important lessons for the measurement of labor market power. The applied literature on labor market power frequently uses the quit elasticity with respect to the wage or related objects as a reduced-form measure of the competitiveness of a labor market (Manning, 2003; Dube et al., 2020; Autor et al., 2023). A higher elasticity suggests more competition.

The elasticity of the quit rate with respect to wages is given by $\frac{d\log(1-F(w))}{d\log w} = \frac{w}{w_u - w}$, following directly from (12). In contrast to what most of the "classical" monopsony literature assumes, this endogenous object is not constant and is instead increasing in w and decreasing in w_u. The latter is shown to increase when there are more firms in the market (Propositions 1 and 2). It follows that, as the market becomes more competitive, the quit elasticity evaluated at a given wage is actually declining. To see why, note that, no matter what M is, an additional dollar of pay needs to come with the same reduction of the quit rate in equilibrium. Since higher M results in a higher level of the quit rate this implies a lower elasticity.

More broadly, in "dynamic monopsony" settings in the search tradition, the labor supply elasticity at the firm level is finite and can be measured. However, in contrast to the neoclassical setting, the elasticity is not a primitive and instead endogeneous with no monotone mapping to allocative efficiency, underemployment, worker well-being, or rent extraction by employers. As we will see below, a shock that drives up employment and wages often reduces the measured labor supply elasticity.

2.4 Noncompete agreements

In this section, we show how the model can straightforwardly be used to theoretically analyze the general equilibrium impact of noncompetes between workers and firms. We revisit noncompetes in section 3.4 for a quantitative assessment.

We assume that k firms have access to a legal technology that allows them to implement and enforce noncompete contracts. We model noncompetes as part of the take-it-or-leave-it offer that is posted by the firm. Instead of just posting a wage, the firm posts a contract that stipulates a wage and a covenant that prohibits the worker from transitioning directly to another firm.[11] In this sense, the noncompete is a voluntary agreement between worker and firm.

From the worker's perspective, signing a noncompete eliminates the option to search for alternative employment opportunities. Consider the wage w_c that makes the worker indifferent between unemployment and working under a noncompete. That wage offers the same value as w_r, the lowest acceptable wage without a noncompete, but will be above it in nominal terms. The reason is that it includes a compensating differential for the foregone option value of on-the-job search. It will also be below the highest wage w_u since it offers only the reservation value.

From the firm's perspective, workers under a noncompete are shielded from outside competition. Consequently, firms with access to noncompetes offer only the lowest acceptable wage, w_c. The user cost of labor is therefore lower for a firm with noncompetes since it allows the firm to pay a lower wage yet have the same turnover as a firm that posts the highest wage w_u. Any firm that can do so will therefore adopt noncompete contracts since these allow the firm to avoid the costs associated with turnover.[12]

The presence of noncompetes then has two direct effects. First, jobs start piling up at the bottom of the job ladder, and there is hence less competition along the interior of the

[11] It is straightforward to limit the scope of these agreements to restrict only transitions to a subset of firms.

[12] This raises the question why not all firms adopt noncompete agreements. Arguably, the required legal resources to set up and enforce such agreements are costly and so not all firms opt in. The choice of contract could straightforwardly be modeled as an upfront investment decision. We also note that, when the hiring costs are heterogeneous, then low-cost firms might not want to adopt noncompetes.

job ladder. As a result, the wage distribution among the remaining employers becomes more compressed to make them indifferent across these wages. These spillovers to firms without noncompetes are documented in a growing empirical literature (Starr et al., 2019; Johnson et al., 2020).

Denoting with ψ the rate at which jobs are offered by the $M-k$ firms without noncompetes, the new wage offer distribution is given by

$$F(w) = \frac{w - w_r}{c(M-k-1)s\psi}. \tag{16}$$

The second effect of noncompete contracts is a decline in the reservation wage. This decline follows directly from the above argument. The firms with noncompetes all offer the lowest possible values, while the remaining firms likewise reduce wages; hence, the lowest acceptable wage falls. This of course further lowers wages according to (16), feeding back into reservation wages and so forth. These factors form the response when ψ is fixed.

We can again analytically characterize reservation wage and highest wage,

$$\begin{aligned} w_r &= b + (1-s)(M-k-1)\psi c - \frac{M-k-1}{M-k}\frac{1-s}{s}(r+\delta)\log\left(\frac{r+\delta+s(M-k)\psi}{r+\delta}\right)c, \\ w_u &= w_r + (M-k-1)s\psi c, \end{aligned} \tag{17}$$

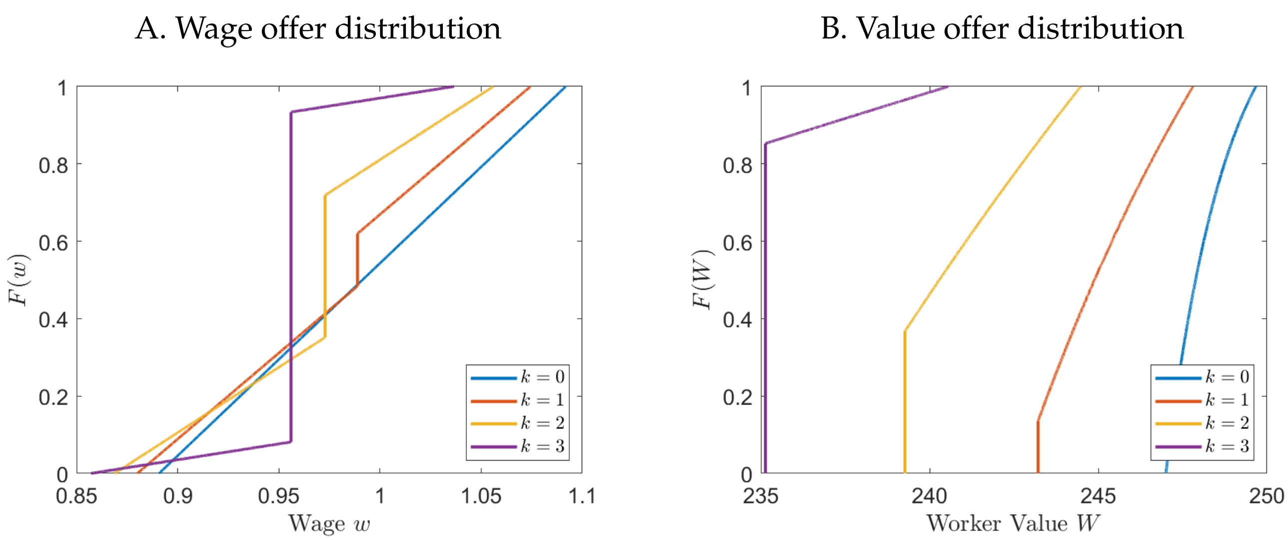
which, given ψ, are both decreasing in k.

To further illustrate this, we plot the equilibrium distribution of posted wages and values in Figure 1. The left panel shows that firms with noncompetes post a mass of wages w_c. These wages are above the reservation wage, reflecting the aforementioned compensating differential. In terms of values, however, these jobs all offer the value of unemployment and are positioned at the bottom rung of the job ladder, as can be seen directly from the right panel.

It follows that, as more and more firms have access to noncompetes, the wage and value distributions shift to the left and conditions for workers deteriorate. The reason is the associated decline in competition for workers and the drop in the reservation wage.

To illustrate these forces, the figure largely shuts down employment effects by picking a low value for the demand elasticity η. In general, however, there is an offsetting procompetitive effect as k increases. The reason is that noncompetes reduce turnover cost and hence increase desired employment, driving up wages. The overall effect on workers

Figure 1: Impact of Noncompetes



Notes: The left figure plots the wage offer distribution as increasingly many firms k have access to noncompetes. The right figure plots the corresponding distribution of values. The parameters correspond to the baseline calibration in Table 4 with $M = 10$ employers.

hence depends on the elasticity of demand in the product market, an issue that we revisit in the quantitative section.

Misallocation We next show that noncompetes introduce misallocation of workers to firms. To see why, consider the user cost of labor at regular employers without noncompetes. Since these firms are indifferent across all wages including the highest one, it is given by $w_u + (r + \delta)c$. For those employers with noncompetes, it is given by $w_c + (r + \delta)c$. We have already shown that $w_c < w_u$. The marginal revenue product of labor under optimal hiring is equated to the user cost of labor. It immediately follows that firms with noncompetes are larger than regular firms, despite operating the same decreasing-returns technology.

This highlights an allocative downside of noncompetes when they are available to only a fraction of employers. These firms face lower turnover cost and hence are inefficiently large relative to their competitors. Noncompetes thus lead to misallocation of labor.

Banning noncompete agreements What happens when noncompetes are banned? We summarize our findings in the following proposition:

PROPOSITION 3. If noncompetes are banned, the highest wage increases, whereas total employment and output fall.

Proof. See Appendix A.4. □

While the highest wage unambiguously rises due to the increase in competition, the effect on mean and reservation wages is ambiguous. The reason is that turnover costs rise, which depresses demand for workers, reducing employment. This has a negative effect on wages and can undo the wage gains that come directly from the rise in competition for workers. In section 3.4, we find a positive impact on mean wages across all the cases we consider.

