Monopsony Makes it Big[*]

Sydnee Caldwell
UC Berkeley & NBER

Arindrajit Dube
UMass-Amherst & NBER

Suresh Naidu
Columbia & NBER

May 26, 2026

Abstract

The literature on imperfect competition in labor markets has expanded rapidly in recent years. This article provides a guide to the field, focusing on the firm-specific ("residual") labor supply elasticity as the definition of a firm's labor market power. We present a general framework showing how this elasticity nests the three widely studied sources of monopsony power: search frictions, preference heterogeneity, and employer concentration. We summarize the empirical estimates of the elasticity of labor supply, highlighting sources of possible heterogeneity. We emphasize that it is difficult to infer elasticities from markdowns (and vice versa) due to the diversity of firm wage-setting practices, illustrating this point using the interaction between monopsony and efficiency wages. We discuss how policy issues in antitrust, labor market regulation, immigration, and macroeconomics interact with monopsony and conclude by listing several areas for future research.

[*]We thank Marco Badilla, Alex Forlenza, James Hickman, Kyle Neary, and Sebastian Puerta for outstanding research assistance. We thank several anonymous referees, participants at the 2024 CEPR Labour Economics Symposium, Emma Du Puy, Steven Durlauf, Joseph Howley, Gina Markov, David Romer, Todd Sorensen, Isaac Sorkin, and Pietro Tebaldi for useful comments.

Contents

1. Introduction

Labor economists have long debated whether (and which) labor markets are best described by perfect competition or by models in which firms have labor market power. In this article, we define labor market monopsony power as a finite, firm-specific labor supply elasticity, denoted η. In a perfectly competitive labor market, each firm faces a flat labor supply curve at the prevailing market wage, so η is infinite. In an imperfectly competitive market, firms face upward-sloping labor supply curves, so η is finite. When η is finite, a firm that wants to recruit or retain more workers must raise the wage.

The firm-specific labor supply elasticity answers a simple question: how much would a firm's employment (L) change if it changed its own wage (w), holding everything else constant? Formally,

$$\eta := \frac{\partial \log L}{\partial \log w}.$$

Researchers often call this elasticity the *residual* or *firm-specific* elasticity. For example, a residual labor supply elasticity of 2 suggests that if a firm cut wages by 10%, it would lose 20% of its workforce. By our definition, a firm has no monopsony power if it faces a flat (infinitely elastic) labor supply curve. A firm has meaningful monopsony power if η is relatively small (for example, below 10).

This definition contrasts with other interpretations of the word monopsony. A literal translation of the word—coined in 1933 when Joan Robinson published "The Economics of Imperfect Competition"—from ancient Greek would be *mono opsonein*, a "single buyer of fish relish".[1] The emphasis on the *mono* has led some past scholars to define monopsony as a scenario in which there is a single employer (or, in the case of oligopsony, a few large employers). But firms may have power over their workers even if there are many other employers in the market; conversely, even a very large employer may have little market power if workers view other jobs as close substitutes. For this reason, the modern literature typically defines monopsony power in terms of the firm-specific elasticity, rather than the number of employers. Robinson herself emphasizes this definition, noting that exploitation arises because the supply of labor is "imperfectly elastic to the unit of control" (Robinson, 1933).[2]

Defining monopsony power as a finite η means that we treat power as a feature of the firm's economic environment rather than an endogenous outcome of firm conduct, such as the wage markdown (the percentage gap between productivity and the wage). This definition allows us to distinguish between the existence of monopsony power and its exercise. As we highlight

[1] See Thornton (2004) for more on the etymology. As a coincidence, the $ops\bar{o}nein$ "buying provisions/fish relish" that Hallward gave Robinson for monopsony could have been instead $ops\bar{o}nion$ as "soldier's pay", as used by Polybius (1889). This would have been a more apt, if counterfactual, origin, as militaries are often classic monopsonies (Asch and Heaton, 2008). More broadly, the Greek New Testament uses $ops\bar{o}nia$ as "wages" in the context of "the wages of sin are death" in Romans 6:23 (Scripture4All Foundation, 2010). A time-traveling ancient Grecian would not be surprised to find monopsony as a term for employer wage-setting.

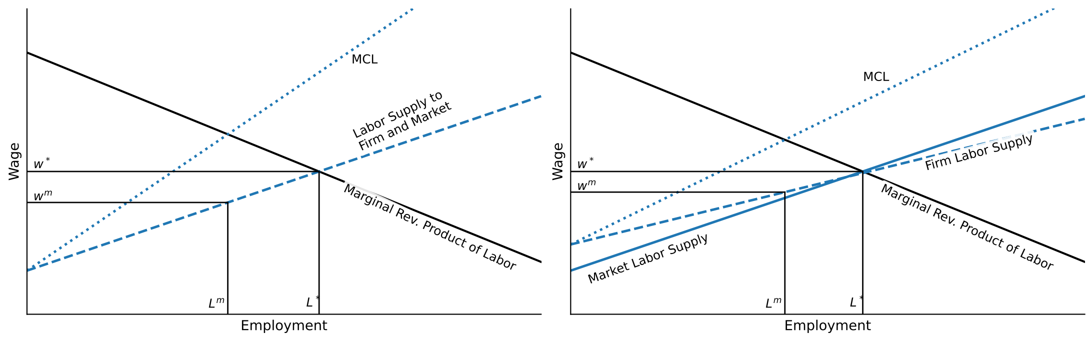
[2] Manning's 2003 book also emphasizes the firm-specific labor supply elasticity. Chapter one opens with "What happens if an employer cuts the wage it pays its workers by one cent? Much of labor economics is built on the assumption that all existing workers immediately leave the firm as that is the implication of the assumption of perfect competition in the labor market... The labor supply curve facing the firm is infinitely elastic." (page 1, Manning 2003).

throughout this review, monopsony power can remain latent (and not affect the wages that a firm offers) for a variety of reasons. For instance, a firm facing a finite η may be subject to a binding minimum wage that leads it to pay workers their full marginal revenue product. In this case, policy constrains the exercise of monopsony power, but monopsony power exists because a wage cut would not result in the immediate loss of all workers. If the policy constraint were removed, the firm might exert its power to lower wages. The residual labor supply elasticity is a causal parameter: it describes how a firm's employment would respond to a change in its own wage, holding all else constant.

What does a finite η imply for wages? Most canonical treatments of monopsony pair upward-sloping labor supply curves with wage posting (Boal et al., 1997; Manning, 2003). This leads firms to optimally trade off employment and payroll costs, and to choose wages below marginal revenue product and employment below the competitive level. We depict this in Panel A of Figure 1; Panel B generalizes the simplest model to a setting with multiple employers; the extent to which workers are diverted to alternative firms captures the intensity of competition.

The basic monopsony model implies a simple wage-setting rule: firms should pay wages that are "marked down" from workers' marginal revenue product by an amount which depends on η. Consider a profit-maximizing human resources (HR) manager who needs to set wages for front-line workers. The manager could estimate η by experimentally varying wages and by regressing log employment in each treatment group on the randomly assigned log wage. The coefficient on log wages would identify η.[3] If labor supply were the only constraint and wages the sole instrument for influencing employment, the optimal wage would satisfy the familiar Lerner rule where

[3]Many MBA programs teach analogous experiments to estimate price elasticities.



(a) Single Employer

(b) Multiple Employers

Figure 1: Market Power in the Labor Market

Notes: Panel (a) illustrates the classic monopsony case, where the firm is constrained only by its upward-sloping labor supply: workers choose between nonemployment and employment at that firm. Panel (b) introduces competition across firms: among workers willing to accept a given wage, some instead choose jobs at alternative employers that they value more. The share of workers who are diverted to other firms captures the intensity of competition, which is shaped by search frictions and the degree of job differentiation. In both panels, L^* denotes the employment associated with perfect competition and L^m denotes the employment the firm chooses. The firm-specific labor supply and market labor supply curves are not drawn on the same scale; the former represents the employment share at firm j, while the latter represents aggregate labor supply across all firms.

we use MRPL to denote the marginal revenue product of labor,

$$w^m = \frac{\eta}{1+\eta}\text{MRPL}. \tag{1}$$

The magnitude of η for a typical firm is a central empirical question.

A large body of empirical work has sought to assess how much monopsony power firms have. As Figure 2 shows, this literature has exploded in recent years. Since its inception, the monopsony literature has undergone several booms and busts. When Joan Robinson published "The Economics of Imperfect Competition" in 1933, there was an initial surge of interest. By the mid-20th century, research on the topic had waned. Monopsony power was often regarded as a theoretical curiosity or as something confined to particular labor markets. Interest began to revive following the surprising lack of disemployment effects of minimum wages found in Card and Krueger

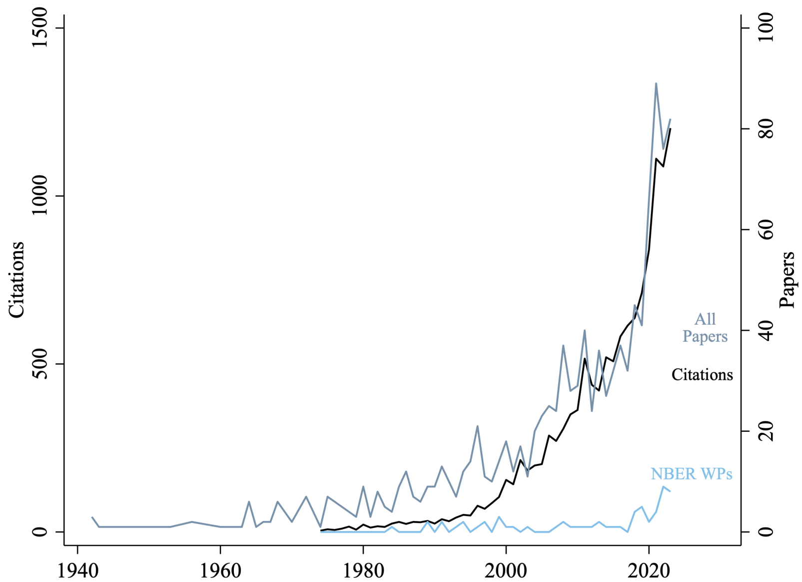
(1995a) and, since the mid-1990s, there have been several surveys, including the 1997 JEL review by Boal et al. (1997), as well as Alan Manning's influential book (2003) and his chapter in the Handbook of Labor Economics (2011).

In 2010, the *Journal of Labor Economics* published papers from a conference on labor market monopsony (Ashenfelter et al., 2010), curating what was, at the time, a sparse literature. By the time the conference was repeated in 2018, the literature on imperfect competition in labor markets had ballooned, not just in labor economics, but also in law and economics, macroeconomics, international trade, and industrial organization. The recent explosion of papers on monopsony reflects—among many factors—the growing availability of administrative matched worker-firm data and the decreasing cost of experimental interventions in real labor markets.

Section 2 presents a unifying framework that nests the leading sources of monopsony power: firm size or concentration, heterogeneity in worker preferences over jobs, and search frictions. We use a discrete-choice model with consideration sets in which workers evaluate only a subset of available jobs and jobs within a set are imperfect substitutes. Within this framework, the firm-specific labor supply elasticity captures the effects of the three central mechanisms.

Section 3 summarizes the evidence on η for firms across a wide variety of geographies and industries. Most estimates place η between 1 and 4, with a median around 2, confirming that monopsony power is widespread and economically meaningful. We organize the review around the different approaches to estimating the firm-specific labor supply elasticity, and we discuss how heterogeneity in existing estimates may reflect differences in methodological approaches or true heterogeneity. Echoing our point that η may not be the only constraint on firms' wage-setting, we find that estimates derived from production functions or methods relying on firm demand shocks tend to be larger than estimates based on experimental and

Figure 2: Research on Monopsony



Note: The black line in this figure plots the number of citations (left y-axis) to papers with the words "monopsony" or "monopsonist" in the title or abstract by year. The gray and blue lines plot the number of papers and National Bureau of Economic Research working papers by year (right y-axis). Data on the number of NBER working papers come from NBER. Data on the number of papers with the words "monopsony" or "monopsonist" in the abstract and data on the number of citations to these papers come from Dimensions.ai, an online aggregator of research data which provides information on both the number of publications each year and the number of times that publications have been cited by other publications in the Dimensions.ai database. Data from NBER were obtained on July 27, 2023 and data from Dimensions.ai were obtained on April 1, 2025.

quasi-experimental wage variation. Much of the recent literature uses quasi-experimental variation and an instrumental variable (IV) approach to estimating η. These approaches rely on assumptions about employers being constrained only by the residual labor supply elasticity. For example, IV estimates based on how employment and wages respond to product demand shocks are larger than estimates from experiments or IV estimates from wage shocks, suggesting that firms may face other constraints in wage-setting, and/or adjust on other margins, like amenities and recruitment effort, besides wages alone.

Section 4 turns from the measurement of monopsony power to its consequences for wage-setting. Recall that our HR manager's optimal wage satisfies the Lerner rule only if labor supply is her sole constraint and the wage is her only instrument. In practice, she may also face mandated wage floors, workers whose effort responds to pay, and instruments beyond wages—any of which can cause wages to diverge from the Lerner prediction. We develop a "monergony" model—a setting in which both the firm-specific labor supply elasticity η and the effort elasticity of wages are finite—that combines monopsony (Manning, 1995) with efficiency wages (Shapiro and Stiglitz, 1984; Bowles, 1985).[4] In this environment, efficiency wage forces can either attenuate or amplify monopsony distortions, even when η is unchanged. We also consider the role of fairness norms, non-wage amenities, strategic interactions among employers, and wage bargaining. As we discuss below, allowing these additional constraints may help resolve some empirical puzzles, for example the potential discrepancy between low residual labor-supply elasticities and the labor share of income.

Section 5 discusses policy domains where monopsony has been relevant.

---

[4]Monergony is a rough translation of "a state of single work", derived from "mono" (single) plus "erg-" (work/effort). We thank Professor Brooke Hawkins for discussing the classical Greek morphology with us. The full model is detailed in Appendix Section A.3.

Section 6 concludes with directions for future research. We emphasize that the evolution of monopsony power over time, the prevalence of collusion, and the details of firm wage-setting are topics that have not received enough attention.

We join a number of recent and contemporaneous surveys of monopsony and wage-setting (Langella and Manning, 2021; Manning, 2021; Ashenfelter et al., 2022; Azar and Marinescu, 2025; Kline, 2025). While these reviews address parts of this terrain, we place particular emphasis on the residual labor supply elasticity. Because η is defined directly from firms' employment responses to wage changes, it does not depend on any particular source of imperfect competition or assumption about how firms set wages. Our framework shows how the three most common sources of monopsony power in the literature—firm size or concentration, heterogeneity in workers' preferences over jobs, and search frictions—can influence η. We also clarify when and how that power translates into wage-setting.

2. The Sources of Monopsony Power

We open the review with a simple framework which builds on recent work in the monopsony and discrete choice literatures and which illustrates different channels through which monopsony power can emerge (Cattaneo et al., 2020; Abaluck and Adams-Prassl, 2021).[5] We show that this framework both nests the three classic channels through which employer monopsony power

[5] A rapidly growing literature in industrial organization, behavioral economics, and quantitative marketing studies random utility models with limited consideration sets (see Honka et al. 2019 for a survey and Barseghyan et al. 2021 for partial identification results and an application to estimating risk preferences). To our knowledge, this approach has not been embraced by labor economists, though it naturally integrates diverse sources of monopsony power. The idea that individuals do not consider all available options is not new. Herbert Simon writes, in his classic article on bounded rationality, "The organism may make its choice within a set of alternatives more limited than the whole range objectively available to it" (Simon, 1955, p. 102).

arises—employer size, idiosyncratic job valuations, and search frictions—and captures additional forces which may influence the residual labor supply elasticity.

2.1 Setup

Suppose worker i obtains flow utility $V_i^j(w_j) = U(w_j) + \epsilon_{ij}$ from working at firm j, where w_j is the wage posted by firm j and ϵ_{ij} is worker i's idiosyncratic preference for firm j. We assume that ϵ_{ij} comes from a continuous distribution and is independently and identically distributed across firms and workers. We use $j = 0$ to denote the option of nonemployment. A non-working individual obtains the value $V_i^0 = U(b_i) + \epsilon_{i0}$ from nonemployment where b_i is nonemployment income, including the monetary value of leisure associated with nonemployment, and ϵ_{i0} is independently drawn from the same distribution as the firm-specific taste shocks.

Workers are only aware of a subset of the jobs in the economy, those in their consideration set, S_i, which always includes nonemployment. This consideration set is a random variable that is independent of a worker's idiosyncratic preferences. The probability that a worker chooses firm j depends on (1) the probability that firm j is in their consideration set, (2) the probability that, of all the jobs in their consideration set S_i, they choose firm j, and (3) the probability that j is also preferred to nonemployment.

The probability that worker i prefers firm j over other firms $k \neq j$ (including nonemployment) is given by:

$$\text{P[i chooses j}|w_j, \mathbf{w_{-j}}, S_i, b_i] = \text{P}\left(j = \left[\arg\max_{k \in S_i} V_i^k(w_k)\right]\right) \tag{2}$$

where $\mathbf{w_{-j}}$ is a vector of wages offered by firms other than j. We assume that idiosyncratic tastes of workers for jobs are independent across firms, and that the inclusion of j in i's consideration set is independent of idiosyncratic tastes, other firms' wages, and outside options. We show in Appendix A that under these independence assumptions, once we condition on any idiosyncratic taste ϵ_{ij} of worker i for firm j, the choice probability is the product of three independent terms: (1) the probability firm j is in a worker's consideration set, (2) the probability j is preferred to other firms in the set (conditional on j being in the set), and (3) the probability j is better than nonemployment. Worker i's unconditional labor supply to firm j is obtained by integrating over ϵ_{ij}, and its elasticity decomposes accordingly:

$$\begin{aligned}\eta_{ij} \quad = \quad & \underbrace{\eta_{ij}^{\text{Consideration}}}_{\text{Elasticity Due to Increased Consideration}} + \\ & \left(\sum\nolimits_{k \in S_i \setminus \{j\}, k>0} \underbrace{\eta_{ijk}}_{\substack{\text{Cross-employer labor} \\ \text{supply elasticity}}} \right) + \underbrace{\eta_{i0}}_{\substack{\text{Market-level elasticity} \\ \text{of labor supply}}} \end{aligned} \tag{3}$$

Equation 3 is the elasticity of worker i's unconditional labor supply s_{ij} with respect to w_j; we derive it formally under the logit-log specialization in Appendix A.1. The decomposition suggests a few points. First, each worker is more elastic to the firm than to the market, except when there are no other competing employers (i.e., the classic monopsony case where the firm is the market). Second, holding other firms' behavior constant, the presence of an additional firm k in a worker's consideration set will tend to make that worker more elastic. Finally, the elasticity of consideration, $\eta_{ij}^{\text{Consideration}}$, is positive whenever an increase in w_j (all else constant) increases the probability that firm j is in the worker's consideration set. Random search models assume that this elasticity is zero; this elasticity is only non-zero under some form of directed search (Wright et al., 2021).[6]

[6]Many models of directed search predict that this elasticity is positive. This implies that

While Equation 3 provides a decomposition of the elasticity of labor supply of worker i to firm j, this may differ from the overall residual labor supply elasticity to firm j if workers are heterogeneous. Consider the perspective of a firm offering w_j, facing a distribution of workers who vary in their consideration sets (S_i), idiosyncratic taste for firm j (ϵ_{ij}), and reservation utility (V_i^0). If the firm cannot wage discriminate across workers, the elasticity of labor supply facing firm j, which we denote η_j, is a weighted average of the worker-specific elasticities η_{ij} in Equation 3. As shown in Appendix A.1, under some simplifying assumptions, we can obtain an expression for the firm-level elasticity of the form:

$$\eta_j = \eta_j^{\text{Consideration}} + \sum_{k \neq j, k>0} \eta_{jk} + \eta_{j0}. \tag{4}$$

Each elasticity in equation 4 measures how the probability of firm j being chosen (or being included in the choice set) responds to the wage it offers (w_j), given the wages offered by other firms $(\mathbf{w_{-j}})$ and given the identities of the other firms in the consideration set.[7] The elasticity $\eta_j^{\text{Consideration}}$ captures how w_j affects the firm's inclusion in the consideration set; η_{jk} captures how w_j affects the share of workers that choose firm j over firm k; and η_{j0} is the market-level labor supply elasticity that measures how a change in firm j's wage changes the share of workers choosing employment over nonemployment.[8] Of course, because these elasticities are functions of the wage level and the underlying distribution of outside options, they are not necessarily constant.

workers who can change their consideration sets in response to firm-specific changes in wages will be more elastic to the firm than those who cannot (Caldwell et al., 2025).

[7]In a more general model, changes in w_j may affect both whether firm j is in the consideration set, and whether other firms are included in the set.

[8]It is important to note that the residual labor supply elasticity as defined here refers to how the employment probability responds to a change in wage, w_j holding other firms' behavior constant. In other words, it does not capture equilibrium effects that may occur in response to such a change in w_j.

The fact that η_j captures both the cross- and within-worker heterogeneity in outside options, as well as whether they are due to search frictions, job differentiation, or the number of employers, makes it a good definition of employer monopsony power. While the relative importance of these forces in different settings remains an open question, Dube et al. (2022) and Caldwell et al. (2025) suggest job differentiation, i.e., variation in non-wage attributes over which workers have heterogeneous preferences, is likely the dominant source of employer monopsony power.

2.2 Sources of Firm Monopsony Power

This framework nests the three workhorse models of monopsony.[9]

Monopsony and Concentration. First, with neither preference heterogeneity nor information frictions, the model nests classic models of monopsony and oligopsony (Robinson, 1933). When there are no other firms, a firm will face an upward-sloping labor supply curve if there is heterogeneity in workers' nonemployment options (which leads to an upward-sloping market supply curve). This heterogeneity in b_i could arise from a variety of channels, including differences in the value of leisure, home production, or migration.

With a single firm j and N workers, with a CDF of unemployment benefits given by $F_b(\cdot)$, the share of labor supply to j would be given by the aggregate extensive margin labor supply function where the probability a worker

[9] Sources of ex post monopsony not discussed in this review include anchoring of beliefs at the current wage (Jäger et al., 2024), firm-specific human capital (Parsons 1972 provides a classic reference, see Bils et al. (2025) for a recent application), and adverse selection in the secondary labor market, as in Acemoglu and Pischke (1998) (with intriguing audit-study evidence in Figure 2 of Kroft et al. 2013 showing that resumes from the already employed have a callback penalty relative to recently unemployed). These have not been as widely used, but are potentially important directions for future research.

chooses to work for firm j (given the wage it offers) is:

$$\mathrm{P}(j|w_j) = \frac{\sum_{i \in N} \mathbf{1}_{w_j \geq b_i}}{N} \approx F_b(w_j).$$

The approximation follows because we are assuming a large number of workers N. Note that we now assume that a worker's consideration set always includes this firm and that there are no idiosyncratic taste shocks. Going beyond a single firm, there is scope for imperfect competition in a market with a small number of firms, even if there are no search frictions or preference heterogeneity. If capacity-constrained firms compete with each other for labor by setting wages, the outcomes mirror those of a typical Cournot model (Kreps and Scheinkman, 1983). If there is an aggregate labor supply curve with elasticity η_0 and heterogeneous firms j, each with employment share s_j, it is well known that each firm's residual labor supply elasticity at Cournot equilibrium can be written as $\eta_j = \frac{\eta_0}{s_j}$. This model has been extensively discussed, and criticized, in industrial organization (Berry et al., 2019).

Preference Heterogeneity. To recover the basic model based on preference heterogeneity, we assume that there are many firms, that workers have homogeneous nonemployment options, which we normalize to zero ($V_i^0 = 0 \ \forall i$), and that each worker's consideration set includes all firms. As we discussed above and show in Appendix A, if workers have idiosyncratic match-specific type-1 extreme value utility shocks with scale parameter σ and $V_i(w_j) = \beta \log(w_j) + \epsilon_{ij}$, then the setup collapses to a standard constant elasticity of substitution (CES) utility function. When the number of options, J , is large, this model delivers an approximately constant labor supply elasticity once we condition on and integrate over the distribution of ϵ_{ij}.[10] The probabil-

[10] The approximation arises because w_j^η appears in both the numerator and denominator. This is a standard approximation when the number of firms is large, but is inappropriate for studying strategic interactions.

ity a worker chooses firm j, given the wages offered by this firm (w_j) and all other firms $(\mathbf{w_{-j}})$ is:

$$\mathrm{P}(j|w_j, \mathbf{w_{-j}}) = \frac{w_j^\eta}{\sum_{k\in J} w_k^\eta} \approx C w_j^\eta.$$

This constant elasticity formulation is tractable and has been used in many parallel analyses on the demand side in fields like macroeconomics, trade, and industrial organization. As Lamadon et al. (2022) show, it is also easy to augment the standard model with firm-specific amenities Q_j, each of which is log-additive with wages w_j and valued equally by all workers, so the probability of a worker choosing j, given the wages and amenities of j and the wages (and amenities) of all other firms becomes:

$$\mathrm{P}(j|w_j, \mathbf{w_{-j}}) = \frac{w_j^\eta Q_j}{\sum_{k\in J} w_k^\eta Q_k}.$$

As Berger et al. (2022) show, it is possible to augment the model so that firms have monopsony power due to both preference heterogeneity and size within a market k, where ψ denotes the cross-market elasticity of substitution from a nested CES supply function. Using s_j to denote the payroll market share of firm j within market k, the firm-specific labor supply curve $l_j = \mathrm{P}(j|w_j, \mathbf{w_{-j}})$ is the share of labor employed by j:

$$\log(l_j) = \left[\frac{1}{\eta}(1 - s_j) + \left(\frac{1}{\psi}\right)s_j\right]^{-1} \log(w_j) + \text{Constant}.$$

This labor supply function has the intuitive property that so long as $\psi < \eta$ (across-market elasticities are less than cross-firm elasticities), the labor supply elasticity is decreasing in a firm's share of payroll. When $s_j = 0$ this reduces to monopsonistic competition with no strategic interactions, with a labor supply elasticity η; when $s_j = 1$ this reduces to a single monopsonist in each labor market facing a market-level labor supply with elasticity ψ.

Berger et al. (2022) also show that this firm-level labor supply function yields a Cournot-like microfoundation for the use of payroll-based HHI as a measure of labor market concentration.

