Chapter 8

Labor market monopsony: fundamentals and frontiers☆

Patrick Kline
UC Berkeley, Berkeley, CA, United States
Corresponding author. e-mail address: pkline@berkeley.edu

Contents

☆ This is the second part of a larger chapter on the topic of "wage setting power" that was initially prepared for the Handbook of Labor Economics conference in Berlin, which was generously funded by the Rockwool Foundation Berlin (RFBerlin). Based on discussions with the editors it was determined to be better for pedagogical reasons to publish the two parts as separate chapters. David Card, Sydnee Caldwell, Daniel Haanwinckel, Attila Lindner, Alan Manning, Damián Vergara, Michael Amiour, Justin Bloesch, Nina Roussille, and Ben Scuderi provided helpful comments on an earlier draft of this chapter that substantially improved the paper. Jordan Cammarota and Jinglin Yang provided outstanding research assistance on this project.

1 Introduction

A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it.

Max Planck (1949)

Understanding the forces governing the determination of wages is a central task of labor economics. For nearly a century, the dominant approach to modeling wage determination has been to approximate labor markets as competitive, with employers treating wages as an external constraint rather than a choice to be optimized. This perspective permeates previous editions of the *Handbook of Labor Economics*, chapters of which, for example, rationalize the explosion of wage inequality in advanced economies as manifestations of complex shifts in underlying supply and demand factors (Katz et al., 1999; Acemoglu and Autor, 2011), interpret the effects of immigration on employment and earnings in terms of market clearing wage adjustments (Borjas, 1999), and emphasize the social costs of distorting putatively competitive wages via legislative mandate (Brown, 1999).

Views in the profession have been changing rapidly. Fueled by the dissemination of large administrative datasets, a growing empirical literature finds that firm heterogeneity plays an important role in wage determination (Kline, 2024). A parallel literature demonstrates that firms respond to idiosyncratic productivity shocks by adjusting wages (Card et al., 2018; Kline et al., 2019; Lamadon et al., 2022; Garin and Silvério, 2023), suggesting that employers have considerable latitude to choose wages that depart from the choices of their peers. Meanwhile, there has been a rekindling of interest in the legal protections against employer wage-setting power enjoyed by workers (Naidu et al., 2018; Posner, 2021). This perfect storm has led to a revival of interest in Robinson (1933)'s theory of monopsony, along with kindred models of search and employer differentiation, as a lens for studying core topics in labor economics (Autor et al., 2023; Borjas and Edo, 2023; Deb et al., 2024). Reflecting on these developments, Card (2022), in his Presidential address to the American Economic Association, argues that "many—or even most—firms have some wage-setting

power." The quest to quantify and formalize the origins of this wage-setting power is now one of the most active frontiers in labor economics.

This chapter reviews the theory of monopsonistic wage setting, its connection to the recent empirical literatures studying the passthrough of economic shocks to wages and employment, and some important challenges the paradigm faces in establishing itself as a coherent framework for analyzing wage inequality. Several high quality reviews of the monopsony literature already exist (Manning, 2011, 2021; Caldwell et al., 2023) and a companion chapter in this Handbook considers oligopsonistic models featuring strategic interactions between firms (Azar and Marinescu, 2024). In contrast to these surveys, the treatment here is organized around empirical and theoretical limitations of the monopsony literature and potential approaches to overcoming those limitations. As such, our focus will be on empirical puzzles and the potential for new tools and insights to resolve those puzzles.

We begin by laying out a modern interpretation of the monopsony framework, where wage-setting power derives from imperfect information about worker outside options. While much attention has been given to macroeconomic models where monopsonistic wage motives form one block of a larger general equilibrium system (Berger et al., 2022; Haanwinckel, 2023; Deb et al., 2024), the focus here will be on the microeconomic tradeoffs faced by a single firm. This "firm's eye" perspective (Mrázová and Neary, 2017) frees us to study the wage setting problem non-parametrically, focusing on the essential microeconomic restrictions of the theory. The core of the model is a mapping between features of the outside option distribution and wages. We study non-parametric shape restrictions on the outside option distribution that ensure this mapping is unique and use these results to develop comparative statics linking shifts in productivity and the outside option distribution to wages, firm size, and profits. We then introduce a menu of tractable parametric specifications of the outside option distribution and establish conditions under which mixtures of these distributions guarantee a unique monopsony wage.