As follows from the proposition, employment and output unambiguously fall. That is, the rise in turnover cost always outweighs the positive effects from a decline in misallocation. We note that this is in line with evidence in Lipsitz and Starr (2022) who find negative employment effects from the 2008 ban on noncompetes for low-skilled workers in Oregon, a study we revisit below.

The second part of the proposition reflects that rent seeking along the job ladder—while gainful for workers— creates wasteful turnover cost. Additionally, in our framework, output is inefficiently low because of product market power. This gets worse when costs rise due to additional turnover.

That said, in the quantitative section, we show that the output and efficiency losses of a US ban on noncompetes are small in comparison with the wage gains to workers. We also caution that our framework omits additional forces that may further dampen or even reverse the efficiency results. For instance, a hiring externality might arise when search is costly and noncompetes reduce the vacancy filling rate. Additionally, on-the-job human capital investment and the allocation of workers to firms may suffer in the presence of noncompetes. We therefore focus on the wage effects of banning noncompetes. We explore an extension that makes the social cost of turnover smaller than the private cost in section 2.5.

Diamond Paradox Diamond (1971) famously argues that, in an equilibrium model of wage posting, no firm should post any wage above the reservation wage, which will hence equal the flow value of unemployment. The wage posting model introduced by Burdett and Mortensen (1998) overcomes this by incorporating competition for workers along the job ladder. What undoes the Diamond equilibrium is a deviation argument that encapsulates job ladder competition. A firm offering marginally higher pay can have discretely lower turnover cost. This ultimately undoes all mass in the wage offer distribution and shifts it outward, so competition among employers leads to gains for workers. The following proposition shows that noncompetes can unravel this.

PROPOSITION 4. If at least all but one firms have noncompetes, then $w_r = w_u = w_c = b$.

Proof. Follows directly from (17) and the noncompete wage $w_c = b + \frac{w_r - b}{1-s}$. □

We have shown in Figure 1 that the introduction of noncompetes can undo competition and reduce wages and worker values, in particular when the demand elasticity is low. Interestingly, Proposition 4 states that, when noncompetes become widespread, the Diamond paradox reappears and wages collapse. That is, when almost all employers can evade competition, the wage distribution becomes degenerate at the flow value of unemployment. Why does the usual deviation argument not undo such an equilibrium? Because workers under a noncompete cannot be poached by definition.

While this, of course, is a stylized result, it demonstrates that noncompetes, if broadly available, are a very powerful tool of rent extraction for employers.

Discussion of assumptions An important aspect of noncompetes, in particular in high-skilled labor markets, is the protection of trade secrets. We focus here on rent extraction and turnover cost, along with misallocation. Starr et al. (2021) report that 14.3% of US workers without a college degree are under a noncompete while 13.3% of workers with annual earnings $< \$40,000$ are covered by one. It seems hard to argue that these millions of low-skilled, low-pay workers have access to important trade secrets, so we omit these considerations here.[13]

[13]Our setting is one of wage posting, which we argue is natural given the emphasis on the low-skilled sector. For a framework that models noncompetes in a sequential auction bargaining setting, see Shi (2023).

Next, it is frequently argued that noncompetes might increase firms' investment in their workers. As discussed above, the hiring cost c includes investment in firm-specific human capital. As already discussed, firms respond to the rising turnover costs that result from a ban on noncompetes. The model hence captures the notion that firms are less willing to invest in their workers in the absence of noncompetes, albeit only along the extensive margin.

Finally, we assume that k firms have access to noncompete contracts that fully restrict the mobility of the worker. One can alternatively restrict the mobility to a fraction of all outside employers. If all firms have access to noncompetes that restrict the direct movement of workers to k other firms, then the equations above for the offer distribution, reservation and highest wage remain exactly the same. Similar arguments apply when noncompetes are imperfectly enforceable which will limit their impact.

2.5 Model Extensions

This section considers several extensions of the model. We show how to extend the model to account for firm heterogeneity to make it amenable to empirical settings where heterogeneity is key, such as merger analysis. We first introduce heterogeneity in productivity only and then turn to the case where firms differ in both productivity and hiring costs. We then show how to introduce convex hiring costs. The final part shows how to adjust the model so as to reduce the social cost of worker turnover. We relegate most details and formal derivations to Appendices.

Heterogeneous Productivity

There are M firms, which differ in terms of productivity x_i with $x_1 > x_2 > .. > x_M$. Firms again choose a distribution of posted wages. In addition, we allow firms to choose a wage-specific contact rate.

In Online Appendix B, we construct the following equilibrium. The support of the wage distribution can be broken into $M-1$ non-overlapping intervals spanning from the reservation wage to the highest wage. Firm M posts only on the highest interval, firm $M-1$ posts on both the highest and second-highest intervals, and so on. Firm 1 posts on all intervals and, in addition, posts a mass of jobs at the reservation wage.

In addition, posted wages on any interval are uniformly distributed. All firms posting wages on a given interval pick the same, interval-specific contact rate. As a result, total employment on that interval is the same across all firms.

What explains those equilibrium features? First, the marginal revenue product is equated across firms. The reason, as before, is that the user cost must optimally be equated within firms and equal to the marginal revenue product and, since all firms post the highest wage, the user cost must be equated across firms.

Smaller firms account for a smaller fraction of jobs, yet these equilibrium features guarantee that their workers face exactly the same amount of outside competition from higher-paying jobs. For any given wage, workers at all firms receive dominating outside offers at the same frequency. This guarantees that all firms face the same user cost across all the wages they post, yet highly productive firms are larger.

This logic rules out an equilibrium where the distribution of posted wages is identical across firms. In this case, smaller firms would face larger turnover cost because of larger outside competition. It also rules out cases where small firms primarily recruit at the bottom of the wage distribution.

We note the resulting size wage discount, the opposite of the empirical picture. The reason is that small firms otherwise face excessively high turnover costs. We therefore turn to an additional extension, heterogeneity in hiring costs.

Heterogeneity in Hiring Costs

If hiring is more extensive at highly productive firms, then the theory can generate a size wage premium. Intuitively, firms with the highest cost are most concerned with worker turnover and hence tend to locate at the top of the job ladder. If the dispersion in costs is large enough, then a positive size–wage relationship arises. In this case, workers will flow toward larger firms with a higher marginal revenue product. Such an assumption seems plausible since a more advanced technology might require more upfront worker training. Blatter et al. (2012) show that, indeed, large, productive firms have sharply higher hiring costs.

We construct an equilibrium with two-dimensional heterogeneity as follows. Again break the support of the wage distribution into $M-1$ intervals. Rank firms according to their hiring cost, with $c_1 > c_2 > .. > c_M$. On the highest wage interval, firm 1 posts uniformly distributed wages as before. However, so does firm 2 and, possibly, firms 3, 4, .. ., with the cutoff depending on their relative costs. These firms no longer pick identical contact rates, however, but instead pick firm-specific ones.

On the second interval, one firm drops out. The cutoff is determined by the desired size of that firm. All but one of the firms posting on the first interval also post on the second. In addition, one or more firms that do not post on the top interval might be added, in order of their hiring cost. This continues until only one firm remains. That firm posts a mass-point at the reservation wage which is at the lower end of the $M-1$th interval.

What gives rise to these features? First, firms that post on the same interval have different contact rates because they trade off outside competition differently. Outside competition must fall by more for firms with lower c to warrant an additional dollar of pay. This puts a restriction on the relative contact rates of firms posting on the same interval because firms with higher cost need to account for more offers. In addition, the relative cost also determines the set of firms that post on any given interval. To see why, an instructive example considers two very high c posting on a given interval. For them to be indifferent across uniformly distributed wages, outside competition must be very low. However, this makes it impossible for a firm that cares little about turnover to also be indifferent across the same wages. As a consequence, that firm locates only on lower-paying intervals.

How, then, can firms with high productivity achieve their desired large scale given these restrictions on the contact rates? They do so by posting wages further down on the job ladder, just as before. This is the force that was already present when firms have heterogeneous productivity. Here, however, what keeps some low c firms additionally out of the high-wage intervals is the indifference requirement within the interval.

Put differently, the hiring cost governs the vertical position of an employer on the job ladder. Productivity, in turn, governs the range of wages it posts in the same way it did before when firms differed only in terms of productivity.

The Online Appendix C presents a simple algorithm that finds the equilibrium in this environment. We put this full-blown model with two-dimensional firm heterogeneity to use in the quantitative exercise in section 3.4.

Convex Adjustment Cost

In our baseline model, firms have two levers to adjust their level of employment: the rate of hiring and the wage they pay these hires. It is worth emphasizing that, in the BM model, this is usually not the case. Instead, firms are endowed with a single job opening and can use only the wage to affect employment (Bilal and Lhuillier, 2021). As such, our model already generalizes the textbook framework.

We assume that the cost of hiring is linear in the number of hires. In the Online Appendix D , we show how to introduce a convex hiring costs into the model. The convexity limits the firms' ability to flexibly use hiring as a tool to manage the size of their workforce employment. Instead, firms rely more heavily on adjusting the wage they pay their hires.