Search Frictions. Finally, without preference heterogeneity, but with the addition of search frictions, the model generates a standard static search model (Mortensen, 2003). To see this, suppose that workers' consideration sets are generated by firms, which send messages to random subsets of workers (see Chapter 1 in Mortensen, 2003). Workers choose their best option from the set of firms from which they received messages. If each firm's offer has an independent and identical chance of being seen by workers, the probability of a worker seeing a particular set of firms is only a function of the size of the set. Using $|S|$ to denote the size of the set and q to denote the offer arrival rate, this probability is: $\mathrm{P}(S) = q^{|S|}(1-q)^{J-|S|}$; the probability a worker receives $|S|$ offers (from any set of firms of size $|S|$) is $\binom{J}{|S|}q^{|S|}(1-q)^{J-|S|}$.

If there are J firms that each randomly send a message to one of N workers, and both J and N are large, the number of offers a worker receives can be approximated by a Poisson distribution with parameter $\lambda = J/N$. The probability a worker with an offer from firm j chooses firm j depends on the probability that firm j has the highest offer. This leads to the following firm-specific supply curve:

$$\mathrm{P}(j|w_j, F(w)) = \sum_{x=0}^{\infty} F(w_j)^x \frac{\lambda^x \exp(-\lambda)}{x!} = e^{-\lambda(1-F(w_j))} \tag{5}$$

This is a standard static search model which has a unique mixed-strategy equilibrium wage distribution, $F(w_j)$, as in Mortensen (2003). An endogenous closed-form solution for F can be obtained using the fact that all firms make identical profits for every wage with positive support in F together with $F(b) = 0$.

Dynamic Considerations. The basic version of our model does not differentiate between changes in employment that arise due to changes in the number of workers joining a firm (recruits) or changes in the number of workers leaving a firm (separations). Many empirical papers measure monopsony power by measuring the impact of a wage change on separations, rather than employment. This focus arises naturally because separation decisions are more readily observable in worker-level data: researchers can identify when an individual worker leaves a firm in response to wage conditions, whereas the overall employment response also depends on recruitment from the external market. A natural concern is that a change in wages may also change the composition of firm applicants or a firm's recruiting standards; these concerns can be difficult to rule out.

Empirical papers which estimate the separation elasticity then approximate the overall firm-specific labor supply elasticity by doubling the absolute value of the separation elasticity. This doubling holds in steady state, when firms are not growing or shrinking, if a number of other assumptions are met, including that the elasticities are constant (Manning, 2003).

In Appendix Section A.2, we outline how a dynamic, continuous version of the consideration set approach nests the canonical Burdett and Mortensen (1998) model, and show how the separation elasticity relates to the residual labor supply elasticity.[11] We also show how the model accommodates the hybrid Bloesch et al. (2026) approach that combines random utility shocks with search.[12] In these models, the offer arrival rate shapes the elasticity of

[11]The key idea is that in a dynamic search model, an employed worker's effective consideration set at any moment consists of their current employer plus—with probability proportional to the offer arrival rate λ—one randomly drawn outside firm. This is a consideration set of size two with deterministic choice (take the higher wage), which is the limit of the discrete choice framework as the idiosyncratic preference variance vanishes.

[12]In that approach, consideration sets remain of size two (as in search), but work-

labor supply facing the firm by influencing how quickly incumbent workers' consideration sets include another firm.[13]

2.3 Consideration Sets

An advantage of the consideration set approach is that it allows a large set of factors to influence the firms workers consider (or receive offers from).[14] It also illustrates how heterogeneity in consideration (and in the responsiveness of consideration to the wage, as captured by $\eta_{ij}^{\text{Consideration}}$) may give firms market power.

Consideration sets may capture the fact that social networks, including family and friends, are important channels for obtaining information about jobs (Granovetter, 1973; Montgomery, 1991; Granovetter, 1995; Bayer et al., 2008; Glitz, 2017; Hensvik and Skans, 2016). Caldwell and Harmon (2019) show that, when a worker's former coworkers are at firms that are growing more, this improves the outside options of the worker, increasing wages and quit rates. Given the large literature which documents homophily (the tendency for people to form social ties with people who are similar to themselves), it seems plausible that such channels could lead firms to have more monopsony power over some groups of workers than others (Black, 1995; Topa, 2011).[15]

Several papers have documented racial differences in call-back rates; con-

ers choose probabilistically between their current firm and the outside offer—yielding a constant-elasticity labor supply function that combines the tractability of random utility with the dynamic structure of search.

[13]In standard directed search models, workers' application decisions trade off the probability of getting a job at a higher wage firm with the value of obtaining that job (Moen, 1997; Wright et al., 2021). There is empirical work documenting a positive link between consideration (as measured by clicks on applications) and wages (Belot et al., 2022).

[14]While closed form implementations of consideration set models do not yield very general function forms, it is easy to obtain numerical results in quantitative applications.

[15]This idea is captured in a reduced form way by Black (1995), who documented how race-specific differences in arrival rates due to the presence of discriminatory employers could yield a racial wage gap, even at non-discriminatory employers.

sideration sets provide a natural way to incorporate between-group differences in offer arrival rates without relying on between-group differences in worker preferences (Bertrand and Mullainathan, 2004; Kline et al., 2022). For instance, if some managers discriminate and are less likely to call back Black applicants than white applicants, Black applicants can have smaller consideration sets even if all workers have the same preferences.

Consideration sets also allow cognitive frictions (limitations in workers' ability or willingness to process information) to influence workers' search behavior. These cognitive frictions arise because it is costly to evaluate job opportunities. Behavioral economists have already integrated the literature on rational inattention with the literature on consideration sets, showing how attention costs or limits to cognitive processing can lead to biased choices within the set of alternatives considered (Caplin et al., 2019).[16] One promising area for future research is to integrate these cognitive and informational constraints more explicitly into labor market models and examine whether (and to what extent) these constraints shape workers' search behavior and preferences.[17]

We expect that the consideration set approach will see more widespread use, both because it naturally embeds the factors discussed above and because online job platforms and application data facilitate measurement—and even experimental manipulation—of consideration sets.[18] A compelling recent example is Roussille and Scuderi (2026), who test models of con-

[16]Rational inattention also provides an alternative micro-foundation for multinomial logit and constant elasticity choice probabilities, as shown by Matejka and McKay (2015). Fosgerau et al. (2020) show that any additive random utility model is representable as a rational inattention model. Maćkowiak et al. (2023) review this literature.

[17]One example of how this could be done comes from Jäger et al. (2024), who consider a monopsony model akin to the Salop and Stiglitz (1977) model of "bargains" and "ripoffs", in which firms set wages anticipating a mix of sophisticated and naive workers. They show that a firm's markdown is increasing in the share of naive workers.

[18]This parallels a literature estimating models of product demand using online search data (Santos et al., 2012).

duct and wage discrimination on an online platform where they can see the choice sets of workers and the bids of firms. Online platforms and other technological changes in job matching may also allow search with recall (either sequential or simultaneous search as in Honka and Chintagunta (2017)), which generates different consideration sets from the more standard no-recall case. Dube et al. (2020) provide an example of where online job platforms can be used to manipulate workers' consideration sets. They use data from a "honeypot" design on Amazon Mechanical Turk ("MTurk"), where each worker is randomized into job offers with experimentally determined wages that only they can see on the platform.[19]

In sum, a variety of mechanisms generate finite firm-specific labor supply elasticities. While the three channels highlighted here are conceptually distinct, they often operate simultaneously in practice, and the relative importance of each is ultimately an empirical question.

3. Estimates of Monopsony Power

The framework in Section 2 showed how concentration, preference heterogeneity, and search frictions can each generate a finite firm-specific labor supply elasticity. A natural next question is: how large is η in practice?

Over the past several decades, a large literature has confirmed that the residual labor supply elasticity is finite, and potentially quite small. This literature has used a variety of approaches—ranging from experimental to structural—to estimate η. In this section, we summarize this evidence. While one emphasis of this review is that η is distinct from firm conduct (for reasons we discuss in Sections 4 and 5), we include a broad range of studies that

[19] In industrial organization, authors have used "capacity constrained" firms as shifters for consideration sets that are orthogonal to tastes (Agarwal and Somaini, 2025). The analogy would be employers who are labor demand constrained, and are thus not hiring even though they are in the market.

Table 1: Mean, Median, and Standard Deviation of Estimates By Estimation Strategy

Estimation Strategy	Average	Std. Dev.	Median	N
Calibrated model	2.19	1.14	1.73	5
Experimental	1.96	1.64	2.13	7
Non-Experimental	1.67	1.17	1.48	26
Production Function	3.02	2.41	2.23	6
Quasi-experimental demand shock	4.14	1.70	3.87	9
Quasi-experimental wage-shock	2.04	2.63	1.42	16

Note: This table summarizes the estimates across studies. The list of included studies is provided in Appendix B.

report estimates of η.

3.1 Empirical Estimates of η

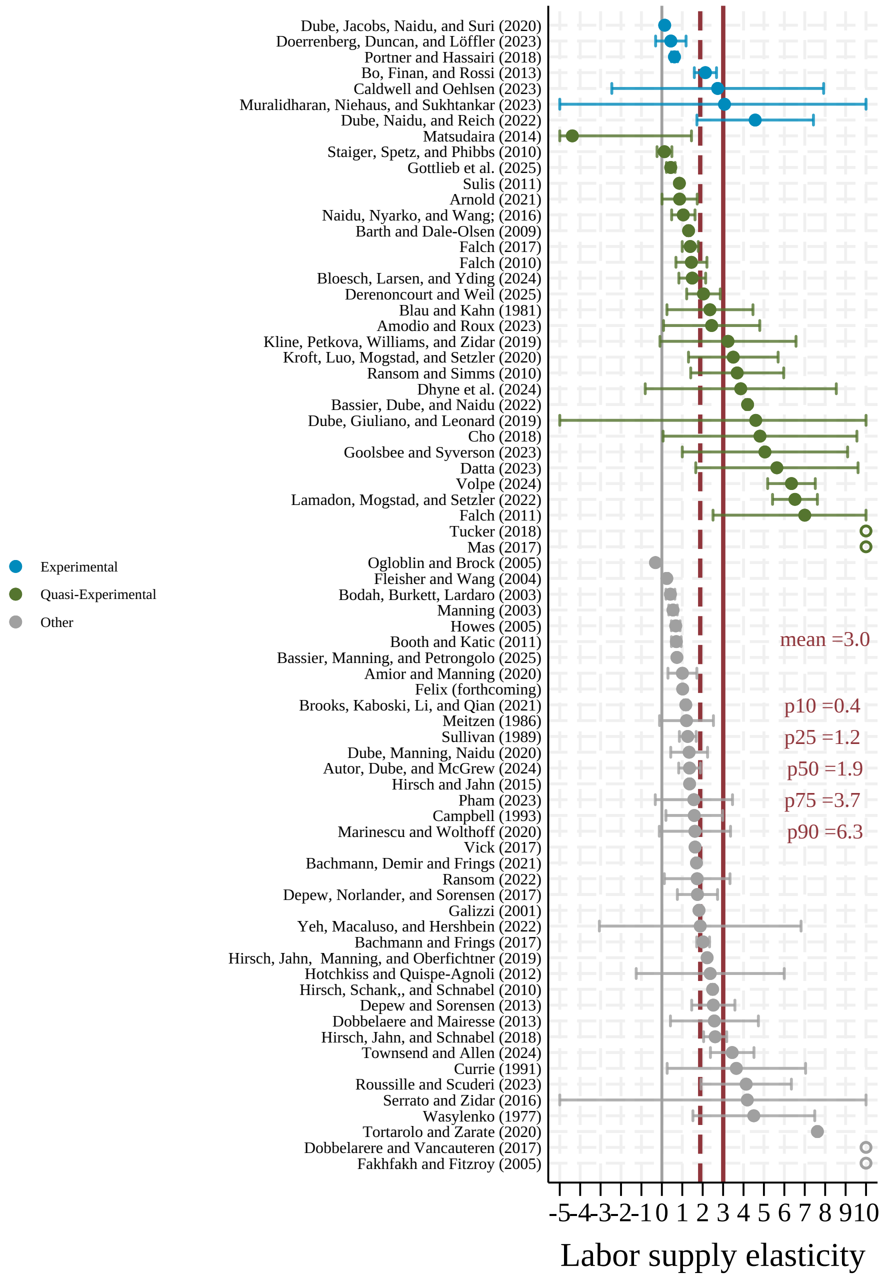
We assemble a list of 73 studies by starting with a previous meta-analysis (Sokolova and Sorensen, 2021), which we supplement with papers from our own review of the literature.[20] The full list is available in Appendix B.1. We classify studies according to the approach they take: (1) experimental, (2) quasi-experimental wage-shock, (3) quasi-experimental demand shock, (4) production function, (5) empirical calibration, or (6) non-experimental.

Figure 3 summarizes the literature to date by plotting the central estimate from each of the studies. A first takeaway is that, despite substantial heterogeneity, estimates are relatively small: the majority are below 6, with a handful of outliers (see Appendix B.2). As reported in Figure 3, the mean estimate is 3.0 and the median is 1.9, consistent with meaningful monopsony power in most settings studied. The natural next question is what drives the heterogeneity across estimates.

One obvious potential explanation is labor market heterogeneity: the pa-

[20] We remove 3 studies from the Sokolova and Sorensen (2021) list that estimate separation elasticities using variation in statutory minimum wages, as these are not firm-specific wage shocks.

Figure 3: Empirical Firm-Specific Labor Supply Elasticities η



Note: Blue estimates at the top are experimental direct estimates of η. Next are green estimates that are calculated based on a quasi-experimental estimand. At the bottom ("Other") are other estimates based on non-experimental variation. Confidence intervals are bottom-coded at -5 and top-coded at 10. Hollow circles indicate point estimates greater than 10. See text for more description.

pers summarized in Figure 3 cover a variety of countries, segments of the labor market (low-wage vs. high-wage), and firms. Market factors such as the number of firms or the extent of information frictions could influence the residual labor supply elasticity. Even within a labor market, firms may differ in the amount of market power that they have, depending on their position in the wage (or job value) distribution. For example, in the theoretical framework with search frictions, a firm initially offering a very low wage would have a higher separation rate than a firm towards the upper end of the distribution; Bassier et al. (2022) provide some evidence consistent with this: separation rates are higher at lower-wage firms.

Another potential explanation is heterogeneity in the time horizons used in each study. If it takes time for workers to move between firms, there may be differences in the short- and long-run residual labor supply elasticities. Bassier and Manning (2025) show that the short-run elasticity is equal to the quit rate times the long-run elasticity, with the long-run elasticity (in the 4-6 range) being more than three times the short-run elasticity (in the 1-2 range) in their context. One potential explanation for the fact that experimental studies tend to yield smaller estimates than quasi-experimental approaches (Table 1) is that these studies tend to rely on shorter time horizons.

But beyond these well-understood sources of heterogeneity, there is a basic concern about identification. One way to read the range of estimates is that specification error and omitted-variables bias may be substantial, making plausibly exogenous variation in firm-level wages crucial. [21]

By this light, the "gold standard" evidence would come from a series of randomized experiments across a variety of jobs, where the researcher (or employer) randomizes wages and ensures that other factors (including ameni-

[21] For example, Sokolova and Sorensen (2021) document that researchers typically obtain very different estimates of the residual labor supply elasticity depending on whether their dependent variable is employment (the typical, "forward" regression) or wages (the "reverse" regression).

ties, recruitment effort, and hiring selectivity) do not change during the course of the experiment.[22] It is not surprising that these experiments are few and far between. Employers may resist efforts to randomize wages due to administrative constraints or perceptions about unfairness. Even if there is sign-off from top-level managers, a researcher has to ensure that lower-level managers do not endogenously respond by changing non-wage benefits, recruitment effort, or hiring standards to ensure a given level of hiring despite the wage change. These concerns must be addressed and ruled out through the design of the experiment.

3.2 Estimates Using Experimental Variation in Wages

Given the requirements involved in running a credible experiment, it is not surprising that few researchers have employed this strategy and that existing RCTs have been conducted in somewhat idiosyncratic labor markets. Dal Bó et al. (2013) analyze a recruitment experiment conducted by the Mexican federal government's Regional Development Program (RDP) which arguably provides the cleanest and most comprehensive example of estimating firm monopsony power. The authors examine the effect of offered wages on both the applicant pool and the rate at which selected candidates accepted offers. Combining both margins, the authors compute a residual labor supply elasticity of 2.15. The joint randomization of offers and wages and the timing of the wage announcements guard against the concern of endogenous recruitment effort.

Other experimental estimates come from gig-worker platforms. These

[22]The ideal time horizon will depend on the contextâa firm considering a temporary wage cut faces the short-run elasticity, while a permanent merger is better evaluated against the long-run elasticity. However, over longer horizons, competing firms may adjust their own wages, making it harder to satisfy the requirement that other firms' behavior is held constant. As we note below, survey experiments offer one way to sidestep this difficulty, as researchers can randomize the time frame of hypothetical job choices without triggering equilibrium responses from competitors.

papers have found, somewhat surprisingly, that even in these markets firms have significant monopsony power. Dube et al. (2020) use an experimental design to measure market power in the crowd-sourcing platform, Amazon Mechanical Turk (or MTurk). The authors posted a total of 5,500 unique image-tagging tasks on MTurk. Workers who completed the screening task could choose whether to complete an additional set of classification tasks for a pay rate which was randomized between $0.05 and $0.15 per image. While workers offered higher rates were more likely to accept the offer, the implied residual labor supply elasticity was only around 0.1. Caldwell and Oehlsen (2023) undertake a set of similar experiments in conjunction with Uber, in which drivers—some of whom could drive for Lyft—were given the opportunity to drive for one week with 10-50% higher earnings per trip. These experiments revealed that drivers who had the opportunity to drive for Lyft were significantly more elastic to Uber. The authors' central estimate of the quit elasticity (around -2) suggests a sizable amount of labor market power in this market.

3.3 Stated Preference and Survey Experiments

While field experiments provide high-stakes evidence of monopsony power, the logistical and ethical hurdles of wage randomization often limit their scope. Reflecting a broader trend in economics toward the use of primary data and "stated-preference" experiments, a growing literature instead uses survey methods—most notably conjoint analysis—to estimate firm-specific labor supply elasticities and valuations of non-wage job characteristics (Wiswall and Zafar, 2018; Maestas et al., 2023). This approach is often the only feasible way to generate standardized estimates of monopsony power across a wide variety of firms and market segments in a single study, as it does not rely on the cooperation of multiple employers to implement identical experimental designs.

A natural concern is that workers' stated preferences in a hypothetical setting may not align with their actual labor-market behavior. However, Mas and Pallais (2017) provide a validation of this methodology; in their study of alternative work arrangements, they find that labor supply elasticities derived from survey experiments are consistent with those from field experiments. Beyond feasibility, the survey approach allows researchers to investigate the "shape" of the residual labor supply curve far from the current wage. Because the survey approach also allows researchers to vary the time horizon of hypothetical job choices (without worrying about competing firms changing their behavior), it is well-suited to examining how elasticities may or may not change over time. Recent applications have used these methods to decompose the sources of employer power, showing that idiosyncratic job differentiation and amenities like "workplace dignity" are key drivers of limited mobility (Dube et al., 2022; Caldwell et al., 2025).

3.4 Leveraging Quasi-Experimental Variation

Because it is difficult to implement randomized designs, much of the evidence on the residual labor supply elasticity comes from quasi-experimental studies. This literature has followed two main approaches. The first approach leverages exogenous variation in wages. The second leverages exogenous shocks to the marginal product of labor, which the researcher uses as an instrument for wages and employment.

Wage shock. Wage shock designs rely on wage changes (almost always increases) that are differentially binding across establishments or firms. A researcher uses this variation to examine how the change in (log) employment responds to the change in (log) wages. Of course, truly exogenous firm-level variation in wages together with labor-supply being the only constraint may be difficult to find outside a deliberate experiment. To show that they credibly identify firm-specific elasticities, the researcher must show that the firm did not respond to the wage shock in other ways, such as by endogenously changing recruitment effort or amenities.

Staiger et al. (2010) conduct the classic wage shock study, examining the effect of mandated wage increases legislated by the Nurse Pay Act of 1990 at Veterans Affairs (VA) hospitals. These pay increases varied by area and, critically, only affected a single employer (a VA hospital) in the local market. The authors find that the mandated wage hikes had a relatively small impact on nurse employment, with an implied residual labor supply elasticity of 0.1. The authors also find positive spillovers on the wages of non-VA hospitals in the area, with spillovers declining in distance to the VA.[23]

More recent papers exploiting this approach include Datta (2024), Emanuel and Harrington (2025), and Dube et al. (2019). Datta (2024) shows that the "Living Wage" laws passed in different London municipalities between 2010 and 2018 raised wages at a large private employer ("the Company"), but at essentially none of its labor market competitors. Datta examines the impact of this wage change on both recruits and separations, finding a separation elasticity of around -1.6 and a recruit elasticity of around 3; the recruit elasticity minus the separation elasticity $(3 - (-1.6))$ suggests an overall residual labor supply elasticity of 4.6.[24] Derenoncourt and Montialoux (2021) leverage differential exposure of stores to the introduction of (highly binding)

[23] The strategy relies on the assumption that the firm (here the VA) is profit-maximizing and that employment is determined by labor supply (i.e., there is not a fixed budget).

[24] Some papers in the monopsony literature examine the impact of wage changes on separation and recruit rates, rather than on employment. Datta argues that the correct definition of the recruit elasticity focuses on the impact on workers who could plausibly be hired by the firm. It is not clear that the sum of the recruit and separation elasticities would produce a residual labor supply elasticity, since it includes the impact on offers that are rejected by workers.

voluntary corporate minimum wages in a difference-in-differences design. They estimate separation elasticities of around -1.0 at major retailers, with an implied labor supply elasticity of 2.0. (They do not find any spillovers to other firms in the local labor markets, a point we will return to below.) Emanuel and Harrington (2025) use a similar exposure design to study a large warehouse employer that adopted a voluntary $15 minimum wage in the aftermath of the pandemic; the bite of this minimum wage varies across the firm's more than 50 stores as a result of differences in regional wages. The authors' store-level separation elasticity of -8.7 is substantially larger in magnitude than estimates in other work; assuming that the overall labor supply elasticity can be found by doubling the separation elasticity, this suggests a residual labor supply elasticity of about 17.[25] Dube et al. (2019) examine the effects of raises enacted by a large retailer in response to a federal minimum wage to workers earning above the minimum. The formula for these pay increases created discontinuities in raises across workers, both within and between stores. They find that once within-store peer effects are accounted for, the wage increase decreased separations moderately, with an implied labor supply elasticity of around 4.6.

Product Demand Shocks. An alternative to using a direct wage shock is to use firm-specific demand shocks; the rent-sharing literature shows that positive firm-specific shocks typically lead firms to raise wages (Card et al., 2013; Garin and Silvério, 2024). It is possible to estimate labor supply elasticities by examining the relative impact of such a shock on wages and employment: $\eta = \frac{d\log(L)}{d\log(w)} = \frac{\gamma^e}{\gamma^w}$ where γ^e, γ^w come from the reduced form and the first stage, respectively:

[25] A pre-pandemic wage increase at the same firm was found to have a smaller (though still sizable) separation elasticity of -4.8.

$$\log(L_j) = \gamma^e Shock_j + \epsilon_j \quad (6)$$

$$\log(w_j) = \gamma^w Shock_j + \nu_j. \quad (7)$$

This strategy requires shocks that are credibly firm-specific and which—to satisfy the exclusion restriction—affect firm employment only through their impact on firm wages. A concern is that firms may respond to a demand shock by increasing wages as well as adjusting other factors (amenities, recruiting effort), that could help the firm expand. It strikes us as particularly important to test for the presence of responses along non-wage margins when using product demand shocks for identification; some of the papers cited below have implemented such tests. In addition, tests for cross-firm wage spillovers (e.g., as in Sharma, 2023) are also particularly useful in these settings as a test of whether the shock is truly firm-specific. For example, many shocks based on aggregate industry or trade variation may affect many firms or may affect market-level labor supply.

There are many examples of researchers using this strategy. For instance, Kroft et al. (2025) ($\eta \approx 4$) use a clear firm-specific demand shock—whether a firm wins the contract for a construction project in a government procurement auction—to estimate labor supply elasticities. The researchers address and find no evidence that firms respond along non-wage margins by examining the effect of the shock on non-wage compensation and safety. Goolsbee and Syverson (2023) use enrollment shocks to higher education institutions to estimate monopsony power in the academic labor market, but are unable to examine changes in recruiting effort, for example. Other researchers use shocks which are technically market-level shocks, but which they argue only affect a small share of the relevant employers in a labor market. For instance, Amodio and de Roux (2023) ($\eta \approx 2.5$) leverage within-sector and within-location variation in foreign exchange rate shocks (based on predetermined export patterns) in Colombia, and Sharma (2023) ($\eta \approx 1.2$ for women, $\eta \approx 2.7$ for men) uses firm-specific demand shocks in the Brazilian textile industry from the end of the Multi-Fiber Arrangement. Frameworks such as Berger et al. (2022) also provide guidance on how market-level demand shocks affect firm wages and employment. Felix (2026) provides one example of how the structure of such a model can be used to estimate implied markdowns. In her setting (Brazil), she estimates that wages were roughly 50% of marginal product (i.e., $w/A \approx 0.5$, implying $\eta \approx 1$ under the standard Lerner rule).

In the absence of quasi-experimental variation, some papers use large firm-level changes in revenue, stock prices, or value-added per worker (Balke and Lamadon, 2022; Seegmiller, 2025; Volpe, 2026) as proxies for changes in the MRPL of workers, relying on covariates and parallel pre-shock trends to argue that the identifying variation is indeed exogenous and firm-specific. These "internal instruments" are useful for identification of labor supply elasticities across a variety of firms and contexts, but may be confounded with other unobserved shocks that are correlated with big changes in sales, value-added, or profits (e.g., corporate restructuring), and can complement more convincingly exogenous (to the firm) events or quasi-experimental shocks.

Non-Experimental Estimates: An older literature (Manning, 2003; Card and Krueger, 1995a) regressed separations/applications on wages, under the assumption that observable covariates were sufficient to absorb confounding heterogeneity. As shown in Table 1, these estimates suggested considerable monopsony power, and served as a springboard for more design-focused approaches. But the differences between the experimental/quasi-experimental and non-experimental estimates in Table 1 suggest that there is indeed scope for omitted-variable bias in purely observational designs.