Non-sequential search models in the tradition of Butters (1977) link the distribution of outside options to the cross-sectional wage distribution. To illustrate the restrictions that search equilibrium can place on outside options, we consider models where the outside option distribution and the cross-sectional wage distribution are mutually determined by search frictions and the distribution of firm productivity. In a first model, both the outside option distribution and the distribution of wages are shown to take a power function form, implying that firms face an isoelastic equilibrium labor supply function. Next, we allow firms to be horizontal differentiated by adding idiosyncratic taste heterogeneity to the model as in Card et al. (2018). This yields a more complex labor supply function that is shown to be well approximated by a "logit-like" specification that only depends on two moments of the cross-sectional wage distribution. The resulting outside option distribution departs more substantially from both the equilibrium wage distribution and the distribution of firm productivity.

Textbook treatments emphasize that monopsonistic wage setting leads firms to become too small and produce inefficiently low output. This conclusion hinges on assumptions regarding the microeconomic forces driving dispersion of worker outside options. If, as in Card et al. (2018), the labor supply curve to the firm reflects worker taste heterogeneity, then match formation will tend to be inefficient because Pareto improving trades are stymied by information problems. However, wage markdowns can be fully efficient if driven entirely by search frictions, a point recently emphasized by Menzio (2024) in the context of consumer search. These divergent conclusions stem from differing assumptions about the productivity of workers in their outside options. Efficiency in a prototypical model with search frictions and taste heterogeneity is studied, illustrating how the cross-sectional relationship between wages and productivity can be used to assess misallocation.

A growing empirical literature studies the effects of idiosyncratic changes to either the distribution of worker outside options (Jäger et al., 2020) or employer productivity (Card et al., 2018; Kline et al., 2019; Lamadon et al., 2022; Garin and Silvério, 2023) on wages and firm size. In both cases, the predictions of the monopsony model are found to hinge critically on shape of the labor supply curve to the firm. In particular, wage passthrough is shown to depend on the "super-elasticity" of labor supply to the firm (i.e., the wage elasticity of the labor supply elasticity), while compensating differentials depend on the local curvature of the supply curve. Heavily utilized isoelastic specifications impose a super-elasticity of zero and restrict the curvature in ways that can dramatically impact qualitative predictions regarding the wage response to labor supply shifts and productivity passthrough. In addition to generating misleading conclusions about wage markdowns, incorrect curvature restrictions yield a distorted view of the likely incidence of mandated employer benefits (Summers, 1989; Gruber, 1994; Finkelstein et al., 2023).

Updating an argument of Bulow and Pfleiderer (1983) regarding cost-price passthrough, productivity-wage passthrough is shown to be insufficient to identify wage markdowns in the absence of additional restrictions. Fortunately, labor supply elasticities and markdowns can typically be recovered (or at least bounded) from joint impacts on firm size and wages. For example, when the labor supply elasticity is constant, instrumenting log wages in a linear model determining log employment will identify the labor supply elasticity. With non-constant elasticities, linear instrumental variables methods will tend to be biased, yielding (at best) a weighted average elasticity (Angrist et al., 2000). When instruments vary continuously, non-parametric instrumental variables methods (Newey and Powell, 2003; Blundell et al., 2007; Santos, 2012; Newey, 2013; Chen et al., 2024) can be applied to estimate or bound elasticities and markdowns at each wage level. Idiosyncratic shocks to firm productivity, firm amenities, or firm-specific outside options are all potentially valid instruments for wages. Instruments that shift the productivity of groups of rival firms are generally invalid without further restrictions because they exert a direct effect on

employment. We review diagnostics for assessing whether instruments contain such group components.

In the basic monopsony model, employers offer all workers the same wage. Models of third-degree wage discrimination are introduced as a compromise between the full-information benchmark—where employers can perfectly observe (and tailor wages to) worker outside options—and the classic wage posting benchmark, where employers are completely unable to discriminate between workers with different outside options. Allowing different wages for different observable types provides an opportunity to study worker-firm sorting. Three explanations for the tendency of high-wage workers to work at high-wage firms are reviewed: that skilled workers have greater labor supply elasticities, that they differentially prefer the amenities of the most productive firms, and that they exhibit supermodular complementarity with the most productive firms. Implications of these stories for the wage structure are discussed and connected to empirical evidence. We then consider a fourth explanation—that wages serve as a screen for worker quality—a hypothesis supported by empirical studies of the relationship between posted wages and the quality of job applicants (Dal Bó et al., 2013; Marinescu and Wolthoff, 2020; Escudero et al., 2024).