While we relegate the details to Online Appendix D , this is technically simple. Firms equate the marginal revenue product to the user cost of the marginal worker. Our baseline framework is captured as the limit where the cost function is linear while the hiring technology of the standard framework constitutes the other, completely inelastic limit.

Social Cost of Turnover

The baseline model operates under the assumption that the training is fully firm specific and that the cost are lost which makes turnover quite costly from a social viewpoint. In this section, we show how the social cost of turnover can be reduced while preserving all the equilibrium conditions in the paper.

To do so, continue to assume that the firms can train workers at cost c. But assume that there is a training company that can train unemployed workers at a cost $\kappa_u c$ and employed workers at a cost $\kappa_e c$ with $\kappa_i \leq 1$. For each worker that it trains, the training company makes a take-it-or-leave-it offer to the employer extracting all the gains. This offer is in equilibrium equal to c, which means that the equilibrium fully corresponds to the baseline model. The one difference is that the social cost of turnover is reduced, by a factor $1 - \kappa_e$.

As κ_e gets smaller, so does the welfare cost of turnover and if κ_e is negative, job-to-job turnover even raises welfare. This is a parsimonious way to capture that the social cost of turnover might be lower than the cost of turnover to the firm. While we do not implement this exercise quantitatively it would be straightforward to do so. The only thing this would

Table 1: Common Parameters

	Value	Moment	Reference
r	0.004	5% annual discount rate	standard
δ	0.029	6% unemployment rate	standard
s	0.302	3.2% job-to-job rate	Moscarini and Thomsson (2007)
α	0.64	direct estimate	Cooper et al. (2007, 2015)

Notes: Monthly calibration. s is picked so as to yield the target job-to-job rate in the model with $M \to \infty$ symmetric firms.

change is the welfare accounting. With $\kappa_e < 1$ we would then typically find smaller welfare cost of turnover-increasing pro-competitive policies such as a ban on noncompetes.

3 Quantitative Applications

We now turn to the quantitative application of the model. We first describe how we calibrate the framework, then offer two separate model validation exercises, and then turn to our main application, a quantitative exploration of the impact of noncompete agreements.

3.1 Calibration

Our calibration strategy is as follows. We fix a set of parameters externally and hold these constant across applications. The remaining parameters are calibrated in an application-specific setting. We calibrate at monthly frequency.

Parameters common to all applications The discount rate is set to $r = .004$ to match an annual discount rate of 5%. The separation rate is set to $\delta = .029$ to match a rate of job loss of 2.9%. We set the curvature of the production function $\alpha = .64$, in line with the estimates of Cooper et al. (2007) and Cooper et al. (2015). We set the relative search efficiency $s = .302$ so as to obtain a monthly job-to-job transition rate of 3.2% (Moscarini and Thomsson, 2007) in a competitive, thick labor market with $M \to \infty$ symmetric firms. These parameters are the same in all applications and are presented in Table 1.

Application-specific parameters. The remaining parameters are application specific. The number of firms M is externally calibrated, separately for each application. Otherwise we proceed in a standard moment-matching fashion. The target moments jointly determine the parameters, but we list them in a way that heuristically points to the most informative moment for each parameter.

We choose c to target the size of the hiring cost relative to the average wage $\mathrm{E}[w]$. We use a value of 2 months of wages but consider as alternative a much higher value of 5 months when we study a ban on noncompetes.[14] We choose the demand shifter $\bar{Q}$ such that the price level is normalized to 1. We set the flow income in unemployment b such that the mean wage is 1. To pick the vector of firm-level productivities, we note that the vector of employment shares can be inverted to the (normalized) vector of productivities given the other parameters. To pick the level of average productivity, we target a job-finding rate of 45% (Shimer, 2005).

Finally, we pick η to target the employment response to shocks reported in the literature, as discussed in the next two subsections. A high demand elasticity results in large positive employment changes in response to any shock that changes the user cost of labor.

3.2 Model Validation: Mergers

We begin with the natural application to mergers where we can use the hospital merger study of Prager and Schmitt (2021) for model validation. This paper has information on both the change in market structure and the resulting response in terms of wages and employment, and so we can ask whether our model can replicate these results. We also discuss how the model can be used to predict the consequences of a proposed merger given market structure.

One of the headline findings in Prager and Schmitt (2021) is that, for the top quartile of concentration-increasing mergers, wages dropped by $4-7\%$.[15] This result is market-wide in that it covers not only the merging hospitals but all hospitals in the commuting zone. They report that the corresponding markets have, on average, 3.3 hospitals and that the

[14] Recall that c captures all costs associated with turnover, including training costs. These values fall into the range reported by Blatter et al. (2012), who measure the overall cost of hiring in Switzerland.

[15] These numbers cover skilled workers and, separately, nursing and pharmacy workers. For unskilled workers, the authors find no change in wages but argue convincingly that unskilled hospital workers have a much larger labor market, which is largely unaffected by the merger.

Table 2: Merger Analysis

	Data (Prager and Schmitt, 2021)	Model
Initial HHI	4580	4580
Δ HHI	2764	2764
Δu	0	0
$\Delta \log(\mathrm{E}[w])$	$(-0.07, -0.04)$	-0.03
Δ log turnover cost	$---$	-0.2
Δ log profits	$---$	0.06
Δ markup	$---$	0.05
Δ log quit ela.	$---$	0.08

Notes: The table report results for simulations of the top quartile of concentration-increasing hospital mergers as reported by Prager and Schmitt (2021).

Herfindahl–Hirschman index (HHI) rises from $4,580$ to $7,344$.

We can directly simulate such a setting in our model, where we think of the firms in the model as all the hospitals in a commuting zone. We assume that there are three employers initially in the market and that two subsequently merge. We do not observe the distribution of employment shares to inform the distribution of productivity but can instead target the initial level of the HHI and the rise resulting from the merger as reported above. To set the demand elasticity η, we target the (un)employment response to the mergers studied in Prager and Schmitt (2021).[16]

The results are reported in Table 2. The model closely captures the results reported in Prager and Schmitt (2021) for the change in and level of concentration and the change in employment, but this is unsurprising since it is targeted. The wage response, however, is not targeted and is at the lower end of the values reported in Prager and Schmitt (2021).

We note that the model-based wage response becomes much larger as we increase the size of the hiring cost to 5 months of average wages. In this case, wages fall by 10.4%.[17] Given that the setting is the health care sector, hiring costs beyond two monthly wages seem plausible. We hence conclude that the model closely aligns with the empirical merger

[16] Their results are noisy and insignificant, and they argue that "nothing in the estimates suggests reductions in employment", so we set that target to zero.

[17] When turnover costs are high, firms compete more fiercely to evade churn. The wage losses from a reduction in outside competition through a merger are hence larger.

study, which validates the quantitative framework.

We can then use the model to compute additional statistics. Profits rise, as do markups in the output market, and turnover costs drop substantially.

The market becomes more concentrated and arguably less competitive with wages falling, yet the equilibrium quit elasticity increases.[18] The reason is again that the decline in competition compresses wages, leading to more local mass. This mirrors our earlier observation in section 2.2 that a decline in competition might increase the equilibrium quit elasticity and cautions against this metric as a direct measure of the competitiveness of the labor market.

We see the model as a natural starting point to assess the labor market consequences of other mergers for which clean ex post data are not (yet) available. The back-of-the-envelope analysis is extremely simple. All that is needed is local employment shares, which translate into a vector of local productivity. The model can then be used to predict the general-equilibrium impact of a merger in the labor market. To do so, assign the combined firm a weighted average of productivity and recompute the equilibrium given a demand elasticity.

Ours is distinct from a neoclassical approach to merger analysis in the labor market along several dimensions. First, we offer a model of the labor supply curve to a firm and hence do not need to resort to ad hoc and somewhat arbitrary assumptions on the labor supply curve to the new, merged firm. Second, in our setup, labor demand responds to rising profits, a form of entry that is shut down in the neoclassical setting. Third, we generalize to cases where demand is not perfectly elastic, which seems important in settings such as the health sector.

We conclude this section by briefly using this merger setting to illustrate some properties of wages in this environment. Figure 2 illustrates the wage offer distribution (and wage distribution) in the model with heterogeneous productivity pre- and post-merger.

The figure shows the mass point at the reservation wage. With three firms pre-merger, it also shows the two intervals above the mass point, with the least productive firm posting only in the top interval.