3.5 Structural Estimates

Finally, a recent literature has adapted structural models from industrial organization to study labor market power. One strand of this literature aims to recover more flexible systems of labor supply facing firms by drawing on tools from structural demand system estimation. A different strand of this literature focuses on directly estimating workers' marginal revenue product and using this to estimate markdowns; the researcher then infers labor supply elasticities from markdowns.[26] Of course, as we discuss in Section 4.1 below, mapping these markdowns to residual labor supply elasticities requires assumptions about firm conduct that may not hold in practice.

Supply Curve Estimation. The first family of approaches leverages the widely used Berry-Levinsohn-Pakes (BLP) estimation of labor supply functions facing firms (1995). These models take as inputs wages and employment (or application) shares at the job by firm by labor market level. For instance, Azar et al. (2022) ($\eta \approx 4.8$) estimate a nested labor supply model on CareerBuilder data, using a variety of instruments for wages, including the number of vacancies, so-called "BLP" instruments (sum of log employment and sum of log employment squared of other firms in the same labor market), as well as the wages in other markets of other firms in the same market, and the average wage (and predicted wage) of the same firm in other markets.

Production or Revenue Function Estimation. A second family of approaches leverages production or revenue function estimation to estimate markdowns, rather than the labor supply elasticities. This approach explicitly models

[26] While markdowns may be of independent interest, one challenge is that researchers do not typically observe total monetary compensation, even in administrative linked employer-employee datasets. Further, the markdown naturally does not take into account non-wage amenities; if these are important, the wage markdown is not a sufficient statistic for the impact of monopsony power on worker welfare.

the firm technology and estimates it using micro-data on firm inputs. Production function parameters can be used to directly calculate the marginal revenue product of labor. Researchers typically compare these estimates to wages to obtain estimates of markdowns and to map these markdowns to the labor supply elasticity using the Lerner rule—i.e., the markdown equals the inverse elasticity (see Section 4.1).

While some research uses quasi-experimental or experimental variation, the most common approach is to proxy productivity using flexible functions of either past investment or a contemporaneously chosen variable input (Olley and Pakes, 1996). The intuition behind this method is that other inputs (or past investment) are monotonic in expected firm productivity, allowing researchers to invert input demand functions to obtain a proxy for productivity. A recent paper using this approach is Yeh et al. (2022), who estimate wage markdowns in the U.S. manufacturing sector from 1976-2014 via production functions. They document that the gap between marginal product and wages narrowed from the late 1970s to the early 2000s but has increased in recent years. The authors estimate markdowns which imply that wages are, on average, 65% of marginal product. Other papers in this literature include Dobbelaere and Mairesse (2013), Brooks et al. (2021), Delabastita and Rubens (2025), and Mertens (2023). Estefan et al. (2026) provide quasi-experimental evidence of how a labor market policy shock affected monopsony power as measured using the production function approach. They find that the introduction of a ban on outsourcing in Mexico led to reductions in markdowns.

A recent paper by Fernald et al. (2026) comprehensively surveys the production function estimation literature and shows that many of the resulting parameter estimates are quite sensitive to modeling choices and measurement error. For example, assuming frictionless labor markets when there is monopsony power, or wage-posting where there is bargaining, can meaningfully change the first-order conditions used to recover parameter estimates. Further, modeling decisions about the production function and the error in the observed data matter a great deal for measuring marginal products. We echo their view that measuring market power going forward should promote "transparency, testing, robustness, and better data"(pg. 75).

Advantages and Disadvantages. Structural identification and estimation of production functions remain active areas of research (Ackerberg et al., 2015; Gandhi et al., 2020). One limitation is that these approaches require strong assumptions. Beyond the standard assumption that firms choose inputs and outputs to minimize costs, there are typically context-specific assumptions regarding firms' information sets and the variable versus fixed nature of inputs. Previous research has shown that differences in assumptions can lead to substantially different estimates (De Roux et al., 2024). For example, non-neutral factor augmenting technologies can bias estimates of labor's marginal product (Raval, 2022; Doraszelski and Jaumandreu, 2018). Disentangling product market power from labor market power requires even more assumptions when physical productivity is not observed, such as in service employment.

One concern in the context of the labor market is that unobserved variation in recruitment costs, amenities, or visibility to workers would also alter firms' employment decisions in a way which would bias production function estimates, even if the researcher can control for productivity.

3.6 Discussion

Even holding fixed the labor market and the time horizon, the identification assumptions required by the different empirical strategies vary in how transparent and credible they are. As we discussed in the Introduction, experimental studies that randomly vary wages, holding all else fixed, identify the average firm-specific labor supply elasticity. For “wage shocks”—such as exogenous policy changes—the employment response traces out the firm-level labor supply curve, requiring minimal assumptions about firm conduct.[27]

Even with credibly exogenous wage variation, a number of issuesâseveral of which we discuss in Section 4, can complicate the interpretation of observed employment responses as labor supply elasticities. A general concern is that headcount employment conflates worker and firm behavior. One simple reason for this is that firms choose how many workers to employ, and may change hiring standards even when exogenously forced to change wages (we develop another example in Section 4.3, where effort responses create a similar wedge). This concern parallels a familiar concern in the labor supply literature: observed hours often reflect employer-imposed constraints rather than worker preferences (Dickens and Lundberg, 1993). In that literature, researchers often turned to flexible hours environments or to settings where they could explicitly model those rigidities. The analogous shift in the monopsony literature has been to focus on well-defined behavioral margins—such as separations, quits, or applications—or to settings where the response of individuals to a wage change can be measured directly.

Failures in identifying assumptions may explain some of the differences in findings across types of studies. Demand-shock studies tend to yield substantially larger estimates than wage-shock studies (Table 1). Demand shocks may also trigger adjustments along non-wage margins, like amenities and

[27] The exclusion restriction holds if the firm does not make other changes—such as to amenities or recruitment efforts—to influence workers’ employment choices. We are interested in average treatment effects, and leave discussion of heterogeneity, e.g., due to non-compliance, to the side. For example, quasi-experimental studies estimate elasticities for compliers whose wages are affected, and for exogenous “demand shocks,” the complier group is instead chosen by the firm through its targeting of wage increases.

recruitment effort, which we discuss below in Section 4, that confound identification of the supply curve. As we show below in Section 4.3, rent-sharing owing to efficiency wages may further complicate matters, though its effect on the recovered η is ambiguous: if firms share rents by raising wages beyond what the supply curve dictates, the wage response γ^w is inflated and $\eta = \gamma^e/\gamma^w$ is biased downward rather than upward. Finally, nominal wage rigidities, either owing to efficiency wages or other reasons, may blunt the pass-through of demand-shocks to wages, attenuating γ^w and resulting in downwardly-biased estimates of η.

Wage-shock approaches, which do not rely on assumptions about how firms react to demand shocks, yield more consistent results: the mean estimate from both experimental and quasi-experimental wage-shock studies is approximately 1.4, compared with over 4 for demand-shock studies. While the similarity between these estimates and those of the experimental approach is encouraging, one potential explanation is that both approaches tend to focus on short horizons. Wage shock estimates still require the assumption that there are no contemporaneous changes in amenities or recruiting effort. In principle these are easier to check in any given context where the firm's wage policy is well-understood.

Additional assumptions are required for the production function approach. Here the researcher must be able to accurately identify production functions and map estimates of the marginal revenue product (and wage) to η. As we discuss in Section 4.3, this mapping relies on additional assumptions about how firms set wages that may not be satisfied in all contexts. As can be seen from Table 1, production function estimates of η tend to be somewhat larger than those using other approaches. One potential explanation is that firms are setting wages to elicit additional effort, as we discuss in Section 4.3, which would lead wages to be higher than would be predicted by the Lerner rule alone.

Finally, calibrated models obviously rely on correct specification of the model. While useful for probing the quantitative implications of monopsony relative to other forces, they require strong assumptions about firm conduct and inter-firm interactions. For example, Berger et al. (2026) strongly advocate for the importance of calibrated general equilibrium models in studying labor market power.

A clear next step is to understand why estimates vary across studies. The variation could reflect failures of identifying assumptions or genuine differences in the residual labor supply elasticity across labor markets; if the latter, different approaches may also be identifying different local average treatment effects. When there is heterogeneity in the response to a treatment, different instruments or identification approaches yield coefficients that are different weighted averages of the heterogeneous treatment effects (Imbens and Angrist, 1994).

In the end, however, despite methodological differences and market heterogeneity, most labor supply elasticity estimates fall within a fairly narrow band. The interquartile range in Figure 3 is 1.2-3.7, consistent with moderate to high monopsony power. While researchers might disagree about what is the most compelling methodology, the evidence, taken together, points strongly towards pervasive monopsony in the labor market. But while the firm-specific labor supply elasticity tells us about the amount of market power the firm has in the labor market, it only indirectly tells us about how firms actually set wages. We turn to the details of firm wage-setting in the next section.

4. Monopsony, Wages, and Constraints Beyond Labor Supply

The evidence summarized in Section 3 indicates that most estimates of η fall between 1 and 4—implying substantial monopsony power. In the simple monopsony framework, labor supply is the only constraint and the wage is the only choice variable. In this section, we consider how other constraints and instruments can interact with monopsony power in setting pay.

4.1 Wage Markdowns and Elasticities

Recall from Section 2 that each firm faces an upward-sloping labor supply curve: the employment share $\mathrm{P}(j|w_j, \mathbf{w_{-j}})$ is increasing in the firm's own wage, with elasticity η_j. Standard monopsony models feature both this upward-sloping supply curve and wage posting (Robinson, 1933; Manning, 2003; Boal et al., 1997). If a firm posts its wage and there are no other complications (e.g., amenities), the optimal wage can be found by maximizing profits, taking the other firms' wages as given (a Nash equilibrium of a simultaneous-move game). For expositional clarity, we assume throughout this section (unless otherwise noted): (i) a unit mass of workers, (ii) constant returns to scale in production, and (iii) a competitive output market. Under these assumptions, the firm's profits are:

$$\pi(w_j, \mathbf{w_{-j}}) = (A - w_j)\mathrm{P}(j|w_j, \mathbf{w_{-j}}) \tag{8}$$

where A is a measure of value-added per worker, $\mathbf{w_{-j}}$ is a vector of wages chosen by other firms, and the price of output has been normalized to 1.[28]

[28]Following the bulk of the literature, we focus on static wage competition between firms; in the conclusion we note that more work on tacit and explicit collusion would be welcome. Constant marginal productivity is not as restrictive as it may seem. Note that any constant

If profit functions are differentiable and concave in a firm's own wages, it is straightforward to rewrite the firm's optimal wage, and hence the Nash equilibrium, as the Lerner rule, which holds for each firm at the Nash equilibrium:

$$\frac{A - w_j}{w_j} = \frac{1}{\eta_j(w_j, \mathbf{w_{-j}})} \tag{9}$$

where $\eta_j(w_j, \mathbf{w_{-j}})$ is the elasticity of labor supply to the firm, holding the vector of all other wages $(\mathbf{w_{-j}})$ constant at the equilibrium values. This equation states that the wage markdown (the percentage gap between productivity and the wage) is equal to the inverse firm-specific labor supply elasticity.[29]

However, a variety of forces may make the link between markdowns and elasticities more complicated than Equation 9 suggests. For instance, when there is heterogeneity in the residual labor supply elasticity across groups of workers in the firm who must be paid the same wage, applying the Lerner rule to the average markdown does not recover the average labor supply elasticity.[30]

To see how heterogeneity in worker productivity complicates the linkage between markdowns and residual labor supply elasticities, suppose there is heterogeneity in marginal products A_i across worker types i (with share s_i employed at j), as well as heterogeneity in residual labor supply elasticities η_j^i to the firm j. Despite this heterogeneity, the firm sets a single wage. In this case, the first-order condition which comes from the firm's profit maxi-

returns production function where all the other inputs are competitively purchased can be re-written in a two-stage manner as having total revenue equal to $p_j l_j$.

[29]We focus on Bertrand, rather than Cournot, models of imperfect competition, with firms picking wages, rather than employment. For most labor markets we suspect wage policies are the primary strategic variable rather than quantities, though this is an empirical question.

[30]A number of new laws require firms to equalize pay across groups of workers with "similar work" (Gentile Passaro et al., 2026).

mization problem is $\sum_i s_i \frac{A_i - w_j}{w_j} \eta_j^i = 1$, which we can rewrite as:

$$\frac{E[A_i] - w_j}{w_j} = \frac{1}{E[\eta_j^i]} - \frac{Cov(\eta_j^i, A_i)}{E[\eta_j^i] w_j}. \tag{10}$$

Here the markdown (the left side of this expression) does not recover the average labor supply elasticity. Rather, it recovers a combination of the average residual labor supply elasticity and its covariance with marginal productivity. This is the simplest illustration of our overarching point that there may be disjuncture between markdowns and labor supply elasticities.

4.2 Non-Wage Considerations

In the simple model, the tight link between residual labor supply elasticities and markdowns implied by Equation 9 also captures the economic incidence of monopsony (lower wages and higher profits). But these implications for welfare break down if firms also influence workers' choices through amenity provision or recruitment effort.

Amenity Provision. Because employers can compensate workers with amenities rather than wages (Sorkin, 2018; Sockin, 2022), the wage markdown is not a sufficient statistic for the impact of monopsony on worker welfare. Instead, the optimal approach for an employer with monopsony power who follows the markdown rule in Equation 9 is to mark down the total value of the job (inclusive of wages and amenities) by the inverse labor supply elasticity. Whether the "markdown" on the total value is greater than or less than the markdown on wages depends on how workers value amenities relative to wages (Volpe, 2026).

There is a recent literature (surveyed in Mas (2025)) which has examined how monopsony power influences amenity provision. This literature typically finds a positive (or at least non-negative) correlation between firm wages and amenities (Maestas et al., 2023; Caldwell et al., 2026b; Mas, 2025). For instance, Lamadon et al. (2022) and Volpe (2026) infer the value of firm-specific amenities using information on firm wages and firm size: high amenity firms are those which are larger than their wage level would otherwise predict. Using this approach, they find that high-amenity firms are also high-wage firms. Dube et al. (2022) find that, in the low-wage labor market, wage increases from the introduction of company minimums were not offset by reduced amenities.[31] The non-negative correlation between wages and amenities suggests that the inequality in utility provided by different firms is similar, or larger than, the inequality in wages provided by different firms.

If firms do not differentiate amenity provision by worker, their optimal choice will depend on the preferences of the marginal—rather than average—worker, generating the "Spence distortion" (Spence, 1975). A standard result in the monopsony literature is that if a firm can set wages differently for two groups of workers, it should apply larger markdowns to the group that is less elastic (Robinson, 1933). This logic applies in a more general model in which firms offer both wages and amenities: the markdown on overall compensation should be larger for the group that is less elastic. Sharma (2023) finds that Brazilian textile firms under-provide amenities valued by women, consistent with the idea that women have lower firm-specific labor supply elasticities. She finds that amenities explain all of the monopsony-induced gender compensation gap among textile workers in Brazil.

Recruitment Effort. The simplest model also assumes that wages are the only tool a firm has to influence employment. If firms can influence workers' mobility through informational channels, such as through job ads or through other "marketing" efforts such as participating in job fairs or ad-

[31] This is consistent with early work by Holzer et al. (1991).

vertising on additional platforms, they have expenditures that do not enter worker utility but make it easier for the firm to recruit workers at a given wage. This is akin to advertising by product market monopolists.[32] More broadly, firms may respond to shocks by increasing (or decreasing) recruiting effort. This can be modeled as a shift of the labor supply curve. Such a response may complicate estimation of labor supply elasticities; this requires movements along the labor supply curve. This concern is particularly relevant for the product demand shock estimates reviewed in Section 3.4: if firms respond to positive shocks by increasing both wages and recruiting simultaneously, the estimated η captures both channels. While the Lerner condition could still hold for wages, the economic incidence of monopsony could be quite different, as resources would be transferred to these recruitment expenditures (either wages of human-resource department workers or intermediate inputs).

4.3 Efficiency wages, labor discipline, and the effort margin

In the basic monopsony model, wages only influence whether and where a worker works. In reality, wages may affect not only the recruitment and retention of workers, but also their productivity on the job. The responsiveness of effort to the wage can constrain firms' wage-setting. Intuitively, if workers exert more effort when pay is higher, a firm which lowers its wage must consider both the fact that fewer workers will be willing to work at that firm and that the remaining workers may exert less effort. There are a variety of models in which wages directly affect productivity, including models of adverse selection, gift exchange/reciprocity, health/psychology (Leibenstein, 1957), firm-specific human capital investment, and effort (Akerlof and

[32] Work by Del Carpio and Fujiwara (2026) and Burn et al. (2022) indicates that the content of job ads affects who applies; a literature in macroeconomics suggests that variation in recruiting intensity is quantitatively important (Gavazza et al., 2018; Bloesch et al., 2026).

Yellen, 1986). Recent work by Emanuel and Harrington (2025) suggests that these factors may be important; in the context of a major warehousing firm, they find that a plausibly exogenous increase in the firm's starting wage decreased separations and increased productivity. This provides direct empirical support for the effort margin, and—as discussed in Section 3—is one channel through which markdown-based estimates of η may diverge from direct measurement.

In the classic efficiency wage model, there is no heterogeneity in workers' outside options (Shapiro and Stiglitz, 1984; Akerlof, 1982; Akerlof and Yellen, 1990). In the labor discipline family of models, firms pay a rent above the outside option to secure effort. In these models, a worker may shirk and (if the firm detects the shirking) be fired. This ties the distribution of outside offers to the productivity of workers. We present a simple model to illustrate the challenges that arise when one combines monopsony and efficiency wage models. We do not view our exercise as the final word, but a preliminary exploration of the issues that arise.

Formally, suppose the probability of being fired if shirking is P, and effort is binary and costs c for all workers. Worker i exerts effort when:

$$e = 1 \iff \underbrace{w - c}_{\text{working}} \geq \underbrace{(1-P) \times w}_{\text{not caught shirking}} + \underbrace{P \times b_i}_{\text{caught shirking}} \iff w \geq \phi + b_i$$

where $\phi \equiv \frac{c}{P}$ is the premium needed to make effort-provision incentive compatible. In contrast to the standard efficiency wage model, the same wage will not satisfy all workers' incentive compatibility constraints at the same time, even if the participation constraints are not binding. Because output depends on effective labor supply (summing over both effort and the number of workers), the marginal cost of *effective labor* now embodies both the participation and effort margins. When the effort margin is relevant, the profit-maximizing wage will tend to exceed the pure monopsony wage, limiting the firm's incentive to fully exploit its monopsony power.

We assume a (smooth) distribution of the (unit mass of) workers' outside options F as given. In our analysis, the effort function $e(w)$ facing the firm depends on whether the firm can screen:

$$e(w) = \begin{cases} \frac{F(w-\phi)}{F(w)} & \text{if screening on } b_i \text{ is infeasible} \\ 1 & \text{if screening on } b_i \text{ is feasible} \end{cases} \tag{11}$$

While the labor supply constraint always binds when screening is feasible, the "no screening" scenario covers two cases. When the labor supply is binding ($L = F(w)$)—Case 2 below—the firm's wage-setting problem looks like that of a monopsonist, but with an MRPL adjusted for effort. However, it is also possible that the firm's profit-maximizing efficiency wage is sufficiently high that the labor supply constraint is slack ($L < F(w)$). In this case—Case 3—labor demand binds instead, and jobs are (randomly) rationed among workers willing to accept the offer at that wage.[33]

In Appendix Section A.3 we present the full version of this model. The model features workers with heterogeneous outside options and illustrates the potential importance of moral hazard considerations by focusing on incentives to supply effort, as in labor discipline models (Shapiro and Stiglitz, 1984; Bowles, 1985). This model draws on the one presented in the open-source CORE economics textbook (CORE, 2024), and differs from a standard monopsony framework in that a higher wage affects both the number of workers employed at the firm and the average amount of effort exerted by hired workers.[34] The model assumes there is a firm with a concave produc-

[33] In all cases, there are some parameterizations in which the firm's optimal wage exceeds the worker's marginal revenue product of labor and in which a firm would prefer to shut down rather than employ workers. For simplicity, we assume that the firm finds it profitable to operate.

[34] We thank David Romer for his extensive feedback and suggestions on developing and

tion function G which maximizes profits.

$$\max_{w,L} G(e(w)L) - wL \text{ subject to } L \leq F(w). \tag{12}$$

In Appendix Section A.3, we show that wages differ depending on whether the firm can screen workers based on their outside options (though the firm still has to pay a single wage).

Full Screening. Figure 4 shows these three cases for a log production function, $G(eL) = a\log(eL)$, and a shifted-power labor supply function, $L(w) = C(w-\mu)^\kappa$ for $w \geq \mu$, $L(w) = 0$ otherwise, where $\mu > 0$ is a parameter of the labor-supply function and C is a positive constant (see Appendix A for details). The log production function ensures independence of the marginal revenue product of labor from effort. The shifted-power labor supply function (used in Card et al., 2018) makes the labor supply elasticity explicitly a function of μ, as well as κ. This functional form also nests the constant elasticity case, which is commonly used in the empirical and structural literature.

Panel A illustrates Case 1 (full screening), where monopsony and labor discipline are combined. As compared to the standard monopsony case, now the firm's inverse supply of effective labor is shifted up by ϕ from $F^{-1}(w)$ to $F^{-1}(w-\phi)$; there is an analogous shift back in the MCL schedule, which is now based on non-shirking workers only. The firm will always turn away workers who are willing to accept the job, but would shirk because their reservation wage is too high. Here there is involuntary (efficient) rationing of jobs, as $F(w-\phi) < F(w)$. As the figure shows, wage setting occurs through solving the monopsony problem but with the modified *non-shirking* MCL, which leads to wages still being set below the marginal product, but higher

characterizing the "monergony" model.

than in the pure monopsony model.

As we show in Appendix Section A.3, the firm's optimal wage in Case 1 takes the form of:

$$w_1 = \frac{\eta(w_1 - \phi)}{1 + \eta(w_1 - \phi)}\text{MRPL} + \underbrace{\frac{\phi}{1 + \eta(w_1 - \phi)}}_{\text{Efficiency Wage Premium}} \tag{13}$$

where $\eta(.)$ is the standard residual labor supply elasticity. We label the second term the "efficiency wage premium." Note this reflects the direct cost of effort as well as a premium to deter shirking. With perfect monitoring, $\phi = c$, and the efficiency wage premium reflects only compensation for the cost of effort. This case provides an illustration of how efficiency wage considerations can limit the monopsony markdown and raise wages.

While the markdown is lower than under monopsony alone, in equilibrium employed workers earn rents above their outside option to deter shirking. There are some workers who would be willing to work at a wage below their marginal product who are nonetheless not hired by the firm. This rationing exists under the assumption that firms can screen on outside options for employment, but still post a single wage. The combination of monopsony and labor discipline allows involuntary unemployment to co-exist with monopsonistic voluntary unemployment and employer labor shortages. Firms would like more non-shirking workers at the going wage, but cannot get many workers because unemployed workers' outside options are sufficiently high that they cannot commit to exert effort at the going wage, but not so high that they do not want to work.

No Screening. In the perfect screening case the incentive compatibility constraint creates a "cliff" at the posted wage: the last worker hired exerts effort and has marginal revenue product strictly greater than the wage, while the first worker not hired would shirk, producing zero, and so would have a marginal revenue product strictly less than the wage. But what happens if the firm has no ability to screen hires on their reservation wage? Here, the optimal wage depends on whether the labor supply constraint is binding. When labor supply binds—shown as Case 2 in Panel B of Figure 4—a firm that wants to hire more workers must raise its wage. When it does this, the marginal hires shirk; however, some of the incumbent workers stop shirking as a result of the wage increase. Overall, this means that the marginal cost of non-shirking labor is still increasing. The firm's optimal wage takes the traditional markdown form, so wages are below marginal product, but the effort-adjusted MRPL_{eff} also accounts for the share of workers who shirk after being hired.

$$w_2 = \frac{\eta(w_2)}{1+\eta(w_2)} \times \underbrace{e(w_2)\text{MRPL}\left(1+\frac{\epsilon(w_2)}{\eta(w_2)}\right)}_{\text{MRPL}_{\text{eff}}} \tag{14}$$

While the markdown appears identical to the standard monopsony model, it is applied to a modified marginal revenue product of labor, which is adjusted for endogenous effort via the effort elasticity $\epsilon(w) = e'(w)w/e(w)$ (which itself depends on $\eta(.)$ and w). In Appendix Section A.3 we derive this term and show that this modified MRPL captures the fact that wage increases affect the supply of effective labor among incumbent workers; in a standard monopsony model, wage increases do not change the productivity of these workers.

In Figure 4, Panel B, this yields an "adjusted" MRPL_{eff} that lies above the MRPL, leading to a wage exceeding the monopsony level. However, as we show in Appendix Section A.3, if a higher wage attracts disproportionately more shirkers, MRPL_{eff} could lie below the MRPL, leading to a lower wage than in standard monopsony. Put differently, efficiency wage concerns could either dampen or amplify the monopsony markdown of wages.

In this variant of the model, there is equilibrium shirking and "termination for cause", as some hired workers do not exert effort. While there is no queue of unemployed workers to get hired, unlike the traditional efficiency wage model, there are equilibrium involuntary separations.

