Monopsony Makes it Big*

Sydnee Caldwell

UC Berkeley & NBER

Arindrajit Dube

UMass-Amherst & NBER

Suresh Naidu

Columbia & NBER

Abstract

The literature on imperfect competition in labor markets has expanded rapidly in recent years. In this survey, we focus on the residual labor supply facing the firm as a core parameter for measuring the a firm's capacity to exercise labor market power. We present a general framework nesting the three widely studied sources of labor market power: search frictions, preference heterogeneity, and concentration. To set the stage for a review of the design-based literature, we examine the partial equilibrium limits of the ideal single-firm experiment in a quantitative simulation, and assess possible biases from spillovers and non-wage adjustments.We then review a wide variety of design-based and model-based estimates of the residual labor supply elasticity, and examine how the literature has dealt with various theoretical considerations surrounding the estimation of the labor supply to the firm. We conclude by highlighting how policy issues in antitrust, labor market regulation, and immigration are illuminated by monopsony and by listing several areas for future research.

*We thank Marco Badilla, Alex Forlenza, and Sebastian Puerta for outstanding research assistance.

Contents

1. Introduction

Labor economists have long been divided over whether (and which) labor markets are best characterized by perfect competition or by models where firms exercise some degree of labor market power. In a perfectly competitive labor market, firms face flat labor supply curves. In an imperfectly competitive labor market they face upward sloping labor supply curves: in order to recruit or retain more workers in a job they must raise wages for all workers in that job. The term monopsony refers to a situation where the labor market is imperfectly competitive and wages are posted (i.e. they are decided before an individual worker and firm meet). The key parameter in the monopsony literature is η, the elasticity of labor supply to a firm. If firms are profit-maximizing, they will set wages so they are marked down from marginal product by $\frac{\eta}{1+\eta}$.

The monopsony literature—and the term monopsony—date back to 1933, when Joan Robinson published "The Economics of Imperfect Competition (Robinson, 1933). While it rolls off the tongue, at least to the initiated, a literal classical Greek translation of monopsony would be *mono opsonein* — "single buyer of delicate fishcakes".[1] The term is, however, generally used to

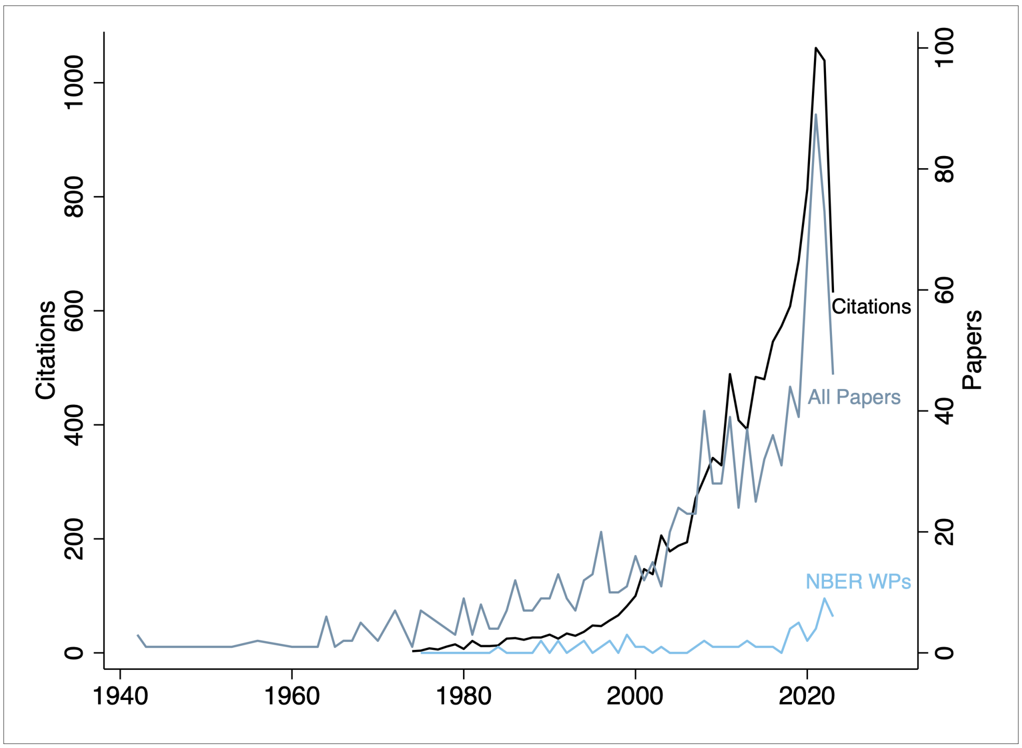
[1] See Thornton (2004) for more on the etymology.

define any situation where there is a single buyer—regardless of the identity of the good they purchase. Since the publication of Robinson's book, there have been several survey articles, including the 1997 JEL survey by (Boal et al., 1997) as well as Alan Manning's influential book (2003) and chapter in the Handbook of Labor Economics (2011). In 2010 the *Journal of Labor Economics* published papers from a conference on monopsony in labor markets (Ashenfelter et al., 2010), gathering what was, at the time, a sparse literature estimating the degree and effects of monopsony power. The conference was repeated in 2018. By then, 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 even some interest from industrial organization.

The recent explosion of papers on monopsony (plotted in Figure 1) reflects both the growing availability of administrative data, specifically matched worker-firm data, and the decreasing cost of experimental interventions in real-world labor markets. Recent empirical work suggests the labor supply elasticity to the firm lies somewhere between 2 and 6, giving firms wide—albeit incomplete—latitude to set wages as they choose. This latitude creates room for new economic forces, some of them firm-specific, to influence wages: norms, media pressure, academic recommendations, HR practices, and managerial biases all become candidate determinants of wage-setting. Some of these constraints emanate from organizational and technological practices inside firms, others are transmitted from management education, media, and still others reflect direct and indirect impacts of government policies.

We begin the review by presenting a fairly general framework that nests various models that generate upwards sloping labor supply functions facing firms. We anchor our model around discrete choice within "consideration sets", allowing for a wide variety of forces—from behavioral biases to search

Figure 1: 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 grey and blue lines plot the number of papers and National Bureau of Economic Research working papers by year. Data on the number of NBER working papers come from NBER. Data on the number of papers and citations come from Dimensions.ai.

frictions to social networks to legal restrictions—to influence the set of jobs workers are choosing from, as well as allowing for imperfect substitutability of jobs within a consideration set. We then discuss how recent papers in this literature emphasize different aspects of this generalized model.

We differ from recent surveys (Langella and Manning, 2021; Manning, 2021; Ashenfelter et al., 2022) in emphasizing three features of monopsony. First, we focus on the residual labor supply facing the firm as a core parameter for measuring the capacity a firm has for exercising labor market power, but not because firms are necessarily setting wages according to the Lerner rule. Instead, the labor supply elasticity facing the firm gives us the scope for employer market power; whether or not employers use it is a different, independently interesting question. We show that this parameter can accommodate the three most common sources of monopsony power in the literature: size or concentration of firms, heterogeneity of workers' preferences over jobs, and search frictions in the labor market. We also show the interpretation of this parameter systematically changes in the presence of non-wage methods of securing labor, for example through the provision of amenities or expending greater recruiting effort—similar to concerns about product quality and advertising in the case of product market competition. Finally, we emphasize how this parameter can be estimated with design-based methods, from experiments to instrumental variables, without requiring implausibly strong assumptions of employer optimization. We conclude by highlighting some policy areas outside of labor economics that have begun to incorporate labor market monopsony.

2. Employer Power and Wages

Standard monopsony models feature two key components: (1) upward sloping firm-specific labor supply and (2) wage posting. Upward-sloping labor supply implies that workers with the same marginal product within a firm differ in their outside options. It also ensures that most workers are infra-marginal: their outside options are worse than their current job. Wage posting means that firms must commit to one wage for a group of workers with the same marginal product (but potentially different outside options). As a result, they trade off the number of workers who will accept the offer of a given wage, with the overall costs of payroll.[2]

2.1 The Standard Model

Figure 2 shows the standard static monopsony model, with a single firm (or job) with a constant average product and a large set of workers with different outside options. The wage chosen by a monopsonist is the familiar one that trades off the revenue generated by an additional worker with the additional payroll costs. Because one wage must be chosen for the job, employing one additional worker raises the wages for all workers, including infra-marginal ones. This creates the standard static monopsony distortion, where some workers who have outside options below their productivity are still not employed by the firm due to the infra-marginal payroll cost increase it would entail.

In the figure below, the planner would choose employment at the point at which worker productivity intersects with the labor supply function. Relative to this level, the firm chooses wages—and therefore employment—that are below the social optimum. If firms can condition wages on workers' outside options, there need not be an employment distortion (see, e.g., Cahuc

[2]There is a separate literature on how individual workers' outside options shape their wage (see, e.g., Caldwell and Harmon, 2018; Jäger et al., 2020). In this review we focus on the classic monopsony case, in which firms "post wages." As previous researchers have noted, surprisingly little is known about the empirical relevance of different wage-setting protocols, and how this varies across markets with different institutional arrangements (such as the extent and form of collective bargaining) (Card, 2022). We leave a broader discussion of the wage-setting literature to future reviews.

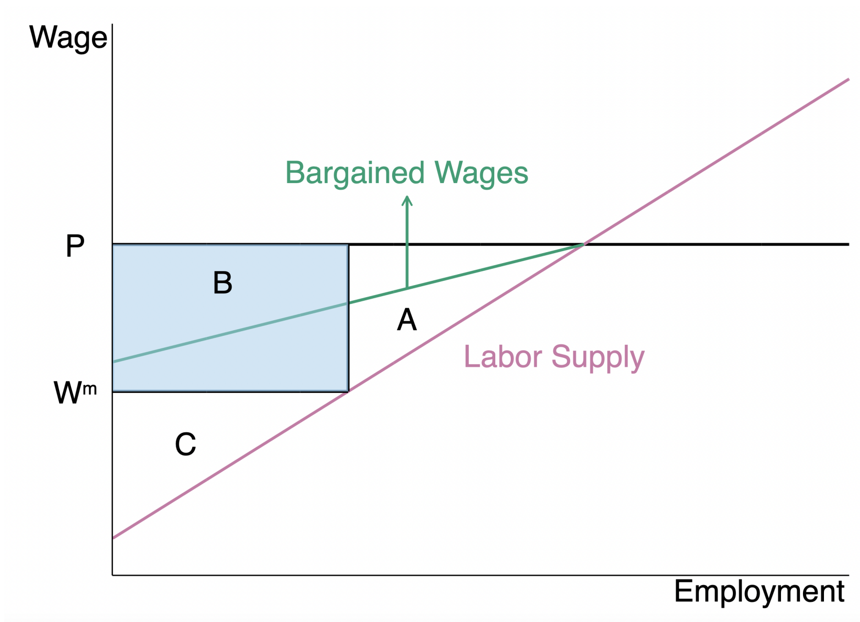
et al., 2006; Jarosch et al., 2019). In that case, employers could make take-it-or-leave it offers to each individual worker, extracting all the surplus (Postel-Vinay and Robin, 2002). On the other hand, if workers could make bids on jobs, workers would extract all the surplus. In Figure 2, we show an equal-split Nash bargain, where workers and firms split the surplus from each individual match. Again, there is no static inefficiency in this case, even though all workers are paid below their marginal product. In frictional search models like the widely used Diamond-Mortensen-Pissarides model with bilateral bargaining, the inefficiency stems from dynamic underinvestment in vacancies by firms, not voluntary underemployment of workers with relatively high outside options (Mortensen, 2003). We use the term monopsony in its classical sense, where firms post wages and the static inefficiency is present.

One reason to focus on the classic monopsony case is that it enables a rich set of theoretical and empirical analogies with the product market. In the industrial organization literature, the monopsonist's wage-setting problem is analogous to the monopolist's pricing problem; there prices are above the social optimum in monopoly pricing, just as wages are below the socially optimum level (Tirole, 1988). Together with the availability of rich, bespoke, datasets on workers and jobs, there has been an increasing literature using the tools of industrial organization to investigate power in labor markets, and we emphasize this in the survey.

There are three sources of monopsony power, each of which highlights different channels through which individual firms may be able to influence wages paid: employer size, search frictions, and heterogeneity in worker preferences. [3] In the simplest monopsony model, there is a single firm. That

[3]Sources of ex-post monopsony not discussed in this review include behavioral anchoring at current jobs Jäger et al. (2022), firm-specific human capital (see e.g. Parsons (1972) among many), and adverse selection in the secondary labor market, as in Acemoglu and Pischke (1998)(with intriguing audit-study evidence in Kroft et al. (2016) showing that re-

Figure 2: Market Power in the Labor Market



Note: This figure depicts a situation where a firm faces an upward-sloping labor supply curve where workers of equal productivity vary in their outside options. If a firm must set a single wage w (i.e. if it cannot wage discriminate), it sets a wage w^m, which is below the marginal product of labor. 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 A. 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.

firms faces an upward sloping labor supply curve as workers differ in their reservation wages; in order to convince more workers to work, it must increase the wages for all workers. A similar phenomenon may occur if firms are "large" within a market. In these models, there is a straightforward mapping from the elasticity of labor supply to the market to the firm-specific labor supply elasticity. Preference heterogeneity can also lead firms—even small firms—to face upward-sloping labor supply curves (Card et al., 2016). These preferences can stem from idiosyncratic valuations by workers of job characteristics—for example over task mix or work activities, personalities of co-workers or managers, or more nebulous preferences like self-image or sense of meaning. [4] Heterogeneity of preferences over jobs may also arise from (possibly rational) inattention of workers to job characteristics (Matejka and McKay, 2015). These idiosyncratic sources of utility are well-known as sources of downward sloping product demand in the industrial organization literature, but have only recently been widely recognized to be an analogous source of labor market power. Finally, information frictions, including search costs faced by workers, can give firms market power.

2.2 Strategic Interactions

The firm's optimal wage markdown can be found by maximizing its profits with respect to the wage:

$$\pi(w_j, w_{-j}) = (p_j - w_j)\mathrm{Pr}(w_j, w_{-j}) \tag{1}$$

sumes 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.

[4] It is known that there is an isomorphism between random-utility models and characteristics models where workers prefer jobs closer in characteristics space to a fixed vector. That is, for every distance function parameterizing a representative worker utility function in characteristics space, there is a distribution of taste shocks that generates the same representative worker labor-supply function to each firm.

where total labor supply is normalized to 1 and $\Pr(w_j, w_{-j})$ is the probability that a worker joins firm j given the wages offered by firm j (w_j) and by other firms (w_{-j}). [5] If the profit functions are quasi-concave and employers post wages, the resulting equilibrium is given by firm-specific Lerner rules:

$$\frac{p_j - w_j}{w_j} = \frac{1}{\eta_j(w_j, w_{-j})} \tag{2}$$

where $\eta_j(w_j, w_{-j})$ is the elasticity of labor supply to the firm. Equation 2 implies the tight link between the residual labor supply elasticity, η_j and the gap between wages and marginal products, or markdowns, when firms set wages constrained only by labor supply.