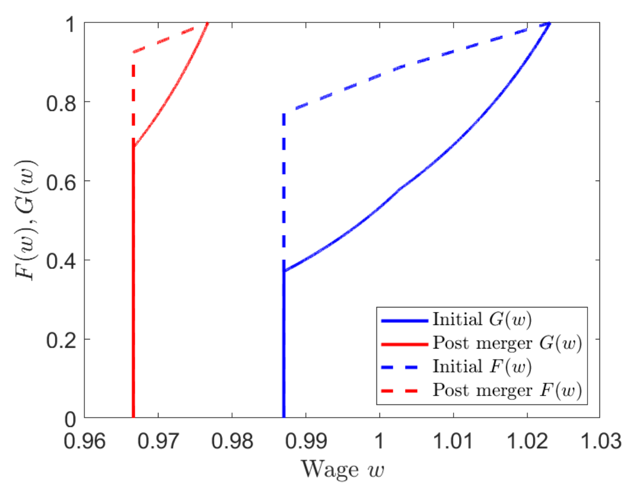
A well-known issue with the BM wage posting model is that its cross-sectional wage

[18] To measure these, we average the local elasticities within each firm and then take an employment-weighted average across firms.

density is increasing. Our framework inherits this issue although the replacement of the hiring cost with a vacancy cost makes it less severe. With homogeneous firms, this leads to a flat instead of increasing wage offer density. With heterogeneous productivity, the wage offer density is decreasing. As follows from section 2.5, when firms also differ in their hiring cost, the model becomes highly flexible and can generate a thin right wage tail. This is particularly true when there are a few firms with high hiring costs, which strikes us as plausible. In either case, we note that we think of the model as one of pay per efficiency unit of labor, and so one would need to remove a worker effect from the empirical distribution to make it comparable with the model.

Finally, as can also be seen in the figure, the variance of log wages is small such that frictional wage dispersion accounts for a small share of overall wage dispersion, in line with Bonhomme et al. (2023). In highly concentrated markets, the distribution is close to degenerate. Nonetheless, pro-competitive policies can have a large impact on the level of wages.

Figure 2: Equilibrium Illustration



Notes: Cumulative distribution function for offered and outstanding wages. The parameters are those underlying the baseline calibration in Table 2. The merger simulation corresponds to the one discussed in this section, mimicking the setting in Prager and Schmitt (2021).

3.3 Model Validation: Noncompetes

We next confront the model with evidence in Lipsitz and Starr (2022) who study the 2008 ban on noncompetes for low-skilled workers in Oregon. This serves as an additional way of validating the model but also provides us with parameter estimates for our counterfactuals in the next section.

We calibrate as discussed in section 3.1. Lipsitz and Starr (2022) report that 14% of employed low-skilled workers were under a noncompete prior to the ban. When picking k and M, we are subject to an integer constraint. Picking $k = 2$ and $M = 15$ gets us closest to the 14% coverage target and HHI levels of 1000–1200, as reported in Jarosch et al. (2023) and Berger et al. (2023a). We assume that the 15 firms are equally productive. Finally, we target an employment reduction following the ban of .8 percentage points[19], which yields an industry demand of $\eta = .19$.[20]

To implement the exercise, we suppose that the labor market is in a steady-state equilibrium. We then compute a counterfactual steady-state equilibrium with $k = 0$ and contrast it with the initial allocation.

The results are presented in Table 4. We have again targeted the employment share and the change in (un)employment. Rising turnover costs following a ban on noncompetes lead to a small decline in labor demand.

The average wage gain of 2.5% in our model closely aligns with the Oregon evidence presented in Lipsitz and Starr (2022). The model also generates a substantial increase in worker mobility as measured by job-to-job transitions, a direct consequence of the ban on mobility restrictions. The Oregon counterpart is comparable but somewhat smaller.

We can use the model to break down the average wage effects into a direct effect deriving from those firms that initially have noncompetes and a spillover effect that arises from the overall increase in competition that affects all firms. In the model, wages increase by 5.7% in the former group. Wages rise by 2% at all other employers, a large spillover effect.

[19]Figure A12 in Lipsitz and Starr (2022) reports increases in unemployment following the ban across all specifications. These are significant and positive, in the range of 1–3 percentage points at short horizons. They remain positive but decline and become insignificant at longer horizons.

[20]This value might appear low. However, labor's cost share is empirically much below one which is the case in the model. In addition, we suspect that in our applications (low-skilled) labor is a complement with other factors of production and so firms cannot substitute in response to a cost shock. For these reasons the empirical demand response to a cost shock is muted and our model picks this up with a low value of η. Finally, the low value is also in line with the city-level estimates of 0.3 in Beaudry et al. (2018).

Table 3: Ban on Noncompetes in Oregon

	Data (Lipsitz and Starr, 2022)	Model
Emp. share non-comp.	0.14	0.148
$\Delta \log(\mathrm{E}[w])$	$0.022 - 0.031$	0.025
Δu	≈ 0.8	0.8
$\Delta \log(\mathrm{jtj})$	$0.12 - 0.18$	0.248
$\Delta \log(w_{nc})$	$---$	0.058
$\Delta \log(w_{rest})$	$---$	0.02

Notes: The results are based on Lipsitz and Starr (2022) where the results for wages are presented in Table 3 and the results for job-to-job transitions in Tables 5 and A9, and the results for the employment rate are based on Figure A12.

Lipsitz and Starr (2022) do not report counterparts for Oregon, but these observations align with the results of Starr et al. (2019) and Johnson et al. (2020), who both highlight substantial spillover effects from noncompete agreements.

Overall, the model aligns closely with the evidence in Lipsitz and Starr (2022).[21]. It is thus a natural laboratory to dig deeper into the economic consequences of a ban on noncompetes, which we do in the next section. We can use the model to report additional statistics that might be hard to measure empirically, and we can consider a wide range of scenarios where clean evidence such as that from Oregon might not (yet) be available.

3.4 Banning Noncompetes

The FTC has proposed a ban on noncompete agreements in the US. While, of course, there might well be exceptions in any future legislation, the proposed policy comes close to a blanket ban, in particular for low-skilled workers. A particular focus of our analysis will be on how the impact of such a policy varies with local labor market conditions.

The same calibration strategy used in section 3.3 yields $M = 10$ and $k = 2$. Given the integer constraints, these values get us closest to our targets for employment concentration and noncompete coverage (1000 (Jarosch et al. (2023) and Berger et al. (2023a)) and 20% (`https://www.ftc.gov/noncompetes`)). If not mentioned otherwise, we assume that firms are symmetric in both hiring cost and productivity. The last two counterfactuals explore

[21] Young (2021) finds that a similar ban in Austria resulted in a smaller change in both mobility and wages.

Table 4: Banning Noncompetes à la FTC

	Baseline	c/E[w]=5	$\eta = 0.5$	$\eta = 5$
Share non-comp.	0.212	0.226	0.224	0.234
$\Delta \log(\mathrm{E}[w])$	0.04	0.05	0.019	0.001
Δu	1.198	1.594	1.592	1.965
$\Delta \log(\text{output})$	-0.008	-0.01	-0.011	-0.013
Δ Utility	-0.009	-0.017	-0.01	-0.01
$\Delta \log(\text{jtj})$	0.354	0.349	0.345	0.335
$\Delta \log(w_{nc})$	0.067	0.118	0.046	0.027
$\Delta \log(w_{rest})$	0.032	0.03	0.011	-0.007

Notes: Counterfactual results for $M = 10$ and $k = 2$ based on recalculating the equilibrium with $k = 0$. Δu reports the percentage-point increase in the unemployment rate. The baseline has $\eta = .19$, as estimated in section 3.3.

the role of heterogeneity.

We import an elasticity of industry demand $\eta = .19$ from the Oregon experiment in the previous subsection. We target a value of the hiring cost of two months of average wages, as before. This constitutes our baseline setting.

We then report results as we vary the demand elasticity η, the hiring cost c, and the number and type of firms with noncompetes k. The quantitative impact of the policy differs quite a bit depending on these parameter values. In practice, a blanket ban on noncompetes would affect many labor markets that differ substantially with respect to these parameters. Thus, instead of offering a single headline number, we present results for a wide range of parameter values and then summarize our key takeaways.

Table 4 reports the results. Under the baseline, average wages increase by 3.9%. This demonstrates that noncompetes powerfully shift rents from workers to firms by reducing competition. The flipside of this is a rise in unemployment slightly above one percentage point, which results from the increase in churn and the rise in turnover costs. This leads to a decrease in output of slightly less than 1%. This tension between redistribution and efficiency shows across all our counterfactuals, suggesting that the misallocation channel is typically dominated by the rise in turnover costs.

We next report the change in utility or welfare that we compute by making use of equation (2). In particular, we compute the consumption-equivalent welfare change across the two allocations and then normalize that by total consumption in the baseline. Total welfare falls, despite a reduction in misallocation, for the reasons already discussed in the theoretical section. First, output falls, yet initial output was already inefficiently low because of product market power. Second, there is additional inefficient worker turnover. We point to the final part of section 2.5 which offers a simple adjustment to the welfare calculations for cases where the private cost of worker turnover exceed the social ones.

We reiterate that, whether workers gain or lose in utility terms hence largely depends on how much they are harmed by these higher prices. In our setting, workers are the sole consumers of an identical good, which makes them bear the full burden of the rise in costs. In settings where the rise in costs are partially borne by other consumers, workers whose noncompetes are outlawed might well also benefit in utility terms.

The increase in churn is strong, with the job-to-job transition rate rising by over 35%. We again break down the wage response into a direct response and a spillover response. The spillovers are even larger here because a larger fraction of the market is initially covered by a noncompete.