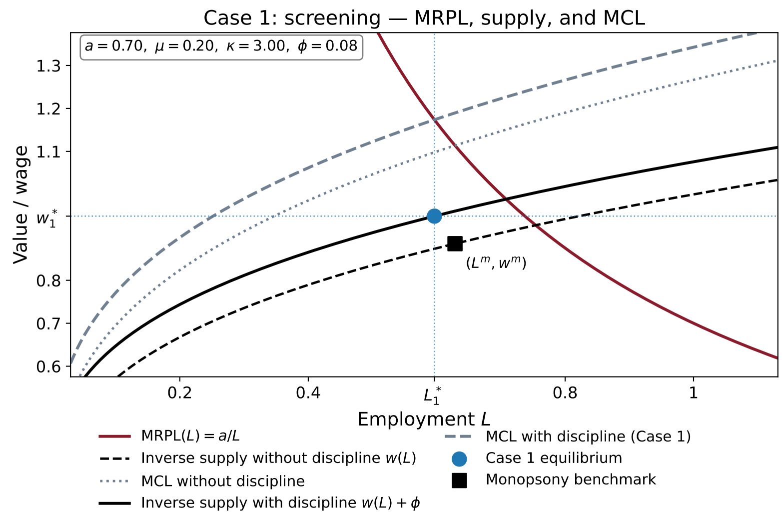
When the labor supply constraint does not bind, the distribution of outside options, as well as the ability to screen on outside options, still matter through the shape of the $e(w)$ function (which depends on the distribution of outside options). We illustrate Case 3 in Panel C of Figure 4. In Appendix Section A.3, we also show that, even in the cases where the labor supply constraint is not binding, η still enters as a determinant of the wage by influencing the shape of the effort supply function facing the firm. The Solow condition becomes:

$$\frac{de(w)}{dw} = \frac{e(w)}{w} \iff \frac{w_3}{w_3 - \phi}\eta(w_3 - \phi) - \eta(w_3) = 1 \tag{15}$$

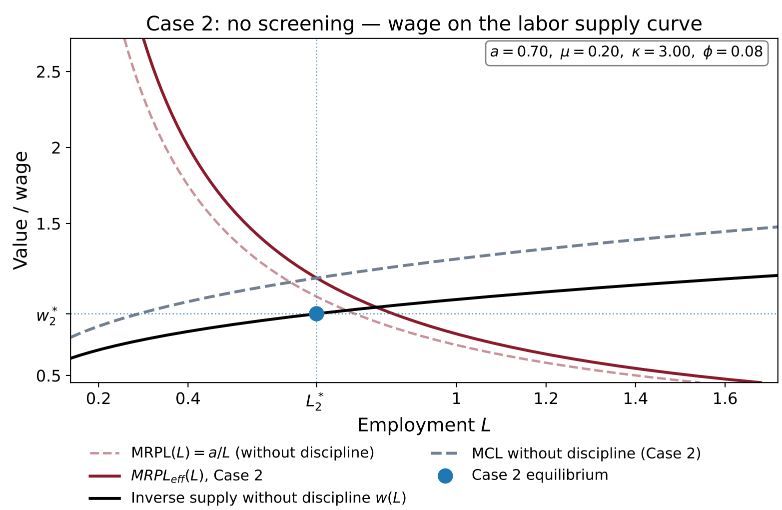
Note that in Case 3, even though the labor supply constraint isn't binding, the elasticity of labor supply, $\eta(.)$ (assessed at different values of the wage), still enters the equilibrium wage condition. This is because the distribution of the outside options of workers matters for both the level of the wage and how changes in the wage affect workers' effort. Via the Solow condition, the responsiveness of effort to wages enters into the firm's optimal wage.[35]

We show in Appendix Section A.3 that when η is constant, $\frac{dw}{d\eta} > 0$, so that an increase in η (e.g., driven by an increase in labor market competition) increases wages even though labor supply is not binding. Workers here are paid their marginal product, and so there is no markdown, but there is involuntary unemployment. There is also no rent-sharing in the sense that

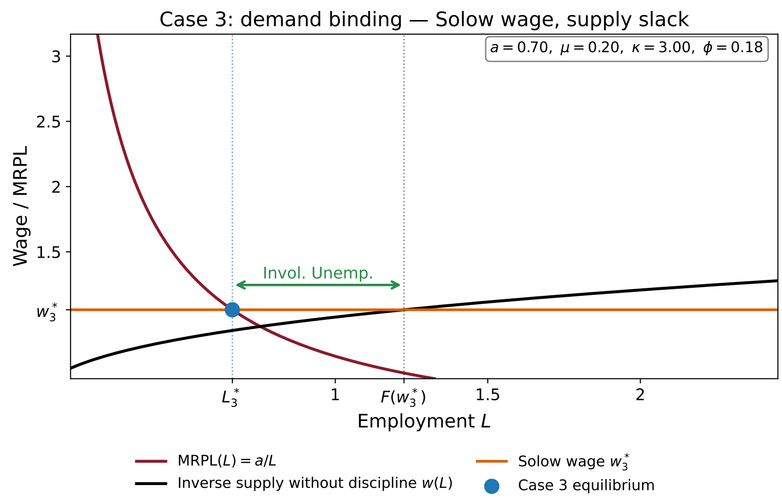
[35] In Appendix Section A.3 we show this model nests the Shapiro-Stiglitz model as a special case as the distribution of outside options approaches a point mass. As $\kappa \to 0$ the equilibrium wage converges to the Shapiro-Stiglitz value of $\mu + \phi$.



(a) Full Screening: Effective Labor Supply Constrained Firms (Case 1)



(b) No Screening: Labor Supply Constrained (Case 2)



(c) No Screening: Labor Demand Constrained (Case 3)

Figure 4: Monopsony and Labor Discipline

Notes: This figure illustrates employment and wages at a firm with a $\log$ production function facing a distribution of outside options (upward-sloping labor supply), and who needs to pay a (constant) rent, c/P, for employed workers not to shirk. The distribution of outside options is parameterized by $F(w) = (w - \mu)^\kappa$, and so the inverse supply curve is given by $w(L) = \mu + L^{\frac{1}{\kappa}}$. Panel (a) shows the case with full screening, where the labor supply of non-shirking workers is given by $F(w - \phi)$ and inverse supply $w(L) = \mu + L^{\frac{1}{\kappa}} + \phi$. Employment L^1 is lower than the monopsony-only level L^m, while wage $w^1 > w^m$. Panels (b) and (c) show the case when screening of hires is not possible. Panel (b) shows that when the labor supply is binding, the wage is marked down on the effort-adjusted MRPL_{eff} based on the same labor supply elasticity $\eta(w)$ as in the standard monopsony case. Panel (c) shows the case where labor supply is not binding.

wages do not respond to changes in firm productivity (unlike in Cases 1 and 2, where the wage is a function of MRPL), as wages are pinned down by the Solow condition, which does not depend on MRPL.

Discussion. Risking the reader's ridicule, we have used monergony as the name for this combination of efficiency wages and monopsony where both effort and labor supply considerations matter. The term better fits a single buyer in the market for *work* (not just any input), and the use of the root erg underscores that the wage is used by the firm to elicit skill and effort—not simply to purchase bodies and hours of "labor power" $L(w)$.

We are not the first to highlight the linkages between monopsony models and efficiency wage models: Manning (2003) and Manning (1995) both discuss this linkage. A model with general efficiency wages (including firm-specific human capital) combined with monopsony is also presented in Emanuel and Harrington (2025). Kline (2025) presents a formula similar to Equation 13 for a generic efficiency wage model. Our treatment is distinct in emphasizing the link between the firm-specific labor supply elasticity and the effort elasticity specific to the labor-discipline model, where it is the gap between the wage paid and the marginal worker's outside option that constrains the supply of non-shirkers. We also highlight how firms' ability to screen workers ex ante based on outside options determines the magnitude of the interaction.

The monergony model also reinforces a point we raised in Section 3.1: when efficiency wage considerations are strong enough that the firm is labor demand constrained (Case 3), headcount employment is determined by the firm's demand for effective labor rather than by the supply curve, and the observed employment response to a wage change need not identify η. Even in the supply-constrained cases, the relevant margin is effective labor rather than headcount. This provides further motivation for focusing on well-defined behavioral margins—such as separations, quits, or applications—that are less contaminated by firm-side selection and effort considerations.

However, this discussion merely scratches the surface of these issues; further research is needed to clarify the relationship between labor discipline and monopsony both theoretically and empirically.

4.4 Other Factors Affecting Firm Wage Setting

While labor supply is a key consideration in firms' wage setting, a variety of other factors may influence whether a firm chooses to post wages consistent with the formula specified in Equation 9 or to follow an alternative wage-setting protocol. These other concerns can move wages either above or below the level implied by Equation 9, depending on context. Thus in at least some contexts, these concerns may dampen the exercise of monopsony power. We have already seen how efficiency wages can play such a role; other examples we discuss below include fairness concerns, limits to optimization, and strategic interactions or collusive practices.

Fairness norms. Firms may consider workers' fairness norms in setting wages. These concerns could reflect horizontal (between workers in a firm or labor market) or vertical equity (e.g., between worker and the owner or a manager) considerations, and may reflect behavioral concerns on the part of both workers and managers.

Workers may care not only about their own wage, but also about how their wage compares to the wages paid to other workers in a peer group (which may be within or outside the firm). There is growing evidence that workers make social comparisons at work (Fehr et al., 1998; Breza et al., 2018). These comparisons may lead workers to respond not only to changes in their own pay but also to changes in the pay of their coworkers. Dube et al. (2019) use plausibly exogenous wage differences within the stores of a national retailer to document that the probability a worker quits depends only weakly on comparisons with the market wage (implying substantial monopsony power) but strongly on the wages of his/her close coworkers. These results are consistent with fairness norms in the form of aversion to being paid less than their similar coworkers. Related evidence is provided by Breza et al. (2019), who find workers are willing to accept lower wages when asked in private than in public, owing to strong informal norms in village labor markets. Goerges and Nosenzo (2020) provide a review of the role of norms, including fairness norms, in the labor market.

It's worth noting that firms' inability to pay workers differently based on their reservation wages—possibly due to fairness norms—is a basis for the monopsony wage setting in Equation 9. At the same time, when there are different types of workers (say by skill type), such fairness norms can shape wage setting in other ways, such as inducing wage uniformity within the firm. The extent to which fairness norms constrain firms' wage-setting depends in part on the span of social comparisons. Horizontal equity concerns can lead to wage uniformity within the firm, as firms may not be able to freely set wages based on group-specific markdowns and productivity. This could explain why firm-specific wage premia (as estimated from models that include both firm and worker fixed effects) tend to be similar across groups of workers within a firm (Abowd et al., 1999; Card et al., 2018). The presence of within-firm social comparisons also offers a potential explanation for the rise of domestic outsourcing: that internal fairness norms make it difficult for firms to fully leverage their monopsony power over certain groups of workers. By outsourcing these workers, a firm can ease this constraint (Goldschmidt and Schmieder, 2017). For instance, a high-wage firm may choose to outsource janitorial workers or cafeteria staff to avoid paying those workers high wages. As Goldschmidt and Schmieder (2017) argue, the pattern of higher outsourcing among "high-rent" firms is consistent with outsourcing being used to lower labor costs and reduce rent-sharing. Many papers have documented a rise in the sorting of low-wage workers to low-wage firms; evasion of internal fairness norms is one potential explanation (Song et al., 2018; Card et al., 2013). Whether these deviations from classic monopsonistic wage setting amplify or reduce the overall wage markdown will differ by context.

While the above examples are of horizontal equity concerns, vertical equity or reciprocity concerns may also constrain monopsony power. For example, such vertical fairness concerns may start mattering for pay setting precisely when market competition is low. Fehr and Schmidt (1999) show that in highly-competitive contexts, fairness norms do not alter behavior, while in low-competition environments (or ones in which workers can "shirk") fairness has some scope to alter individual behavior. It is possible that monopsony power, defined as a low η, may be correlated with norms of fairness that restrain the exercise of that power, weakening the link between monopsony power and markdowns.

Weakening vertical equity norms may play an important role in the fall of rent-sharing documented by Acemoglu et al. (2025). They show that when companies moved to hire MBA-trained managers, this led to lower wages (especially for blue-collar workers) and increased quits, higher profits, and no change in productivity. This suggests that MBA-trained managers exercised monopsony power that was previously present, but was presumably restrained by internal workplace norms, where the supervisor had risen from the ranks of the workforce.

Finally, there are other more direct constraints on wages that can limit monopsony power. These include statutory minimum wages (Dube and Lindner, 2024), and also pay floors arising from sectoral collective bargaining (Card and Cardoso, 2022; Jäger et al., 2025). Moreover, fairness or social perceptions can affect reservation wages even without a legal minimum wage (Falk et al., 2006), possibly enforced through worker social sanctions (Breza et al., 2019). We discuss unions and minimum wages further in Section 5.

Limits to optimization. Organizational factors may also shape how firms exercise their market power in the labor market when setting wages. For instance, firms that operate in many markets sometimes pay the same wages in all markets, even when those markets differ in market conditions, such as the number of other firms or the wages paid by other firms (Hazell et al., 2025). This means wages can be either greater or lower than the local monopsonistic optimum. In one stark (but not unusual) example, Coviello et al. (2022) find that a major national retailer facing state-level minimum wage changes modified neither its compensation policy (except to comply with the local minimum wages) nor its product pricing—both of which were set uniformly at the national level. Firms also typically pay "round number" wages (Dube et al., 2026). In a competitive market, small deviations from the optimal wage result in significant profit losses, pushing firms towards precise wage setting. However, with monopsony power, the relationship between profits and wages becomes flatter around the optimal wage level. This flatter envelope means that even moderate deviations from the optimal wage might not reduce profits by a meaningful amount (Akerlof and Yellen, 1986). With monopsony power, firms' incentives to optimize, rather than satisfice, in wage-setting are blunted. Indeed, Dube et al. (2026) show theoretically and empirically that such mis-pricing is much more likely to exist in markets with some monopsony power. Whether such mispricing reduces or increases the wage markdown depends on the context, and would benefit from further research.

Strategic Interactions and Spillovers. Finally, while Equation 9 suggests that the elasticity to the firm is driven by both the wages at firm j and at other firms, most models of monopsony power assume that there are no strategic interactions in wage-setting: firms do not consider how changes to their own w_j in turn influence wages $\mathbf{w_{-j}}$ at other firms (Burdett and Mortensen, 1998; Card et al., 2018). This simplifies the analysis significantly. In contrast, models of oligopsony assume that firms do consider how their own wages affect competitors. So a rival's wage increase raises firm j's optimal wage; when firms internalize this complementarity, the equilibrium wage lies below the atomistic-firm benchmark, meaning that strategic interactions *amplify*, rather than mute, the exercise of monopsony power.

While such strategic interactions could have important implications for measuring labor market power, how can we assess their empirical relevance? Berger et al. (2022) structurally estimate an oligopsony model using the association between firms' local labor market share and their employment and wage responses to state corporate tax changes; they argue their evidence is consistent with a sizable amount of strategic wage setting. However, what about more direct estimates?

One way is to look at what happens when a single firm changes its wage: do competitors' behavior respond?[36] In a model of monopsonistic competition (where there are many firms), the answer would be "no," while in an oligopsonistic model, the answer would be "yes." The empirical evidence on the importance of strategic interactions based on cross-firm wage spillovers is mixed. Early work found evidence of wage spillovers in the market for nurses, studying spillovers from VA-hospital wage increases on neighboring non-VA hospitals (Staiger et al., 2010). By contrast, Derenoncourt and Weil

[36]While numerous papers have documented spillovers from sectoral collective bargaining (Bassier, 2024; Hermo, 2025; Demir, 2026), spillovers from a wage increase covering a set of employers are, of course, conceptually distinct from spillovers from a single firm.

(2025) find that large low-wage employers' adoption of voluntary wage floors had no effect on the wages of others operating in the same labor market, even though these policies substantially raised pay for the company's own workers. Their estimates rule out anything but a small amount of strategic interaction, and are at odds with the conclusions reached by Berger et al. (2022). Similarly, Roussille and Scuderi (2026) test for strategic wage-setting on a tech platform in a model that nests different types of firm conduct and do not find evidence for strategic interactions. In contrast, Jungerman (2024) leverages administrative data from France to show that wage changes of labor-market "leaders" are passed along to other firms; this fact is used to motivate a quantitative model where the monopsony power of large granular firms depresses human capital accumulation.

When thinking about strategic interactions, it is also important to consider the relevant scale at which wages are set. There is evidence suggesting that many firms set wages at a level broader than a local labor market. For instance, Hazell et al. (2025) show that many employers post the same wage in many labor markets for the same job, rather than customizing to local labor market conditions. Hjort et al. (2026) show that multinationals transmit minimum wage increases in their headquarter countries to their subsidiaries abroad.

Cross-employer interactions can also take the form of collusion. While anecdotes abound, the most common reference in economics is still Adam Smith's (1776) oft-quoted "We rarely hear, it has been said, of the combinations of masters, though frequently of those of the workman. But whoever imagines, upon this account, that masters rarely combine, is as ignorant of the world as of the subject." There have been some documented cases of employer collusion to set wages in the tech sector (Gibson, 2025; Herrera-Caicedo et al., 2025), but it is not clear what the overall prevalence is. Delabastita and Rubens (2025) study labor market collusion in the historical

Belgian coal sector. Relatedly, while unions have been extensively studied in collective bargaining environments, the role of employer associations in maintaining collusive wage standards is also understudied (Martins, 2020). Sharma (2025) presents evidence of labor-market collusion within an employers' association in the Indian textile sector.

As we discuss in the conclusion, more work on the extent to which there are strategic interactions in wage-setting—or to which firms collude to set pay—is needed to understand the broader relevance of these channels.

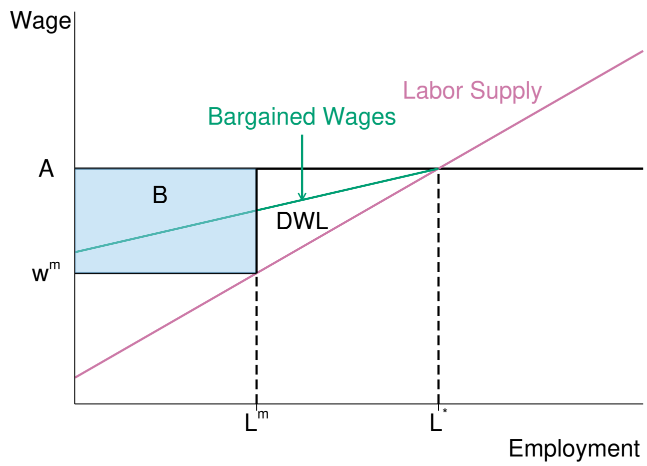
4.5 Beyond Wage Posting

The one-to-one relationship between residual labor supply elasticities, wages, and markdowns (Equation 9) also breaks down if firms do not post wages. Figure 5 shows how a firm facing an upward-sloping supply curve would set wages if it opted for wage posting or if it opted to individually negotiate wages with each worker.[37] Again, assuming the simplest monopsony model, a firm which posts wages will choose wages w^m, which are "marked down" from A according to the Lerner rule. However, if the firm can bargain with workers individually, any wage between A and the individual worker's outside option (as given on the labor supply curve) will be admissible. For instance, the blue line on the figure shows wages that result from efficient bilateral bargaining between each individual worker and the firm.

As the Figure illustrates, wage-setting protocols other than wage posting can lead a firm facing an upward-sloping labor supply curve to pay workers less than their marginal product. Only some of these would lead firms to hire

[37] As Card (2022) discusses, there is little evidence on the prevalence of different wage-setting mechanisms. Caldwell et al. (2026a) provide evidence from Germany that suggests the majority of firms are willing to differentiate pay between workers they perceive to have the same productivity; the fact that many workers fail to negotiate (even when they have an outside offer) limits the extent to which this practice may generate meaningful variation in pay.

Figure 5: Wage Posting vs Other Wage Policies



Note: This figure depicts a situation where a firm faces an upward-sloping labor supply curve where workers of equal productivity A vary in their outside options. If a firm cannot wage discriminate and must set a single wage w, it sets a wage w^m, which is below the marginal product of labor, A. The firm earns rents B. Thus the monopsony wage is below marginal product, transferring B from workers to firms, and causes deadweight loss given by the triangle DWL. The bargaining line is the wage for each worker assuming individual bargaining under perfect information: the wage is a fixed share of the gap between marginal product and outside option.

fewer workers than they would in a perfectly competitive market (see, e.g., Jarosch et al., 2024, for a model without such distortions). Wage posting as in the traditional monopsony model entails inefficiently low employment, because the firm offers the same wage to all workers and must weigh the cost of raising wages on inframarginal workers when it expands.

The Figure also illustrates that the labor supply elasticity (or more generally the distribution of outside options) does not reveal the gap between individual workers' marginal product and their wage. With wage posting, the deadweight loss and the split of the surplus are determined by η. Outside of wage posting, the residual labor supply elasticity—as a measure of the dispersion of workers' private information about outside options—may be relevant for efficiency in more general mechanisms.[38] Separately, even when the allocation is efficient, the residual labor supply elasticity can shape the division of surplus. In the constant MPL case (as depicted in Figure 5), the distribution of outside options—which also governs the residual labor supply elasticity—determines the amount of total surplus to be split between employer and workers in an efficient bargain, even when the split is determined by an outside bargaining parameter.

Summing up this section, while some definitions of monopsony have hewed to profit-maximizing posted wages, we think this is unnecessarily restrictive. Future work could integrate more realistic models of firm behavior with the omnipresence of heterogeneity in workers' outside options, as measured by the elasticity of labor supply facing the firm.

[38] Bulow and Roberts (1989) show that optimal auctions allocate to the seller with highest marginal revenue, and the analogue for the labor market would be the employer allocating the job to the lowest marginal cost worker.

5. Implications of Firm Monopsony Power

What are the implications of monopsony power for policy-making and institutional design? Naturally, an increase in labor market power (i.e., a decrease in the residual labor supply elasticity) can have undesirable impacts on both wages and hiring. Moreover, even if an increase in labor market power is not reflected in immediate changes in wages and working conditions, there may be reasons for concern. For instance, employers may not be able to commit not to use market power in the future. For these reasons, it is valuable to understand how policies attenuate or counteract labor market power. In this section, we provide some examples of policies which interact with firm monopsony power.

Competition Policy. Several authorities that regulate competition in the product market have begun to more actively police firms' behavior in the labor market. For example, antitrust authorities have begun to screen mergers based on harms due to labor market competition (Prager, 2025; Naidu et al., 2018; Marinescu and Hovenkamp, 2019). There is direct empirical evidence that mergers have decreased wages and employment in contexts where they dramatically increase concentration (see, e.g., Arnold, 2025; Prager and Schmitt, 2021; Benmelech et al., 2022; Arnold et al., 2026; Lagaras, 2025). For instance, Prager and Schmitt (2021) find that hospital mergers increased concentration and led to decreased wage growth, especially in areas with weak unions. However, because firms have monopsony power for a variety of reasons—not just size and concentration—merger review is unlikely to be the preferred instrument for combating monopsony power (Naidu and Posner, 2022). Antitrust authorities have a variety of tools—beyond merger review—to promote competition and to police anti-competitive behavior. Anti-competitive behavior includes unilateral conduct, such as vertical restraints (no-poach, no-hire, and non-compete agreements), as well as horizontal conduct such as collusion and cartelization.

The legal system can facilitate monopsony by allowing voluntary contracts that restrict ex post mobility. These restrictions function by narrowing workers' consideration sets, effectively lowering the residual labor supply elasticity. One prominent example is franchise agreements, which prohibit units from recruiting or hiring workers from other units within the same chain (Krueger and Ashenfelter, 2022).[39] The Federal Trade Commission has recently focused on "unfair" restraints on competition, and recently proposed a ban on non-compete agreements (covenants not to compete) for most workers.[40] Starr (2026) shows that non-compete agreements are surprisingly widespread, even in settings where they are unenforceable and even in settings such as the fast food industry where there were no plausible concerns regarding theft of intellectual property.[41] The rise of "fissured" labor markets and outsourcing has introduced the possibility of vertical restraints in which employers via temp agencies or gig platforms contractually prohibit clients from making direct offers to workers (Weil, 2014); understanding the prevalence of such restraints is an important area for future research.

Recent policy and legal discussions have proposed widening the objective of antitrust policy beyond maximizing economic welfare (Wu and Cashman, 2018). These objectives include economic output, ideas of fairness, or simply maintaining a competitive process as required by law. Even when the normative objectives are non-economic, the residual labor supply elas-

[39] See also work by Callaci et al. (2024b) and Callaci et al. (2024a).

[40] Federal courts, as of the time of writing, have challenged the FTC's authority to impose such a ban.

[41] There has been an active and growing recent literature on the theory and evidence around non-competes that we do not have space to consider here (see, e.g., Prescott et al., 2016; Shi, 2023; Johnson et al., 2025; Starr et al., 2021; Krueger and Ashenfelter, 2022; Gottfries and Jarosch, 2026).

ticity can still be informative. Indeed, perspectives that emphasize that antitrust should prioritize maintaining a competitive process (Werden, 2024; Steinbaum and Stucke, 2020), regardless of economic incidence, may find the elasticity of residual labor supply particularly useful, even when it does not map neatly into the monopsonistic markdown.

While a great deal of progress in competition policy has come from arbitraging ideas of competition in the product market into the labor market, there remain important differences, some of which we have discussed in Section 4. For example, in the presence of wage-rigidities as in the monergony model, mergers that increase labor market power may not manifest as wage cuts, but instead as layoffs, which do not have a ready product-market analogue.

Issues of market definition loom large in antitrust policy. Scholars have used geographic boundaries, or commuting-zone by industry or occupation labels, to define labor markets, and have calculated concentration measures using shares of employment (Rinz et al., 2018), flows of employment (Arnold, 2025), payroll (Berger et al., 2022), and vacancies (Azar et al., 2020). Many labor markets are indeed quite local, with workers not applying to jobs very far from where they live (Manning and Petrongolo, 2017). However, the surprising uniformity of wages within many firms—and the prevalence of national wage-setting policies—calls into question whether local concentration is the only, or even the best, lens through which to assess how changes in concentration may have affected firms' wage-setting power and conduct. Issues of market definition in the labor market may also sometimes need to account for involuntary unemployment and demand-constrained firms, which may alter the appropriate measure of concentration.

Minimum Wages and Unions. In the presence of labor market power, policies that mandate higher wages (such as a minimum wage policy) may also improve welfare (Stigler, 1946). A key empirical prediction of a monopsonistic competition model is that increasing the minimum wage could increase employment (for a range of minimum wages).

The fact that, when firms have monopsony power, increases in the minimum wage can increase employment would seem to suggest that the size and sign of the employment response to minimum wage policy may help shed light on the size of the residual labor supply elasticity. In their recent review of the minimum wage evidence, Dube and Lindner (2024) document that, while the empirical literature has produced a range of estimates—including estimates which suggest the minimum wage decreases employment (Clemens et al., 2018; Jardim et al., 2022)—most of the literature has yielded relatively small employment effects (Cengiz et al., 2019, Derenoncourt and Montialoux, 2021): the median own-wage elasticity of employment in the minimum wage literature is around -0.13, indicating a very modest impact. Following Card and Krueger (1995b), researchers have often appealed to labor market power and firm heterogeneity as an explanation for positive employment or limited disemployment effects (Manning, 2021).

This interpretation may well be sensible, but it is useful to recognize that monopsony power is neither necessary nor sufficient for a small disemployment (or small employment) effect. For example, Harasztosi and Lindner (2019) find modest disemployment effects in their study of a large minimum wage increase in Hungary, which they rationalize with low product demand elasticities and input substitution elasticities in a competitive model. Even in models with monopsonistic features, the extent and sign of the employment effect and welfare effects depend on model parameters (Berger et al., 2022).[42]

[42] A "robust monopsony regulation" approach is suggested in Guo and Shmaya (2025); applying this to the labor market would imply that a mixture of payroll subsidies (or EITC) and minimum wages is the regret-minimizing policy for regulating a monopsonist, with the share of the former increasing in the welfare weight put on firm profits.