Many monopsony models explicitly assume that there are no strategic interactions in firm wage-setting, implying that $\partial w_{-j}/\partial w_j = 0$. This can be justified—as we do in the model in Section 3—by assuming that there are many small firms. While this may be appropriate in some cases, researchers in the monopsony literature are often interested in settings in which there are a small number of firms. In this type of environment, strategic interactions may become important. Empirical studies in the industrial organization literature have shown this in the cases of prices: once firms interact, detailed specifications of market structure become important for the analysis. In that literature (see, e.g., Weyl and Fabinger, 2013), the effect of market structure and interactions is sometimes summarized by "conduct parameters". For instance, in a Nash equilibrium in wages with (possibly heterogeneous) firms paying an identical wage w we can express the Lerner rule for wages as:

$$\frac{p_j - w}{w} = \frac{\theta_j}{\eta^M} \tag{3}$$

[5] Constant marginal productivity is not as restrictive as it may seem. Note that any constant 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 $p_j l_j$

where θ_j is a "conduct parameter" and $\eta^M = \frac{w}{\sum_j l_j(w)} \frac{d(\sum_j l_j(w))}{dw}$ is an aggregate labor supply elasticity. [6] In the case of a single firm, this is simply the standard monopsony mark-down equation.

Equation 3 shows that strategic interactions can affect the mapping between firm-specific labor supply elasticities and mark-downs. A recent strand in the literature, beginning with Berger et al. (2019a), has used macroeconomic models to explore the quantitative importance of these channels (see also Gottfries and Jarosch, 2023; Jarosch et al., 2019). In the Berger et al. (2019a) model workers first choose a labor market—based on the expected utility of jobs in that market—and then choose a firm within that market. Employers set wages both to attract workers from other firms in their market and to attract workers to their market. The model generates an elegant measure of monopsony power: the payroll HHI index. However, Berger et al. (2019a) note that that partial equilibrium estimates of the residual supply elasticity may be poor proxies for the equilibrium degree of market power when other firms best-respond with their wages. Their framework has since been adapted to study the effect of minimum wages (Berger et al., 2022b), merger guidelines (Berger et al., 2023), gender gaps (Sharma, 2023), and trade liberalization (Felix, 2021).

To date, there is little evidence that strategic interactions are quantitatively important. A paper hewing closely to our approach is Roussille and Scuderi (2021), which uses data from an online platform in which workers (primarily software engineers) post "ask wages" and firms can make bids on them to test between different models of firm conduct. They find lit-

[6] In the case of perfect monopoly or complete collusion $\theta_j = 1$; with symmetric Cournot competition it is equal to $\frac{1}{J}$. If firms are asymmetric, $\theta_j = s_j$, firm $j's$ share of employment. As the number of firms gets large—and these shares go to 0—the expression collapses to the perfectly competitive case. With differentiated Bertrand competition (e.g. random utility), $\theta_j = \frac{1}{p\frac{dl_j}{dw_j} + p\sum_{j \neq k} \frac{dw_k}{dw_j}\frac{dl_j}{dw_k}}$ summarizing the own- and cross- effects of a change in firm $j's$ wages on their employment l_j. This again collapses to 0, i.e. perfect competition. when the variance of the random component of utility goes to 0.

tle evidence for strategic interactions; the data best fit a model in which firms compete via monopsonistic competition. The results are also consistent with recent work by Derenoncourt et al. (2021), who find that voluntary and announced wage increases by large employers, including Walmart and Amazon, do not seem to lead other local firms to change their wages, suggesting little scope for strategic interactions or spillovers in response to unilateral firm policies. Finally, Sharma (2023) tests for and rejects spillovers from wage increases at individual firms induced by idiosyncratic demand shocks.[7]

The existing evidence suggests that—at least in the markets studied so far—strategic interactions may not have a large impact on equilibrium wages. One potential explanation for the lack of evidence on strategic interactions is that defining the scope of the labor market is intrinsically difficult, and the relevant scope may be more limited than market concepts typically used.[8] Another potential explanation comes from the emerging literature on "behavioral firms". For instance, two recent papers have shown that the largest employers do not set wages that vary across establishments with different labor market conditions (Hjort et al., 2020; Hazell et al., 2022). Firms also exhibit other "coarse" wage setting behavior: for example, they typically offer wages in round numbers (Dube et al., 2019a). This is hard to rationalize with a purely competitive model: it seems unlikely that the distribution of worker productivities would exhibit similar bunching. The fact that firms behave in this way in setting prices has been well documented in the indus-

[7]The paper is similar to Staiger et al. (2010b) who explore spillovers from the Veteran's Administration hospitals setting wages on nearby hospitals.

[8]For instance, Manning and Petrongolo (2017) document that labor markets are more local than standard commuting zone definitions would imply. Le Barbanchon et al. (2021) and Caldwell and Danieli (2021) highlight heterogeneity in willingness to commute by gender. Nimczik (2017) provides a novel way to define a worker's labor market, based on worker networks. Given the importance of market definition for merger review, we expect this to be an active area of research.

trial organization literature. Just as in the product market, mis-optimization persists because firms have wage-setting power: perfectly competitive labor markets would force non-optimizing firms to pay the market wage or go out of business. To a point, the presence of labor market power can allow firms to ignore the some actions of its competitors when setting wages.

2.3 Moving from Labor Supply Elasticities to Markdowns

One reason empirical researchers often estimate the residual labor supply facing the firm is to estimate the "markdown", the gap between wages and marginal product. In the absence of strategic interactions, imposing a labor supply elasticity of $\eta = 3$ on equation 2, for instance, would suggest that workers' wages are 25% below their marginal product.

While monopsony might be ubiquitous, it may also be countervailed by other labor market phenomena. One challenge with inferring markdowns from residual labor supply elasticities is that—putting aside the behavioral factors mentioned in the previous sections—firms often face internal and external constraints on wage-setting. [9]Formal constraints could include those imposed by minimum wages, or by agreements negotiated with unions or workers' councils (Bhuller et al., 2022). Less formal constraints could be those imposed by "fairness" norms within a firm, which could constrain firms' ability to differentiate pay across workers (see, e.g., Dube et al., 2019b; Giupponi and Machin, 2022), or by pressure placed on them by threat of unionization, consumers, policymakers, or the general public. Suggestively, the introduction of voluntary, corporate minimum wages by many large U.S. employers (such as Walmart and Amazon) starting in 2014 was preceded

[9]The monopsony model we consider already imposes one constraint on firm wage-setting: that firms do not condition pay on workers' outside options. While there is a long and growing theoretical and empirical literature on how changes in individual workers' outside options shape their wages, that is outside the scope of this review.

by worker protests and pressure campaign as part of the "Fight for Fifteen" movement. A literature in macroeconomics has highlighted the importance of a wide variety of constraints, including fairness, in generating downward nominal wage rigidities.(Bewley, 1999; Kaur, 2019). These constraints may lead firms to set higher wages than equation 2 would suggest.

A second challenge is that, in some settings, the amount of effort workers exert or their unobserved productivity may depend on their wage. Firms may use their wage policies to screen or provide workers with incentives; these could be seen as internal constraints on firm wage-setting. If we represent these constraints as generating a link between wages and labor productivity, which we call "effective labor" $e(w)$, the firm's problem is to choose w to maximize $\pi(w) = F(e(w_j) \times P(w_j, w_{-j})) - w \times P(w_j, w_{-j})$. where we use F as a stand-in for the firm's revenue function The first-order condition on the wage does not yield the simple markdown on marginal product, but instead $\frac{(\epsilon_e+\eta)}{(1+\eta)} = \frac{w}{f'(.e(w)} = \frac{w}{MPL}$ where ϵ_e is the elasticity of effective labor with respect to wage. A larger ϵ_e implies a smaller gap between wage and marginal product, even when the firm labor supply is inelastic. Recent work suggests that ϵ_e might be sizable: Emanuel and Harrington (2020) document a large productivity-wage elasticity among manufacturing workers at a large Fortune 500 company.

Even beyond these challenges, there is growing evidence that individual human resource managers and executives are influential in wage-setting. Researchers linking management practices scores to data on wage premia have found that better managed firms seem to pay higher wages (Bloom and Van Reenen, 2007; Bender et al., 2018). And a recent paper by Acemoglu et al. (2022) provides direct evidence that individual managers are influential in pay determination. The authors of that paper use U.S. and Danish linked employer-employee data to document "manager effects" in wage-setting, and to show that managers with business-school degrees—

increasingly hired by companies starting in the 1980s—pay lower wages than managers without such degrees (holding fixed the identity of the firm and the quality of the workers). It is hard to square their results with a simple model in which all firms simply set prices according to equation 2. Better understanding wage setting practices at companies—and how they are mediated by labor market power—strikes us as a fruitful and exciting frontier of research.

3. A Unifying Framework of Labor Supply With Consideration Sets

We next present a general framework which nests the three core sources of employer monopsony power: idiosyncratic job valuations by workers, search frictions, and employer size. The key conceptual innovation is the introduction—building on recent work in the discrete choice literature—of "consideration sets" (Cattaneo et al., 2020; Abaluck and Adams-Prassl, 2021). Specifically, we assume that workers can only see a random subset of the jobs available, and then choose from among this set. Choice sets are a natural way of capturing the role search frictions play: workers do not see all the offers available at the same time. While existing models of optimal search yield very specific formulations of consideration sets, empirical work has considered both behavioral factors and social networks as determinants of the sets of jobs workers know about. This approach can nest both formulations. [10]

[10] In the Burdett-Mortensen model, workers see only their current firm and one other (Burdett and Mortensen, 1998). In directed search, workers choose sets to apply to based on expected wages, and so the probability of applying to a firm is increasing in the wage (Wright et al., 2021). In models involving networks, the consideration set would depend on an agent's neighbors in the network, which could be summarized by degree or other network properties depending on the model (Mortensen and Vishwanath, 1994; Caldwell and Harmon, 2019).

3.1 Setup

Suppose worker i obtains flow utility $U(w_j) - b_i + \epsilon_i^j$ from working at firm j, where w_j is the wage offered by firm j, b_i is the worker's outside option, and ϵ_j^i is worker i's idiosyncratic preference for firm j. Workers are only aware of a subset of the jobs in the economy, those in their consideration set. The probability a worker chooses firm j depends on both the probability firm j is in their consideration set, and the probability that—given a consideration set—they choose firm j. For simplicity, we assume that the probability a firm is in a worker's consideration set is exogenous. Formally, the probability that the worker chooses firm j is:

$$\mathrm{P}[\text{choose } \mathbf{j} | w_j, w_{-j}] = \sum_{S \in \mathcal{S}(j)} \mathrm{P}[j = \left[\arg\max_{k \in S} V^i(w_k) | S \right] \mathrm{P}(S) \tag{4}$$

where $\mathcal{S}(j)$ is the set of all consideration sets that include firm j.

Before worker i observes her consideration set, her expected utility is:

$$E[V^i] = \sum_{S \in \mathcal{S}(j)} E[\max_{k \in S} U(w_k) | S] \mathrm{P}(S) - b_i \tag{5}$$

where the interior expression says that, conditional on seeing the consideration set S, she picks the firm—out of all firms k that are in S—which maximizes her utility. The elasticity of her labor to firm j is

$$\eta^j = \sum_{S \in 2^J} \alpha_{S(j)} \eta^j(S) \tag{6}$$

where $\alpha_{S(j)} = P(S|j)$ is the probability that the choice set is S given that the worker chose j, given by Bayes' rule.[11] $\eta^j(S)$ is the elasticity of supply to j when the consideration set is S.

[11] $\alpha_{S,j} = P(S)P(j|S) / \sum_{S \in 2^J} P(S)P(j|S) = P(j, S)/P(j)$.

3.2 Special Cases of Random Utility, Search, and Concentration

It is easy to see how this model nests the three workhorse models of monopsony. In order to recover the basic model based on preference heterogeneity, we can assume that workers have homogenous outside options ($b_i = 0$) normalized to zero, and that all workers have the same consideration set, consisting of all firms. If workers have idiosyncratic match-specific type-2 extreme value utility shocks with variance $\frac{1}{\sigma}$, and $U(w) = \beta \log(w)$ the setup collapses to that presented in Card et al. (2018) and delivers the constant residual labor supply elasticity:

$$Pr(j|w_j, w_{-j}) = Pr(j|w_j) = \frac{w_j^{\beta/\sigma}}{\sum_{j=1}^{J} w_i^{\beta/\sigma}}$$

When J is large, the denominator can be treated as constant by each firm, and the labor supply elasticity $\eta = \beta/\sigma$.[12]

Without preference heterogeneity, but with the addition of search frictions, the model collapses to a standard static search model. To see this, suppose that workers' consideration sets are generated by firms, which send messages that arrive 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 firm offers have identical chances of being seen by workers, the probability of a worker seeing a set is only a function of the size of the set: $Pr(S) = Pr(|S|) = \binom{J}{|S|} q^{|S|}(1-q)^{J-|S|}$. If the number of firms J and the number of workers N are large, this can be approximated by a Poisson distribution with parameter $\lambda = J/N$. The labor supply curve to firm j is

[12]The set-up in Berger et al. (2019b) can be mapped into this framework assuming that the consideration sets $S(j)$ are individual labor markets, chosen by workers, and themselves valued by logit preferences. Lamadon et al. (2019) augment the standard model with common firm-specific amenities valued by all workers.

then the cumulative probability of getting x offers times the probability that employer j has the highest offer out of those x, given by:

$$Pr(j|w_j, F(w)) = \sum_{x=1}^{\infty} F(w_j)^x \frac{\lambda^x exp(-\lambda)}{x!} = e^{-\lambda(1-F(w_j))} \tag{7}$$

This is a standard search model. As is standard, a closed form solution for the (non-degenerate) wage offer distribution, F, can be obtained using the fact that all firms make identical profits for every wage with positive support in F, and that $F(b) = 0$.