Role of Local Labor Market Features

We begin the heterogeneity analysis by investigating the role of the training cost by increasing its value from 2 to 5 months of training. This results in larger effects all around; in particular, the wage gains from a ban are even larger, while employment, output, and welfare fall even more. The reason, as before, is that training costs are effectively the frictions that determine the size of rents. When these are large, noncompetes redistribute even more. This carries a first important lesson on the heterogeneity in the impact of a ban. We should expect the response to be particularly forceful in a setting where rents are large.

We now turn to the role of product demand by raising the value of the demand elasticity η to .5 and then to 5. The corresponding rows show that procompetitive policies increase wages on a large scale only when demand is inelastic. Otherwise, these wage gains evaporate since the rising turnover costs cannot be passed into prices and hence result in lower labor demand, undoing the procompetitive effects of the policy on wages. The demand elasticity does not affect the welfare losses because these depend solely on the increase in churn and the increase in the product market wedge relative to the reduction in misallocation.

We continue the heterogeneity analysis in Table 5. We ask how the impact of a ban

Table 5: Banning Noncompetes: Heterogeneity

	k=5	k=c/E[w]=5	High	Low
Share non-comp.	0.513	0.528	0.186	0.207
$\Delta \log(\mathrm{E}[w])$	0.113	0.168	0.069	0.011
Δu	3.208	4.602	0.912	0.933
$\Delta \log(\text{output})$	-0.022	-0.032	-0.007	-0.003
Δ Utility	-0.022	-0.039	-0.008	-0.004
$\Delta \log(\text{jtj})$	1.066	1.018	0.261	0.297
$\Delta \log(w_{nc})$	0.126	0.198	0.092	-0.028
$\Delta \log(w_{rest})$	0.1	0.136	0.064	0.019

Counterfactual results based on recalculating the equilibrium with $k = 0$. Δu reports the percentage-point increase in the unemployment rate. The last column reports results based on the full model with two-dimensional firm heterogeneity, where the least productive firms use noncompetes. The results for the most productive firms with noncompetes appear in the second-to-last column.

changes when a larger share of workers is under a noncompete. We suspect that many labor markets have no noncompetes while they are ubiquitous in others, so we report results when the initial coverage is 50%. The results confirm the theoretical observations regarding the Diamond Paradox since the impact of a ban is far larger, with wages rising by over 11 log points. Worker churn rises sharply, and so we correspondingly see larger output, employment, and welfare reductions. The spillover effects are very strong, with wages rising by almost the same amount in firms that have no noncompetes to begin with.

To illustrate that a ban on noncompetes can sharply raise wages in settings that arguably describe many labor markets well, we next consider the setting with high training costs and widespread use of noncompetes (column 2). In this case, wages rise by almost 17 log points.

Finally, the last two columns focus on firm heterogeneity. We consider the extended version of the model, presented in section 2.5, in which firms differ in both their hiring costs and level of productivity. We assume that productivity is log-normally distributed, and we pick the dispersion such that the 90–10 log difference is 0.651, in line with Table 1 of Syverson (2004). We assume that the productivities of the M firms are equi-spaced.

Turning to the hiring cost, Table 1 of Blatter et al. (2012) suggests that these broadly scale in the following log-linear fashion with firm size, $\log(c) = \tilde{c}_0 + 0.13 \log(n)$. When market share and hiring cost are small, employment in the model satisfies $\log(n) \propto \frac{\log(x)}{1-\alpha}$. This results in the following relation between firm productivity and hiring costs:

$$\log(c(x)) = c_0 + \frac{0.13}{1-\alpha}\log(x). \tag{18}$$

We use this formulation for the hiring cost and consider two cases: a case in which the two most productive firms have access to noncompete contracts (labeled "High") and the contrasting case in which the two least productive firms have access to noncompetes ("Low").[22] In both cases, the initial employment share with noncompete contracts is approximately 20%.

The results show that when the most productive firms have access to noncompetes, wages increase more than under the baseline model. In contrast, noncompetes suppress wages less when they are used by the least productive employers. That is, when noncompetes are used by the firms with the most desirable jobs, as we suspect they are, then they are even more extractive. Interestingly, the welfare and output losses are particularly small relative to the wage gains in this case.

In contrast, when noncompetes are initially used by the least productive firms, wage gains are small, in particular relative to the welfare loss. We emphasize that wages at these firms might even fall because they no longer need to pay the compensating differential. We also note that this describes a scenario is which the market becomes more competitive and average wages increase yet the treated firms lower their wages. This cautions against simple difference-in-difference designs comparing firms that use noncompetes with those that do not following a ban and highlights the role of general-equilibrium effects.

In summary, noncompetes decrease competition and lower wages, as the theoretical section on noncompetes suggested. The wage benefits to workers of a ban depend on the demand elasticity because a ban raises turnover costs. This reduces labor demand, potentially undoing the procompetitive effects of a ban on wages. The wage gains are otherwise large when a large fraction of the market is covered by noncompetes, when hiring costs are large, and when the most productive firms use noncompetes.

Banning noncompetes is hard to justify on pure efficiency grounds. The reason is that

[22] It is possible that firms with low hiring costs would not want to adopt noncompetes since the required compensating differential might exceed the gains in terms of turnover cost. This is easiest to see for a firm with $c_i \to 0$. In all of our exercises, all firms rationally adopt noncompetes.

aggregate output and employment usually contract, albeit usually in an order of less than 1%. This reflects rising turnover costs, which are partially offset by a reduction in misallocation. We again caution that our analysis omits several forces that might reduce or even reverse the output and welfare losses. In either case, the size of the wage gains relative to these losses suggests that a ban on noncompetes is an effective way of redistributing from firms to workers.

4 Conclusion

We generalize the canonical dynamic monopsony model of Burdett and Mortensen (1998) and make it amenable to the structural analysis of anticompetitive practices by firms in the labor market. We show formally how market structure affects wages and employment and that noncompetes have the potential to sharply reduce wages by eliminating competition among the job ladder.

Our quantitative model is used to show that a ban on noncompetes would increase wages by 2–10%, depending on local conditions. It would typically lead to a small decline in output and welfare due to higher turnover cost.

We demonstrate that the model can fruitfully be used for merger analysis since it has predictions for the response of employment, wages, and output that are grounded in a theory of the firm level labor supply curve. It might also be used to assess the consequences of other anticompetitive practices in the labor market, such as no-poaching agreements like the ones studied in Krueger and Ashenfelter (2022) or cases where firms collude on wage limits.

References

Autor, David, Arindrajit Dube, and Annie McGrew, "The Unexpected Compression: Competition at Work in the Low Wage Labor Market," Working Paper 31010, National Bureau of Economic Research March 2023.

Bagga, Sadhika, "Firm market power, worker mobility, and wages in the US labor market," *Journal of Labor Economics*, 2023, *41* (S1), S205–S256.

Beaudry, Paul, David A Green, and Ben M Sand, "In search of labor demand," *American Economic Review*, 2018, *108* (9), 2714–2757.

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

Berger, David W, Kyle F Herkenhoff, and Simon Mongey, "Anatomy of monopsony: Search frictions, amenities, concentration," *mimeo*, 2023.

__ , Thomas Hasenzagl, Kyle F Herkenhoff, Simon Mongey, and Eric A Posner, "Merger Guidelines for the Labor Market," *mimeo*, 2023.

Bilal, Adrien and Hugo Lhuillier, "Outsourcing, inequality and aggregate output," *mimeo*, 2021.

__ , Niklas Engbom, Simon Mongey, and Giovanni L Violante, "Firm and worker dynamics in a frictional labor market," *Econometrica*, 2022, *90* (4), 1425–1462.

Blatter, Marc, Samuel Muehlemann, and Samuel Schenker, "The costs of hiring skilled workers," *European Economic Review*, 2012, *56* (1), 20–35.

Bonhomme, Stéphane, Kerstin Holzheu, Thibaut Lamadon, Elena Manresa, Magne Mogstad, and Bradley Setzler, "How much should we trust estimates of firm effects and worker sorting?," *Journal of Labor Economics*, 2023, *41* (2), 291–322.

Burdett, Kenneth and Dale Mortensen, "Wage Differentials, Employer Size, and Unemployment," *International Economic Review*, 1998, *39* (2), 257–73.

Cahuc, Pierre, Fabien Postel-Vinay, and Jean-Marc Robin, "Wage Bargaining with On-the-Job Search: Theory and Evidence," *Econometrica*, 03 2006, *74* (2), 323–364.

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

Cooper, Russell, John Haltiwanger, and Jonathan L Willis, "Search frictions: Matching aggregate and establishment observations," *Journal of Monetary Economics*, 2007, *54*, 56–78.

__ , __ , and __ , "Dynamics of labor demand: Evidence from plant-level observations and aggregate implications," *Research in Economics*, 2015, *69* (1), 37–50.

Diamond, Peter A, "A model of price adjustment," *Journal of economic theory*, 1971, *3* (2), 156–168.

Dube, Arindrajit, Jeff Jacobs, Suresh Naidu, and Siddharth Suri, "Monopsony in online labor markets," *American Economic Review: Insights*, 2020, *2* (1), 33–46.