On the other hand, the existence of labor market power implies other testable predictions about the effects of the minimum wage—many of which have been found to be consistent with the empirical evidence. For instance, a number of recent papers have tested whether (as monopsony models predict) the employment effect of the minimum wage differs in more and less concentrated labor markets (assuming that more concentrated markets have more monopsony power). Azar et al. (2024), Corella (2020), and Wiltshire (2025) find that the employment effects of minimum wages are more positive in more concentrated labor markets.

Models in which firms have monopsony power typically predict that an increase in the minimum wage will lead to a reallocation of employment across firms, shifting workers from lower-productivity to higher-productivity firms; Dustmann et al. (2022) (German national minimum wage), Rao and Risch (2026) (U.S. state minimum wages), Bassier (2023) (South African sectoral wage floors), and Demir (2026) (German sectoral wage floors) show empirical evidence consistent with this prediction. Search-based models predict that an increase in the minimum wage will lead to a reduction in workers' separations (as workers are re-allocated away from lowest-wage and highest-turnover firms); there is widespread evidence for this prediction (Cardoso and Portugal, 2005; Brochu and Green, 2013; Dube et al., 2016). Models in which firms have monopsony power also make novel predictions about the direction of labor-labor substitution that are borne out in recent work (Datta and Machin, 2024).

More generally, as Dube and Lindner (2024) point out, the minimum wage evidence resists a simple, labor-market-only framework. Monopsony models can rationalize modest employment responses and reductions in worker turnover, but explaining the near-complete pass-through of higher wage costs to prices alongside small employment effects likely requires richer models—ones that incorporate product-market pricing power and behavioral features. The observed rise in productivity following minimum-wage increases suggests that "monergony"-style mechanisms may play an important role in understanding the channels of adjustment (Coviello et al., 2022). Future work can also shed light on how additional margins such as recruitment and amenities respond to a mandated wage floor in the presence of monopsony power. For example, endogenous recruitment may shrink the employment response toward zero and mitigate any loss of profits—providing another channel (besides price pass-through) that could help rationalize findings of near-zero employment and profit effects (Rao and Risch, 2026).

Similar dynamics apply to collectively bargained wage floors (Jäger et al., 2025). Many recent papers have found that markdowns are lower in unionized firms (Azkarate-Askasua and Zerecero, 2025).[43] However, unions influence recruitment policies and amenities as well as setting rigid wages (Corradini et al., 2025; Arold et al., 2025). How collective bargaining shapes labor market outcomes in the presence of monopsony and non-wage instruments is a promising direction for future research (Lagos, 2025).

Immigration. The presence of monopsony power may also shape how immigration affects the labor market. The canonical competitive model has long been used to study the labor market effects of immigration, and is summarized in the recent handbook chapter by Dustmann and Schönberg (2025). But recent work has suggested that imperfect competition in the labor market may represent an additional channel (beyond aggregate labor supply or labor demand) through which immigrants may affect native wages. If employers cannot perfectly wage discriminate between natives and migrants

[43] Dodini et al. (2026) find that the union premium in Norway is increasing in local labor market concentration, and a number of papers have found that the correlation between wages and concentration is smaller in unionized labor markets (Benmelech et al., 2020). In the presence of monopsony profits, employers are willing to spend considerable resources to resist unionization (Wang and Young, 2024).

(an assumption which may be more reasonable in some settings than others), and migrants have a lower elasticity to the firm, then increases in the relative supply of migrants will decrease the average labor supply elasticity to the firm (Amior and Manning, 2026; Borjas and Edo, 2023; Gyetvay and Keita, 2023; Amior and Stuhler, 2026). As Amior and Manning show, this lower elasticity can increase the degree of labor market power faced by natives—even if immigration has only a weak (or even positive) effect on wages.

In addition, temporary work visas—which restrict the mobility of temporary migrants—remain one of the few areas of modern labor markets with legally protected monopsony power. Recent papers have studied monopsony power in guest worker programs linked to rich administrative data in rich countries (Kroft et al., 2026; Townsend and Allan, 2025; Naidu et al., 2016). For example, in the restrictive context of the UAE, Naidu et al. (2016) document that a reform that made it easier for workers to move upon contract expiration raised the firm-specific labor supply elasticity from 1 to 2.5.

One of the most studied work visa programs is the H-1B visa program for skilled workers in the United States (Doran et al., 2022). Whether the ex-post monopsony power granted to firms has a significant impact on the wages and employment of these workers is an unresolved question: Hunt and Xie (2019) suggest that, in the United States, resident status does not change workers' wages, while Wang (2021) suggests it increases wages by 5.5%.[44]

While the preceding discussion has focused on the exercise of monop-

[44]Kim and Pei (2022) leverage the randomness in the H-1B lottery system to generate exogenous variation in the HHI across H-1B labor markets, and find that wages for new migrants are significantly lower in high-concentration labor markets. Depew et al. (2017) use H-1B and L1 worker data from six Indian IT firms to look at job mobility, and find that workers switch firms and return to India in sync with the U.S. business cycle. In sum, whether the H-1B visa program is a source of monopsony power remains an open question deserving of more research, as is research on less skilled visa categories like H2-A (agriculture) and H2-B (unskilled non-agriculture).

sony power in markets with immigrant labor, more research is needed to determine whether firms use all the monopsony power they have, and whether there is an optimal degree of employer power in the guest worker context. This context is one in which some of the ex ante recruitment expenditures and training costs of migration are borne by employers, and are presumably defrayed by ex post mobility restrictions. What differential firm targeting of recruitment and amenities to migrants does to native wages in models with labor market power remains an open question. Similarly, how immigration interacts with other constraints on employer wage-setting in frictional labor markets remains unexplored: immigrants may weaken fairness constraints on wage-setting or lower outside options, which would change wages for native workers independently from effects on marginal products or markdowns.

Monopsony and the Macroeconomy. Finally, labor market power has implications for macroeconomic policymaking. Monopsony power varies with the business cycle: many studies have documented that quit rates decline in recessions and that separation elasticities become smaller in magnitude, implying that firms have more monopsony power (Bassier et al., 2022; Webber, 2022; Hirsch et al., 2018).

From a theoretical perspective, tighter labor markets can lead to greater job-to-job flows in job ladder models like Burdett and Mortensen (1998), where a higher offer arrival rate in a tighter labor market would also tend to increase the separation elasticities (see also Moscarini and Postel-Vinay (2018)). Autor et al. (2023) find that an exceptionally tight labor market in the aftermath of the COVID-19 pandemic led to a much more competitive market in terms of elevated quits (the "Great Reshuffle"), increased separation elasticities, a reinvigorated job ladder, and reallocation of non-college workers toward better paid (and possibly higher productivity) firms. This led to a substantial reduction in wage inequality over the 2021-2023 period. Overall, this early set of evidence points to possibly important interactions between macroeconomic conditions and labor market competition.

On-the-job search and wage posting, along with fairness constraints, also produce novel insights about the nature of wage rigidity and unemployment fluctuations (Fukui, 2020). While macroeconomic measures of labor market tightness have focused on involuntary unemployment, along with vacancies as the primary ingredients, the work above suggests that including measures of separation rates and elasticity of labor supply in the calculation of slack would yield a more accurate measure of labor market tightness. Finally, while the presence of involuntary unemployment may appear at odds with the presence of monopsony power, this need not be true. In Section 4, we provided an example where firms screen hires based on reservation wages in the presence of unobserved effort and monopsony power. Further work along the "monergony model" of Section 4.3—along with a more realistic partial screening ability among firms—seems like a promising avenue for explaining how nominal wage rigidity and involuntary unemployment can co-exist with monopsony power.

Adding other constraints on wage-setting can help square low residual labor-supply elasticities with the observed labor share of income. Multiplying both sides of the markdown rule by L/Y implies that the labor share of income should be the markdown times the output elasticity of labor. The macro output elasticity range of 0.81 to 0.63 from Vollrath (2024), and a residual labor supply elasticity of 3 (the average from Figure 3) would imply labor shares between 0.6 and 0.47. The lower end of this range is well below current labor shares, suggesting that markdowns due to monopsony alone cannot be the full story of wage-setting. Given the uncertainty in the aggregate output elasticity of labor, this statistic alone doesn't discriminate between monopsony and other models of the labor market.[45]

6. Conclusion and Directions for Future Research

As this review documents, researchers have employed a wide range of empirical strategies to estimate the residual labor supply elasticity. There are several promising avenues for future research. One direction is to produce additional direct estimates of the residual labor supply elasticity, perhaps by replicating gig-economy experimental designs in other low-wage markets where researcher interventions are ethically and practically feasible. Another approach would be to leverage the survey methods often used to elicit workers' preferences over amenities (Maestas et al., 2023; Mas, 2025). Section 3 provided examples in which carefully designed surveys helped clarify distinct sources of monopsony power. Additional research combining surveys with administrative datasets could further illuminate both the magnitudes and underlying sources of monopsony power. Surveys, along with larger administrative datasets, also give the power to trace out parts of the residual labor supply curve that are far away from the margin of current wages, allowing both more group-specific heterogeneity and more flexible functional forms than the widely used constant elasticity specification.

One important question is the economic incidence of monopsony once employer and worker investments are considered. In the product market, patents both provide firms with monopoly power and incentivize investment in research. It is not clear whether, in practice, monopsony power incentivizes entrepreneurs to enter and invest or incentivizes employers to supply general human capital. The profits to monopsony power may be only quasi-rents, eaten up by ex ante investments, recruitment expenditures, and

[45] Indeed, the primary methodology Vollrath (2024) draws on implicitly assumes competitive input markets, and will therefore be biased under monopsony.

fixed costs like land and intellectual property. On the other side of the market, workers may sort into more competitive labor markets and may be willing to take lower amenities or higher housing costs to avoid higher monopsony power. Some work in this direction can be found in Kahn and Tracy (2024) as well as the unpublished dissertation by Bamford (2022), both of which embed local monopsonistic labor markets in a model of spatial equilibrium with housing costs and show that housing costs capitalize some of the benefits of additional competition.

In addition, the evolution of monopsony power over time remains an underexplored area. While historical studies exist (Delabastita and Rubens, 2025; Naidu, 2010; Boal, 1995), the broad evolution of labor-market power throughout the 20th century remains largely unmapped. This likely reflects the fact that matched worker-firm data and exogenous variation in firm-specific wages are difficult to obtain in historical contexts, despite the fact that voluntary separations were regularly measured by the Bureau of Labor Statistics since at least the 1950s (Pencavel, 1972). But even for recent decades, little is known about how the residual labor supply elasticity has changed, both for the whole population and for specific subgroups of workers (see Depew and Sørensen (2013) for a firm-specific exception). How wage-setting mechanisms that interact with the residual labor supply elasticity, including external pay constraints like unions and regulation as well as internal constraints like norms and HR practices, have changed over time is another important area for future research (see Paker et al. (2023) for an examination of nominal pay-setting and monopsony in 17th century England).

The interplay between monopsony power and other constraints on firm behavior also deserves further attention. In practice, firms simultaneously face upward-sloping labor supply and effort responses; understanding how these forces jointly shape wages—the problem we label monergony in Section 4—remains an open theoretical and empirical challenge.

Finally, while wage posting is the wage-setting model associated with textbook models of monopsony, we have covered only a small share of the diversity of wage-setting practices in this paper. More research is needed on the details of wage-setting protocols, and the extent to which firms wage discriminate or bargain individually with workers (Card, 2022; Caldwell et al., 2026a). These practices are likely to evolve as platforms, remote work, and artificial agents evolve. At the same time, new methods in automatic application screening and algorithmic human resources will change the instruments employers have to alter the effective labor supply to the firm. That said, in light of predictions made by economists who thought the Internet would eliminate search frictions only to find them stubbornly persistent, we caution against thinking that the latest technology is likely to fundamentally alter the balance of power between workers and employers. While we think imperfect competition will remain a robust paradigm for understanding the labor market, technological changes could alter the mechanisms through which firms exercise monopsony power. For instance, algorithms may make it easier for firms to identify which workers are at risk of leaving their firm or to identify the effort exerted by individual workers. At the same time, technologies may make it easier for outside firms to contact and attempt to poach workers from other firms. Understanding how these developments affect the power held by firms (and the wages they pay to workers) is an important question for future research.

References

Abaluck, Jason and Abi Adams-Prassl, "What Do Consumers Consider Before They Choose? Identification from Asymmetric Demand Responses," *The Quarterly Journal of Economics*, 2021, *136* (3), 1611–1663.

Abowd, John M, Francis Kramarz, and David N Margolis, "High Wage Workers and High Wage Firms," *Econometrica*, 1999, *67* (2), 251–333.

Acemoglu, Daron, Alex He, and Daniel Le Maire, "Eclipse of Rent-Sharing: The Effects of Managers' Business Education on Wages and the Labor Share in the US and Denmark," *American Economic Review*, 2025. Forthcoming.

_ and Jörn-Steffen Pischke, "Why Do Firms Train? Theory and Evidence," *The Quarterly Journal of Economics*, 1998, *113* (1), 79–119.

Ackerberg, Daniel A, Kevin Caves, and Garth Frazer, "Identification Properties of Recent Production Function Estimators," *Econometrica*, 2015, *83* (6), 2411–2451.

Agarwal, Nikhil and Paulo Somaini, "Demand Analysis under Latent Choice Constraints," *The Review of Economic Studies*, 2025. Forthcoming.

Akerlof, George A, "Labor Contracts as Partial Gift Exchange," *The Quarterly Journal of Economics*, 1982, *97* (4), 543–569.

_ and Janet L Yellen, *Efficiency Wage Models of the Labor Market*, Cambridge University Press, 1986.

_ and _ , "The Fair Wage-Effort Hypothesis and Unemployment," *The Quarterly Journal of Economics*, 1990, *105* (2), 255–283.

Amior, Michael and Alan Manning, "Monopsony and the wage effects of migration," *The Economic Journal*, 2026, *136* (674), 402–439.

— and Jan Stuhler, "Immigration, Monopsony and the Distribution of Firm Pay," 2026. Working Paper.

Amodio, Francesco and Nicolas de Roux, "Measuring Labor Market Power in Developing Countries: Evidence from Colombian Plants," *Journal of Labor Economics*, 2023.

Arnold, David, "Mergers and Acquisitions, Local Labor Market Concentration, and Worker Outcomes," 2025. Working Paper.

— , Kevin S. Milligan, Terry Moon, and Amirhossein Tavakoli, "Job Transitions and Employee Earnings After Acquisitions: Linking Corporate and Worker Outcomes," *The Review of Economics and Statistics*, 2026. Forthcoming.

Arold, Benjamin W, Elliott Ash, W Bentley MacLeod, and Suresh Naidu, "Worker Rights in Collective Bargaining," 2025.

Asch, Beth and Paul Heaton, "Monopsony and labor supply in the army and navy," Technical Report 2008.

Ashenfelter, Orley C, Henry Farber, and Michael R Ransom, "Labor Market Monopsony," *Journal of Labor Economics*, 2010, *28* (2), 203–210.

Ashenfelter, Orley, David Card, Henry Farber, and Michael R Ransom, "Monopsony in the Labor Market: New Empirical Results and New Public Policies," *Journal of Human Resources*, 2022, *57* (3), S1–S10.

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

Azar, José A, Steven T Berry, and Ioana Marinescu, "Estimating Labor Market Power," Technical Report, National Bureau of Economic Research 2022.

Azar, José and Ioana Marinescu, "Monopsony power in the labor market," in Christian Dustmann and Thomas Lemieux, eds., *Handbook of Labor Economics*, Vol. 5, Elsevier, 2025, pp. 761–827.

_ , Emiliano Huet-Vaughn, Ioana Marinescu, Bledi Taska, and Till Von Wachter, "Minimum Wage Employment Effects and Labour Market Concentration," *Review of Economic Studies*, 2024, *91* (4), 1843–1883.

_ , Ioana Marinescu, Marshall Steinbaum, and Bledi Taska, "Concentration in US Labor Markets: Evidence from Online Vacancy Data," *Labour Economics*, 2020, *66*, 101886.

Azkarate-Askasua, Miren and Miguel Zerecero, "Union and Firm Labor Market Power," 2025. Working Paper.

Balke, Neele and Thibaut Lamadon, "Productivity Shocks, Long-Term Contracts, and Earnings Dynamics," *American Economic Review*, 2022.

Bamford, Iain, *Essays in Urban Economics*, Columbia University, 2022.

Barseghyan, Levon, Maura Coughlin, Francesca Molinari, and Joshua C Teitelbaum, "Heterogeneous Choice Sets and Preferences," *Econometrica*, 2021, *89* (5), 2015–2048.

Bassier, Ihsaan, "Firms and Inequality When Unemployment Is High," *Journal of Development Economics*, 2023, *161*, 103029.

_ , "Collective Bargaining and Spillovers in Local Labor Markets," 2024. Working Paper.

_ and Alan Manning, "Estimating labour market power: the long and short of it," Technical Report, Centre for Economic Performance 2025. Working paper.

_ , Arindrajit Dube, and Suresh Naidu, "Monopsony in movers: The elasticity of labor supply to firm wage policies," *Journal of Human Resources*, 2022, *57* (S), S50–s86.

Bayer, Patrick, Stephen L Ross, and Giorgio Topa, "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes," *Journal of Political Economy*, 2008, *116* (6), 1150–1196.

Belot, Michèle, Philipp Kircher, and Paul Muller, "How Wage Announcements Affect Job Search—A Field Experiment," *American Economic Journal: Macroeconomics*, October 2022, *14* (4), 1–67.

Benmelech, Efraim, Nittai K Bergman, and Hyunseob Kim, "Strong Employers and Weak Employees: How Does Employer Concentration Affect Wages?," *Journal of Human Resources*, 2020, pp. 0119–10007R1.

_ , _ , and _ , "Strong employers and weak employees: How does employer concentration affect wages?," *Journal of Human Resources*, 2022, *57* (S), S200–S250.

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

_ , _ , and _ , "Labor Market Power: From Micro Evidence to Macro Consequences," *Journal of Economic Perspectives*, 2026, *40* (1), 93–114.

Berry, Steven, James Levinsohn, and Ariel Pakes, "Automobile Prices in Market Equilibrium," *Econometrica: Journal of the Econometric Society*, 1995, pp. 841–890.

_ , Martin Gaynor, and Fiona Scott Morton, "Do Increasing Markups Matter? Lessons from Empirical Industrial Organization," *Journal of Economic Perspectives*, 2019, *33* (3), 44–68.

Bertrand, Marianne and Sendhil Mullainathan, "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *American Economic Review*, 2004, *94* (4), 991–1013.

Bils, Mark, Bariş Kaymak, and Kai-Jie Wu, "Robinson meets roy: Monopsony power and comparative advantage," Technical Report, National Bureau of Economic Research 2025.

Black, Dan A, "Discrimination in an Equilibrium Search Model," *Journal of Labor Economics*, 1995, *13* (2), 309–334.

Bloesch, Justin, Birthe Larsen, and Anders Yding, "Monopsony with Recruiting," 2026.

Bó, Ernesto Dal, Frederico Finan, and Martín A Rossi, "Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service," *The Quarterly Journal of Economics*, 2013, *128* (3), 1169–1218.

Boal, William M, "Testing for Employer Monopsony in Turn-of-the-Century Coal Mining," *The RAND Journal of Economics*, 1995, pp. 519–536.

_ , Michael R Ransom et al., "Monopsony in the Labor Market," *Journal of Economic Literature*, 1997, *35* (1), 86–112.

Borjas, George J and Anthony Edo, "Monopsony, Efficiency, and the Regularization of Undocumented Immigrants," Technical Report, National Bureau of Economic Research 2023.

Bowles, Samuel, "The Production Process in a Competitive Economy: Walrasian, Neo-Hobbesian, and Marxian Models," *The American Economic Review*, 1985, *75* (1), 16–36.

Breza, Emily, Supreet Kaur, and Nandita Krishnaswamy, "Social Norms as a Determinant of Aggregate Labor Supply: Evidence from Decentralized Spot Markets," Technical Report, National Bureau of Economic Research 2019.

_ , _ , and Yogita Shamdasani, "The Morale Effects of Pay Inequality," *The Quarterly Journal of Economics*, 2018, *133* (2), 611–663.

Brochu, Pierre and David A Green, "The impact of minimum wages on labour market transitions," *The Economic Journal*, 2013, *123* (573), 1203–1235.

Brooks, Wyatt J, Joseph P Kaboski, Yao Amber Li, and Wei Qian, "Exploitation of Labor? Classical Monopsony Power and Labor's Share," *Journal of Development Economics*, 2021, *150*, 102627.

Bulow, Jeremy and John Roberts, "The Simple Economics of Optimal Auctions," *Journal of Political Economy*, 1989, *97* (5), 1060–1090.

Burdett, Kenneth and Dale T Mortensen, "Wage Differentials, Employer Size, and Unemployment," *International Economic Review*, 1998, pp. 257–273.

Burn, Ian, Patrick Button, Luis Munguia Corella, and David Neumark, "Does ageist language in job ads predict age discrimination in hiring?," *Journal of Labor Economics*, 2022, *40* (3), 613–667.

Caldwell, Sydnee and Emily Oehlsen, "Gender, outside options, and labor supply: Experimental evidence from the gig economy," 2023. University of California, Berkeley Working Paper.

— and Nikolaj Harmon, "Outside Options, Bargaining, and Wages: Evidence from Coworker Networks," *UC Berkeley Department of Economics Working Paper*, 2019.

— , Ingrid Haegele, and Joerg Heining, "Why Workers Stay: Pay, Beliefs, and Attachment," Working Paper 33445, National Bureau of Economic Research February 2025.

— , — , and — , "Bargaining and Inequality in the Labor Market," *Quarterly Journal of Economics*, 2026, *141* (1), 315–371.

— , — , and — , "Firm Pay, Amenities, and Inequality," *UC Berkeley Department of Economics Working Paper*, 2026.

Callaci, Brian, Matthew Gibson, Sérgio Pinto, Marshall Steinbaum, and Matt Walsh, "The effect of franchise no-poaching restrictions on worker earnings," *Review of Economics and Statistics*, 2024, pp. 1–35.

— , SÃ©rgio Pinto, Marshall Steinbaum, and Matt Walsh, "Vertical Restraints and Labor Markets in Franchised Industries," in Benjamin Elsner and Solomon W. Polachek, eds., *Big Data Applications in Labor Economics, Part B*, Emerald Publishing Limited, 12 2024.

Caplin, Andrew, Mark Dean, and John Leahy, "Rational Inattention, Optimal Consideration Sets, and Stochastic Choice," *The Review of Economic Studies*, 2019, *86* (3), 1061–1094.

Card, David, "Who Set Your Wage?," *American Economic Review*, 2022, *112* (4), 1075–90.

— , Ana Rute Cardoso, Joerg Heining, and Patrick Kline, "Firms and Labor Market Inequality: Evidence and Some Theory," *Journal of Labor Economics*, 2018, *36* (S1), S13–S70.

_ and Alan B Krueger, *Myth and Measurement: The New Economics of the Minimum Wage-Twentieth-Anniversary Edition*, Princeton University Press, 1995.

_ and _ , "Time-Series Minimum-Wage Studies: A Meta-Analysis," *The American Economic Review*, 1995, *85* (2), 238–243.

_ and Ana Rute Cardoso, "Wage Flexibility Under Sectoral Bargaining," *Journal of the European Economic Association*, 2022, *20* (5), 2013–2061.

_ , Jörg Heining, and Patrick Kline, "Workplace Heterogeneity and the Rise of West German Wage Inequality," *The Quarterly Journal of Economics*, 2013, *128* (3), 967–1015.

Cardoso, Ana Rute and Pedro Portugal, "Contractual Wages and the Wage Cushion Under Different Bargaining Settings," *Journal of Labor Economics*, 2005, *23* (4), 875–902.

Carpio, Lucia Del and Thomas Fujiwara, "Spillovers and Scalability in Job Ad Experiments: Evidence from Gender-Neutral Language," 2026. Working Paper.

Cattaneo, Matias D, Xinwei Ma, Yusufcan Masatlioglu, and Elchin Suleymanov, "A Random Attention Model," *Journal of Political Economy*, 2020, *128* (7), 2796–2836.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer, "The Effect of Minimum Wages on Low-Wage Jobs," *The Quarterly Journal of Economics*, 2019, *134* (3), 1405–1454.

Clemens, Jeffrey, Lisa B Kahn, and Jonathan Meer, "The Minimum Wage, Fringe Benefits, and Worker Welfare," Technical Report, National Bureau of Economic Research 2018.

CORE, *The Economy 2.0: Microeconomics*, CORE, 2024.

Corella, Luis F Munguía, "Minimum Wages in Monopsonistic Labor Markets," *IZA Journal of Labor Economics*, 2020, *9* (1).

Corradini, Viola, Lorenzo Lagos, and Garima Sharma, "Collective Bargaining for Women: How Unions Can Create Female-Friendly Jobs," *The Quarterly Journal of Economics*, 2025, *140* (3), 2053–2105.

Coviello, Decio, Erika Deserranno, and Nicola Persico, "Minimum Wage and Individual Worker Productivity: Evidence from a Large US Retailer," *Journal of Political Economy*, 2022, *130* (9), 2315–2360.

Datta, Nikhil, "Local monopsony power," Technical Report, Centre for Economic Performance 2024. Working paper.

— and Stephen Machin, "Government Contracting and Living Wages >Minimum Wages," 2024. Working paper.

Delabastita, Vincent and Michael Rubens, "Colluding against workers," *Journal of Political Economy*, 2025, *133* (6), 1796–1839.

Demir, Gökay, "Labor Market Frictions and Spillover Effects from Publicly Announced Sectoral Minimum Wages," 2026. Working Paper.

Depew, Briggs and Todd A Sørensen, "The Elasticity of Labor Supply to the Firm over the Business Cycle," *Labour Economics*, 2013, *24*, 196–204.

—, Peter Norlander, and Todd A Sørensen, "Inter-Firm Mobility and Return Migration Patterns of Skilled Guest Workers," *Journal of Population Economics*, 2017, *30*, 681–721.

Derenoncourt, Ellora and Claire Montialoux, "Minimum Wages and Racial Inequality," *The Quarterly Journal of Economics*, 2021, *136* (1), 169–228.

— and David Weil, "Voluntary minimum wages: The local labor market effects of national retailer policies," *The Quarterly Journal of Economics*, 2025, *140* (3), 1901–1958.