Finally, with neither preference heterogeneity nor information frictions, the model nests the classic model of monopsony (Robinson, 1933). Without heterogeneity in outside options (e.g. the value of leisure, home production, or migration) each firm would face a perfectly elastic labor supply curve: even a single employer would not be able to pay less than workers' outside options. With heterogeneity in outside options b_i, however, there is an upward-sloping labor supply curve to the market. In this case, classical single-employer monopsony or Cournot models of imperfect competition may apply (Tirole, 1988). For example, with a single firm j and I workers the share of labor supply to j would be given by the aggregate extensive margin labor supply function:

$$Pr(j|w_j) = \frac{\sum_{i \in I} \mathbf{1}_{w_j \geq b_i}}{|I|}.$$

3.3 Flows and Dynamics

Empirical papers in the monopsony literature often 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 quits). This is because there is a steady state relationship between the separation and recruit elasticities—namely that $\eta^R = \eta^q$—which holds in a broad class of dynamic models, provided that there is no impact on transitions to unemployment. [13]Further, there is a tight link between η^q_j and η^R_j: the recruit-weighted quit elasticity is equal to the recruit-weighted recruit elasticity (see Appendix A.3 for the derivation). If each elasticity is itself a constant, this implies that the elasticities are equal, and that the overall labor supply elasticity can be calculated by multiplying the quit elasticity by -2. [14]

We can derive recruitment and separation elasticities in the context of our model if we incorporate search dynamics. We do so by embedding the above framework into dynamic discrete choice problem, where worker i's value of a job at firm j is given by:

$$V_t^j(w_j) = U(w_j) + \epsilon_i^j + \beta \left(\sum_{S \in \mathcal{S}(j)} E[\max_{l \in S} U(w_l) | S] P(S|j) \right) \tag{8}$$

where $\mathcal{S}$ is the set of possible consideration sets for a worker at firm j.[15]For

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

[14]There are few settings in which it is possible to estimate both the recruitment and separation elasticities in the same setting. Emanuel and Harrington (2020) estimate both and find that they are roughly equal; Datta (2022) find that the recruitment elasticity is significantly higher than the separation elasticity. Because this derivation is based on the assumption that a firm is in steady state, and that firm wage changes only affect where workers work—not *whether* they do so—whether this approximation is appropriate may depend on the setting. The approximation does not, however, require the steady state to be one in which the firm is optimizing. But because a researcher must decide how to define the pool of potential recruits, there are more degrees of freedom in estimating recruitment elasticities.

[15]One important restriction we impose is that the probability firm k is in worker i's consideration set is—conditional on the identity of her current firm—independent of the wage offered by firm k. This means that while a firm that changes its wage will see a change in the number of recruits and quits, this operates only through the conditional choice probabilities. This does not require all firms to appear in a worker's consideration set with the

example, a special case (and inspiration) for this model is Langella and Manning (2021). We obtain this case by setting S equal to the current firm (i.e. the worker has no choice of firms) with probability $1-\lambda$ and the set of all firms with probability λ. Note that we simplify the problem by assuming that—apart from idiosyncratic preferences—workers are homogeneous.

The probability a worker chooses firm k in the next period depends on (1) the probability firm k is in her consideration set and (2) the probability that, conditional on firm k being in her consideration set, she chooses firm k. We use $Z(k|S)$ to denote the probability worker i chooses firm k, conditional on observing consideration set S. Given the functional form we assumed for $U(w)$ this can be written as:

$$Z(k|S,\mathbf{w}) = \exp(u(w_k)/\sum_{l\in S}\exp(u(w_l)).$$

The probability a worker chooses firm k given the vector of wages **w** is

$$P(k|\mathbf{w},j(i)) \quad = \quad \sum_{S\in\mathscr{P}}\left[Z(k|S,\mathbf{w})\times P(S|j(i))\right].$$

If workers are myopic, this setup yields an overall elasticity to firm j:[16]

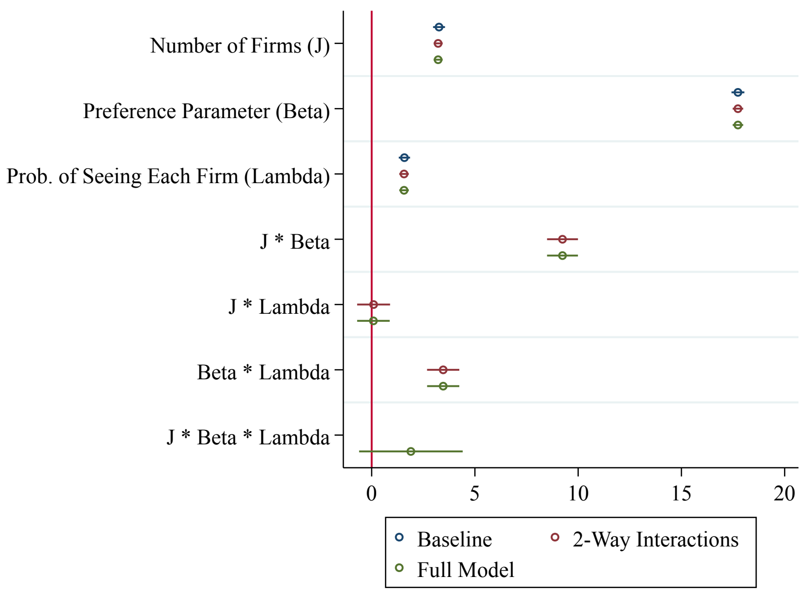
$$\eta^j = \frac{\sum_{S\in\mathscr{P}(i)} P(S|j(i))\times Z(k|S,\mathbf{w})\times(1-Z(k|S,\mathbf{w}))}{\sum_{S\in\mathscr{P}(i)} P(S|j(i))\times Z(k|S(i),\mathbf{w})}\times\left(u'(w_k)\times w_k\right). \quad (9)$$

In order to illustrate how this elasticity captures each of the three sources of monopsony power emphasized above, we simulate the model under a

same probability, however. Below, we consider the possibility that firms vary in "visibility" v_j, which influences the probability they appear in a worker's consideration set. Visibility could vary across firms for a variety of reasons, including differences in recruiting practices or brand recognition, which we discuss in Section 4.2.

[16]Assuming workers are myopic allows us to ignore the possibility that workers' choices today are influenced by their design to have specific firms in their consideration set in the future. Appendix A.2 provides more details on the derivation under the assumption of myopia.

Figure 3: Relationship Between η^j and Sources of Monopsony Power



Note: This figure depicts the standardized coefficients from regressions of η^j on the parameters associated with each simulation: N (the number of firms), β (the preference parameter), and λ (the probability of seeing any given firm). We simulate the model 2200 times. In all simulations we assume that all firms post the same wage, 10, and that—for simplicity—there are no exogenous transitions to unemployment or other employers (i.e. "godfather" shocks) .

range of different parameters. Specifically, we find the steady state of the model for labor markets which vary in the number of firms (J between 5 and 50), the probability a worker sees the offer from a given firm (λ between 0.005 and .4), and in how workers trade off wages and non-wage amenities ($u(w) = \beta \log w$, where β ranges from 0.5 to 10). In all simulations we assume that all firms post wages of 10, which is just a normalization for interpreting β. We then calculate the elasticity by changing the wages at a single firm and dividing the percentage change in employment by the percentage change in wages. Under this range of parameters, η^j ranges from very small numbers (close to 0) to very large numbers (close to 23). For instance, under the assumption that $\lambda = .005$, $\beta = 4$, and $J = 30$, we obtain an elasticity of around 3. The mean elasticity in the simulation is around 7; this is primarily useful to keep in mind as a reference for the comparative statics below.

We use simple regression models to summarize the broad patterns of how the 3 key parameters affect the labor supply elasticity. Figure 3 presents standardized coefficients from two models. In the first model (in blue) we regress the standardized elasticity on the number of firms (J), measure of search frictions (λ), and the job-differentiation preference parameter (β). To make it easier to interpret the magnitude of the coefficients, we normalize each of J, λ and β by subtracting the mid-point of the range, and dividing by half the range. The standardized variables, therefore, all range from $-.5$ to $.5$. The coefficients can be interpreted as the impact of moving from the bottom to top of the range.

As the coefficients illustrate, labor supply elasticities are higher (firms have less monopsony power) when there are more firms. Varying market concentration from 5 to 50 firms[17] raises the residual labor supply elasticity by approximately 2 around a sample mean of 7, when evaluated at the mid-

[17]Recall that concentration measured via a HHI can be converted to an effective number of firms with $N = 1/HHI$.

point of other parameter ranges. In presence of a moderate degree of search and preference based friction in the model, fairly sizable changes in firm size yields a fairly modest change in the labor supply elasticity. As expected, reducing search frictions by raising the arrival rate (λ) raises the labor supply elasticity; but again, when jobs are differentiated, increasing the offer visibility rate from approximately 0.01 to 0.4 raises the labor supply elasticity by only around 1. In contrast, when workers' weight on wages is high as compared to the idiosyncratic preferences (so β goes from 0.5 to 10), the labor supply elasticity increases sizably by around 9.

To show the interplay between the three forces of monopsony, we estimate two additional regression models. The second model (in red) includes two- way interactions between J, β, and λ; the third model (in green) also includes the three-way interaction. As the second model indicates, there are positive pairwise interactions between search frictions and preferences, and concentration and search frictions, and preferences and size, but not between search frictions and size. For example, if search frictions are low (so λ is large), an increase in the wage component relative to idiosyncratic preference (β) leads to a larger increase in the labor supply elasticity. The same applies to an increase in market size. The absence of any significant interaction between J and λ is both because there is no unemployment in this model and because additional firms do not alter the probability of seeing an offer from any particular firm (λ is exogenous). The third model shows that there are positive interactions among all three sources of monopsony, so that each margin of improved labor market competition enhances the impact of the other two.

The simulation shows that in a market that is already beset by sizable monopsony power due to job differentiation and or search friction (so β and λ are low), the incremental impact of reducing concentration will be modest, relative to markets that are already competitive on these two dimensions.

But in thick, laissez-faire labor markets, where J is already high, the effects of reduced search frictions or improved recruitment technology (raising λ) will be larger than in concentrated labor markets.In contexts where amenities are fixed, and a large share of compensation (e.g. gender preferences over jobs as in (Sharma, 2023)), β will be low and the effects of search frictions and concentration will be smaller than in contexts where wages (or endogenous compensation) are highly valued. Or increases in labor market tightness, e.g. λ, will generate bigger increases in wages where concentration is already low and preference heterogeneity is low (e.g. low-wage labor markets with many small firms). Exploring the distribution of and interactions between different sources of monopsony power is an important area for future research.

4. Non-Wage Determinants of Labor Supply Elasticities

Most of the empirical and theoretical literature focuses on the link between wages and labor supply. However, firms may offer workers both wages and non-wage amenities. They may also incur costs—in ways that do not directly enter the worker's utility function—to improve recruitment or retention. Incorporating both of these elements can alter the interpretation of existing estimates of firm monopsony power.

4.1 Amenities

Because the value of a job is more than just the wage, monopsony power may influence more than the wage, and estimates of monopsony may themselves be interpreted differently. Maestas et al. (2018) documented that non-wage amenities represent an important share of compensation. The standard framework for interpreting the joint distribution of wages and amenities is hedonic equilibrium (Smith, 1776; Rosen, 1974). Under this theory, all else equal, jobs with worse amenities are associated with higher wages, which provide a “compensating differential”. In a competitive market, the assumption of hedonic equilibrium motivates a number of policy relevant estimates, ranging from the value of a statistical life and workers compensation for injuries, to court award decisions in employment lawsuits. The extent to which workers value a variety of non-wage job characteristics is partially revealed by the higher wage paid.[18]

Labor market monopsony modifies, but does not completely overturn, the implications of hedonic equilibrium. When employers have labor market power, the value of the jobs that they offer is lower than in the competitive benchmark, but the mix of amenities and wages is still efficient. So while the overall level of employment is inefficiently low, the composition of compensation is not necessarily altered. A model integrating search based monopsony with amenities is Hwang et al. (1998), where firms can vary both in productivity as well as the cost of providing amenities. In this model, jobs vary in both the level of utility as well as the composition of wages and amenities. Many other papers have since incorporated both wages and amenities into monopsonistic labor markets (Lang and Majumdar, 2004; Dey and Flinn, 2005; Bonhomme and Jolivet, 2009; Sullivan and To, 2014; Lamadon, 2014; Taber and Vejlin, 2020).

Because frictional labor market models generate predictions about flows, they enable researchers to obtain new estimates of compensating differentials, inferring the value of amenities from the ratio of the effect of ameni-

[18]A Roy (1951) explanation of AKM might imply that all the firm-specific premia are explained by compensating differentials, but then this would not explain the responsiveness of employment flows to AKM firm-premia.

ties on firm-specific labor supply to residual labor supply elasticity with respect to wages (Gronberg and Reed, 1994).[19] Recent work by Sorkin, exploiting flows of workers across firms unexplained by wages, suggests that there may be substantial variation in the overall value provided by different firms (2018). Sorkin (2018) uses the PageRank algorithm, applied to job-to-job transitions in the LEHD, to obtain a centrality based measure of total firm value. He shows that about 70% of AKM fixed effects in earnings are explained by compensating differentials, with the rest being firm rents. Taber and Vejlin (2020) estimate the contribution of amenities to wage inequality in Denmark and find that eliminating amenities would raise wages by about 20%.

In general, the presence of endogenous non-wage amenities can also affect the estimation of labor supply elasticities. For some questions, it will be only of interest how the amenities alter the responsiveness of employment to wages. For other questions we will want to know how responsive employment is to "total job utility", including the value of non-pecuniary job amenities. If amenities are additively separable with wages—and a fixed characteristic of firms—it is straight-forward to incorporate them into the analysis (Lamadon et al., 2019).[20] If firms change amenities along with wages, however, monopsonistic firms may alter the composition of pay and amenities away from the socially optimal level, while still being efficient for the marginal worker, in order to recruit and retain (productive) workers at lower cost as in Spence (1975). Further, it is important to distinguish between the labor supply elasticities to the firm with respect to wages and with respect

[19] Lavetti and Schmutte (2016) also adjust hedonic wages for monopsony power, and recover valuation of a statistical life.

[20] The authors identify amenities up to scale by the gap between firm-size and the employment level predicted by the firm wage given labor supply elasticity identified by shocks to marginal product that do not also affect amenities (by assumption). Formally, if $U(w_j, A_j) = \beta log(w_j) + log(A_j) + \epsilon_j^i$, with a variance of random utility shocks of $\frac{1}{\eta}$, then $log(A_j) = log(l_j) - \frac{\beta}{\eta} log(w_j)$ is the level of amenities.

to utility. If the latter is significantly different from the former and amenities are a large share of job value. then the welfare costs of monopsony may be substantially larger than implied by residual supply elasticities with respect to the wage. Further, if firms, or jobs within a firm, vary in the costs of delivering amenities, then even exogenous shocks to wages may be confounded by endogenous changes in amenities. Without directly observing amenities, controlling for these confounds may be difficult.