— **, Laura Giuliano, and Jonathan Leonard**, "Fairness and frictions: The impact of unequal raises on quit behavior," *American Economic Review*, 2019, *109* (2), 620–663.

Elsby, Michael WL and Axel Gottfries, "Firm dynamics, on-the-job search, and labor market fluctuations," *The Review of Economic Studies*, 2022, *89* (3), 1370–1419.

Gouin-Bonenfant, Émilien, "Productivity Dispersion, Between-Firm Competition, and the Labor Share," *Econometrica*, 2022, *90* (6), 2755–2793.

Jarosch, Gregor, Jan Sebastian Nimczik, and Isaac Sorkin, "Granular search, market structure, and wages," *mimeo*, 2023.

Johnson, Matthew S, Kurt Lavetti, and Michael Lipsitz, "The labor market effects of legal restrictions on worker mobility," *mimeo*, 2020.

Jungerman, William, "Dynamic Monopsony and Human Capital," *mimeo*, 2023.

Krueger, Alan B and Orley Ashenfelter, "Theory and evidence on employer collusion in the franchise sector," *Journal of Human Resources*, 2022, *57* (S), S324–S348.

Lentz, Rasmus and Dale T Mortensen, "Labor market friction, firm heterogeneity, and aggregate employment and productivity," *mimeo*, 2012.

Lipsitz, Michael and Evan Starr, "Low-wage workers and the enforceability of noncompete agreements," *Management Science*, 2022, *68* (1), 143–170.

Manning, Alan, *Monopsony in motion: Imperfect competition in labor markets*, Princeton University Press, 2003.

— , "Chapter 11 - Imperfect Competition in the Labor Market," in David Card and Orley Ashenfelter, eds., *David Card and Orley Ashenfelter, eds.*, Vol. 4 of *Handbook of Labor Economics*, Elsevier, 2011, pp. 973 – 1041.

— , "Monopsony in Labor Markets: A Review," *ILR Review*, 2021, *74* (1), 3–26.

Moscarini, Giuseppe and Kaj Thomsson, "Occupational and Job Mobility in the US," *The Scandinavian Journal of Economics*, 2007, *109* (4), 807–836.

Potter, Tristan, Bart Hobijn, and André Kurmann, "On the Inefficiency of Non-Competes in Low-Wage Labor Markets," *mimeo*, 2022.

Prager, Elena and Matt Schmitt, "Employer consolidation and wages: Evidence from hospitals," *American Economic Review*, 2021, *111* (2), 397–427.

Robinson, Joan, *The economics of imperfect competition*, Macmillan, 1933.

Shi, Liyan, "Optimal regulation of noncompete contracts," *Econometrica*, 2023, *91* (2), 425–

463.

Shimer, Robert, "The cyclical behavior of equilibrium unemployment and vacancies," *American economic review*, 2005, *95* (1), 25–49.

— , "On-the-job search and strategic bargaining," *European Economic Review*, May 2006, *50* (4), 811–830.

Starr, Evan, James J Prescott, and Norman Bishara, "Noncompetes in the US labor force," *Journal of Law and Economics*, 2021.

— **, Justin Frake, and Rajshree Agarwal**, "Mobility constraint externalities," *Organization Science*, 2019, *30* (5), 961–980.

Syverson, Chad, "Product substitutability and productivity dispersion," *Review of Economics and Statistics*, 2004, *86* (2), 534–550.

Young, Samuel, "Noncompete clauses, job mobility, and job quality: Evidence from a low-earning noncompete ban in austria," *mimeo*, 2021.

APPENDIX

A Proofs

A.1 Proof of Equilibrium Characterization

To prove uniqueness, we proceed similarly to the standard BM model. First, no two firms offer wages at a mass point, as this would give either firm an incentive to change those wage offers to just above the mass point to reduce turnover costs. Second, there can be no mass point strictly above the reservation wage. To see this, note first that no other firm will offer a wage just to the left of the mass point, as such a firm would save on turnover costs by paying a wage just above the mass point. The firm offering wages at the mass point could therefore reduce the wage slightly without increasing turnover, which contradicts the assumption that the initial mass point was optimal. Third, there cannot be a gap in the distribution, as a firm paying a wage just above the gap would have unchanged turnover but lower wage costs if those wage offers were reduced to just above the start of the gap.

We have already established that the user cost of labor must be equated across all wages offered by a firm. We next prove that the user cost is the same for all firms. To do so, note that all firms have the option to pay the highest posted wage w_u with user cost $w_u + c\,(r+\delta)$. We next show that all firms pay the highest wage and that the support is connected below the highest wage. Assume that firm i posts wage w' but no wages on some interval (w', w''), with $w'' > w'$. Since there can be no gap in the support, as we already established, (and a positive measure of jobs has to be posted by at least two firms), another firm j offers a measure of jobs in some region (w', w'''), with $w''' > w'$. It follows that turnover in this region above w' falls by more for firm i than for firm j, and hence, the user cost is reduced more for i than j. Since the user cost for j must be constant over this region, the user cost must be lower at w''' for firm i than at w'. Firm i would therefore optimally make higher wage offers, contradicting the initial conjecture. It follows that the support is connected below w_u and all firms post w_u.

A.2 Proof of Proposition 1

Differentiate (3) with respect to w, imposing symmetry. Then, integrate up using the fact that $W\left(w_r\right) = U$ to get that

$$W\left(w\right) - U = \int_{w_r}^{w} \frac{1}{r+\delta+Ms\psi\left(1-F\left(\tilde{w}\right)\right)} d\tilde{w}.$$

Use again that $W\left(w_r\right) = U$ to get that

$$\begin{aligned}
w_r &= b + (1-s)\psi M \int_{w_r}^{\infty} (W(\tilde{w}) - W(w_r)) dF(\tilde{w}) \\
&= b + (1-s)\psi M \int_{w_r}^{\infty} W'(\tilde{w})(1-F(\tilde{w})) d\tilde{w} \\
&= b + (1-s)(M-1)\psi c - \frac{M-1}{M} c \frac{1-s}{s} (r+\delta) \log\left(\frac{r+\delta+sM\psi}{r+\delta}\right) \\
&= b + (1-s)c\left(1-\frac{1}{M}\right)\left(\lambda - \frac{r+\delta}{s}\log\left(\frac{r+\delta+s\lambda}{r+\delta}\right)\right) \qquad (19)
\end{aligned}$$

where the last line uses the definition of the contact rate $\lambda \equiv M\psi$. It follows immediately that the reservation wage is increasing in M. We have already established that $w^u = w_r + c\left(1-\frac{1}{M}\right)s\lambda = b + c\left(1-\frac{1}{M}\right)\left(\lambda - (1-s)\frac{r+\delta}{s}\log\left(\frac{r+\delta+s\lambda}{r+\delta}\right)\right)$, so this carries over to the highest wage. Next, flow balance implies that

$$
\begin{aligned}
G(w) &= \frac{F(w)}{(1+s\lambda/\delta(1-F(w)))}, \\
g(w) &= f(w)\frac{1+s\lambda/\delta}{(1+s\lambda/\delta(1-F(w)))^2}.
\end{aligned}
$$

The average wage solves

$$
\begin{aligned}
E(w) &= \int_{w_r}^{w_u} wf(w)\frac{1+s\lambda/\delta}{(1+s\lambda/\delta(1-F(w)))^2}dw \\
&= b + c\left(1-\frac{1}{M}\right)\times \\
&\quad \left(\lambda-(1-s)\frac{r+\delta}{s}\log\left(\frac{r+\delta+s\lambda}{r+\delta}\right)+\delta\left(1-\frac{\delta+s\lambda}{s\lambda}\log\left(\frac{\delta+s\lambda}{\delta}\right)\right)\right)
\end{aligned}
$$

The large term in brackets in the third line is positive, and we have already established that w_r and $(w_u - w_r)$ are increasing in M, so mean wages are increasing in M. Since employment and total output are unchanged by assumption, this immediately implies that average profits fall.

A.3 Proof of Proposition 2

Use the expression for the lowest wage in (19) and that $M(\alpha x)^{\frac{1}{1-\alpha}}$ equals some constant κ for all M to shut down mechanical productivity effects from moving around M. Substitute both into the labor market clearing condition (14) to get that λ solves

$$
\frac{\lambda}{\lambda+\delta}-\kappa\left(\frac{1}{b+(r+\delta)c+c\frac{M-1}{M}\left(\lambda-\frac{1-s}{s}(r+\delta)\log\left(\frac{r+\delta+s\lambda}{r+\delta}\right)\right)}\right)^{\frac{1}{1-\alpha}}=0. \tag{20}
$$

Differentiating the left-hand side with respect to λ gives

$$
\begin{aligned}
&\frac{\delta}{(\delta+\lambda)^2}+\left(1-(1-s)\frac{r+\delta}{r+\delta+s\lambda}\right)\frac{1}{1-\alpha}c\frac{M-1}{M}\kappa\times \\
&\quad\left(\frac{1}{b+(r+\delta)c+c\frac{M-1}{M}\left(\lambda-\frac{1-s}{s}(r+\delta)\log\left(\frac{r+\delta+s\lambda}{r+\delta}\right)\right)}\right)^{\frac{1}{1-\alpha}}.
\end{aligned}
$$

This is positive since $s > 0$ and total employment $N > 0$ (hence the term in large brackets is positive). Next, the left-hand side of (20) is positive as $\lambda \to \infty$ and negative as $\lambda \to 0$. It follows that an equilibrium value for λ exists and is unique.