Dickens, William T. and Shelly J. Lundberg, "Hours Restrictions and Labor Supply," *International Economic Review*, 1993, *34* (1), 169–192.

Dobbelaere, Sabien and Jacques Mairesse, "Panel Data Estimates of the Production Function and Product and Labor Market Imperfections," *Journal of Applied Econometrics*, 2013, *28* (1), 1–46.

Dodini, Samuel, Kjell G Salvanes, and Alexander Willén, "The Dynamics of Power in Labor Markets: Monopolistic Unions Versus Monopsonistic Employers," 2026.

Doran, Kirk, Alexander Gelber, and Adam Isen, "The Effects of High-Skilled Immigration Policy on Firms: Evidence from Visa Lotteries," *Journal of Political Economy*, 2022, *130* (10), 2501–2533.

Doraszelski, Ulrich and Jordi Jaumandreu, "Measuring the Bias of Technological Change," *Journal of Political Economy*, 2018, *126* (3), 1027–1084.

Dube, Arindrajit, Alan Manning, and Suresh Naidu, "Monopsony and Employer Misoptimization Explain Why Wages Bunch at Round Numbers," *American Economic Review*, 2026. Forthcoming.

— and Attila Lindner, "Minimum wages in the 21st century," *Handbook of Labor Economics*, 2024, *5*, 261–383.

— , 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), 620–63.

_ , Suresh Naidu, and Adam D Reich, "Power and Dignity in the Low-Wage Labor Market: Theory and Evidence from Wal-Mart Workers," Technical Report, National Bureau of Economic Research 2022.

_ , T William Lester, and Michael Reich, "Minimum Wage Shocks, Employment Flows, and Labor Market Frictions," *Journal of Labor Economics*, 2016, *34* (3), 663–704.

Dustmann, Christian and Uta Schönberg, "Linking Empirical Evidence to Theory: A Framework for Understanding Immigration's Labor Market Effects," in "Handbook of Labor Economics," Vol. 6, Elsevier, 2025, pp. 447–548.

_ , Attila Lindner, Uta Schönberg, Matthias Umkehrer, and Philipp Vom Berge, "Reallocation Effects of the Minimum Wage," *The Quarterly Journal of Economics*, 2022, *137* (1), 267–328.

Emanuel, Natalia and Emma Harrington, "The Payoffs of Higher Pay: Labor Supply and Productivity Responses to a Voluntary Firm Minimum Wage," *Working Paper*, 2025.

Estefan, Alejandro, Roberto Gerhard, Joseph P Kaboski, Illenin O Kondo, and Wei Qian, "Outsourcing Policy and Worker Outcomes: Causal Evidence from a Mexican Ban," 2026. Working Paper.

Falk, Armin, Ernst Fehr, and Christian Zehnder, "Fairness perceptions and reservation wagesâthe behavioral effects of minimum wage laws," *The Quarterly Journal of Economics*, 2006, *121* (4), 1347–1381.

Fehr, Ernst and Klaus M Schmidt, "A Theory of Fairness, Competition, and Cooperation," *The Quarterly Journal of Economics*, 1999, *114* (3), 817–868.

_ , Erich Kirchler, Andreas Weichbold, and Simon Gächter, "When Social Norms Overpower Competition: Gift Exchange in Experimental Labor Markets," *Journal of Labor Economics*, 1998, *16* (2), 324–351.

Felix, Mayara, "Trade, Labor Market Concentration, and Wages," 2026. Working Paper.

Fernald, John, Amit Gandhi, Dimitrije Ruzic, and James Traina, "Micro and Macro Perspectives on Production-Based Markups," *arXiv preprint arXiv:2604.13224*, 2026.

Fosgerau, Mogens, Emerson Melo, Andre De Palma, and Matthew Shum, "Discrete Choice and Rational Inattention: A General Equivalence Result," *International Economic Review*, 2020, *61* (4), 1569–1589.

Fukui, Masao, "A Theory of Wage Rigidity and Unemployment Fluctuations with On-the-Job Search," *Job Market Paper, Massachusetts Institute of Technology*, 2020.

Gandhi, Amit, Salvador Navarro, and David A Rivers, "On the Identification of Gross Output Production Functions," *Journal of Political Economy*, 2020, *128* (8), 2973–3016.

Garin, Andrew and Filipe Silvério, "How responsive are wages to firm-specific changes in labour demand? Evidence from idiosyncratic export demand shocks," *Review of Economic Studies*, 2024, *91* (3), 1671–1710.

Gavazza, Alessandro, Simon Mongey, and Giovanni L Violante, "Aggregate Recruiting Intensity," *American Economic Review*, 2018, *108* (8), 2088–2127.

Gibson, Matthew D, “Employer market power in Silicon Valley,” *The Economic Journal*, 2025. Forthcoming.

Glitz, Albrecht, “Coworker Networks in the Labour Market,” *Labour Economics*, 2017, *44*, 218–230.

Goerges, Luise and Daniele Nosenzo, “Social Norms and the Labor Market,” *Handbook of Labor, Human Resources and Population Economics*, 2020, pp. 1–26.

Goldschmidt, Deborah and Johannes F Schmieder, “The Rise of Domestic Outsourcing and the Evolution of the German Wage Structure,” *The Quarterly Journal of Economics*, 2017, *132* (3), 1165–1217.

Goolsbee, Austan and Chad Syverson, “Monopsony power in higher education: A tale of two tracks,” *Journal of Labor Economics*, 2023, *41* (S1), S257–S290.

Gottfries, Axel and Gregor Jarosch, “Dynamic Monopsony with Large Firms,” 2026. Working Paper.

Granovetter, Mark, *Getting a Job: A Study of Contacts and Careers*, University of Chicago Press, 1995.

Granovetter, Mark S, “The Strength of Weak Ties,” *American Journal of Sociology*, 1973, *78* (6), 1360–1380.

Guo, Yingni and Eran Shmaya, “Robust monopoly regulation,” *American Economic Review*, 2025, *115* (2), 599–634.

Gyetvay, Sam and Sekou Keita, “Competition, Mobility and Immigration,” 2023. Working paper.

Harasztosi, Péter and Attila Lindner, "Who Pays for the Minimum Wage?," *American Economic Review*, 2019, *109* (8), 2693–2727.

Hazell, Jonathon, Christina Patterson, Heather Sarsons, and Bledi Taska, "National Wage Setting," 2025. Working Paper.

Hensvik, Lena and Oskar Nordström Skans, "Social Networks, Employee Selection, and Labor Market Outcomes," *Journal of Labor Economics*, 2016, *34* (4), 825–867.

Hermo, Santiago, "Collective Bargaining Networks and the Propagation of Shocks," 2025. Working Paper.

Herrera-Caicedo, Alejandro, Jessica Jeffers, and Elena Prager, "Collusion Through Common Leadership," Technical Report, National Bureau of Economic Research 2025.

Hirsch, Boris, Elke J Jahn, and Claus Schnabel, "Do Employers Have More Monopsony Power in Slack Labor Markets?," *ILR Review*, 2018, *71* (3), 676–704.

Hjort, Jonas, Xuan Li, and Heather Sarsons, "Across-country wage compression in multinationals," *American Economic Review*, 2026, *116* (1), 52–87.

Holzer, Harry J., Lawrence F. Katz, and Alan B. Krueger, "Job Queues and Wages*," *The Quarterly Journal of Economics*, 08 1991, *106* (3), 739–768.

Honka, Elisabeth, Ali Hortaçsu, and Matthijs Wildenbeest, "Empirical Search and Consideration Sets," in "Handbook of the Economics of Marketing," Vol. 1, Elsevier, 2019, pp. 193–257.

_ and Pradeep Chintagunta, "Simultaneous or Sequential? Search Strategies in the US Auto Insurance Industry," *Marketing Science*, 2017, *36* (1), 21–42.

Hunt, Jennifer and Bin Xie, "How Restricted Is the Job Mobility of Skilled Temporary Work Visa Holders?," *Journal of Policy Analysis and Management*, 2019, *38* (1), 41–64.

Imbens, Guido W. and Joshua D. Angrist, "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 1994, *62* (2), 467–475.

Jäger, Simon, Christopher Roth, Nina Roussille, and Benjamin Schoefer, "Worker beliefs about outside options," *The Quarterly Journal of Economics*, 2024, *139* (3), 1505–1556.

_ , Suresh Naidu, and Benjamin Schoefer, "Collective bargaining, unions, and the wage structure: An international perspective," in Christian Dustmann and Thomas Lemieux, eds., *Handbook of Labor Economics*, Vol. 6, Elsevier, 2025, pp. 229–372.

Jardim, Ekaterina, Mark C Long, Robert Plotnick, Emma Van Inwegen, Jacob Vigdor, and Hilary Wething, "Minimum-Wage Increases and Low-Wage Employment: Evidence from Seattle," *American Economic Journal: Economic Policy*, 2022, *14* (2), 263–314.

Jarosch, Gregor, Jan Sebastian Nimczik, and Isaac Sorkin, "Granular search, market structure, and wages," *Review of Economic Studies*, 2024, *91* (6), 3569–3607.

Johnson, Matthew S, Kurt Lavetti, and Michael Lipsitz, "The labor market effects of legal restrictions on worker mobility," *Journal of Political Economy*, 2025, *133* (9), 2735–2793.

Jungerman, William, "Dynamic Monopsony and Human Capital," 2024. Working Paper.

Kahn, Matthew E and Joseph Tracy, "Monopsony in Spatial Equilibrium," *Regional Science and Urban Economics*, 2024, *104*, 103956.

Kim, Seohee and Alison Pei, "Monopsony in the High-Skilled Migrant Labor Market-Evidence from the H-1B Petition Data," *Available at SSRN 4010152*, 2022.

Kline, Patrick, "Labor market monopsony: Fundamentals and frontiers," in Christian Dustmann and Thomas Lemieux, eds., *Handbook of Labor Economics*, Vol. 6, Elsevier, 2025, pp. 655–728.

_ , Evan K Rose, and Christopher R Walters, "Systemic Discrimination Among Large US Employers," *The Quarterly Journal of Economics*, 2022, *137* (4), 1963–2036.

Kreps, David M and Jose A Scheinkman, "Quantity Precommitment and Bertrand Competition Yield Cournot Outcomes," *The Bell Journal of Economics*, 1983, pp. 326–337.

Kroft, Kory, Fabian Lange, and Matthew J Notowidigdo, "Duration Dependence and Labor Market Conditions: Evidence from a Field Experiment," *The Quarterly Journal of Economics*, 2013, *128* (3), 1123–1167.

_ , Isaac Norwich, Matthew J Notowidigdo, and Stephen Tino, "The Labor Market Return to Permanent Residency," Technical Report, National Bureau of Economic Research 2026.

_ , Yao Luo, Magne Mogstad, and Bradley Setzler, "Imperfect Competition and Rents in Labor and Product Markets: The Case of the Construction Industry," *American Economic Review*, 2025, *115* (9), 2926–2969.

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.

Lagaras, Spyridon, "M&As, employee costs and labor reallocation," *Journal of Finance*, 2025. Forthcoming.

Lagos, Lorenzo, "Union Bargaining Power and the Amenity-Wage Tradeoff," Technical Report, National Bureau of Economic Research 2025.

Lamadon, Thibaut, Magne Mogstad, and Bradley Setzler, "Imperfect competition, compensating differentials, and rent sharing in the US labor market," *American Economic Review*, 2022, *112* (1), 169–212.

Langella, Monica and Alan Manning, "Marshall Lecture 2020: The Measure of Monopsony," *Journal of the European Economic Association*, 2021, *19* (6), 2929–2957.

Leibenstein, Harvey, "The Theory of Underemployment in Backward Economies," *Journal of Political Economy*, 1957, *65* (2), 91–103.

los Santos, Babur De, Ali Hortaçsu, and Matthijs R Wildenbeest, "Testing Models of Consumer Search Using Data on Web Browsing and Purchasing Behavior," *American Economic Review*, 2012, *102* (6), 2955–2980.

Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt, "Rational Inattention: A Review," *Journal of Economic Literature*, 2023, *61* (1), 226–273.

Maestas, Nicole, Kathleen J Mullen, David Powell, Till Von Wachter, and Jeffrey B Wenger, "The value of working conditions in the United States and implications for the structure of wages," *American Economic Review*, 2023, *113* (7), 2007–2047.

Manning, Alan, "How Do We Know That Real Wages Are Too High?," *The Quarterly Journal of Economics*, 1995, *110* (4), 1111–1125.

_ , *Monopsony in Motion: Imperfect Competition in Labor Markets*, Princeton University Press, 2003.

_ , "Imperfect Competition in the Labor Market," *Handbook of Labor Economics*, 2011, *4*, 973–1041.

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

_ and Barbara Petrongolo, "How Local Are Labor Markets? Evidence from a Spatial Job Search Model," *American Economic Review*, 2017, *107* (10), 2877–2907.

Marinescu, Ioana and Herbert Hovenkamp, "Anticompetitive Mergers in Labor Markets," *Indiana Law Journal*, 2019.

Martins, Pedro S, "What do employers' associations do?," Technical Report, IZA Discussion Papers 2020.

Mas, Alexandre, *Non-Wage Amenities*, Vol. 6, Elsevier,

_ and Amanda Pallais, "Valuing Alternative Work Arrangements," *American Economic Review*, 2017, *107* (12), 3722–59.

Matejka, Filip and Alisdair McKay, "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model," *American Economic Review*, 2015, *105* (1), 272–98.

Mertens, Matthias, "Labor Market Power, Superstar Firms, and Between-Firm Wage (in) Equality," *International Journal of Industrial Organization*, 2023.

Moen, Espen R, "Competitive Search Equilibrium," *Journal of Political Economy*, 1997, *105* (2), 385–411.

Montgomery, James D, "Social Networks and Labor-Market Outcomes: Toward an Economic Analysis," *The American Economic Review*, 1991, *81* (5), 1408–1418.

Mortensen, Dale, *Wage Dispersion: Why Are Similar Workers Paid Differently?*, MIT press, 2003.

Moscarini, Giuseppe and Fabien Postel-Vinay, "The Cyclical Job Ladder," *Annual Review of Economics*, 2018, *10* (1), 165–188.

Naidu, Suresh, "Recruitment Restrictions and Labor Markets: Evidence from the Postbellum US South," *Journal of Labor Economics*, 2010, *28* (2), 413–445.

_ and Eric A Posner, "Labor Monopsony and the Limits of the Law," *Journal of Human Resources*, 2022, *57* (S), S284–S323.

_ , _ , and Glen Weyl, "Antitrust Remedies for Labor Market Power," *Harvard Law Review*, 2018, *132* (2), 536–601.

_ , Yaw Nyarko, and Shing-Yi Wang, "Monopsony Power in Migrant Labor Markets: Evidence from the United Arab Emirates," *Journal of Political Economy*, 2016, *124* (6), 1735–1792.

Olley, G Steven and Ariel Pakes, "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 1996, *64* (6), 1263–1297.

Paker, Meredith M, Judy Z Stephenson, and Patrick Wallis, "Nominal Wage Patterns, Monopsony, and Labour Market Power in Early Modern England," *The Economic History Review*, 2023.

Parsons, Donald O, "Specific Human Capital: An Application to Quit Rates and Layoff Rates," *Journal of Political Economy*, 1972, *80* (6), 1120–1143.

Passaro, Diego Gentile, Fuhito Kojima, and Bobak Pakzad-Hurson, "Equal pay for Similar work," *American Economic Review*, 2026, *116* (3), 977–1013.

Pencavel, John H, "Wages, Specific Training, and Labor Turnover in US Manufacturing Industries," *International Economic Review*, 1972, pp. 53–64.

Polybius, *The Histories of Polybius*, London and New York: Macmillan and Co., 1889. 2 Volumes.

Prager, Elena, "Antitrust Enforcement in Labor Markets," 2025.

_ and Matt Schmitt, "Employer Consolidation and Wages: Evidence from Hospitals," *American Economic Review*, 2021, *111* (2), 397–427.

Prescott, James J, Norman D Bishara, and Evan Starr, "Understanding Noncompetition Agreements: The 2014 Noncompete Survey Project," *Mich. St. L. Rev.*, 2016, p. 369.

Rao, Nirupama L and Max Risch, "Who's Afraid of the Minimum Wage? Measuring the Impacts on Independent Businesses Using Matched US Tax Returns," *The Quarterly Journal of Economics*, 2026, *141* (1), 373–427.

Raval, Devesh, "Testing the Production Approach to Markup Estimation," *Review of Economic Studies*, 2022.

Rinz, Kevin et al., "Labor Market Concentration, Earnings Inequality, and Earnings Mobility," *Center for Administrative Records Research and Applications Working Paper*, 2018, *10.*

Robinson, Joan, *The Economics of Imperfect Competition*, Palgrave Macmillan, 1933.

Roussille, Nina and Benjamin Scuderi, “Bidding for Talent,” Technical Report, Working paper 2026.

Roux, Nicolás De, Marcela Eslava, Santiago Franco, and Eric Verhoogen, “Estimating Production Functions in Differentiated-Product Industries with Quantity Information and External Instruments,” Technical Report, National Bureau of Economic Research 2024.

Salop, Steven and Joseph Stiglitz, “Bargains and Ripoffs: A Model of Monopolistically Competitive Price Dispersion,” *The Review of Economic Studies*, 1977, *44* (3), 493–510.

Scripture4All Foundation, “Greek Interlinear New Testament: Romans Chapter 6,” https://www.scripture4all.org/OnlineInterlinear/NTpdf/rom6.pdf 2010. Accessed: 2026-05-26.

Seegmiller, Bryan, “Valuing Labor Market Power: The Role of Productivity Advantages,” 2025. Working Paper.

Shapiro, Carl and Joseph E Stiglitz, “Equilibrium Unemployment as a Worker Discipline Device,” *The American Economic Review*, 1984, *74* (3), 433–444.

Sharma, Garima, “Monopsony and Gender,” *Unpublished Manuscript*, 2023.

_ , “Collusion Among Employers in India,” 2025. Working Paper.

Shi, Liyan, “Optimal regulation of noncompete contracts,” *Econometrica*, 2023, *91* (2), 425–463.

Simon, Herbert A, “A Behavioral Model of Rational Choice,” *The Quarterly Journal of Economics*, 1955, pp. 99–118.

Smith, Adam, *The Wealth of Nations*, Vol. 2 1776.

Sockin, Jason, "Show Me the Amenity: Are Higher-Paying Firms Better All Around?," 2022. CESifo Working Paper.

Sokolova, Anna and Todd Sorensen, "Monopsony in Labor Markets: A Meta-Analysis," *ILR Review*, 2021, *74* (1), 27–55.

Song, Jae, David J Price, Fatih Guvenen, Nicholas Bloom, and Till Von Wachter, "Firming Up Inequality," *The Quarterly Journal of Economics*, 2018, *134* (1), 1–50.

Sorkin, Isaac, "Ranking Firms Using Revealed Preference," *The Quarterly Journal of Economics*, 2018, *133* (3), 1331–1393.

Spence, A Michael, "Monopoly, Quality, and Regulation," *The Bell Journal of Economics*, 1975, pp. 417–429.

Staiger, D, Joanne Spetz, and C Phibbs, "Is There Monopsony in the Labor Market ? Evidence from a Natural Experiment," *Journal of Labor Economics*, 2010, *28* (2).

Starr, Evan, "The Economics of Noncompete Clauses," *Journal of Economic Perspectives*, 2026, *40* (1), 139–166.

_ , James J Prescott, and Norman Bishara, "Noncompetes in the US Labor Force," *Journal of Law and Economics*, 2021.

Steinbaum, Marshall and Maurice E Stucke, "The Effective Competition Standard," *The University of Chicago Law Review*, 2020, *87* (2), 595–623.

Stigler, George J, "The Economics of Minimum Wage Legislation," *The American Economic Review*, 1946, *36* (3), 358–365.

Thornton, Robert J, "Retrospectives: How Joan Robinson and BL Hallward Named Monopsony," *Journal of Economic Perspectives*, 2004, *18* (2), 257–261.

Topa, Giorgio, "Labor markets and referrals," in A. Bisin Benhabib and MO Jackson, eds., *Handbook of social economics*, Vol. 1, Elsevier, 2011, pp. 1193–1221.

Townsend, Wilbur and Corey Allan, "How Restricting Migrants' Job Options Affects Both Migrants and Existing Residents," 2025. Working Paper.

Vollrath, Dietrich, "The elasticity of aggregate output with respect to capital and labor," *American Economic Journal: Macroeconomics*, 2024, *16* (4), 470–504.

Volpe, Oscar, "Job Preferences, Labor Market Power, and Inequality," Technical Report, Working Paper 2026.

Wang, Sean and Samuel Young, "Unionization, Employer Opposition, and Establishment Closure," 2024. Working Paper.

Wang, Xuening, "US Permanent Residency, Job Mobility, and Earnings," *Journal of Labor Economics*, 2021, *39* (3), 639–671.

Webber, Douglas A, "Labor Market Competition and Employment Adjustment over the Business Cycle," *Journal of Human Resources*, 2022, *57* (S), S87–S110.

Weil, David, "The Fissured Workplace: Why Work Became So Bad for So Many and What Can Be Done to Improve It," in "The Fissured Workplace," Harvard University Press, 2014.

Werden, Gregory J, "The Competitive Process Standard," *Antitrust Law Journal*, 2024, *86* (2), 579–645.

Wiltshire, Justin C, "Walmart Supercenters and Monopsony Power: How a Large, Low-Wage Employer Impacts Local Labor Markets," 2025. Working Paper.

Wiswall, Matthew and Basit Zafar, "Preference for the Workplace, Investment in Human Capital, and Gender," *The Quarterly Journal of Economics*, 2018, *133* (1), 457–507.

Wright, Randall, Philipp Kircher, Benoît Julien, and Veronica Guerrieri, "Directed Search and Competitive Search Equilibrium: A Guided Tour," *Journal of Economic Literature*, March 2021, *59* (1), 90–148.

Wu, Tim and Marc Cashman, *The Curse of Bigness*, Penguin Random House, 2018.

Yeh, Chen, Claudia Macaluso, and Brad Hershbein, "Monopsony in the US Labor Market," *American Economic Review*, 2022, *112* (7), 2099–2138.

A. Theoretical Appendix

A.1 Microfoundations for Firm-Specific Labor Supply Elasticity

Recall that worker i's (unconditional) labor supply to firm j—the probability that i chooses j, integrated over ϵ_{ij} and $\{\epsilon_{ik}\}_{k\neq j}$—is given by:

$$\text{P[i chooses j}|w_j, \mathbf{w_{-j}}, S_i] = \text{P}\left(j = \left[\arg\max_{k\in S_i} V_i^k(w_k)\right]\right) \tag{16}$$

where $\mathbf{w_{-j}}$ is a vector of wages offered by firms other than j. We use $k^* := \arg\max_{k\in S_i} V_i^k(w_k)$ to denote the best job for i.

Conditioning on the j-specific independent taste shock for i, ϵ_{ij}, in this expression allows Equation 16 to be expressed as a product of pairwise choice probabilities, including non-employment:[46]

$$\begin{aligned}\text{P[i chooses j} \mid w_j, \mathbf{w_{-j}}, \epsilon_{ij}, S_i] \quad &= \quad \underbrace{\text{P}(V_i^j(w_j) > V_i^0|w_j, \epsilon_{ij}, b)}_{\text{prob. j better than unemp}} \\ &\times \underbrace{\prod_{k\in S_i\setminus\{j\,\cup o\}} \text{P}\big(V_i^j(w_j) > V_i^k(w_k) \bigm| w_j, \mathbf{w_{-j}}, \epsilon_{ij}, S_i\big)}_{\text{pairwise choice probabilities}}\end{aligned}$$

Log-differentiating the conditional probability above yields a conditional elasticity, which is an intermediate object—not the labor supply elasticity itself:

[46] The independence of the probability of choosing the other jobs or the outside option requires conditioning on the j-specific taste shock ϵ_{ij}. Integrating over this shock would lead to dependence. The conditioning here is purely calculational—both ϵ_{ij} and $\{\epsilon_{ik}\}_{k\neq j}$ are random variables. Integrating back over ϵ_{ij}, as in Equation 17, recovers worker i's unconditional labor supply.

$$\eta_{ij}(w_j|S_i,b_i,\epsilon_{ij}) = \underbrace{\sum_{k\neq j,k>0} \eta_{ijk}(w_j|S_i,b_i,\epsilon_{ij})}_{\text{Competition with other firms in } S_i} + \underbrace{\eta_{ij0}(w_j|S_i,b_i,\epsilon_{ij})}_{\text{Extensive margin from non-employment}}$$

This conditional elasticity is a stepping stone in the calculations, not the labor supply elasticity. To recover worker i's labor supply to firm j, we first integrate over ϵ_{ij} (holding S_i fixed):

$$\text{P[i chooses j} \mid w_j, \mathbf{w_{-j}}, b_i, S_i] = \int \text{P[i chooses j} \mid w_j, \mathbf{w_{-j}}, \epsilon_{ij}, S_i, b_i]\, dF_\epsilon(\epsilon_{ij}|i) \tag{17}$$

Consideration elasticity: We now incorporate endogenous S_i into the residual labor-supply elasticity decomposition. In general, modeling how a set of options responds to the price of a single one is a complicated problem. To illustrate our main points, we take a much-simplified approach, and assume an own elasticity, leaving out microfoundations due to directed search, for example. We assume that firm j is in the consideration set S_i with probability $\pi_{ij}(w_j)$, which has an elasticity

$$\eta^{\pi}_{ij} := \frac{w_j}{\pi_{ij}} \frac{\mathrm{d}\pi_{ij}}{\mathrm{d}w_j}. \tag{18}$$

Note that we assume that π_{ij} only depends on w_j (no cross partials, and no dependence on ϵ_{ij}), so that consideration sets are independent of both other firms' wages and idiosyncratic tastes.