A recent example of how how to integrate the monopsony literature with the literature on non-wage amenities comes from Dube et al. (2022). This paper uses a conjoint experiment—a standard technique in the literature on non-wage amenities (see, e.g., Maestas et al., 2018)—in which the authors asked workers, incumbent Walmart employees, about their willingness to quit in response to changes in wages and amenities. By allowing workers to choose between staying or leaving their current firm—rather than asking workers to choose between hypothetical jobs—along with a question about the time since the last dominated offer, the authors are able to estimate the quit elasticity incorporating both search as well as random tastes. These data reveal that wages and amenities are complements in these workers' utility, which implies that amenities are under-provided, as well as that exogenous increases in wages will induce firms to also raise amenities; estimates leveraging the variation in the bite of Walmart's voluntary minimum wage support this prediction.

While more research at the intersection of compensating differentials and monopsony is needed, two general take-aways are 1) worker values for amenities cannot readily be read off the observed variation in the labor market, and the 2) the degree and costs of monopsony are likely to be larger than what would be observed with wages alone. As has been recognized in product markets for a long time, quality is a key dimension of competition, and its amenity analogue in the labor market may be even more so. In addition, as we show below, with endogenous amenities, accounting for heterogeneity in non-wage job characteristics is necessary in order to recover the labor supply elasticity.

4.2 Recruiting and Retention Effort

Both wages and amenities provide utility value to workers. However, firms may also use other channels to attract or retain workers. For instance, they may take actions to affect the probability a worker sees their job (e.g., by using head hunters or temp agencies). They may also take actions to reduce the probability one of their workers leaves (e.g., by blocking job search sites or by limiting the information provided in a reference). While some of these actions may positively affect workers' option value—by increasing the probability an outside worker can choose to join the firm—they do not change the utility a worker associates with a given consideration set. A distinctive prediction recruitment effort makes is that Slutsky symmetry does not hold (Abaluck and Adams-Prassl, 2021): cross-elasticities of residual labor-supply functions need not be equal.

An existing literature has shown that fast-growing firms grow more quickly primarily by filling vacancies faster, not by posting more vacancies (Davis et al., 2013; Faberman and Menzio, 2018). Factors such as "advertising expenditures, screening methods, hiring standards, and compensation packages – influence job-filling rates through effects on applications per vacancy, applicant screening times, and acceptance rates of job offers" (Davis et al., 2013). Of course, the efficacy of recruitment may depend both on factors the firm can and cannot (easily) control. While firms may be able to improve the content of their job ads, expand their human resource departments, or spend more on external recruiting services, other factors such as their location relative to the relevant pool of workers or the social network of their employees (which they may mine for referrals) may be hard, if not impossible, to change. In a model many of these factors would function similarly.

Endogenous recruitment effort may reconcile empirical puzzles in the monopsony literature, including the apparent inconsistency between empirical estimates of the labor supply elasticity and the labor share of net value added (Bloesch and Larsen, 2023).[21] Further, endogenous recruitment effort can account for the discrepancy between the firm-size wage correlation (generally low, particularly at the top) and estimated labor supply elasticities. At the aggregate level, changes in firm recruiting effort may explain countercyclical movement in the vacancy yield over time (Gavazza et al., 2018).[22]Finally, the presence of recruitment costs also has implications for how a firm profits respond to paying a higher wage; namely, savings on these costs is another margin of adjustment which has implications for wage setting (including wage dispersion) and for policy.

The fact that employers may vary in recruiting effort—and that they may change this effort endogenously—complicates the relationship between wages and labor supply (Manning, 2006). For instance, if a firm changes its hiring standards coincident with changing its wage, the simple relationship between the number of new hires and the level of wages may not provide the labor supply elasticity. A researcher could address this concern through obtaining information on both the quantity and quality of applicants (and

[21]As Bloesch and Larsen (2023) note, with a labor supply elasticity of 4, and realistic depreciation rates and product market power, a simple monopsony model implies the labor share of net value added in the overall U.S. economy would be close to 60%, a full 20 percentage points below the observed value. But if firms are spending some of the monopsony profits on intermediate inputs that are being used for recruitment effort, the net value added measure falls so that the labor share is back to realistic levels. Of course, there are other potential explanations. One is that firms are not using all of their monopsony power.

[22]In Appendix B, we show that the assumption of exogenous vacancies yields a "Dogit" specification of labor-supply facing the firm, and this can micro-found a regression of log(wages) on market-level vacancy HHI , the specifications presented in Azar et al. (2018). By using the vacancy HHI, rather than the employment or payroll HHI, this model avoids the criticism that there is (in equilibrium) a size-wage premium, which leads to bias in the employment HHI.

potentially the quantity and quality of incumbent separators). One of the experiments we discuss below does precisely this (Dal Bó et al., 2013); there is a literature which has attempted to use online job boards to measure the relationship between firm-specific labor supply and wages (Banfi and Villena-Roldan, 2019; Belot et al., 2022).[23] Most of the existing empirical evidence has focused on how firms change recruitment effort. How—and whether—firms change retention effort is an interesting area for future research. To the extent that firms only invest in the recruitment margin, it may be possible to estimate labor supply elasticities by focusing only on the separation margin, or vice versa.

Another issue is that the steady-state equivalence of recruitment and separation elasticities researchers typically rely on breaks when firms can raise recruitment rates through changes in recruiting intensity, but have to rely on higher wages alone to lower quit rates. Further, employment levels and labor supply elasticities may vary across firms for reasons other than productivity or wages. We illustrate some of these challenges in the next section.

5. Measuring Monopsony

In most empirical papers on monopsony, the goal is to estimate the residual labor supply elasticity, and to use this to infer employers' wage-setting power. Researchers use a variety of strategies to estimate these elasticities, ranging from reduced form strategies that rely on firm-specific productivity shocks to quantitative macroeconomic multi-market models to production-function based models of labor supply. While this parameter may not yield

[23] One limitation of these studies is that the set of workers who obtain jobs via online job boards may differ from the larger pool of workers in the market. Further, it is empirically rare for firms to provide wages in job ads (Hazell and Taska, 2018); if workers have imperfect information about the wages offered by particular firms, they may be less responsive to wage changes than these estimates would imply (Jäger et al., 2022; Caldwell et al., 2023b).

an estimate of the "markdown"—for the reasons discussed in Section 2—when firms post wages this elasticity is still informative about the amount of monopsony power an employer has.

5.1 Simulating a Firm-Level Experiment

A criticism of the design-based approach is that firm-level experiments are inevitably partial equilibrium, and do not account for reactions either at the firm experiencing the shock, or the responses of other competitors in the labor market. While we do not propose to resolve this debate here, we can use the framework developed in Section 2 to get a preliminary sense of how important these considerations might be in practice.

For simplicity we assume that there are a large number of small firms, each of which post a wage of 10 and each of which provide an amenity value of 1. While stylized, we prefer this approach to one in which we specify (1) the underlying distribution of firm productivity, (2) how firms trade off wages, amenities and recruiting in moving along the labor supply curve, and (3) how firms' wages interact with each other. In our baseline simulation we assume that workers are—apart from their idiosyncratic preferences—homogeneous, that there are no godfather or separation shocks, that $u(w) = 3\log w$, that the probability a worker not at firm j sees firm j is 0.01.[24]

We estimate the firm-level elasticity by dividing the percent change in employment at the treated firm by the percent change in wages at that firm. We also calculate the average cross-elasticity of other firms' employment by dividing the percentage change in employment at these firms by the percentage change in outside wages faced by each firm (averaged across initial market shares). Without endogenous changes in amenities or recruitment

[24] Incorporating godfather or separation shocks changes the quantitative, but not qualitative findings.

effort—and given that we have assumed there is no unemployment—this elasticity is the opposite of the own-wage elasticity at the treated firm. The first row of Table 1 shows that, under our baseline parameters, the true elasticity of labor supply to firm j is 2.23. This is the partial equilibrium response a researcher would estimate by comparing employment before and after a known, exogenous wage change.

We first explore situations in which the treated firm responds by changing its behavior. A firm—even one committed to an experimental evaluation—may adjust the level of amenities in order to offset the costs of the wage increase. The second row of Table 1 presents elasticities that we computed by comparing the baseline steady state to one in which the treated firm has indeed doubled its wage, but has also halved the value of its non-wage amenities. As this table indicates, we obtain a smaller elasticity in this scenario if we do not take the change in amenities into account. If the firm responds by reducing its recruitment effort, we also obtain a smaller elasticity. If the amenity and recruitment effort respond in the same direction as the treated firm wage, then long-run elasticities would be larger than the short-run elasticities, following the Le Chatelier principle.

We then explore situations in which outside firms respond to the treated firm's actions. This may be more likely to occur if the treatment is visible or long-run. The sign of firms' response depends on whether firms are strategic complements or strategic substitutes. We assign half of the control firms to be "spillover" firms, and first consider three scenarios. We consider scenarios in which the spillover firms increase their wages, amenities, or recruitment effort by (a completely arbitrary) 5%. They may do this in order to mitigate their employment losses in the face of the treatment firm's wage increase. The Table shows that, relative to the baseline scenario, we see smaller labor supply elasticities at the treated firm. In addition, we would obtain different estimates of the outside-wage elasticity, depending on whether we used data on spillover or non-spillover control firms. If the outside firms decreased wages or amenities, or decreased their recruitment effort in response to the treated firm's wage change, we would obtain larger estimates.

This approach illustrates that in the presence of endogenous responses, by either the firm itself or its competitors, the ideal experiment does not identify the residual labor supply elasticity. However, in our example, the magnitude of the bias in estimating the own-elasticity is modest, between 5% and 40%; concretely, if the true labor supply elasticity is around 2, the estimated elasticity without would be roughly between 1 and 3.This makes us cautiously optimistic about the role of firm-level experiments in eliciting approximately accurate estimates of the labor supply elasticity. In particular, if the bias is solely due to wage spillovers (i.e., not from treatment or control firms altering amenities or recruitment), in our simulations it is at most 13%; we note that in this case the direction of bias is positive, meaning monopsony power would be under-estimated. A final implication of these simulations is that if the researcher is using a difference-in-differences estimate of the ideal parameter, the estimated elasticities would be generally be biased upwards, as workers move from control firms to the treatment firm. Difference-in-differences estimates can be significantly biased by SUTVA violations, highlighting the complications of constructing control groups from observational data alone. Of course, this is only one simulation, with a number of simplifying assumptions; additional research is required to determine whether the magnitude if the bias is likely to be generally small.

5.2 Experimental Labor Supply Elasticities

When it comes to estimating labor supply elasticities, confounding variation in observational wages is—completely unsurprisingly—a major concern. A recent meta-analysis by Sokolova and Sorensen (2021) documents that re-

Table 1: Firm-Level Experiments

	Own-Elasticity	Cross-Elasticity: Spillover	Cross-Elasticity: Non-Spillover
	(1)	(2)	(3)
Ideal Experiment	2.23	-2.23	
With Wage Spillovers	1.94	1.65	-2.12
With Amenity Response			
At Treated Firm	2.12	-2.12	
At Outside Firms	1.91	1.82	-2.28
With Recruiting Response			
At Treated Firm	1.27	-0.14	
At Outside Firms	1.77	2.49	-2.89
Number of Firms	1	25	25

Note: This table presents estimates of own-firm and cross-firm elasticities computed by simulating the model described in Section 3 assuming there are 51 firms that each post a wage of 10 and an amenity value of 1. Workers are homogeneous except for their idiosyncratic preferences. For simplicity we assume there are no separation or godfather shocks and that the probability a worker sees a given firm is (for each firm) $\lambda = 0.005$. We calculate elasticities by taking the difference in employment between two steady states, and dividing by the change in wages. The elasticities in row 1 allow control firms to raise wages by 10% in response to treatment firms, to allow wage spillovers. Rows 3 and 4 allow the treatment firm and the control firms to reduce or increase amenities to offset the treatment firm wage increase, and rows 5 and 6 do the same for recruitment effort. The elasticities in column 1 are calculated using the changes in employment and wages at the treated firm. The elasticities in column 2 are calculated using the changes in employment at the "spillover firms" (which change their amenities or recruitment effort). The elasticities in column 3 are calculated using the changes in employment at the "non-spillover firms" whose amenities and recruitment effort is unchanged.

searchers typically obtain very different estimates of the residual labor supply elasticity, depending on whether they use a specification where employment (or employment flows) are the dependent variable or whether wages are the dependent variable. As is well-known, when there is omitted variables bias, these "forward" and "reverse" regressions can be used to bound the true parameter. In particular, if there is an omitted variable—such as endogenous non-wage amenities or recruiting effort–that generates a negative bias in the relationship between employment and wages, then the coefficient in this regression is a lower bound of the true elasticity and the coefficient in the reverse regression (wages on employment) yields an upper bound on the elasticity. The range of estimates in Sokolova and Sorensen (from a lower "best practice" bound of 7 to an upper bound of 100) suggests the omitted variables bias is large (2021). This highlights the importance of identifying plausibly exogenous variation in firm-level wages.

By this light, the "gold standard" evidence would come from a series of randomized experiments across a variety of jobs, where wages are randomized, holding amenities and recruitment effort constant, and with changes in employment as the outcome. The resulting percentage change in employment across treated and control jobs, divided by the percentage change in wages, would recover the labor supply elasticity facing the firm.

It is not surprising that these experiments are few and far between. Because there are many constraints on firm wage setting practices, discussions with employers to randomize wages can be met with resistance. Employers may worry about administrative constraints or perceptions about unfairness. In response to a wage change, lower-level managers may endogenously respond by changing non-wage benefits or recruitment effort. These concerns could be addressed through the design of the experiment. However, if a firm is "large" in a market, experimental variation in wages could induce equilibrium responses by both other workers and other firms; as shown in the previous section, this makes the reduced form elasticities heterogeneous and difficult to interpret.