Next, note that the left-hand side of (20) is increasing in M. To see this, note that $\lambda - \frac{1-s}{s}(r+\delta)\log\left(\frac{r+\delta+s\lambda}{r+\delta}\right) > 0$ for $\lambda > 0$. It follows from this and the previous paragraph that the equilibrium contact rate λ is declining in M. Hence, unemployment rises.

We have that firm-level profits can be written as $x_i N_i^{\alpha} - N_i\left(\alpha x_i N_i^{\alpha-1}\right)$ because the user cost of labor is equated to the marginal product. It follows that firm-level profits are $\Pi_i = (1-\alpha)\,Y_i$ while economy-wide profits are $\Pi = (1-\alpha)\,Y = (1-\alpha)\left[xM^{1-\alpha}\right]N^{\alpha}$. The term in squared brackets is constant, and so total profits are down.

The user cost of labor at the highest wage is given by $w_u + c(r+\delta)$, which must equal the marginal revenue product. Since the price is exogenous, it follows that the highest wage rises with M. We have examples showing that the effect on the reservation wage and the mean wage is ambiguous.

A.4 Proof of Proposition 3

We prove that the highest wage increases by contradiction. Assume it remained unchanged. Use (17) and substitute for w_r, which then gives an (implicit) expression for the total amount of employment offers by regular firms $(M-k)\psi$ given by,

$$w_u = b + \frac{M-k-1}{M-k}c\left(\left(M-k\right)\psi - \frac{1-s}{s}(r+\delta)\log\left(\frac{r+\delta+s(M-k)\psi}{r+\delta}\right)\right)$$

Since the term inside the brackets increases in $(M-k)\psi$, it follows that $(M-k)\psi$ decreases as k decreases (since $\frac{M-k-1}{M-k}$ is decreasing in k). Therefore, for a given highest wage, total employment at firms without noncompetes falls as k decreases. Thus, total output falls, too, and the marginal revenue product of labor rises since employment is lower and the price is higher. This yields the contradiction since the user cost of labor is unchanged given an unchanged highest wage. The same argument applies to a falling highest wage. It follows that the equilibrium value of ψ and the highest wage must rise.

To prove that output falls, again proceed by contradiction. Suppose that it was unchanged. We established in the main text that firms with noncompetes are larger than regular firms. To keep output the same, it then must be that firms that do not initially use noncompetes must increase employment (since employment after the ban is equalized across firms). It then again follows from the above equation that the highest wage rises. Thus, the user cost of firms that did not initially use noncompetes increases. In turn, their marginal revenue product falls since employment and sales share rise. This immediately yields a contradiction. The same argument applies to rising output. Equilibrium output thus falls.

Finally, the marginal revenue product of labor is equalized after a ban that eliminates all misallocation. In conjunction with falling output, this implies that employment falls, too.

ONLINE APPENDIX

B Heterogeneous Firms – Details

We proceed similar to before in that we boil down the model to characterizing the reservation wage and the contact rate, the latter now being firm specific. All other objects can then be constructed from that.

B.1 Constructing Equilibrium

Posit the following equilibrium. There are M endogenous intervals on $[w_r, w_u]$ with the first interval degenerate at w_r and the remaining intervals nondegenerate on $(w_{i-1}, w_i]$ for $i > 1$. Firm k posts jobs uniformly on all intervals $i = k : M$. For instance, firm M posts only on the highest interval $(w_{M-1}, w_u]$, while firm 1 posts on all intervals. Firm 1 is also the only firm posting a mass of jobs at the reservation wage.

In turn, all wages posted on some given interval attract workers at the common rate ψ_i. That is, firms choose the same contact rates for wages in identical intervals. That rate satisfies, for $i > 1$,

$$\psi_i = \frac{\psi}{i-1} \frac{w_i - w_{i-1}}{w_u - w_r}. \tag{21}$$

As a consequence, a worker with a wage w on the interval $(w_{i-1}, w_i]$ receives job offers from a higher interval at rate $\sum_{j=i+1}^{M} (j-1)\psi_j$, while she receives job offers with a higher wage from the same interval from outside employers at rate $(i-1)\psi_i \frac{w_i - w}{w_i - w_{i-1}}$. Plugging in for ψ_i, the workers are poached at rate

$$\begin{aligned}
(i-1)\frac{\psi}{i-1}\frac{w_i - w_{i-1}}{w_u - w_r}\frac{w_i - w}{w_i - w_{i-1}} + \sum_{j=i+1}^{M}(j-1)\psi_j &= \psi\left(\frac{w_i - w}{w_u - w_r} + \frac{w_u - w_i}{w_u - w_r}\right) \\
&= \psi \frac{w_u - w}{w_u - w_r}.
\end{aligned}$$

This guarantees that the user cost of labor, given by

$$m = w + c\left(r + \delta + s\psi\frac{w_u - w}{w_u - w_r}\right)\Bigg(= w_u + c(r+\delta),$$

is equated across all posted wages.

Contact Rates

How do we solve for the cutoffs? Fix a highest wage w_u. We use the definition of m as the marginal revenue product of workers $p\alpha x_i N_i^{\alpha-1}\left(1-\frac{1}{\eta}\frac{x_i N_i^{\alpha}}{\sum x_j N_j^{\alpha}}\right)$ and that optimal total employment at all firms satisfies the usual first-order condition, $m = w_u + c(r+\delta)$. This gives a system of M equations for each firm i's employment N_i given w_u,

$$\left(\frac{\sum x_j N_j^{\alpha}}{\bar{Q}}\right)^{-1/\eta} \alpha x_i N_i^{\alpha-1}\left(1-\frac{1}{\eta}\frac{x_i N_i^{\alpha}}{\sum x_j N_j^{\alpha}}\right) = w_u + c\,(r+\delta)\,.$$

This can be solved to recover the level of employment at each firm. Using the level of employment at each firm, we can calculate the job offer arrival rates of each firm. The flow balance at firm M (or, equivalently, on the highest wage interval) requires

$$\delta N_M = \psi_M\left(1-\sum_j N_j\right) + s\psi_M\left(\frac{\lambda}{\lambda+\delta} - M N_M\right)$$

The left-hand side is the outflows due to job loss; the right-hand side is the hires from unemployment and employment. To understand the last term in brackets, notice that this is just total employment outside the wage interval where firm M posts (all firms post there, and the interval accounts for employment N_M, i.e., the total employment of firm M, at each firm). Why do the worker flows at that wage interval exactly net out? Because for every worker at firm M who makes contact with an outside offer from the same wage interval at rate $\psi_M(M-1)$, there are M workers at outside firms that contact firm M at rate ψ, which exactly cancels out.

Proceed recursively for the other wage intervals to have that flow balance at interval k requires

$$(N_k - N_{k+1})\left(\delta + s\sum_{j=k+1}^{M}\psi_j j\right) = \psi_k\left(1-\sum_j N_j\right) + s\psi_k\left(\sum_j N_j - \sum_{j=k}^{M} j\left(N_j - N_{j+1}\right)\right)$$

where $N_{M+1} = 0$. The first term on the left-hand side is total employment of firm k on interval k; the rest follows directly from the previous equation. The outflow captured by the left-hand side now contains job-to-job flows to higher-paying wage brackets. The term in large brackets on the right-hand side gives total employment in the wage intervals below. Solve for the contact rate,

$$\psi_k = \frac{\left(\delta + s\sum_{j=k+1}^{M} \psi_j j\right)\left(N_k - N_{k+1}\right)}{1 - \sum_j N_j + s\left(\sum_j N_j - \sum_{j=k}^{M} j\left(N_j - N_{j+1}\right)\right)},$$

where we use that $\sum_{j=k}^{M} j\left(N_j - N_{j+1}\right) = kN_k + \sum_{j=k+1}^{M} N_j$.