Combining this with 17, we can get a complete expression for the worker-specific supply to firm j as:

$$s_{ij} = \int \pi_{ij}(w_j) \text{P[i chooses j} \mid w_j, \mathbf{w_{-j}}, \epsilon_{ij}, S_i, b_i] \, dF_\epsilon(\epsilon_{ij}|i) \tag{19}$$

The corresponding worker-specific elasticity, still an intermediate object, is:

$$\eta_{ij}(w_j|b_i, \epsilon_{ij}) = \underbrace{\eta^\pi_{ij}}_{\text{Consideration}} + \underbrace{\sum_{k \neq j, k>0} \eta_{ijk}(w_j|S_i, b_i, \epsilon_{ij})}_{\text{Competition with other firms in } S_i} + \underbrace{\eta_{ij0}(w_j|S_i, b_i, \epsilon_{ij})}_{\text{Extensive margin from non-employment}}$$

For each worker i we now have a elasticity that depends on tastes, the distribution of consideration sets, as well as outside options, that we can aggregate to get the labor supply facing the firm. Integrating over the distribution of workers i, we finally get the overall labor supply facing the firm as:

$$s_j = \int \int \pi_{ij}(w_j) \text{P[i chooses j} \mid w_j, \mathbf{w_{-j}}, \epsilon_{ij}, S_i, b_i] \, dF_\epsilon(\epsilon_{ij}|i) dF(i) \tag{20}$$

We can then get a general expression for the elasticity to firm j

$$\eta_j = \frac{w_j}{s_j} \frac{ds_j}{dw_j} = \tilde{E}_i[\eta^\pi_{ij}] + \sum_{k \neq j, k>0} \tilde{E}_i[\eta_{jk}] + \tilde{E}_i[\eta_{j0}] \tag{21}$$

where $\tilde{E}_i[\mathbf{f}] \equiv \int \int \frac{P_{ij}}{s_j} \pi_{ij} \mathbf{f} dF_\epsilon(\epsilon_{ij}|i) dF(i)$.

While this expression is very general, and summarizes the three main forces driving monopsony, it is not clear how one would operationalize this in a parametric model. For this reason, we now specialize to the common log utility, logit error specification in the next section.

Logit-Log Specialization

Assume $\epsilon_{ik} \sim$ iid Gumbel$(0, \sigma_i)$ and $U_k(w_k) = \beta \log w_k$ for all firms k, and $U_o(b_i) = \beta \log(b_i)$ for non-employment. Integrating (17) over ϵ_{ij} yields the multinomial logit:

$$s_{ij} := P[i \text{ chooses } j \mid w_j, \mathbf{w_{-j}}, b_i, S_i] = \frac{w_j^{\beta/\sigma_i}}{b_i^{\frac{\beta}{\sigma_i}} + \sum_{k \in S_i \setminus \{o\}} w_k^{\beta/\sigma_i}}. \tag{22}$$

Define the outside-option share analogously:

$$s_{i0} := \frac{b_i^{\frac{\beta}{\sigma_i}}}{b_i^{\frac{\beta}{\sigma_i}} + \sum_{k \in S_i \setminus \{o\}} w_k^{\beta/\sigma_i}}. \tag{23}$$

Write $\alpha_i := \beta/\sigma_i$ (worker-specific β_i can be accommodated without loss).

Derivative of the conditional choice probability

Holding the consideration set constant, and differentiating (22) with respect to w_j gives:

$$\frac{\mathrm{d}s_{ij}}{\mathrm{d}w_j} = s_{ij}\,(1 - s_{ij})\,\frac{\alpha_i}{w_j} = s_{ij}\left(s_{i0} + \sum_{k \neq j, k>0} s_{ik}\right)\frac{\alpha_i}{w_j}, \tag{24}$$

where the second equality uses $1 - s_{ij} = s_{i0} + \sum_{k \neq j, k>0} s_{ik}$.

Aggregate share and its elasticity

The share of workers at firm j is given by:

$$s_j = \int \pi_{ij}(w_j)\, s_{ij}(w_j) dF(i), \tag{25}$$

where $dF(i)$ integrates over worker heterogeneity (including consideration probabilities π_{ik} as well as outside options b_i or preference heterogeneity σ_i).

Differentiating and multiplying by w_j/s_j we get the own-wage labor supply elasticity to firm j:

$$\begin{aligned}
\eta_j &= \frac{w_j}{s_j}\frac{\mathrm{d}s_j}{\mathrm{d}w_j} \\
&= \frac{w_j}{s_j}\int \left[\frac{\mathrm{d}\pi_{ij}}{\mathrm{d}w_j}\, s_{ij} \;+\; \pi_{ij}\,\frac{\mathrm{d}s_{ij}}{\mathrm{d}w_j}\right] dF(i) \\
&= \frac{1}{s_j}\int \pi_{ij}\, s_{ij}\left[\underbrace{\frac{w_j}{\pi_{ij}}\frac{\mathrm{d}\pi_{ij}}{\mathrm{d}w_j}}_{\eta^{\pi}_{ij}} \;+\; \underbrace{(1-s_{ij})\,\alpha_i}_{\text{from (24)}}\right] dF(i).
\end{aligned} \tag{26}$$

Expanding $1 - s_{ij} = s_{i0} + \sum_{k\neq j, k>0} s_{ik}$:

$$\boxed{\eta_j = \frac{1}{s_j}\int \pi_{ij}\, s_{ij}\left[\underbrace{\eta^{\pi}_{ij}}_{\text{consideration}} \;+\; \underbrace{\alpha_i \sum_{k\neq j, k>0} s_{ik}}_{\text{competition across firms}} \;+\; \underbrace{\alpha_i\, s_{i0}}_{\text{exit to outside}}\right] dF(i).} \tag{27}$$

Three-part decomposition

Define the component elasticities:

$$\eta_j^{\text{consideration}} = \frac{1}{s_j}\int \pi_{ij}\, s_{ij}\, \eta^{\pi}_{ij}\; dF(i), \tag{28}$$

$$\sum_{k\neq j}\eta_{jk} = \frac{1}{s_j}\int \pi_{ij}\, s_{ij}\left(\sum_{k\neq j, k>0} s_{ik}\right)\alpha_i\; dF(i). \tag{29}$$

$$\eta_{j0} = \frac{1}{s_j}\int \pi_{ij}\, s_{ij}\; s_{i0}\; \alpha_i\; dF(i), \tag{30}$$

So the overall labor supply elasticity to firm j can be written as:

$$\eta_j = \eta_j^{\text{consideration}} + \sum_{k \neq j, k>0} \eta_{jk} + \eta_{j0}. \tag{31}$$

Interpretation

- **Consideration margin** ($\eta_j^{\text{consideration}}$): A higher wage draws more workers into considering firm j. This operates before any preference comparison.
- **Competition Across Firms** ($\sum_{k \neq j, k>0} \eta_{jk}$): A higher wage poaches workers from competing firms k. This is the intensive (job-switching) margin.
- **Exit to Non-employment** (η_{j0}): A higher wage pulls workers from nonemployment into firm j. This is the extensive (participation) margin.

A.2 Dynamic search as a limit of discrete choice

While our model incorporates the idea that workers' consideration sets may respond to wages, it is more common in the monopsony literature to invoke random search. Random search models focus on the rate at which workers obtain offers from the firm distribution of wages $F(w)$ (at rate λ), and the rate at which they exogenously lose jobs (rate δ). Denote the share unemployed by $u(t)$.

The connection to our framework is direct: in a dynamic search model, an employed worker's effective consideration set at any instant consists of their current employer together with—with probability $\lambda \Delta t$ in a small interval Δt—a single randomly drawn outside firm. The choice between the current and the outside firm is deterministic when there is no preference heterogeneity: the worker takes the higher-paying job.

Writing this in our notation, the dynamic labor supply facing a single firm is given by:

$$\begin{aligned}
\mathrm{P}(j|w_j,\mathbf{w_{-j}},t &+ \Delta t) = (1-\Delta t\delta)\mathrm{P}(j,t)\\
&+ \Delta t\lambda\Big(\sum_{k\neq j}\big[\mathrm{P}(k,t)f(j)\mathrm{P}(k\to j|w_j,\mathbf{w_{-j}}) - \mathrm{P}(j,t)f(k)\mathrm{P}(j\to k|w_j,\mathbf{w_{-j}})\big]\Big)\\
&+ \Delta t\lambda u(t)f(j)
\end{aligned}$$

where $\mathrm{P}(k,t)$ denotes the share of the total workforce at firm k at time t (so that $\sum_k \mathrm{P}(k,t)+u(t)=1$), and $f(j)$ is the probability that a random offer is drawn from firm j. The term $f(j)$ ensures that only offers specifically from firm j contribute to its recruitment (and similarly $f(k)$ for the separation flow to firm k). Note that no additional factor of $(1-u(t))$ appears in the job-to-job flow terms because the employment level is already captured in $\mathrm{P}(k,t)$. For simplicity, let all workers have a common reservation wage b. Because the probability that a worker transitions from k to j is a step function:

$$\mathrm{P}(k\to j|w_j,\mathbf{w_{-j}}) = \mathbf{1}_{w_k<w_j},$$

the cross-firm elasticity is effectively zero, except in the zero probability case of $w_k = w_j$. However, firms face upward-sloping supply curves because there is an asymmetry between recruits and separations: recruits are pulled from the firms below j in the job ladder while separations are lost to firms above j in the job ladder. By toggling the wage, an employer changes their rank in the wage distribution, altering the relative flows of recruits and separations, which changes the labor supply to the firm.

We can obtain this labor supply elasticity by taking limits in Δt and approximating sums by integrals. Letting $p(w,t)$ denote per-firm employment (i.e., the number of workers at a firm offering wage w) at time t, this gives us the following equation for the evolution of $p(w,t)$ from Burdett and Mortensen

(1998):

$$\frac{dp(w,t)}{dt} = \lambda\big(G(w,t) - p(w,t)(1 - F(w))\big) - \delta p(w,t)$$

while unemployment evolves as:

$$\frac{du(t)}{dt} = \delta(1 - u(t)) - \lambda u(t)$$

Here $G(w,t)$ denotes the total pool of workers who would accept an offer at wage w:

$$G(w,t) \equiv u(t) + \int_b^w f(w')p(w',t)\,dw'$$

which comprises all unemployed workers (who accept any offer) plus the mass of employed workers currently earning below w. Note that $f(w')p(w',t)$ gives the density of employed workers across wages, since $p(w',t)$ is per-firm employment and $f(w')$ is the density of wage offers.

The steady-state firm-level employment is given by:

$$p(w) = \frac{\lambda G(w)}{\lambda(1 - F(w)) + \delta}$$

where we write $G(w)$ as the steady-state $G(w,t)$. Unemployment is given by:

$$u = \frac{\delta}{\delta + \lambda}$$

G and p are jointly determined in steady state (since $G(w)$ depends on $p(w')$ for $w' < w$, and p depends on G). This joint determination yields per-firm employment levels that are in general nonlinear functions of (log) wages, and hence yield non-constant elasticities of labor supply facing the firm. Note that the flow of separations is $s(w) = \lambda(1 - F(w)) + \delta$, while the recruit inflow is $\lambda((1 - u)G(w) + u)$. In steady state, these flows are equal, which yields the steady-state firm-level labor supply, $p(w)$. This also highlights an important aspect of the residual labor supply elasticity: that it is the sum of the recruit and (negative) separation elasticities.

The steady-state labor supply elasticity. Log-differentiating the steady-state expression $p(w) = \frac{\lambda(u+(1-u)G(w))}{s(w)}$ with respect to $\log w$, where $s(w) = \lambda(1 - F(w)) + \delta$, yields:

$$\eta(w) \equiv \frac{d \log p(w)}{d \log w} = \underbrace{\frac{(1-u)wG'(w)}{u+(1-u)G(w)}}_{\eta^R(w)} + \underbrace{\frac{w\lambda f(w)}{s(w)}}_{-\eta^S(w)} \tag{32}$$

The first term is the recruit elasticity $\eta^R(w)$—how the recruitable pool responds to the wage—and the second is $-\eta^S(w)$, the negative of the separation elasticity, since a higher wage reduces the separation rate by moving the firm up the offer distribution. Both components are wage-dependent. Since in steady state $G'(w) = f(w)p(w)$ and $p(w) = \lambda(u + (1-u)G(w))/s(w)$, the recruit elasticity can be written as $\eta^R(w) = \frac{wf(w)p(w)}{u+(1-u)G(w)} = \frac{\lambda wf(w)}{s(w)}$, which implies $\eta^R(w) = -\eta^S(w)$—the doubling relationship holds at every point in the wage distribution (not just at a symmetric equilibrium), because it is a consequence of the steady-state flow balance. The total elasticity is therefore:

$$\eta(w) = \frac{2\lambda wf(w)}{s(w)} = \frac{2\lambda wf(w)}{\lambda(1-F(w))+\delta}$$

This expression shows clearly how monopsony power varies across the wage distribution. At the bottom of the distribution (near the reservation wage b), $s(w)$ is large (many firms offer more, so separations are frequent) and $f(w)$ may be small, yielding a low elasticity and substantial monopsony power. At the top, $s(w) \approx \delta$ (only exogenous separations remain), so the elasticity can be very high. This illustrates how a single structural model can generate wide variation in monopsony power across the wage distribution.

Many empirical papers in the monopsony literature estimate the amount of monopsony power held by firms by estimating the wage elasticity of separations and using this to calculate the total labor supply elasticity to the firm (comprising both the elasticity of separations and recruitment). This is because there is a steady state relationship between the separation and recruit elasticities—namely that $\eta^R = -\eta^S$—which holds in a broad class of dynamic models, provided that there is no wage impact on transitions to unemployment (Manning, 2003).[47] where the elasticities are evaluated at the equilibrium values. The key property ensuring $\eta^R = -\eta^S$ is steady-state balance, so the marginal effect of firm j raising its wage on its recruit inflow (through the transition probability $Pr(k \to j|w_j, w_k)$) is the mirror image of the effect on its quit outflow (through $Pr(j \to k|w_j, w_k)$). As shown above, in the Burdett-Mortensen model this relationship holds throughout the wage distribution as a consequence of steady-state flow balance. If each elasticity is a constant, this implies that the recruit and separation elasticities are equal in absolute value, and that the overall labor supply elasticity can be calculated by multiplying the separation elasticity by -2.

Search with random utility. Generally, however, search models do not lead to constant elasticity labor supply functions. One way to get a constant elasticity labor supply function in a search model is by combining the search structure of consideration sets with random utility draws, as in Bloesch et al. (2026), albeit with a degenerate endogenous wage distribution. Suppose, as in standard search, employed workers only have consideration sets of at most two firms—their current employer and one outside offer. Adding type-1 extreme value preference shocks with scale parameter $1/\eta$ to log-wage util-

[47] To see this, note the probability of transitioning from firm j to firm k is equal to $\mathrm{P}(k|j, \boldsymbol{w})$ where $\boldsymbol{w}$ is a vector of wages. Recruits are given by $R_j = \sum_{k \neq j} \mathrm{P}(j|k, \boldsymbol{w})\mathrm{P}(k)N$ and separations are given by $(1 - \mathrm{P}(j|j, \boldsymbol{w}))\,(\mathrm{P}(j)N) = s_j N_j$. In steady state, the number of workers joining a firm (from all firms) must equal the number leaving: $R_j = s_j N_j$. Log differentiating this expression yields the familiar expression: $\eta_j^N = \eta_j^R - \eta_j^S$

ity, the probabilities of a worker transitioning from k to j (a recruitment event for firm j) and from j to k (a separation from firm j) are:

$$P(k \to j | w_j, \mathbf{w_{-j}}) = \frac{w_j^\eta}{w_j^\eta + w_k^\eta} \qquad \text{and} \qquad P(j \to k | w_j, \mathbf{w_{-j}}) = \frac{w_k^\eta}{w_j^\eta + w_k^\eta}$$

where the logistic form follows from the standard result for pairwise choice under type-1 extreme value shocks.

In their model, Bloesch et al. (2026) show that, integrating over the distribution of vacancies offered and worker wages, these pairwise consideration sets generate a constant labor supply elasticity, equal to the sum of the recruitment elasticity and the (negative of) the separation elasticity. We refer readers to their paper for further details.

A.3 Static Monergony

In this appendix we present the full monergony model introduced in Section 4—a combination of efficiency wages and monopsony in which both the effort elasticity and the residual labor supply elasticity to the firm are finite.

We illustrate the role of efficiency-wage considerations in a monopsony setting using a labor-discipline model (Shapiro and Stiglitz, 1984; Bowles, 1985). If effort is non-contractible and hard to monitor, a higher wage raises the cost of job loss and thereby reduces shirking. We show how the distribution of outside options shapes wage setting even when effort elicitation matters. When the firm can screen on outside options, wage setting is governed by the elasticity of the *effective* labor supply of non-shirkers, and workers receive an "efficiency wage premium" above the classic monopsony wage to deter shirking. Here the firm always exercises monopsony power and is "labor-hungry," but there is also involuntary rationing: some workers willing to take the job at the posted wage are turned away because they would shirk. When screening is not available, some workers shirk in equilibrium. The firm may be (i) supply-constrained, when both effort elicitation and participation matter, leading to a markdown, or (ii) demand-constrained, when effort extraction is sufficiently costly and the optimal wage is sufficiently high that the supply constraint does not bind, and jobs are randomly rationed. Even then, the Solow wage depends on the shape of the outside-option distribution.

Model setup: Workers have heterogeneous outside options b_i with smooth CDF $F()$ and density $f()$. In standard monopsony, the firm posts w and all with $b_i \leq w$ work and produce the same output. In our setting, each worker sees w and her private b_i and chooses effort $e_i \in \{0, 1\}$. The firm may be able to detect b_i (depending on the full screening versus no-screening case); if it knows b_i it can condition hiring on b_i, but cannot wage discriminate. Importantly, it cannot observe effort, but uses a monitoring technology that detects shirking ($e_i = 0$) with probability $P \in (0, 1)$. A detected shirker is fired and receives b_i, so there is limited liability. A non-shirker produces one unit of effective labor at cost c.

Given the monitoring technology, a risk-neutral worker with outside option b_i exerts effort only if

$$\underbrace{w - c}_{\text{working}} \geq \underbrace{(1 - P) \times w}_{\text{not caught shirking}} + \underbrace{P \times b_i}_{\text{caught shirking}}$$

which can be rewritten as:

$$w \geq \frac{c}{P} + b_i \tag{33}$$

The worker only accepts the offer if individual rationality (participation constraint) holds (i.e., if accepting the offer is weakly better than taking the outside option b_i):

$$\max\{\underbrace{w-c}_{\text{work}}, \underbrace{(1-P)\times w + P\times b_i}_{\text{shirk}}\} \geq b_i$$

Since $P < 1$, if the incentive compatibility constraint is satisfied, the participation constraint is also satisfied. Let $\phi = \frac{c}{P}$ be the no-shirking premium that solves the no-shirking constraint for a worker, so that:

$$Pw + (1-P)b_i - c \geq b_i \implies w \geq b_i + \phi$$

We use L to denote the total labor hired by the firm. For a firm offering w, the total supply of effective labor (which we denote $E(w)$) depends on both whether workers accept the offer and whether they do not shirk (shirkers contribute nothing to production). Note that $E(w)$ denotes the maximum effective labor available at wage w; realized effective labor may be lower if the firm is demand constrained. Further, for notational simplicity we do not condition all expressions on F.

The total supply of effective labor, assuming a unit mass of labor, is given by:

$$E(w) = \int \mathbf{1}_{w \geq b_i + \phi} dF(b_i) = F(w-\phi) \tag{34}$$

The average effective labor per worker hired depends on whether the firm can screen on b_i:

$$e(w) = \begin{cases} \frac{F(w-\phi)}{F(w)} & \text{if screening on } b_i \text{ infeasible } (L \leq F(w)) \\ 1 & \text{if screening on } b_i \text{ feasible } (L \leq F(w-\phi)) \end{cases} \tag{35}$$

With full screening, the firm rationally hires only workers with $b_i \leq w - \phi$ (non-shirkers). Without screening, only workers with $b_i \leq w - \phi$ provide effort, but workers with $b_i \leq w$ accept the job.

If $G(E)$ is a smooth, increasing, and concave production function, the firm's problem is either :

$$\max_{w,L} G(e(w)L) - wL \quad \text{s.t.} \quad L \leq F(w).$$

$$\max_{w,L} G(e(w)L) - wL \quad \text{s.t.} \quad L \leq F(w - \phi).$$

if screening is feasible. $e(w)$ is as defined in (35). In all cases, there are some parameterizations in which the firm would prefer to shut down; for simplicity, we assume the firm finds it profitable to operate.

Explicit Solution for Main Text Examples We next present explicit solutions for the three cases presented graphically in the main text. We deliberately chose tractable functional forms. In particular we assume a log production function: $G(e(w)L) = a\log(e(w)L)$. This ensures that the MRPL schedule $G'(e(w)L) = a/(e(w)L)$ depends on effective labor alone, so the MRPL curve does not shift across cases, facilitating a graphical presentation. Further, we suppose labor supply to the firm is a shifted-power function, as in Card et al. (2018), which can be microfounded as a random utility labor supply function with deterministic utility from a firm paying wage w_j given by $\kappa\log(w_j - \mu)$, where μ is a parameter of the labor supply function and workers have firm-specific idiosyncratic preferences (ϵ_{ij}) which follow a type-1 extreme value distribution. The resulting labor supply function to a firm is given by[48]:

$$L(w) = F(w) = \begin{cases} C(w-\mu)^{\kappa}, & \text{if } w \geq \mu, \\ 0, & \text{if } w < \mu, \end{cases} \qquad \kappa > 0, (w-\mu)^{\kappa} \leq \frac{1}{C}$$

[48]While these microfoundations are inconsistent with the risk-neutral worker utilities from the previous section, we merely need the shifted-power functional form for our illustrative calculations and closed form expressions. An alternative approach would be to directly assume the distribution of outside options have the shifted power form.

where $\frac{1}{C} = \sum_k (w_k - \mu)^\kappa$ is a normalization term that includes w_j and ensures that $F(w)$ is a distribution. Following Card et al. (2018) we assume the set of k firms is large so that the effect of the wage on the denominator C is negligible over the relevant range. Without loss of generality we set $C = 1$ and suppress it henceforth. The corresponding density of reservation utilities is

$$f(w) = F'(w) = \kappa(w - \mu)^{\kappa - 1} \quad \text{for } w > \mu$$

and the elasticity of labor supply is:

$$\eta(w) = \kappa \frac{w}{w - \mu}.$$

The log production function ensures independence of the marginal revenue product of labor from effort. The shifted-power labor supply function (used in Card et al. (2018)) makes the labor supply elasticity explicitly a function of μ. There is a simple comparison with Shapiro and Stiglitz (1984). It also nests the constant elasticity case, which is commonly used in the empirical and structural literature. A limitation of the constant elasticity case is that it does not allow the elasticity to be different at a wage $w > \mu$ versus at $w = \mu$, and as we discussed in the main text and show below, this obscures the relationship between the Shapiro-Stiglitz model and our model.

As discussed in the text, there are three relevant cases, which we summarize in Table 2. For the figures in the text, we fix $a = .7$, $\kappa = 3$, $\mu = 0.2$, and $\phi = 0.08$ for Cases 1 and 2, and $\phi = 0.18$ for Case 3.

Table 2: **Monergony Cases**

	Full Screening	No Screening
Labor supply constrained	**Case 1:** $e(w) = 1$ $L = F(w - \phi)$	**Case 2:** $e(w) = \dfrac{F(w-\phi)}{F(w)}$ $L = F(w)$
Labor demand constrained	—	**Case 3:** $e(w) = \dfrac{F(w-\phi)}{F(w)}$ $L < F(w)$

Note: When screening on b_i is possible (Case 1), only workers who will exert effort are hired, so there is (efficient) rationing of jobs. Without screening (Cases 2 and 3), some hired workers shirk, of whom a share P are fired. In Case 3, the firm hires a random subset of workers willing to work (random rationing).

Case 1: Full screening is feasible (labor supply binds in terms of effective labor). With full screening, the firm never hires workers who would shirk (those with $b_i \in (w - \phi, w]$).[49] Hence the total amount of labor equals the total amount of effective labor $L = E$. Employer screening leads jobs to be rationed (efficiently) as there are workers willing to work at the offered wage of w who are not hired because they have too high a b_i (as they would shirk). This is also a novel prediction of why employers may spend resources screening workers even as they are "labor hungry."

The firm's problem reduces to

$$\max_{w} \; \Pi(w) = G(E(w)) - wE(w), \qquad E'(w) = f(w - \phi).$$

The first-order condition for the profit-maximizing wage, w_1, is:

$$G'(E)E'(w_1) - \big(E(w_1) + w_1 E'(w_1)\big) = 0 \iff G'(E) = \frac{E(w_1)}{E'(w_1)} + w_1 = \frac{F(w_1 - \phi)}{f(w_1 - \phi)} + w_1.$$

[49]The firm would increase profits by firing a shirker, since that would not affect total production, while lowering costs.

Rearranging,

$$w_1 \;=\; G'(E) \;-\; \frac{F(w_1-\phi)}{f(w_1-\phi)}. \tag{36}$$

Let $\eta(x) \equiv \dfrac{x\,f(x)}{F(x)}$ denote the labor supply elasticity evaluated at x. Since $\dfrac{F(w-\phi)}{f(w-\phi)} = \dfrac{w-\phi}{\eta(w-\phi)}$, (36) implies

$$w_1 \;=\; G'(E) - \frac{w_1-\phi}{\eta(w_1-\phi)} \iff w_1 \;=\; \frac{\eta(w_1-\phi)}{1+\eta(w_1-\phi)}\underbrace{G'(E)}_{\text{MRPL}} + \underbrace{\frac{\phi}{1+\eta(w_1-\phi)}}_{\text{Efficiency Wage Premium}} \tag{37}$$

This is the standard monopsony markdown, but applied to *effective* labor E, plus an additional "Efficiency Wage Premium" term $\phi/(1+\eta)$ arising from the no-shirking requirement. Intuitively, the firm exercises monopsony power over E (supply $F(w_1-\phi)$), and all hires exert effort, so $G'(E)$ is the MRPL of effective labor.

In Figure 4, this case can be shown as a shift back in the firm-specific labor supply, and hence the MCL curve, as compared to the classic monopsony case, because now a wage of w yields $E = F(w-\phi)$ instead of $F(w)$ workers. For a given number of effective workers, the MRPL is the same, $G'(E)$, because all hired workers exert effort when employers can perfectly screen based on outside options. Wages are higher than under classic monopsony. The full screening case provides a clear illustration of how incentive compatibility constraints can offset monopsony wage-setting.