Thus far, deliberate RCTs implementing this strategy have thus been conducted in somewhat idiosyncratic labor markets. Dal Bó et al. (2013) arguably provide the cleanest and most comprehensive example of estimating firm level market power. They analyze a recruitment experiment conducted by the Mexican federal government's Regional Development Program (RDP). The joint randomization of offers and wages, along with the timing of the wage announcements, guard against a number of the concerns of endogenous recruitment effort that we highlighted in the previous section. This experimental design allowed the authors to assess the impact of wages on labor supply, or more accurately, recruitment. The authors examine the effects on both the applicant pool and the rate at which selected candidates accepted offers. Combining both margins, the authors compute an arc elasticity of labor supply of 2.15.[25]

Other experimental estimates come from gig-worker platforms. Dube et al. (2018) use an experimental design to measure market power in the crowd-sourcing platform, Amazon Mechanical Turk (or M-Turk). The authors posted a total of 5,500 unique image-tagging tasks on MTurk that included a very short survey and a screening task. Respondents were then offered a choice of completing an additional set of classification tasks for a pay rate which was randomized between \$0.05 and \$0.15. Workers could accept or reject the offer. When workers on M-Turk were offered at the high end of the range (\$0.15), around 79 percent of them accepted the task. But when they were offered at the low end (\$0.05), 73 percent of them still accepted the task. So, while the acceptance rate was rising with the reward, the im-

[25] In addition, the sites offering higher salary saw improved quality of candidates in terms of test scores, past earnings, and occupational profile. The recruitment estimates measuring monopsony are, therefore, conditional on any impact a higher wage has on quality.

plied labor supply elasticity was slightly smaller than 0.1, a surprisingly low amount. The findings do not appear to be a feature of this particular experimental setup. A re-analysis of 5 prior wage experiments on MTurk (that had been conducted for purposes other than measuring market power) suggested a (precision weighted) labor supply elasticity of 0.14. The results suggests that the online platform (contrary to what one may have thought) is characterized by high degree of labor market power. One possible reason for this may be that the tasks on the platform are highly idiosyncratic, leading to sizable differentiation in jobs. Another might be that it is difficult for workers to process information about pay and job quality in spite of their being posted in a centralized fashion, leading to search frictions.

Caldwell and Oehlsen (2018) undertook a series of experiments in conjunction with Uber, where a set of Uber drivers were given the opportunity to drive for one week with 10-50% higher earnings for each trip. The timing and generosity were both randomly varied from driver to driver. Moreover, while some drivers had access to a competing ride-share company, other drivers did not. By analyzing how Uber drivers leave the other competing ride-share employer and shift to Uber in response to the (randomly varied) additional pay at Uber, the authors can calculate a residual labor supply elasticity facing the firm. Their central estimate is around -2, suggesting a sizable degree of labor market power. This elasticity did not vary appreciably between men and women.

The designs used in the gig economy could be replicated by researchers themselves in markets where wages are low enough that it is feasible for a researcher to intervene in labor markets independently. [26]However, one lim-

[26]For instance, Hardy and McCasland (2023) conduct a field experiment randomly placing unemployed young people as apprentices in small firms in Ghana. Treated firms experienced increases in firm size of about half a worker, and a 10 percent increase in profits. This suggests these firms were, indeed, labor supply constrained. More of this type of supplemental evidence would be useful is providing a better understanding of what types of constraints faced by the firm are binding.

itation of the experimental literature—unlikely to be solved by producing estimates in a greater variety of empirical settings—is that some of the constraints on firm wage-setting described in section 2.3 are less likely to be operative in the settings in which it is feasible to run experiments. One notable aspect of these experimental evidence is that it is largely a job (or task) based variation in wages in a context where there production is largely separable and not joint across workers. None of the workers in these four experiments were working side by side with other similar workers at a workplace, which made it feasible to vary their wages. It seems organizationally difficult–if not impossible–to vary wages across similar co-workers at a worksite given concerns about fairness. Even more importantly, the estimates from such variation are unlikely to yield a measure of monopsony power as they likely additionally reflect animus against unequal treatment among peers. For both experimental and quasi-experimental evidence in the context of joint production, the typical unit of treatment is a workplace (establishment) within a firm, or the firm itself. Naturally, such a design creates practical challenges, since the number of treated units has to be sufficiently large to allow informative inference.

One way forward might be to elicit labor supply elasticities using the survey methodology often used to elicit workers' preferences over amenities (see Maestas et al., 2018, for one such example). While researchers often prefer to use data on revealed preferences, rather than stated preferences, such experiments could yield useful insights—especially in light of the financial and logistical difficulties associated with field experiments in this area.

5.3 Leveraging Quasi-Experimental Variation

Without readily implementable experimental variation, researchers have relied on other sources of plausibly exogenous variation in firm or job wages to trace out labor supply elasticities. Some of these papers have used exogenous variation in wages ("wage shock"). Others have relied on exogenous shocks to marginal products of labor, and use this as an instrument for firm wages and employment, with the resulting estimates tracing out the labor supply elasticity from variation in firm labor demand ("demand shock").

5.3.1 Wage shock

The classic "wage shock" study was conducted by Staiger et al. (2010a), who examine the effect of mandated wage increases legislated by the Nurse Pay Act of 1990 at Veteran's Affairs (VA) hospitals. These pay increases varied by area, and critically, affected only affected a single employer (VA hospital) in the local labor market. This approximates an experimental design where different VAs in different markets were randomly mandated to increase wages by different amount. As the mandated wage hikes raised VA wages relative to area hospitals, the authors use that instrumented variation in wages to examine change in nurse employment. They find a surprisingly small change in 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. These spillovers from a single firm's increase in wages are consistent with market power, though the interpretation of the labor supply elasticity from this variation depends on whether the spillovers reflect strategic interactions among employers. Overall, the small size of the estimated labor supply elasticity could reflect a combination of very thin labor market for nurses in these areas, the short run nature of the supply response, or strategic response by competitors.

Datta (2022) studies a large, private employer that provides services to various greater London municipalities in the UK. Datta uses the fact that different municipalities passed "Living Wage" laws that mandated municipal contractors to pay a minimum wage; and they did so at different times between 2010 and 2018. These laws ended up raising wages at the Company, but for essentially none of its labor market competitors. As in the VA example, this is a single workplace in a local labor market to raise wages at plausibly exogenous times. However, different from the VA example, this firm is situated in a fairly thick labor markets for low-wage service jobs. Datta leverages the variation to study the impact on both separations as well as recruitment (namely, the rate at which vacancies are both posted and filled). Considering the impact on (incumbent) workers' separations, Datta finds a separation elasticity of around -1.6. He also estimates how applications, offers, and rejections respond to wages in order to construct a "recruit" elasticity: namely, he defines the recruits to be those who are offered a job (but may or may not take it). He finds this recruit elasticity to be around 3. Taking the sum of the separation and recruit elasticity implies an overall labor supply elasticity of around 4.6; if firm conduct fully reflects this monopsony power, wages would be marked down by around 18 percent in this market.[27]

Emanuel and Harrington (2020) use a similar design to study a warehouse employer that adopted a voluntary $15 minimum wage in the aftermath of the pandemic. As in the case of municipal contractors and the VA, the bite of this minimum wage varies across the more than 50 stores as a result of differences in regional wages. The authors' store-level separation response of -9.6 is substantially larger than the other work. Applying the approximation of doubling the separations elasticity, the authors' findings imply a markdown of only around 5 percent.[28]

Dube et al. (2019b) examine the effects of raises enacted by a large retailer in response to a federal minimum wage increase. These raises went to not

[27] Datta's definition of recruit elasticity is different from a hires elasticity; for this reason it is not clear that the sum of the recruit and separation elasticity would produce a residual labor supply elasticity, since it includes impact on offers that are rejected by workers.

[28] The authors also find substantial productivity gains from higher wage.

just minimum wage workers but also those who were initially earning somewhat above the minimum. Critically, the raises were done in a coarse fashion where the new wage was a step function of the old wage for many workers, leading to discontinuities in raises across coworkers. Depending on their initial salary, two workers earning within 1 cent of each other could see raises that were 2 percent different: some were "lucky" while others were "unlucky." Given the small number of workers per store, it was possible for store level wages to have risen more or less entirely due to having more or less "lucky" workers. This allowed the authors to use both (within-store) worker level variation as well as store-level variation in wages to study the impact of raises on quits. The authors argue that while the store-level variation plausibly identifies monopsony power, the individual level variation mixes monopsony power as well as peer effects (due to fairness concerns). Consistent with this interpretation, they find large separation response when using within-store variation in wages across peers (around -6). But when using variation across stores, the separation elasticities are only around -2 (though somewhat imprecise).

In all four papers above, identification comes from either a government mandated, or a headquarter-mandated, wage increase that is differentially binding across local establishments or firms. In many cases, this may provide a reasonable approximation to a randomized wage increase across workplaces that can be used to study employment flows and levels. However, they are subject to the caveats discussed above: if local managers (or even headquarters in the case of a legally mandated wage hike) can vary amenities or recruitment effort in response to the wage increase, the measured labor supply effects may not only measure the workers labor supply elasticity to compensation but also change in compensation or ease of finding out about the position. This does not, of course, make the reduced form labor supply response to the mandated wage uninteresting; however, it suggests that additional evidence on these margins would be useful when translating the reduced form labor supply estimates to the underlying theoretically-grounded estimands.

One additional challenge with comparing individual case studies is that the separation or labor supply elasticities are expected to vary widely depending, not just on the market structure, but also the location of the firm 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 much higher quit elasticity than a firm towards the upper end of the distribution. We currently do not have experimental or quasi-experimental labor supply elasticities from different firms in the same market. However, a number of papers have estimated observational separation elasticities using a matching approach. For example, Bassier et al. (2022) use a matched event study design to track observationally similar workers who leave the same origin firm and move to destination firms that differ in terms of overall average wages. As expected, workers moving to higher wage firms see a larger raise. By considering how that wage differential affects the workers' separation behavior at the destination firms, the authors estimate a separation elasticity of around 2, or a labor supply elasticity of around 4. While the matching approach may better able to deal with worker-level omitted variables than purely cross sectional analysis, even if workers had been randomly placed at high versus low wage firms, a key challenge is that these firms may differ on other margins. This includes non-wage compensation as well as recruitment strategies that firms may; the authors do attempt to account for other likely non-wage amenities, but the observational nature of the analysis places clear limitations on this exercise.[29]

[29] As our previous discussion pointed out, this challenge could remain even when the firm wage is randomized if other margins such as amenities or recruitment effort could respond to the wage.

5.3.2 Product demand shocks

A number of papers have used demand shocks that primarily affects an individual firm without affecting rivals in the relevant labor market. As expected from the literature on rent sharing (Card et al., 2013; Garin and Silverio, 2022), firms typically raise wages following a positive demand shock. The idiosyncratic component of product demand can be used as an instrument for firm-level wages in a regression of employment on wages. Taking the standard model seriously, we can first estimate a a "rent-sharing" specification for the first stage using a firm j level demand shock as the instrument:

$$log(wage_j) = \gamma^w Shock_j + \nu_j \tag{10}$$

Subsequently, we can evaluate the effect of the same shock on employment as the reduced form:

$$log(employment_j) = \gamma^e Shock_j + \epsilon_j \tag{11}$$

Putting the two together, we can obtain an estimate of the labor supply elasticity elasticity as $\eta = \frac{\gamma^e}{\gamma^w} \approx \frac{dlog(employment)}{dlog(wage)}$. Importantly, just as with the wage variation, the shock must be *firm-specific*; therefore, labor market-wide shocks are poor candidates, unless they happen to only affect one or a small number of firms in a local market, or unless one leans more heavily on the model structure for identification. Another important assumption is the lack of confounding firm-level changes (e.g., amenities) in response to the shock, although this is partly testable.

Kroft et al. (2020) use government procurement auctions as a firm-level demand shifter for construction projects; using both difference-in-differences and RDD, the wage and employment responses imply labor supply elasticity of around 4. They consider amenity response in the form of non-wage compensation, safety and rule out such compensating differentials. Amodio and de Roux (2023) use foreign exchange rate shocks along with an exporting firm's exposure to that country's currency based on pre-determined export patterns using data from Colombia. Although this is a market level shock, the presumption is that the firm-level shocks are only affecting a small share of relevant employers in the labor market, and the authors rely on within-sector and within-location variation for identification. They find a residual labor supply elasticity of around 2.5. Sharma (2023) uses firm-specific demand shock in the Brazilian textile industry from the end of the Multi-Fiber Arrangement to estimate wage and separation responses: key to identification is that the shock only affected a small number of firms. She finds substantially smaller labor supply elasticity for women (around 1.2) than men (around 2.7). This greater monopsony power seems to arise through female-favored amenities, including shorter commute and safety.

Some papers have explicitly modeled how firm wage and employment may respond to a market level demand shock and have used this structure to estimate implied mark downs. One example is Felix (2021), who studies the estimate of trade on concentration and wages using tariff shocks from Brazil's trade liberalization. She structures her analysis based on the framework in Berger et al. (2022a), using the nested CES model interpret labor and wage response at the firm level (within a market) as well as market level (along with estimates of concentration) to estimate that wages in the Brazilian setting were around 50% of marginal product. She finds that trade competition reduced wages by increasing labor market power (via increased concentration), offsetting the wage gains from the trade-induced reallocation of activity to productive firms.

Goolsbee and Syverson (2019)estimate the link between employment and wages using a measure of product demand (a lag of the number of undergraduate applications to an university) as an instrument for faculty employment. The authors' estimates for tenure track faculty suggest a moderate amount of monopsony power (residual labor supply elasticity of around 5), but their estimates suggest non-tenure track faculty face a highly competitive market.

There are a number of factors to consider when employing a demand shifter as an instrument for wages. Naturally, it is difficult to isolate an idiosyncratic demand shock in local labor market; different papers have done this more or less convincingly. But even leaving that aside, increased product demand is likely to not only raise a firm's wages but also amenities or other expenditures in recruitment as a firm attempts to expand production. In that case, attributing all the employment change resulting from a product demand shock to the associated change in wages resulting from the same shock would tend to overstate the residual labor supply elasticity. More generally, it strikes us a particularly important to test for the presence of responses along non-wage margins when using product demand shocks for identification—tests that some of the papers have implemented. Tests for cross-firm spillovers (e.g., as in Sharma, 2023) are also particularly useful in these settings.

5.4 Structural Estimates

A recent literature has adapted structural models from industrial organization to study labor market power. The first family of approaches leverage the widely used Berry-Levinsohn-Pakes (BLP) estimation of labor-supply functions facing firms. (2004) 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) estimate a nested labor supply model on Burning Glass data, recovering firm-specific labor supply elasticities of 4.8. They use 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 other instruments such 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.

A second family of approaches leverages production (and revenue) function estimation. Production function parameters can be used to directly calculate the marginal (revenue) product of labor, and then compare that to wages to obtain estimates of markdowns, which, under the assumption of monopsony, map into estimates of firm-specific labor supply elasticities. This approach explicitly models the firm technology, and estimates it using micro-data on firm inputs. One key concern is unobserved productivity, which would create biases in observed coefficients as it is correlated with both inputs and output. While some research uses quasi-experimental or experimental variation, by far the most frequent strategy is to proxy productivity using flexible functions of either past investment or a contemporaneously chosen variable input (1992). The intuition is that other inputs (or past investment) is monotonic in (expected) productivity, and the demand for that input can be inverted to obtain a proxy for productivity.