We now turn to solving for the distribution of workers over wages and the job-to-job transition rate. Denote as w a wage in region k with rank $F_k(w) = \frac{w - w_{k-1}}{w_k - w_{k-1}}$; the CDF of wages then solves

$$\left(\delta + s\psi_k k\left(1 - F_k(w)\right) + s\sum_{j=k+1}^{M} j\psi_j\right) G(w)(1-u) = \left(\psi_k k F_k(w) + \sum_{j=1}^{k-1} j\psi_j\right) u.$$

Rearranging gives the CDF and corresponding density

$$\begin{aligned}
G(w) &= \left(\frac{\delta + s\sum_{j=1}^{M} j\psi_j}{\delta + s\psi_k k\left(1 - F_k(w)\right) + s\sum_{j=k+1}^{M} j\psi_j} - 1\right)\frac{\delta}{s\sum_{j=1}^{M} j\psi_j}, \\
g(w) &= s\psi_k k f_k(w)\frac{\delta + s\sum_{j=1}^{M} j\psi_j}{\left(\delta + s\psi_k k\left(1 - F_k(w)\right) + s\sum_{j=k+1}^{M} j\psi_j\right)^2}\frac{\delta}{s\sum_{j=1}^{M} j\psi_j}.
\end{aligned}$$

The job-to-job transition rate is therefore

$$\begin{aligned}
&\frac{\delta\left(\delta + s\lambda\right)}{s\lambda}\sum_{k=1}^{M}\left(\log\left(\frac{\delta + s\sum_{j=k}^{M} j\psi_j}{\delta + s\sum_{j=k+1}^{M} j\psi_j}\right)\frac{k-1}{k}\right. \\
&\left.- \left(\frac{1}{\delta + s\sum_{j=k+1}^{M} j\psi_j} - \frac{1}{\delta + s\sum_{j=k}^{M} j\psi_j}\right)\left(\delta\frac{k-1}{k} + s\sum_{j=k+1}^{M}\psi_j\left(1 - \frac{j}{k}\right)\right)\right).
\end{aligned}$$

From the perspective of the worker, the job offer arrival rate is $\lambda = \sum_{j=1}^{M} j\psi_j$ with the associated wage offer distribution

$$\tilde{F}(w) = \frac{\psi_1}{\lambda}1(w \geq w_r) + \sum_{j=2}^{M}\frac{j\psi_j}{\lambda}\frac{\min\{\max\{w, w_{j-1}\}, w_j\} - w_{j-1}}{w_j - w_{j-1}}.$$

Wage Cutoffs

We still need to solve for all the wage cutoffs. Pick the constant $\psi = \frac{w_u - w_r}{sc}$. Then, we have from (21) that $w_2 = w_r + \psi_2 sc$, $w_3 = w_2 + 2\psi_3 sc$ and so, recursively,

$$w_k = w_r + \sum_{i=2}^{k} (i-1)\psi_i sc. \tag{22}$$

Thus, we can express the highest wage in terms of the reservation wage and the contact rates that we already solved for.

For the reservation wage, proceed as in A.2 but without imposing symmetry. As always, the difference between the reservation wage and flow value of unemployment covers the foregone option value of search when accepting a job,

$$w_r - b = (1-s)\sum_{i=2}^{M} \int_{w_{-i}}^{w_i} i\psi_i (W(\tilde{w}) - W(w_r)) dF(\tilde{w}).$$

The usual integration-by-parts algebra gives

$$w_r - b = (1-s)\sum_{i=2}^{M} \int_{w_{-i}}^{w_i} \frac{i\psi_i(1-F(\tilde{w})) + \sum_{j=i+1}^{M} \psi_j j}{r + \delta + s\left(\sum_{j=i+1}^{M} \psi_j j + i\psi_i \frac{w_i - \tilde{w}}{w_i - w_{i-1}}\right)} d\tilde{w}.$$

The term in brackets in the denominator is, as usual, the rate at which the worker leaves for a higher-paying job. Therefore,

$$w_r = sb + (1-s)w_u - (1-s)c\sum_{i=2}^{M} (r+\delta)\frac{i-1}{i}\log\left(\frac{r+\delta+s\sum_{j=i}^{M}\psi_j j}{r+\delta+s\sum_{j=i+1}^{M}\psi_j j}\right)$$

Plug this into (22), evaluated at $k = M$, to get the expression for the highest wage

$$w_u = b + \sum_{i=1}^{M}(i-1)\psi_i c - \frac{1-s}{s}c\sum_{i=2}^{M}(r+\delta)\frac{i-1}{i}\log\left(\frac{r+\delta+s\sum_{j=i}^{M}\psi_j j}{r+\delta+s\sum_{j=i+1}^{M}\psi_j j}\right)$$

The model can then easily be solved by iterating on the highest wage until convergence.

C Heterogeneous Hiring Costs – Details

This section presents a simple algorithm for solving the model with heterogeneous hiring costs. The main difficulty relative to the baseline model is that the user cost is not the same across firms. We can, however, still utilize a similar approach to solve the model. We start by guessing the user costs of labor for each firm. Given the user cost of labor, we can then find each firm's employment by solving a system of M equations. In particular, using the fact that optimal employment at firm j, N_j, equates its user cost of labor to the marginal revenue product given in expression (10), we recover the level of employment at each firm.

Posit, similar to the model with heterogeneous productivity only, the following equilibrium. There are M endogenous intervals on $[w_r, w_u]$ indexed by $i = \{1, 2, ..., M\}$. The first interval is degenerate at w_r and the remaining intervals nondegenerate on $(w_{i-1}, w_i]$ for $i > 1$. We solve the model starting from the highest interval M. The highest wage is equal to the user cost of the firm with the highest hiring cost minus $(r+\delta)c_1$, so it directly follows given a guess for the user cost.

To solve for the distribution below on interval i, proceed as follows. Denote by K_i the set of firms that still have positions to fill. That is, they do not hire all their desired workers N_i on higher intervals. This, of course, means that K_M contains all employers.

Out of the K_i firms, the k firms with highest cost post on $(w_{i-1}, w_i]$ if

$$\frac{1}{k-1}\sum_{j\leq k}\frac{1}{c_j} \leq \frac{1}{c_{k+1}} \text{ and } \frac{1}{k-2}\sum_{j\leq k-1}\frac{1}{c_j} > \frac{1}{c_k}.$$

These conditions guarantee that firms with lower cost do not want to offer a wage on this interval or above and that firms with weakly higher cost do. Denote by $\psi_{j,i}$ the rate of offers of the j'th firm (ranked by their hiring cost) on interval i. These have to ensure that the user cost for all k firms posting on that interval is constant across the interval. For the user cost to be constant, we need

$$\sum_{j\neq z, j\leq k}\frac{\psi_{j,i}}{w_i - w_{i-1}} = 1/c_z \rightarrow (k-1)\sum_{j\leq k}\frac{\psi_{j,i}}{w_i - w_{i-1}} = \sum_{j\leq k} 1/c_j \;\forall z \leq k.$$

The offer rate for any firm z that is both in K_i and belongs to the k firms posting on interval

i can therefore be solved for in terms of the length of the interval w_{i-1} as

$$\frac{\psi_{z,i}}{w_i - w_{i-1}} = \frac{1}{k-1} \sum_{j \leq k} 1/c_j - 1/c_z.$$

Denote by ψ_i the rate at which workers receive offers in interval i, which is given by

$$\psi_i = \sum_{j \leq k} \psi_{j,i}.$$

Denote the steady-state unemployment rate by u. Given the unemployment rate and the employment in higher intervals, total employment in interval i, E_i, satisfies the flow balance

$$\psi_i \left(u + s \left(1 - u - \sum_{j \geq i} E_j \right) \right) = \left(\delta + s \sum_{j > i} \psi_j \right) E_i.$$

The employment in this interval for firm z is given by $\frac{\psi_{z,i}}{\psi_i} E_i$. We pick the length of the interval, w_{i-1}, such that the level of employment is equal to the desired level derived above for one of the k firms (and all other firms have a smaller level of employment than their desired level). The user cost for any firm $j \leq k$ to post wages in the interval $(w_{i-1}, w_i]$ but not above it can be calculated as

$$w_i + \left(r + \delta + s \sum_{j > i} \psi_j \right) c_j.$$

We have now calculated all relevant objects for interval i and can then return to the next interval.

The firm that remains post an additional mass point of jobs at the reservation wage such that the remaining firm also reaches its desired size.

Using the equations above, we get an implied user cost and a reservation wage. We then update the user costs until the user cost equals the guess and the lowest wage is equal to the reservation wage.

D Convex Adjustment Cost

Assume that there are M firms with homogeneous productivity, each facing a convex hiring cost function $C(H) = \bar{c}\frac{H^{1+\gamma}}{1+\gamma}$, where H represents the number of hires. The parameter γ controls the convexity of the hiring cost function, with $\gamma = 0$ corresponding to the baseline model where hiring costs scale linearly and $\gamma \to \infty$ corresponding to the limit case when adjusting hiring becomes prohibitively expensive, as in Burdett and Mortensen (1998). To calculate the firm value in this extended model, we can use the same approach as in the baseline model replacing the constant hiring cost c with the marginal hiring cost $C'(H)$. We further use that the job-to-job transition rate is given by $\frac{M-1}{M}\frac{\delta}{s\lambda}(\delta + s\lambda)\ln\left(1 + \frac{s\lambda}{\delta}\right) - \frac{M-1}{M}\delta$, which implies that total hires H at each firm solves

$$H(\psi) = \psi\frac{\delta}{\delta + M\psi} + \frac{M\psi}{\delta + M\psi}\frac{M-1}{M^2}\left(\frac{\delta}{s\lambda}(\delta + s\lambda)\ln\left(1 + \frac{s\lambda}{\delta}\right) - \delta\right).$$

Plugging this into the expression for labor market clearing condition (14) yields an implicit expression for the equilibrium offer rate ψ. The rest of the analysis is identical.