We can also see why the bottom-left cell of Table 2 is empty. Since employers can hire based on outside options, but cannot condition the wage on the outside option, they will always be (effective) labor supply constrained, and the traditional efficiency-wage condition, or Solow condition, will never bind. Suppose the number of hired workers (L) were less than the number of non-shirkers willing to work at the profit-maximizing wage w, so that

$L < F(w-\phi)$. An employer who can screen could then lower the wage to w' such that $L = F(w'-\phi)$ and still hire L non-shirkers, thereby raising profits; this contradicts the claim that w was profit-maximizing. Furthermore, it never pays for a firm to hire more than $F(w-\phi)$ workers at wage w. If $L > F(w-\phi)$ at the optimum, some workers are shirking; a profit-maximizing employer with the ability to screen can fire an expected shirker, reducing costs without harming output. This contradicts the claim that L is profit-maximizing. Therefore, with full screening, profit maximization implies the effective labor supply constraint is always binding: $L = F(w-\phi)$.

Adding Functional Form Assumptions: With shifted-power labor supply, the binding supply constraint implies

$$L_1 = F(w_1-\phi) = (w_1-\phi-\mu)^{\kappa}, \qquad e(w_1) = 1, \qquad E_1 = L_1.$$

Profit as a function of w is

$$\begin{aligned}\Pi_1(w) &= G(E_1(w)) - wL_1(w) = a\ln\Big((w-\phi-\mu)^{\kappa}\Big) - w(w-\phi-\mu)^{\kappa} \\ &= a\kappa\ln(w-\phi-\mu) - w(w-\phi-\mu)^{\kappa}.\end{aligned}$$

The first-order condition $d\Pi_1/dw = 0$ is

$$\frac{a\kappa}{(w_1-\phi-\mu)} - (w_1-\phi-\mu)^{\kappa} - \kappa w_1(w_1-\phi-\mu)^{\kappa-1} = 0.$$

The marginal revenue product of labor at the optimum is

$$\text{MRPL}_1 = G'(E_1) = \frac{a}{E_1} = \frac{a}{(w_1-\phi-\mu)^{\kappa}},$$

which is graphically illustrated with parameter values in Figure 4a.

Case 2: No screening, and labor supply binds. Now suppose the firm cannot condition hiring on μ, and the labor supply constraint binds. In this case effective labor is given by $E(w) = F(w - \phi)$, while employment is $L = F(w)$ as all workers with outside option worse than w will choose to work at the firm. The firm's problem is

$$\max_{w} \; G(F(w - \phi)) - wF(w).$$

Here, a higher wage raises effective labor through both participation and effort. Since $E'(w) = f(w - \phi)$, the first-order condition for w_2 (the optimal wage in the second case) is:

$$G'(E(w_2))\, f(w_2 - \phi) \;-\; \big(F(w_2) + w_2 f(w_2)\big) \;=\; 0.$$

Note that the MCL term is exactly the same as in a monopsony problem, while the MRPL is now modified by effort extraction. We can write the wage as a standard monopsony markdown applied to an effort-adjusted marginal revenue product of labor, MRPL_{eff}:

$$\text{MRPL}_{\text{eff}}(w_2) \;\equiv\; G'\big(F(w_2 - \phi)\big) \cdot \frac{f(w_2 - \phi)}{f(w_2)}, \qquad w_2 \;=\; \frac{\eta(w_2)}{1 + \eta(w_2)}\, \text{MRPL}_{\text{eff}}(w_2).$$

The first factor, $G'\big(F(w - \phi)\big)$, captures that, holding employment fixed, effective labor is scarcer when some workers shirk, so the marginal product per effective unit is higher. The second factor, $f(w_2 - \phi)/f(w_2)$, captures the *conversion rate* from a higher wage to additional effective labor through the no-shirking threshold. Relative to the classic monopsony case $E = F(w)$, these two forces can offset or reinforce each other, so the net shift in MRPL_{eff} (and hence in wages/employment) is theoretically ambiguous.

In our shifted-power labor supply example, without screening, the firm hires all workers with $b_i \le w$, so

$$L_2 = F(w_2) = (w_2 - \mu)^\kappa, \qquad e(w_2) = \frac{F(w_2 - \phi)}{F(w_2)} = \left(\frac{w_2 - \phi - \mu}{w_2 - \mu}\right)^\kappa.$$

Effective labor is again just the mass of workers who exert effort:

$$E_2 = e(w_2)L_2 = F(w_2 - \phi) = (w_2 - \phi - \mu)^\kappa.$$

Profits are

$$\begin{aligned}\Pi_2(w) &= G(E_2(w)) - wL_2(w) = a \ln\Big((w - \phi - \mu)^\kappa\Big) - w(w - \mu)^\kappa \\ &= a\kappa \ln(w - \phi - \mu) - w(w - \mu)^\kappa.\end{aligned}$$

The first-order condition $d\Pi_2/dw = 0$ is

$$\frac{a\kappa}{(w_2 - \phi - \mu)} - (w_2 - \mu)^\kappa - \kappa w_2 (w_2 - \mu)^{\kappa - 1} = 0,$$

or equivalently

$$\frac{a\kappa}{(w_2 - \phi - \mu)} = (w_2 - \mu)^{\kappa - 1}\big[(w_2 - \mu) + \kappa w_2\big].$$

This equation implicitly defines the Case-2 efficiency wage w_2 as a function of (a, μ, κ, ϕ). We show this case in Figure 4b, where we keep all the parameters as in Figure 4a.

Case 3: No screening, and labor demand binds. The labor supply constraint need not bind when effort (hence productivity per worker) rises with the wage. If the efficiency-wage choice of w (i.e., the choice of w ignoring $L \le F(w)$) yields excess labor, the constraint is slack and the unconstrained solution is optimal.

Because the firm cannot condition hiring on b_i, the average effective labor per hired worker is $e(w_3) = F(w_3 - \phi)/F(w_3)$; this is the share of workers who do not shirk, among workers who are willing to accept the offer. The amount of effective labor employed at the firm is $e(w_3)L_3$. The firm's unconstrained problem picks w_3 and L_3 to solve:

$$\max_{w_3, L_3} \; G(e(w_3)L_3) - w_3 L_3.$$

with first-order conditions:

$$G'(eL_3)\, e'(w_3)\, L_3 - L_3 = 0, \qquad G'(eL_3)\, e(w_3) - w_3 = 0.$$

Using w_3 to denote the firm's profit-maximizing wage, and combining the two first-order conditions gives the Solow condition

$$\frac{e'(w_3)}{e(w_3)} = \frac{1}{w_3},$$

which, using $e(w_3) = F(w_3 - \phi)/F(w_3)$ and defining $\eta(x) \equiv x f(x)/F(x)$, is

$$\eta(w_3 - \phi)\, \frac{w_3}{w_3 - \phi} - \eta(w_3) = 1, \tag{38}$$

which gives us a solution:

$$w_3 = \frac{\phi(\eta(w_3) + 1)}{1 - \eta(w_3 - \phi) + \eta(w_3)}$$

The wage w_3 is pinned down by the Solow condition and depends only on the outside-option distribution (via η) and ϕ, not on $G(\cdot)$. Given this "Solow wage," w_3, employment satisfies

$$G'(e(w_3)L_3)\, e(w_3) = w_3,$$

so productivity shifts (G') move L_3 at a fixed w_3 (quantity adjustment).

Even without a monopsony markdown, the Solow wage depends on F through $\eta(\cdot)$. Intuitively, raising w_3 increases the mass of non-shirkers $F(w_3-\phi)$ and, with random rationing, the share of non-shirkers among those willing to work ($F(w)$); both enter $e(w_3)$. Hence the shape of F disciplines the optimal wage via the Solow condition. Note that under constant elasticity η, the Solow wage $w_3 = (1+\eta)\,\phi$ is rising in η: a more competitive labor market yields a higher wage even though there is no monopsonistic markdown.

With log production and shifted-power labor supply, but without a labor supply constraint, the firm solves

$$\max_{w_3,L_3}\; G(e(w_3)L_3)-w_3L_3 = \max_{w_3,L_3}\; a\ln(e(w_3)L_3)-w_3L_3, \qquad e(w_3) = \left(\frac{w_3-\phi-\mu}{w_3-\mu}\right)^{\kappa},$$

subject to $L_3 \leq F(w_3)$, but with the constraint slack at the optimum.

The first-order conditions are

$$G'(eL_3)\,e'(w_3)L_3 - L_3 = 0, \qquad G'(eL_3)\,e(w_3) - w_3 = 0.$$

Combining them yields the Solow condition

$$\frac{e'(w_3)}{e(w_3)}\,w_3 = 1.$$

Since

$$\ln e(w) = \kappa\big[\ln(w-\phi-\mu) - \ln(w-\mu)\big],$$

we have

$$\frac{d\ln e(w)}{dw} = \kappa\left(\frac{1}{w-\phi-\mu} - \frac{1}{w-\mu}\right) = \frac{\kappa\phi}{(w-\phi-\mu)(w-\mu)}.$$

Thus the Solow condition becomes

$$\frac{\kappa\phi}{(w_3-\phi-\mu)(w_3-\mu)}\,w_3 = 1 \qquad \iff \qquad \kappa\phi w_3 = (w_3-\phi-\mu)(w_3-\mu).$$

The relevant (positive root) solution is

$$w_3(\phi) = \mu + \frac{(1+\kappa)\phi + \sqrt{(1+\kappa)^2\phi^2 + 4\kappa\phi\mu}}{2}, \tag{39}$$

The first-order condition with respect to L_3 is

$$G'(e(w_3)L_3)\,e(w_3) = w_3 \quad \Longrightarrow \quad \frac{a}{e(w_3)L_3}\,e(w_3) = \frac{a}{L_3} = w_3,$$

so

$$\boxed{L_3 = \frac{a}{w_3(\phi)}, \qquad e(w_3)L_3 = \left(\frac{w_3-\phi-\mu}{w_3-\mu}\right)^{\kappa}\frac{a}{w_3}.}$$

This is the case illustrated in Figure 4c with the same parameters as Figure 4b, except that ϕ is increased to ensure that the equilibrium is in the labor-demand constrained region.

Shapiro-Stiglitz Limit. Further observe from Equation 39 that as κ goes to 0, the solution will be a wage equal to the Shapiro-Stiglitz wage $\mu+\phi$, and will be demand constrained so long as the total amount of labor (normalized to 1) is sufficiently large. That is, the condition for involuntary unemployment is $MRPL(L_3) = \mu + \phi > a = MRPL(1)$.

Constant Elasticity Case. Suppose $\mu = 0$, so $\eta = \kappa$ is constant. Then Equation 39 reduces to $(1+\eta)\phi$. This case makes it easy to see that as the elasticity of labor supply becomes larger, wages increase, despite labor supply not binding.

A.4 Transitions From Case 3 to Case 2

When firms cannot condition hiring on μ, they may be constrained by either labor demand (Case 3) or labor supply (Case 2). We next consider how a firm may transition between the two cases.

The supply constraint $L \leq F(w)$ is slack iff $L_3 \leq F(w_3)$, i.e.

$$\frac{a}{w_3(\phi)} \;\leq\; \big[w_3(\phi) - \mu\big]^{\kappa}.$$

At the boundary $L_3 = F(w_3)$ we have

$$a^*(\phi) = w_3(\phi)\big[w_3(\phi) - \mu\big]^{\kappa},$$

so for a given efficiency cost ϕ the global optimum is

$$\begin{cases} \text{Case 3 (demand-constrained Solow solution)}, & a \leq a^*(\phi), \\ \text{Case 2 (supply-constrained interior solution)}, & a > a^*(\phi). \end{cases}$$

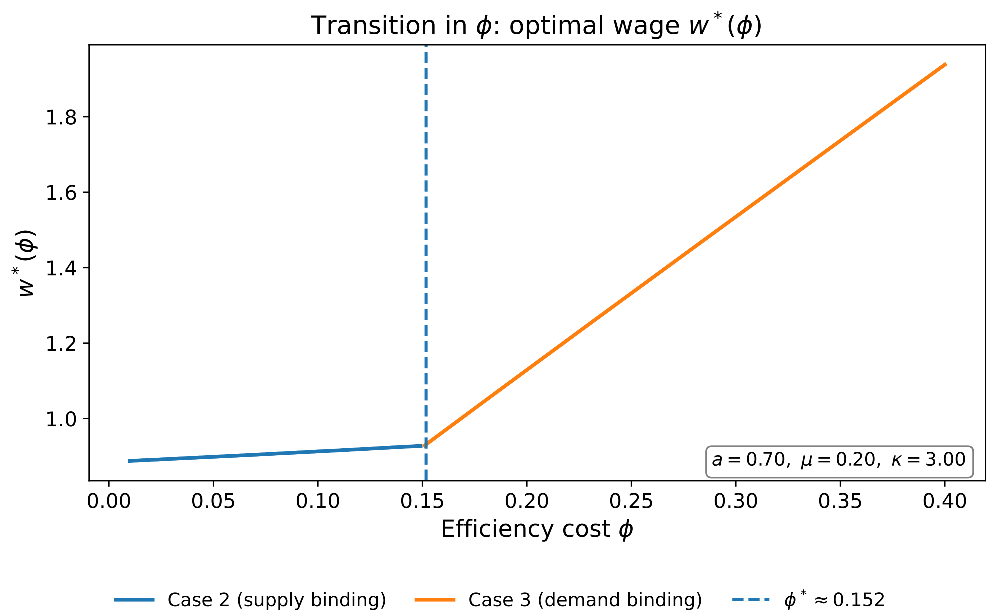
Equivalently, for given (a, μ, κ) there exists a cutoff ϕ^* (defined implicitly by $a = w_3(\phi^*)\big[w_3(\phi^*) - \mu\big]^{\kappa}$) such that the global optimum switches from the supply-constrained Case 2 for $\phi < \phi^*$ to the demand-constrained Solow solution (Case 3) for $\phi > \phi^*$.

$$\begin{cases} \text{Case 3 (demand-constrained Solow solution)}, & \phi \geq \phi^*(a), \\ \text{Case 2 (supply-constrained interior solution)}, & \phi < \phi^*(a). \end{cases}$$

The global optimum switches from the supply-constrained interior solution (Case 2) for $\phi < \phi^*$ to the demand-constrained Solow solution (Case 3) for $\phi > \phi^*$. We show this transition in the assumed parameter values in Figure 6, along with wages and employment.

We can also show how the global optimum switches from Case 3 to Case 2 as labor demand increases, by varying a. Figure 7 shows this transition occurs as labor demand increases, with the transition happening from a Case 3 fixed-wage, no-rent-sharing equilibrium with involuntary unemployment to a Case 2 rent-sharing equilibrium with no involuntary unemployment (but equilibrium shirking).

Remarks. These three cases are based on polar assumptions of full or no screening, and are meant to be a first pass at studying how to combine labor discipline and monopsony. More generally, if firms have some (partial) ability to screen, the logic of the full screening case—where firms are labor hungry and yet turn away some workers, and firms' markdown is limited by efficiency wage considerations—may be relevant for a broad segment of the labor market. Analyzing that possibility is a useful direction for future work.



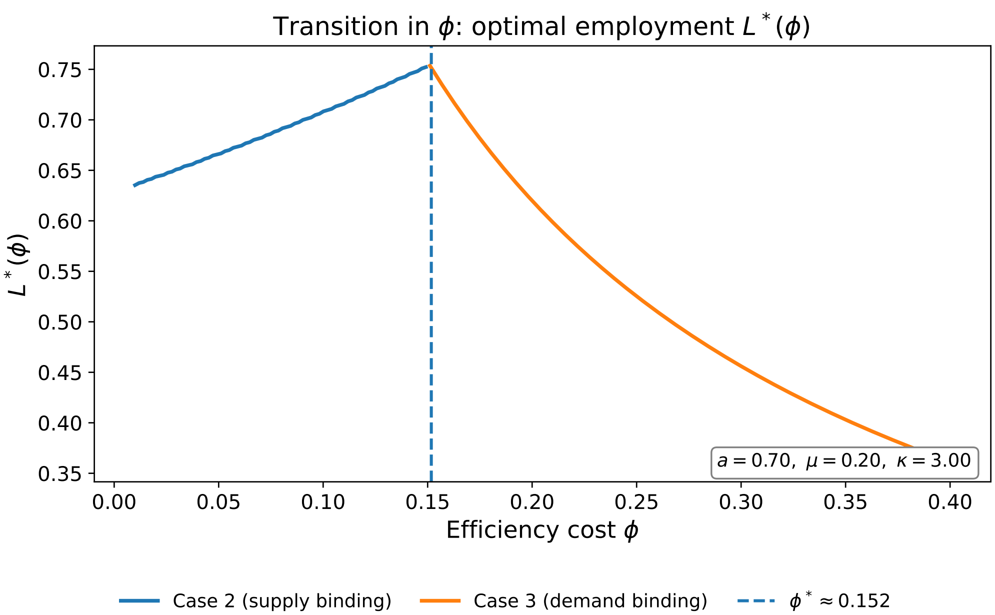


Figure 6: Employment and wages without screening for values of ϕ, showing transition from labor supply constrained case to labor demand constrained case

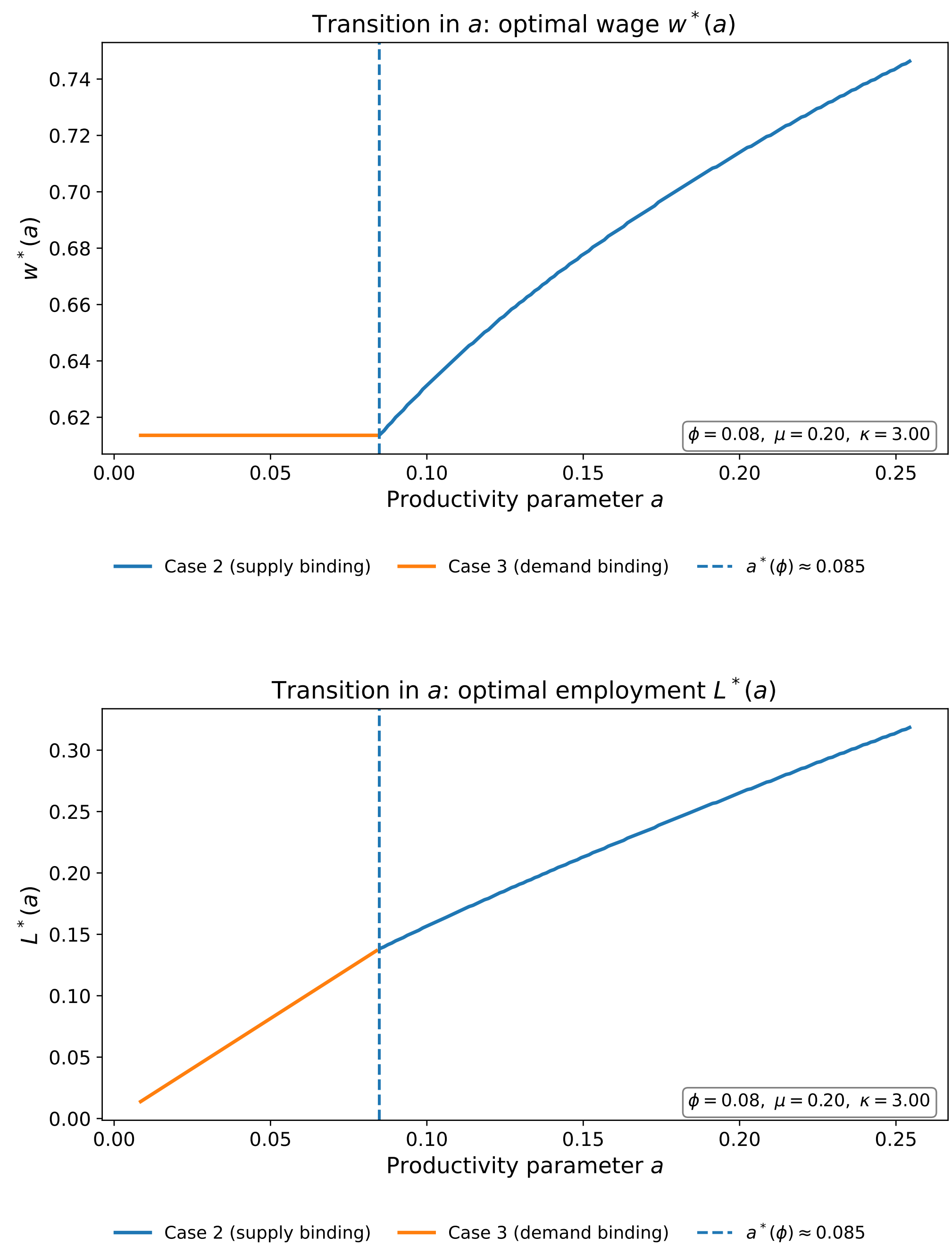


Figure 7: Employment and wages without screening for values of a, showing transition from labor supply constrained case to labor-demand constrained case.

B. Empirical Appendix

B.1 Studies Included in the Forest Plot

Study	Classification	Estimate	Standard Error
Bachmann and Frings (2017)	Calibrated model	2.022	0.160
Bachmann, Demir and Frings (2021)	Calibrated model	1.696	0.030
Dube, Manning, Naidu (2020)	Calibrated model	1.337	0.460
Ransom (2022)	Calibrated model	1.730	0.820
Serrato and Zidar (2016)	Calibrated model	4.188	4.795
Bo, Finan, and Rossi (2013)	Experimental	2.134	0.277
Caldwell and Oehlsen (2023)	Experimental	2.736	2.647
Doerrenberg, Duncan, and Löffler (2023)	Experimental	0.440	0.380
Dube, Jacobs, Naidu, and Suri (2020)	Experimental	0.140	0.020
Dube, Naidu, and Reich (2022)	Experimental	4.572	1.456
Muralidharan, Niehaus, and Sukhtankar (2023)	Experimental	3.070	6.960
Portner and Hassairi (2018)	Experimental	0.620	0.080
Amior and Manning (2020)	Non-Experimental	1.007	0.361
Autor, Dube, and McGrew (2024)	Non-Experimental	1.358	0.271
Bassier, Manning, and Petrongolo (2025)	Non-Experimental	0.738	0.004
Bodah, Burkett, Lardaro (2003)	Non-Experimental	0.420	0.111
Booth and Katic (2011)	Non-Experimental	0.709	0.122
Brooks, Kaboski, Li, and Qian (2021)	Non-Experimental	1.175	0.052
Campbell (1993)	Non-Experimental	1.590	0.710
Currie (1991)	Non-Experimental	3.650	1.730
Depew and Sorensen (2013)	Non-Experimental	2.520	0.540
Depew, Norlander, and Sorensen (2017)	Non-Experimental	1.744	0.502
Fleisher and Wang (2004)	Non-Experimental	0.245	0.046
Galizzi (2001)	Non-Experimental	1.820	0.040
Hirsch and Jahn (2015)	Non-Experimental	1.360	0.035
Hirsch, Jahn, Manning, and Oberfichtner (2019)	Non-Experimental	2.220	0.039
Hirsch, Jahn, and Schnabel (2018)	Non-Experimental	2.612	0.289
Hirsch, Schank,, and Schnabel (2010)	Non-Experimental	2.489	0.050
Hotchkiss and Quispe-Agnoli (2012)	Non-Experimental	2.370	1.850
Howes (2005)	Non-Experimental	0.680	0.108
Manning (2003)	Non-Experimental	0.547	0.107
Marinescu and Wolthoff (2020)	Non-Experimental	1.620	0.890
Meitzen (1986)	Non-Experimental	1.207	0.673
Ogloblin and Brock (2005)	Non-Experimental	-0.313	0.008
Roussille and Scuderi (2023)	Non-Experimental	4.130	1.130
Sullivan (1989)	Non-Experimental	1.270	0.210
Vick (2017)	Non-Experimental	1.630	0.020
Wasylenko (1977)	Non-Experimental	4.505	1.522
Dobbelaere and Mairesse (2013)	Production Function	2.574	1.099
Felix (forthcoming)	Production Function	1.020	0.008
Pham (2023)	Production Function	1.570	0.965
Tortarolo and Zarate (2020)	Production Function	7.620	0.043
Townsend and Allen (2024)	Production Function	3.446	0.544
Yeh, Macaluso, and Hershbein (2022)	Production Function	1.880	2.520
Amodio and Roux (2023)	Quasi-experimental demand shock	2.439	1.203
Bloesch, Larsen, and Yding (2024)	Quasi-experimental demand shock	1.490	0.335
Cho (2018)	Quasi-experimental demand shock	4.808	2.422
Dhyne et al. (2024)	Quasi-experimental demand shock	3.865	2.387
Goolsbee and Syverson (2023)	Quasi-experimental demand shock	5.051	2.066
Kline, Petkova, Williams, and Zidar (2019)	Quasi-experimental demand shock	3.240	1.700
Kroft, Luo, Mogstad, and Setzler (2020)	Quasi-experimental demand shock	3.500	1.120
Lamadon, Mogstad, and Setzler (2022)	Quasi-experimental demand shock	6.520	0.560

Volpe (2024)	Quasi-experimental demand shock	6.350	0.595
Arnold (2021)	Quasi-experimental wage-shock	0.870	0.440
Barth and Dale-Olsen (2009)	Quasi-experimental wage-shock	1.310	0.043
Bassier, Dube, and Naidu (2022)	Quasi-experimental wage-shock	4.200	0.054
Blau and Kahn (1981)	Quasi-experimental wage-shock	2.356	1.075
Datta (2023)	Quasi-experimental wage-shock	5.635	2.028
Derenoncourt and Weil (2025)	Quasi-experimental wage-shock	2.040	0.420
Dube, Giuliano, and Leonard (2019)	Quasi-experimental wage-shock	4.600	9.000
Falch (2010)	Quasi-experimental wage-shock	1.449	0.387
Falch (2011)	Quasi-experimental wage-shock	7.000	2.293
Falch (2017)	Quasi-experimental wage-shock	1.390	0.200
Gottlieb et al. (2025)	Quasi-experimental wage-shock	0.437	0.109
Matsudaira (2014)	Quasi-experimental wage-shock	-4.384	2.977
Naidu, Nyarko, and Wang; (2016)	Quasi-experimental wage-shock	1.052	0.292
Ransom and Simms (2010)	Quasi-experimental wage-shock	3.691	1.162
Staiger, Spetz, and Phibbs (2010)	Quasi-experimental wage-shock	0.127	0.185
Sulis (2011)	Quasi-experimental wage-shock	0.859	0.012

Note: We define outlier observations as having a labor supply elasticity of more than 15 or less than -10 or a standard error greater than 15.

B.2 Outliers

Study	Classification	Estimate	Standard Error
Fakhfakh and Fitzroy (2005)	Non-Experimental	18.182	6.667
Dobbelarere and Vancauteren (2017)	Production Function	11.427	50.400
Tucker (2018)	Quasi-experimental demand shock	15.550	5.150
Mas (2017)	Quasi-experimental wage-shock	18.347	8.629