The relevance of this approach for monopsony is that it allows the marginal revenue product of labor to be calculated directly and compared to the wage, allowing markdowns to be recovered without having to directly estimate labor supply functions. Dobbelaere and Mairesse (2013) is the earliest paper we know of to implement this approach, albeit using panel variation alone to identify production function parameters. A recent paper in this literature is Yeh et al. (2022) who estimate wage markdowns in the U.S. manufacturing sector from 1976-2014 via production functions. They find that the gap between marginal product and wages has been decreasing from the late 1970s to the early 2000s, but increasing in recent years, albeit in a matter uncorrelated with concentration. The authors estimate a markdown of 1.53, implying that wages are only 65% of marginal product (consistent with a labor supply elasticity of 1.85).

Using the production function methodology with rich historical Belgian cost and revenue data at the mine-level between 1845 and 1910, Delabastita and Rubens (2022) estimate markdowns in the mining labor market. This paper has the virtue of being in a historical empirical setting prior to antitrust policy, and can look at the effect of the formation of a coal cartel in 1897, and find that this cartel, while having only a limited effect on prices (as coal is very tradable) had a large effect on wage markdowns, with wages going from 66% of marginal product in the early period (residual supply elasticity of 2) to less than 50% after the cartel formed (residual supply elasticity of 1).

Another approach developed by De Loecker and Eeckout estimates markups by comparing sales to estimated marginal costs, using cost-function estimates similarly obtained via a control function approach. This approach is applied to the German labor market by Mertens (2022)., while Brooks et al. (2021) apply this approach to Chinese and Indian manufacturing data. Like other papers in this literature, both estimate significant markdowns.

Criticisms of the structural identification of production functions abound, along with innovative responses (Ackerberg et al., 2015; Gandhi et al., 2020). But the crucial assumption is that firms are choosing inputs and outputs to maximize profits, alongside more context-specific assumptions as to the information available to firms and the variable vs fixed nature of inputs used to construct the proxy. But differences in assumptions can lead to very different estimates. For example, non-neutral factor augmenting technologies can bias estimates of labor's marginal product (Raval, 2022), not to mention unobserved variation in recruitment costs or visibility would alter the employment decisions of the firm in a manner that would bias production function estimates, even if productivity is controlled for. Similarly while wages are often used as stand-ins for total compensation, markdowns might be quite different when computed net of amenity costs and benefits. Finally, disentangling labor market power from product market power outside of tradable manufacturing sectors, where output quantities (let alone quality-adjusted quantities) are difficult to observe, requires further assumptions.

It also tempting to take estimates of markdowns and to use these to infer the labor supply elasticity facing the firm. However, as Dobbelaere and Mairesse (2013) and Mertens (2022) recognize, this is not necessarily the only interpretation of markdowns: they could result from a variety of labor market institutions, including firm-union bargaining, or even firm-worker bargaining, with potentially different efficiency implications. Just as it is difficult to infer markdowns from labor supply elasticities, it is difficult to infer labor supply elasticities from markdowns.

5.5 A Checklist for Estimation of Residual Labor Supply Functions

In this section, we summarize the discussion and provide some guidelines for what kind of variation will recover the degree of monopsony power facing the firm.

1. Who decided to change wages, and at what level do they vary? Are there agency issues inside the firm that may generate confounding effects?

2. How are you measuring changes in labor supply? Is the treatment population well-defined? Do the data distinguish between voluntary vs involuntary separations? Are there compositional effects?

3. Are there strategic interactions between firms? Are there ways of assessing if rivals or competitors respond to each other?

4. Are there effects on the morale or productivity of incumbent workers (treatment or control)?

As we emphasize in Section 2, there are many reasons firms may not set wages using the simple relationship described in equation 2. Before directly linking residual labor supply elasticities to markdowns, a researcher should also comment on (1) the relevance of different wage-setting protocols and (2) internal and external constraints on firm wage-setting.

6. Policy Implications

We have highlighted the importance of the residual labor supply elasticity, η, as a summary measure of labor market power. This parameter is empirically identifiable using modern research designs, encompasses the major sources of market power, varies in ways that can be tested across contexts, and, does not depend on any assumptions about firm optimization. Despite, or indeed because of this, η may be a better guide to policy than e.g. the passthrough rate. As we discussed above, the passthrough rate is sufficient to recover firm and worker surplus when firms set wages to maximize profits. But given the messiness of labor market conduct, a more robust assumption might be that we can measure firms' capacity to exercise labor market power, rather than their exercise of it.

6.1 Competition Policy

Labor market competition has been of recent direct interest to antitrust scholars and policymakers.(Naidu et al. (2018)Marinescu and Hovenkamp (2019)).The bread-and-butter of competition policy (and many expert consultants) is merger review, as mergers can lower the level of labor market competition. Building in part on the recent empirical literature, antitrust authorities have begun to incorporate labor market harms into their criteria for assessing anti-competitive behavior. For example, antitrust authorities have begun screening mergers based on harms due to labor market competition; the most recent Federal Trade Commission (FTC) proposal to revise the horizontal merger guidelines includes a new point "When a Merger Involves Competing Buyers, the Agencies Examine Whether It May Substantially Lessen Competition for Workers or Other Sellers". A small but growing empirical literature has documented negative effects of mergers on wages and employment (see, e.g., Arnold, 2019; Prager and Schmitt, 2021; Benmelech et al., 2022; Arnold et al., 2022; Lagaras, 2019).

However, antitrust authorities have a variety of tools—beyond merger review—to police a anti-competitive behavior, including collusion and cartelization (horizontal conduct), as well as practices that reduce competition (single firm conduct). While legal restrictions on employer competition are relatively rare, the legal system can facilitate monopsony by allowing voluntary contracts that restrict ex-post mobility. Decreases in transactions costs that have led to widespread outsourcing in the labor market (Weil 2006) have also led to new forms of vertical restraints along the value chain, where for example employers of workers via temp agencies or gig platforms are contractually prohibited from making direct offers to those employers. Most prominent among these are franchise agreements, including those studied by Krueger and Ashenfelter (2022).

A form of single firm conduct in the labor market that has garnered a considerable amount of recent attention is non-compete agreements (or covenants not to compete) (see, e.g., Prescott et al., 2016; Shi, 2018; Johnson et al., 2020). Non-compete clauses, contractual restrictions written between workers and firms that restrict the range of post-employment employers, have proliferated in the past few decades, leading the FTC to propose a ban on non-competes. Starr et al. (2021) show that non-competes are surprisingly widespread, even when non-enforceable.

Economists have long played important roles in both public and private regulation of product market competition. The recent expansion to labor markets is welcome. Following Naidu and Posner (2018), we caution against an over-reliance on antitrust doctrine alone as a panacea for pervasive monopsony in the labor market. Owing to long-term relationships with workers and widespread fairness norms, researchers should not assume that patterned and rigid wage setting is necessarily the result of employer collusion, or that worker welfare maps simply into estimates of η. Antitrust agencies are currently set up to police egregious anti-competitive practices by large or coordinated businesses, but there is considerable labor market power outside that environment. Wage regulation, unions, and macroeconomic conditions likely play at least as important a role as competition policy in restraining monopsony power.

6.2 Minimum Wages and Unions

Economists have long recognized that in the presence of labor market power, policies that mandate higher wages (such as a minimum wage policy) may be welfare improving (Stigler, 1946). A key empirical prediction of a model of monopsonistic competition is that employment effects could be positive for minimum wages up to a point. This would seem to suggest that a 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. While the empirical literature has produced a range of estimates, including more negative ones (e.g., Clemens et al., 2018), a sizable part of the literature has yielded relatively small employment effects (e.g., Cengiz et al. (2019), Derenoncourt and Montialoux (2021)). Following Card and Krueger (1995), researchers have often appealed to labor market power as an explanation for limited employment effect (see, e.g., Manning, 2021).

This interpretation may well be sensible, but it is useful to recognize that monopsony power is neither necessary or sufficient for a small disemployment effect. For example, Harasztosi and Lindner (2019) find a very modest disemployment effect in their study of a large minimum wage increase in Hungary, but explain it through low product demand elasticities and input substitution elasticities in a competitive model.

While the size of the employment effect of minimum wages is a rather coarse way to assess the extent of labor market power, there are other testable predictions that labor market power makes about minimum wage effects. When it comes to testing for labor market power, instead of looking at how overall low-wage employment responds to a minimum wage shock, it is more informative but assess how the employment response varies across firms. Under monopsony, a higher minimum wage can lead to reallocation of employment across firms, shifting activity from lower productivity to higher productivity firms. This was shown empirically in the context of the German minimum wage by Dustmann et al. (2022) find that underneath the small overall employment effect, there are sizable shifts away from lower-productivity to higher-productivity firms. In the context of collectively bargained minimum wages in South Africa, Bassier (2023) shows a similar reallocation towards higher productivity firms; Demir (2022) finds similar results in the German context for sectoral minimum wages. These findings of re-allocations from various types of wage mandates yield much more informative tests of the presence of monopsony than the pattern of overall employment response to minimum wage shocks. Similarly, models of dynamic monopsony and search frictions predict reductions in workers' separations as workers are re-allocated away from lowest-wage and highest turnover firms. A number of papers have corroborated this prediction empirically (Cardoso and Portugal, 2005; Brochu and Green, 2011; Dube et al., 2016).

Recently, a number of papers have tested whether (as the theory predicts) the employment effect of the minimum wage is heterogeneous by the extent of market power in the local labor market. Azar et al. (2019) find that employment effects of minimum wages are substantially more positive in more commuting zones where the low-wage labor market is more concentrated. Similarly, Wiltshire (2021) shows that when concentration in the low-wage labor market rose from the entry of Walmart super-centers, the effect of minimum wages was more likely to be positive. Overall, the focus on mechanisms and predictions other than the overall employment effect strikes us as a productive direction for the literature to better understand how monopsony power mediates the impact of minimum wage policy. Ultimately, the extent of market power and its exercise has implications for the setting of optimal minimum wages. As expected, this is a complicated undertaking since it involves both the efficiency and distributional considerations, and how monopsony power mediates both (for two very different examples of recent work in this arena, see Berger et al. (2022b); Vergara (2022)).[30]

Union wage setting makes predictions similar to the minimum wage, where the wage chosen is the outcome of the union's collective bargain with employer(s). The counterfactual for union wages is thus not the competitive wage, but rather the monopsonist one. Consistent with this, Dodini et al. (2021) finds that the union premium in Norway is monotonically increasing in local labor market concentration, and a number of papers have found that the effect of concentration is lower in unionized labor markets (Benmelech et al., 2020). But beyond wages, unions can also influence amenities as well as recruitment and retention practices, and the interaction of these details of collective bargaining agreements terms with monopsony is a promising di-

[30] A "robust monopsony regulation" approach is suggested in Guo and Shmaya (2019), who show a mixture of payroll subsidies (or EITC) and minimum wages is regret-minimizing, with the share of the former increasing in the welfare weight put on firm profits.

rection for future research (Lagos, 2019). Further, in the presence of monopsony profits, employers are willing to spend considerable resources in order resist unionization (Wang and Young, 2022).

6.3 Immigration

Much of the analysis of the effects of immigration on native workers in the labor market has been conducted under the perfectly competitive hypothesis. Even in the perfectly competitive model, immigrants are more than a shock to labor supply, and immigrants can affect the labor demand curve for native workers, with recent empirical evidence suggesting small net effects.

A much more recent literature has considered the implications of labor market monopsony for immigration and migration policy. The additional margin suggested by monopsony is that immigrants may alter the native residual elasticity facing firms, either increasing or decreasing monopsony power of firms vs-a-vs natives. Amior and Manning (2020) extend the traditional constant-returns model of immigration, where the long-run impact of immigration cannot lower the level of native wages once all factors adjust, to incorporate labor market monopsony. In their model, some migrants have a lower residual labor supply elasticity than natives, but employers can't discriminate between immigrants and natives within skill-groups. So an increase in immigration, which would have no effect on native wages in the neoclassical long-run, winds up reducing native wages because it amplifies the average monopsony power employers have over workers. The authors show that native wages do fall in labor market cells exposed to immigrant shocks where the immigrants are more likely to be immobile/inelastic.

Borjas and Edo (2023) empirically consider the effect of the 1981 migrant legalization on native wages in France, explicitly testing the hypothesis that regularization, by increasing mobility of immigrant workers (see also Amuedo-Dorantes and Bansak 2011), reduced the monopsony power of employers over both native born workers and immigrants. They find that the regularization raised wages and employment for both low-skill natives and immigrant workers, consistent with monopsonistic labor markets.

This strand of argument relies on employers being unable to discriminate between natives and immigrants, an assumption that likely only holds true to a partial degree and in some contexts. If institutions effectively segregate immigrant workers from natives, allowing different wages to be paid, then the effect of immigration on native wages may operate all through the effects on labor demand (which again can be positive or negative depending on the technology), making the analysis of the impact of immigration on native wages very similar to the neoclassical case, albeit with more surplus captured by employers than under monopsony.

At an extreme example of immigrant labor market segregation are guest worker programs, which are one of the few remaining areas of modern labor markets with legally protected monopsony power. Foreign workers are admitted as temporary participants in the local labor market, with restrictions on mobility that grant employers ex-post monopsony power. But whether this is true in practice, for example in the United States, remains an open question. Hunt and Xie (2019) and Wang (2021) find divergent effects of visa-restricted job mobility on wages, however, with the former finding no wage effects of permanent residency and the latter findings a 5.5% increase in wages.[31]

[31] Kim and Pei (2022) leverage the randomness in the H1-B lottery system to generate exogenous variation in the HHI across H1-B labor markets, and find that wages for new migrants are significantly lower in high-concentration labor markets. Depew et al. (2017) use H1-B and L1 worker data from 6 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 H1-B visa program is a source of monopsony power remains a open question deserving of more research, as is research on less skilled visa categories like H2-A (agriculture) and H2-B (non-agriculture but not skilled).

In the quite restrictive India-UAE context, Naidu et al. (2016) study restrictions on migrant workers in the UAE, finding that a mobility reform raised the relevant labor supply elasticity (which can only be non-zero at the end of a contract) from 1 to roughly 2.5. They also used this shock to incumbent worker mobility as a marginal productivity of labor shock to new workers, and use wage and employment changes for new recruits to recover a separate estimate of monopsony power for recruitment in source countries, with a recruitment elasticity of 1. Naidu et al. (2023) build on this using experimental estimates of marginal treatment effects for potential migrants, concluding that improved labor market standards for guest workers would increase migrant supply, realizing a much larger share of the potential migration surplus.

7. Conclusion and Directions for Future Research

Monopsony is but the beginning of a research program investigating how employers set wages, and how non-market and market forces interact in the labor market. While we mentioned several open areas for research throughout the review, we conclude by listing four additional areas we believe deserve further study.

First, one important question is whether monopsony is inefficient once employer investments are considered. For instance, patents both provide firms with monopoly power and incentivize investment in research. It is not clear whether there are analogous examples in which monopsony power incentivizes entrepreneurs to invest or firms to invest in general training. The profits to monopsony power may be only quasi-rents, eaten up by ex-ante investments, recruitment expenditures, and fixed costs.

Second, Adam Smith famously noted, “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" (1776). This remains largely true to this day. While there are some prominent examples of employer collusion, our evidence on whether (and when) this occurs is relatively sparse. There is little empirical or theoretical work on collusion in the labor market, or on how repeated interactions between employers may facilitate this collusion (for one such paper, see Gibson 2021). A recent literature in industrial organization has focused on the detection of collusion (e.g. in procurement auctions). Empirical tools for detecting collusion in labor markets could develop along similar lines.

Third, the evolution of monopsony power remains an under-explored area. While there are historical examples of egregious monopsony power (Naidu, 2010; Boal, 1995), the extent of monopsony prior to the early 21st century remains a mystery. This likely reflects the fact that matched worker-firm data and exogenous variation in firm-specific wages is difficult to obtain in historical contexts. But even for recent decades, little is known about how η has changed, both for the whole population as well as specific subgroups of workers (although see (Depew and Sørensen, 2013) for a firm-specific exception). Relatedly, understanding the evolution of possible countervailing constraints on monopsony power is critical for assessing the role of market power on the wage distribution.

Finally, while we have focused on wage-posting as the model of wage-setting associated with classical monopsony, more research is needed on the details of wage-setting protocols (Card, 2022; Caldwell et al., 2023a), and the extent to which firms wage discriminate or bargain individually with workers. There is likely heterogeneity in wage-setting practices, both across firms and across types of workers. New methods in algorithmic human resources, including AI-based predictions of individual workers labor market prospects (Vafa et al., 2023), may also change how firms set wages, and how competitive the labor market is. While we think imperfect competition will remain a robust paradigm for understanding the labor market , the specific ways in which it is exercised may change along with technology and management.

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.

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," Technical Report, National Bureau of Economic Research 2022.

_ 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.

Amior, Michael and Alan Manning, "Monopsony and the Wage Effects of Migration," 2020.

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

Amuedo-Dorantes, Catalina and Cynthia Bansak, "The impact of amnesty on labor market outcomes: A panel study using the legalized population survey," *Industrial Relations: A Journal of Economy and Society*, 2011, *50* (3), 443–471.

Arnold, David, "Mergers and Acquisitions, Local Labor Market Concentration, and Worker Outcomes," 2019.

_ , Kevin Milligan, Terry S Moon, and Amirhossein Tavakoli, "Firm Growth and Employee Earnings After Acquisitions: Linking Corporate and Worker Outcomes," 2022.

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.

Azar, José A, Steven T Berry, and Ioana Marinescu, "Estimating labor market power," Technical Report, National Bureau of Economic Research 2022.

Azar, José, Emiliano Huet-Vaughn, Ioana Marinescu, Bledi Taska, and Till Von Wachter, "Minimum Wage Employment Effects and Labor Market Concentration," Technical Report, National Bureau of Economic Research 2019.

_ , Ioana Marinescu, and Bledi Taska, "Concentration in US Labor Markets: Evidence From Online Vacancy Data," Technical Report, National Bureau of Economic Research Working Paper 24395, https://www.nber.org/papers/w24395.pdf 2018.

Banfi, Stefano and Benjamin Villena-Roldan, "Do high-wage jobs attract more applicants? Directed search evidence from the online labor market," *Journal of Labor Economics*, 2019, *37* (3), 715–746.

Barbanchon, Thomas Le, Roland Rathelot, and Alexandra Roulet, "Gender Differences in Job Search: Trading off Commute Against Wage," *The Quarterly Journal of Economics*, 2021, *136* (1), 381–426.

Bassier, Ihsaan, "Firms and inequality when unemployment is high," *Journal of Development Economics*, 2023, *161*, 103029.

_ , 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.

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.

Bender, Stefan, Nicholas Bloom, David Card, John Van Reenen, and Stefanie Wolter, "Management Practices, Workforce Selection, and Productivity," *Journal of Labor Economics*, 2018, *36* (S1), S371–S409.

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," Technical Report, Working Paper 2019.

_ , _ , and _ , "Labor Market Power," *American Economic Review*, 2022, *112* (4), 1147–93.

Berger, David W, Kyle F Herkenhoff, and Simon Mongey, "Labor market power," Technical Report, National Bureau of Economic Research 2019.

_ , _ , and _ , "Minimum wages, efficiency and welfare," Technical Report, National Bureau of Economic Research 2022.

_ , Thomas Hasenzagl, Kyle F Herkenhoff, Simon Mongey, and Eric A Posner, "Merger Guidelines for the Labor Market," Technical Report, National Bureau of Economic Research 2023.

Berry, Steven, James Levinsohn, and Ariel Pakes, "Differentiated products demand systems from a combination of micro and macro data: The new car market," *Journal of political Economy*, 2004, *112* (1), 68–105.

Bewley, Truman F, *Why Wages Don't Fall During a Recession*, Harvard University Press, 1999.

Bhuller, Manudeep, Karl Ove Moene, Magne Mogstad, and Ola L Vestad, "Facts and fantasies about wage setting and collective bargaining," *Journal of Economic Perspectives*, 2022, *36* (4), 29–52.

Bloesch, Justin and Birthe Larsen, "New vs. Classical Monopsony," 2023.

Bloom, Nicholas and John Van Reenen, "Measuring and explaining management practices across firms and countries," *The quarterly journal of Economics*, 2007, *122* (4), 1351–1408.

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.

Bonhomme, Stéphane and Grégory Jolivet, "The pervasive absence of compensating differentials," *Journal of Applied Econometrics*, 2009, *24* (5), 763–795.

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

Brochu, Pierre and David A Green, "The impact of minimum wages on quit, layoff and hiring rates," Technical Report, IFS Working Papers 2011.

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.

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

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

Caldwell, Sydnee and Emily Oehlsen, "Monopsony and the Gender Wage Gap: Experimental Evidence from the Gig Economy," Technical Report, Working Paper 2018.

_ and Nikolaj Harmon, "Outside Options, Bargaining, and Wages: Evidence from Coworker Networks," Technical Report, Working Paper 2018.

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

_ and Oren Danieli, "Outside Options in the Labor Market," Technical Report, Harvard University Working Paper 2021.

_ , Ingrid Haegele, and Joerg Heining, "Bargaining and Inequality in the Labor Market," *UC Berkeley Department of Economics Working Paper*, 2023.

_ , _ , and _ , "Wage Premia and Worker Search," *UC Berkeley Department of Economics Working Paper*, 2023.

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.

_ , _ , Jörg Heining, and Patrick Kline, "Firms and Labor Market Inequality: Evidence and Some Theory," 2016. NBER Working Paper No. 22850.

_ and Alan B Krueger, "Time-series minimum-wage studies: a meta-analysis," *The American Economic Review*, 1995, *85* (2), 238–243.

_ , 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.

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.

Datta, Nikhil, “Local Monopsony Power,” 2022.

Davis, Steven J, R Jason Faberman, and John C Haltiwanger, “The establishment-level behavior of vacancies and hiring,” *The Quarterly Journal of Economics*, 2013, *128* (2), 581–622.

Delabastita, Vincent and Michael Rubens, “Colluding Against Workers,” 2022.

Demir, Gökay, “Labor market frictions and spillover effects from publicly announced sectoral minimum wages,” *Ruhr Economic Papers*, 2022.

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.

— , Clemens Noelke, David Weil, and Bledi Taska, "Spillover Effects from Voluntary Employer Minimum Wages," Technical Report, National Bureau of Economic Research 2021.

Dey, Matthew S and Christopher J Flinn, "An equilibrium model of health insurance provision and wage determination," *Econometrica*, 2005, *73* (2), 571–627.

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," 2021.

Dube, Arindrajit, Alan Manning, and Suresh Naidu, "Monopsony and Employer Mis-Optimization Account for Round Number Bunching in the Wage Distribution," *Unpublished manuscript*, 2019.

— , Jeff Jacobs, Suresh Naidu, and Siddarth Suri, "Monopsony Power in Crowdsourcing Markets," 2018. Unpublished manuscript.

— , 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, 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," *Working Paper*, 2020.

Faberman, R Jason and Guido Menzio, "Evidence on the relationship between recruiting and the starting wage," *Labour Economics*, 2018, *50*, 67–79.

Felix, Mayara, "Trade, labor market concentration, and wages," *Job Market Paper*, 2021.

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 Silverio, "How Does Firm Performance Affect Wages? Evidence from Idiosyncratic Export Shocks," 2022. Review of Economic Studies.

Gaudry, Marc and Marcel Dagenais, "The use of Box-Cox transformations in regression models with heteroskedastic autoregressive residuals," *Cahier de recherche*, 1978, (7814).

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

Gibson, Matthew, "Employer market power in Silicon Valley," 2021.

Giupponi, Giulia and Stephen Machin, “Labour market inequality,” *IFS Deaton Review*, 2022, *15*.

Goolsbee, Austan and Chad Syverson, “Monopsony power in higher education: A tale of two tracks,” Technical Report, National Bureau of Economic Research 2019.

Gottfries, Axel and Gregor Jarosch, “Dynamic monopsony with large firms and an application to non-competes,” 2023.

Gronberg, Timothy J and W Robert Reed, “Estimating workers’ marginal willingness to pay for job attributes using duration data,” *Journal of Human Resources*, 1994, pp. 911–931.

Guo, Yingni and Eran Shmaya, “Robust Monopoly regulation,” *arXiv preprint arXiv:1910.04260*, 2019.

Harasztosi, Péter and Attila Lindner, “Who pays for the minimum wage?,” *American Economic Review*, 2019, *109* (8), 2693–2727.

Hardy, Morgan and Jamie McCasland, “Are small firms labor constrained? experimental evidence from ghana,” *American Economic Journal: Applied Economics*, 2023, *15* (2), 253–284.

Hazell, Jonathon and Bledi Taska, “Posted Wage Rigidity,” Technical Report 2018.

_, Christina Patterson, Heather Sarsons, and Bledi Taska, “National wage setting,” Technical Report, National Bureau of Economic Research 2022.

Hjort, Jonas, Xuan Li, and Heather Sarsons, “Across-Country Wage Compression in Multinationals,” Technical Report, National Bureau of Economic Research 2020.

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.

Jäger, Simon, Benjamin Schoefer, Samuel Young, and Josef Zweimüller, "Wages and the Value of Nonemployment," *The Quarterly Journal of Economics*, 2020, *135* (4), 1905–1963.

_ , Christopher Roth, Nina Roussille, and Benjamin Schoefer, "Worker Beliefs About Outside Options," Technical Report, National Bureau of Economic Research 2022.

Jarosch, Gregor, Jan Sebastian Nimczik, and Isaac Sorkin, "Granular Search, Market Structure, and Wages," Technical Report, National Bureau of Economic Research 2019.

Johnson, Matthew S, Kurt Lavetti, and Michael Lipsitz, "The labor market effects of legal restrictions on worker mobility," *Available at SSRN 3455381*, 2020.

Kaur, Supreet, "Nominal wage rigidity in village labor markets," *American Economic Review*, 2019, *109* (10), 3585–3616.

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.

Kroft, Kory, Fabian Lange, Matthew J Notowidigdo, and Lawrence F Katz, "Long-term unemployment and the Great Recession: the role of composition, duration dependence, and nonparticipation," *Journal of Labor Economics*, 2016, *34* (S1), S7–S54.

__ , Yao Luo, Magne Mogstad, and Bradley Setzler, "Imperfect competition and rents in labor and product markets: The case of the construction industry," Technical Report, National Bureau of Economic Research 2020.

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," *Available at SSRN 3560620*, 2019.

Lagos, Lorenzo, "Labor Market Institutions and the Composition of Firm Compensation: Evidence from Brazilian Collective Bargaining," *Job Market Paper*, 2019.

Lamadon, Thibaut, "Productivity Shocks, Dynamic Contracts, and Income Uncertainty," *manuscript, University of Chicago*, 2014.

__ , Magne Mogstad, and Bradley Setzler, "Imperfect competition, compensating differentials and rent sharing in the US labor market," Technical Report, National Bureau of Economic Research 2019.

Lang, Kevin and Sumon Majumdar, "The pricing of job characteristics when markets do not clear: theory and policy implications," *International Economic Review*, 2004, *45* (4), 1111–1128.

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

Lavetti, Kurt and Ian M Schmutte, "Estimating compensating wage differentials with endogenous job mobility," 2016.

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," Technical Report, National Bureau of Economic Research 2018.

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

_ , "A generalised model of monopsony," *The Economic Journal*, 2006, *116* (508), 84–100.

_ , "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," *Ind. LJ*, 2019, *94*, 1031.

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," *Superstar Firms, and Between-Firm Wage (In) Equality*, 2022.

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

Mortensen, Dale T and Tara Vishwanath, "Personal Contacts and Earnings: It Is Who You Know!," *Labour Economics*, 1994, *1* (2), 187–201.

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," Technical Report, Working paper, Columbia University 2018.

_ , _ , 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.

_ , _ , and _ , "The Benefits and Costs of Guest Worker Programs: Experimental Evidence from the India-UAE Migration Corridor," Technical Report, National Bureau of Economic Research 2023.

Nimczik, Jan Sebastian, "Job Mobility Networks and Endogenous Labor Markets," 2017.

Olley, Steven and Ariel Pakes, "The dynamics of productivity in the telecommunications equipment industry," 1992.

Parsons, Donald O, "Specific human capital: An application to quit rates and layoff rates," *Journal of political economy*, 1972, *80* (6), 1120–1143.

Postel-Vinay, Fabien and Jean-Marc Robin, "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity," *Econometrica*, 2002, *70* (6), 2295–2350.

Prager, Elena 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.

Raval, Devesh, “Testing the production approach to markup estimation,” *Review of Economic Studies*, 2022.

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

Rosen, Sherwin, “Hedonic prices and implicit markets: product differentiation in pure competition,” *Journal of political economy*, 1974, *82* (1), 34–55.

Roussille, Nina and Benjamin Scuderi, “Bidding for Talent,” Technical Report, Working paper 2021.

Roy, Andrew Donald, “Some thoughts on the distribution of earnings,” *Oxford economic papers*, 1951, *3* (2), 135–146.

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

Shi, Liyan, “Restrictions on Executive Mobility and Reallocation: The Aggregate Effect of Non-Competition Contractsâ€,” Technical Report, Technical Report 2018.

shin Hwang, Hae, Dale T Mortensen, and W Robert Reed, “Hedonic wages and labor market search,” *Journal of Labor Economics*, 1998, *16* (4), 815–847.

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

Sokolova, Anna and Todd Sorensen, “Monopsony in Labor Markets: A Meta-Analysis,” 2018.

_ and _ , “Monopsony in labor markets: A meta-analysis,” *ILR Review*, 2021, *74* (1), 27–55.

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).

Staiger, Douglas O, Joanne Spetz, and Ciaran S Phibbs, “Is there monopsony in the labor market? Evidence from a natural experiment,” *Journal of Labor Economics*, 2010, *28* (2), 211–236.

Starr, Evan, James J Prescott, and Norman Bishara, “Noncompetes in the US Labor Force,” *Journal of Law and Economics*, 2021.

Stigler, George J, “The economics of minimum wage legislation,” *The American Economic Review*, 1946, *36* (3), 358–365.

Sullivan, Paul and Ted To, “Search and nonwage job characteristics,” *Journal of Human Resources*, 2014, *49* (2), 472–507.

Taber, Christopher and Rune Vejlin, “Estimation of a roy/search/compensating differential model of the labor market,” *Econometrica*, 2020, *88* (3), 1031–1069.

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

Tirole, Jean, *The Theory of Industrial Organization*, MIT press, 1988.

Train, Kenneth E, *Discrete choice methods with simulation*, Cambridge university press, 2009.

Vafa, Keyon, Susan Athey, and David M Blei, "Decomposing Changes in the Gender Wage Gap over Worker Careers," 2023.

Vergara, Damián, "Minimum Wages and Optimal Redistribution," *arXiv preprint arXiv:2202.00839*, 2022.

Wang, Sean and Samuel Young, "Unionization, Employer Opposition, and Establishment Closure," 2022.

Wang, Xuening, "US permanent residency, job mobility, and earnings," *Journal of Labor Economics*, 2021, *39* (3), 639–671.

Weyl, E Glen and Michal Fabinger, "Pass-through as an economic tool: Principles of incidence under imperfect competition," *Journal of Political Economy*, 2013, *121* (3), 528–583.

Wiltshire, Justin C, "Walmart supercenters and monopsony power: How a large, low-wage employer impacts local labor markets," *Job market paper*, 2021.

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.

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 Individual-Specific Elasticities: Static Case

The elasticity of labor supply to the firm is $\frac{\partial P_j}{\partial w_j} \times \frac{w_j}{P_j}$. The probability of choosing firm j is defined in equation 4.

$$\frac{\partial P_j}{\partial w_j} = \frac{\partial}{\partial w_j} \sum_{S \in \mathcal{S}(j)} \mathrm{P}[j = \left[\arg\max_{k \in S} V^i(w_k)|S\right] \mathrm{P}(S)$$

Note that the probability of seeing a given consideration set does not depend on the wage. Therefore we can take the derivative inside the summation and simply calculate—for each consideration set—the derivative of the choice probability with respect to the wage. This has a familiar solution. Recall that $\Pr[j = [\arg\max_{k \in S} V^i(w_k)|S] = \exp(u(w_j))/\sum_{k \in S} \exp(u(w_k))$. The derivative of this with respect to w_j is $u'(w_j) \times P_{j|S} \times (1 - P_{j|S})$.

Returning to the original expression, we can then write

$$\begin{aligned} \frac{\partial P_j}{\partial w_j} &= \sum_{S \in \mathcal{S}(j)} \left\{ \frac{\partial}{\partial w_j} \mathrm{P}[j = \left[\arg\max_{k \in S} V^i(w_k)|S\right] \right\} \mathrm{P}(S) \\ &= \sum_{S \in \mathcal{S}(j)} u'(w_j) P_{j|S}(1 - P_{j|S}) P(S) \end{aligned}$$

The elasticity is found by multiplying by w_j and dividing by the original probability

$$\eta_{ij} = \frac{w_j \times \sum_{S \in \mathcal{S}(j)} u'(w_j) P_{j|S}(1 - P_{j|S}) P(S)}{\sum_{S \in \mathcal{S}(j)} P_{j|S} \times P(S)}.$$

It is somewhat easier to interpret this expression if we define $\alpha_{S(j)} := P(S|j)$ to be the probability that the choice set was S given that the worker chose j.

By Bayes rule this is

$$\alpha_{S,j} := P(S|j) = \frac{P_{j|S} \times P(S)}{\sum_{S \in 2^J} P_{j|S} \times P(S)} = P(j,S)/P_j.$$

Using this terminology, we can then write

$$\begin{aligned} \eta_{ij} &= \sum_{S \in \mathcal{S}(j)} \left(w_j \times u'(w_j)(1 - P_{j|S})\right) \times \frac{P(S)P_{j|S}}{\sum_{S \in \mathcal{S}(j)} P_{j|S} \times P(S)}) \\ &= \sum_{S \in \mathcal{S}(j)} \eta_{ij}(S) \times \alpha_{S,j} \end{aligned}$$

where $\eta_{ij}(S)$ is the elasticity of worker i to firm j conditional on having consideration set S.

A.2 Individual-Specific Elasticities: Dynamic Case

We can derive recruitment and separation elasticities in the context of our model if we incorporate search dynamics. We assume worker i's value of a job at firm j is given by:

$$V_t^j(w_j) = U(w_j) + \epsilon_i^j + \beta \left(\sum_{S \in \mathcal{S}(j)} E[\max_{l \in S} U(w_l)|S] P(S|j) \right) \tag{12}$$

where $\mathcal{S}$ is the set of possible consideration sets for a worker at firm j.[32]The probability a worker chooses firm k in the next period depends on (1) the probability firm k is in her consideration set and (2) the probability that, conditional on firm k being in her consideration set, she chooses firm k. We use $Z(k|CS)$ to denote the probability worker i chooses firm k, conditional

[32]A special case (and inspiration) for this model is Langella and Manning (2021). We obtain this case by setting S equal to the current firm (i.e. the worker has no choice of firms) with probability $1 - \lambda$ and the set of all firms with probability λ. Note that we simplify the problem by assuming that—apart from idiosyncratic preferences—workers are homogeneous.

on observing consideration set S. Given the functional form we assumed for $U(w)$ this can be written as:

$$Z(k|S,\mathbf{w}) = \exp(u(w_k)/\sum_{l\in S}\exp(u(w_l)).$$

The probability a worker chooses firm k given the vector of wages **w** is

$$P(k|\mathbf{w},j(i)) \quad = \quad \sum_{S\in\mathscr{P}}\left[Z(k|S,\mathbf{w})\times P(S|j(i))\right].$$

One important restriction we impose is that the probability firm k is in worker i's consideration set is—conditional on the identity of her current firm—independent of the wage offered by firm k. This means that while a firm that changes its wage will see a change in the number of recruits and quits, this operates only through the conditional choice probabilities. This does not require all firms to appear in a worker's consideration set with the same probability, however. Below, we consider the possibility that firms vary in "visibility" v_j, which influences the probability they appear in a worker's consideration set. Visibility could vary across firms for a variety of reasons, including differences in recruiting practices or brand recognition, which we discuss in Section 4.2.

We obtain the elasticity of worker i to firm j first by considering how a change in w_j affects choice probabilities:

$$\frac{\partial P(k|w,j(i))}{\partial w_k} \quad = \quad \sum_{S\in\mathscr{P}(i)} P(S|j(i))\times\frac{\partial Z(k|S,\mathbf{w})}{\partial w_k}.$$

As in standard discrete choice models, conditional on the consideration set, the derivative of the choice probabilities is simply $Z(k|S)\times(1-Z(k|S))\times$

$u'(w_k)$ (Train, 2009). We then obtain the overall elasticity to firm j:

$$\eta^{ij} = \frac{\sum_{S\in\mathscr{P}(i)} P(S|j(i)) \times Z(k|S,\mathbf{w}) \times (1 - Z(k|S,\mathbf{w}))}{\sum_{S\in\mathscr{P}(i)} P(S|j(i)) \times Z(k|S(i),\mathbf{w})} \times (u'(w_k) \times w_k)\,.$$

The denominator of this expression is simply the expected steady state market share among ex ante homogeneous workers. We could calculate the numerator by simulating histories of such individuals, and by taking the expected value of $Z(k|S) \times (1 - Z(k|S))$ after drawing new consideration sets from these individuals from $\mathscr{P}(i)$.

A.3 Recruitment-Weighted Elasticities in the Dynamic Model

Since $\sum_k P(k|j,\boldsymbol{w}) = 1$, we must have $\frac{d\sum_k Pr(k|j,\boldsymbol{w})}{dw_j} = 0$. This implies that $\frac{-dP(j|j,\boldsymbol{w})}{dw_j} = \frac{d\sum_{k\neq j} P(k|j,\boldsymbol{w})}{dw_j}$. We can then write:

$$\begin{aligned}
\frac{-dP(j|j,\boldsymbol{w})}{dw_j} &= \frac{d\sum_{k\neq j} P(k|j,\boldsymbol{w})}{dw_j} \\
&= \frac{-d\sum_{k\neq j} P(j|k,\boldsymbol{w})\frac{P(k)}{P(j)}}{dw_j} \\
&= -\sum_{k\neq j} \frac{dP(j|k,\boldsymbol{w})}{dw_j}\frac{P(k)}{P(j)}
\end{aligned}$$

where the second line follows as a result of Bayes rule ($P(k|j,w) \times P(j) = P(j|k,w) \times P(k)$). The third line follows from the second because $\sum_{k\neq j} d\frac{\frac{P(k)}{P(j)}}{dw_j} = 0$ in steady state.

Using this relationship in the calculation below we get that recruit-weighted quit elasticity is equal to the recruit-weighted recruit elasticity. We start by writing out the definition of the recruit-weighted quit elasticity:

$$\sum_j R_j \epsilon_j^Q = \sum_j R_j \frac{d(1-P(j|j,\boldsymbol{w}))}{dw_j} \frac{w_j}{1-P(j|j)}$$
$$= \sum_j \left(\frac{R_j}{1-P(j|j)} \left(\frac{-dP(j|j,\boldsymbol{w})}{dw_j} w_j \right) \right)$$

We then note that the number of recruits to firm j is equal to the number of people who choose firm j, less the number of people who pick firm j who were already there:

$$\sum_j R_j \epsilon_j^Q = \sum_j \left(\frac{P(j) - P(j|j,\boldsymbol{w})P(j)}{1-P(j|j,\boldsymbol{w})} \left(\frac{-dP(j|j,\boldsymbol{w})}{dw_j} w_j \right) \right).$$

Some simple algebra follows:

$$\sum_j R_j \epsilon_j^Q = \sum_j \left(\frac{P(j)(1-P(j|j,\boldsymbol{w}))}{1-P(j|j,\boldsymbol{w})} \left(\frac{-dP(j|j,\boldsymbol{w})}{dw_j} w_j \right) \right)$$
$$= \sum_j P(j,\boldsymbol{w}) \frac{-dP(j|j,\boldsymbol{w})}{dw_j} w_j$$
$$= \sum_j P(j,\boldsymbol{w}) \frac{-d\sum_{k\neq j} P(j|k,\boldsymbol{w}) \frac{P(k,\boldsymbol{w})}{P(j,\boldsymbol{w})}}{dw_j} w_j$$

where we arrive at the third line by noting that $P(j|j,w) = \sum_{k\neq j} P(j|k,w) \times \frac{P(k,w)}{P(j,w)}$. The latter is simply the number of recruits to firm j from other firms.

We can therefore rewrite this as

$$\sum_j R_j \epsilon_j^Q = \sum_j \frac{-d\sum_{k\neq j} P(j|k,\boldsymbol{w})P(k,\boldsymbol{w})}{dw_j} w_j$$
$$= -\sum_j \left(\frac{dR_j}{dw_j}\right)\left(\frac{R_j}{R_j}\right) w_j$$
$$= -\sum_j R_j \epsilon_j^R.$$

If the recruit and quit elasticities are each constant, this expression implies that they are equal. This expression allows us to multiply the quit elasticity by -2 to estimate the overall labor supply elasticity.

B. Microfounding Vacancy HHI Wage Regressions

In this Appendix, we use the assumption that variation in visibility is exogenous to micro-found a wage HHI in vacancies. The employment HHI is an equilibrium object, but the vacancy HHI is more likely to be driven by variation in firm recruiting costs that are independent of productivity and wages. Suppose employers have, exogenously given, vacancies V_j that summarize how visible they are to workers. Let $s_j^v = v_j/\sum_k v_k$ be the share of vacancies, and assume that workers' consideration sets are such that there is a fifty-fifty chance they: (1) choose firm j by default according to s_j^v or (2) choose between all firms according to their CES utility function. [33] The resulting labor supply to firm j is of the form:

$$Pr(\text{choose j—}V_j, V_{-j}, w_j, w_{-j}) = \frac{s_j^V}{2} + \frac{1}{2}\frac{w_j^\epsilon}{\sum w_k^\epsilon} \tag{13}$$

[33] This is a variant of the "Dogit" model from transportation economics, and is analogous to the frequently used notion of locals and tourists in product pricing models (Gaudry and Dagenais, 1978). Locals are price sensitive, while tourists just take the first thing they see.

and the resulting elasticity is decreasing in the share of vacancies:

$$\eta_j(s_j^V) = \epsilon \frac{w_j^\epsilon}{s_j^V \sum w_i^\epsilon + w_j^\epsilon}$$

This predicts that, holding productivity fixed, firms with more vacancies will have lower wages, and be smaller in terms of employment. With heterogeneous firms, a Taylor approximation yields a relationship between share weighted average wages and the Herfindahl of vacancies within a market. The relationship $\frac{w_j}{p_j} = \frac{\eta_j(s_j)}{1+\eta_j(s_j)}$can be log-linearized in s_j and then aggregated to give a linear equation relating average wages to the Herfindahl of vacancies, $H^v = \sum(s_j^V)^2$ and average productivity.

$$E[log(w_j)] = \sum s_j^V log(w_j) \approx -C \times \sum(s_j^V)^2 + \sum s_j^V log(p_j) = -CH^v + E[log(p_j)].$$

We assumed that vacancies (as a proxy for a broad notion of job visibility) are exogenous. However, a key concern remains adequately controlling for $E[p]$: instruments that change H^v are also likely to change other moments of the distribution of shares, and thus the overall aggregate labor demand shock.

Now it could be that vacancies are endogenous, but so long as there is exogenous cross-market firm-level variation in the costs of posting a vacancy, the Hausman instruments used in Azar et al. will be valid.