A longstanding critique of the monopsony model is that it presumes firms are able to commit to posting wages, which seems at odds with the observation that bargaining behavior is prevalent in some settings. For instance, Van Reenen (2024) remarks that "even at the macro level, it is unclear that wage posting is a better approximation than bargaining in many countries." Models of bargaining with incomplete information offer a potentially fruitful means of resolving this tension between the two modeling paradigms. A simple class of models where firms commit to a maximum wage and then engage in ex-post bargaining is introduced and shown to exhibit first-order conditions isomorphic to those in the monopsony model, while also providing a rationalization of wage dispersion within the firm for equivalent workers. In this model, productivity shifts not only affect average wages but also wage dispersion within the firm. This model is itself shown to be a limiting case of the more general "double auction" framework of Chatterjee and Samuelson (1983), which features hiring inefficiencies that stem from the presence of private information on both sides of the labor market. Recent empirical work corroborates the importance of private information for wage setting, finding that both within-firm wage inequality and average wage levels respond to changes in pay transparency (Mas, 2017; Baker et al., 2023; Cullen and Pakzad-Hurson, 2023). The double auction framework suggests that some productivity shifters may shift the wage demands of workers, which potentially invalidates their use as instruments. It also offers a potential explanation for differences in wage passthrough between groups of workers that have equivalent labor supply elasticities.

Finally, we discuss some puzzles that arise in monopsonistic interpretations of two types of passthrough. The first puzzle is that monopsonistic interpretations of productivity-wage passthrough often yield sizable markdowns that

imply firms are implausibly profitable, a problem that has also been noted by Bloesch et al. (2024). Allowing non-constant elasticities can help to reconcile this puzzle, as can accounting for firm adjustment costs. Both of these extensions have implications for the proper measurement of wage markdowns. The introduction of recruiting costs is shown to present additional difficulties for estimation of labor supply curves and some directions for future work in this area are suggested. Another sort of puzzle concerns the strong passthrough of minimum wages to product prices, which has long been cited as a challenge to monopsonistic interpretations of minimum wage results (Welch, 1995; Brown, 1995, 1999). Reviewing the mechanics of minimum wage passthrough in the monopsony model, we show that firm heterogeneity can generate positive market average passthrough with mild employment responses even when employment and prices are negatively related at each individual firm. Cross-sectional heterogeneity is a less plausible explanation for case studies of narrowly defined sectors (e.g., fast food) where disemployment effects have been negligible but passthrough has been shown to be strong. One explanation for strong passthrough in such settings is that wage increases lead to improvements in service quality that customers value. Another is that in inflationary environments, minimum wage hikes may trigger product price increases that would have occurred anyway, suggesting the puzzle is ephemeral.

2 The basic model

This section introduces the theory of monopsonistic wage-setting using a stylized model that will serve as a foundation for the rest of the chapter. Section 2.1 provides some historical background on the monopsony literature and the empirical motivation for such models. Section 2.2 introduces the model formally, describing the firm's optimization problem along with necessary and sufficient conditions for a unique wage to arise. Key concepts such as wage markdowns and exploitation are defined and the monopsony interpretation of labor shortages is reviewed. Section 2.3 considers some comparative statics that introduce the reader to topics considered in greater depth later in the chapter. Section 2.4 reviews some simple functional forms for the distribution of outside options and discusses conditions under which mixtures of these distributions yield a unique wage. We then study how some of these distributions can emerge from simple equilibrium search models. Section 2.5 provides an introduction to the welfare issues surrounding monopsonistic models. Working through a stylized model of match formation with search frictions, we explore conditions under which wage markdowns are compatible with efficient allocations of workers to firms.

2.1 Background

The term "monopsony" is evocative of settings with a single dominant employer. As Boal and Ransom (1997) recount, the monopsony moniker—which was suggested to Joan Robinson by classics scholar B.L. Hallward—seems to have

contributed to the theory's tepid reception for much of the 20th century, during which it was presumed that monopsonistic wage setting pertained primarily to highly specialized settings such as company towns. However, the central idea of monopsony is simply that firms must raise their wages to grow large. This tradeoff between firm size and average labor costs is arguably relevant for all firms, regardless of the number of local competitors they face.

Indeed, one of the best documented facts in empirical labor economics is the firm size wage premium (Moore, 1911; Brown and Medoff, 1989; Brown et al., 1990; Oi and Idson, 1999). Analyzing the earnings changes accompanying worker switches between employers in U.S. Social Security records from 2007 to 2013, Bloom et al. (2018) estimate that moving from a small firm with 10-50 employees to a medium sized firm with 1,000-2,500 employees raises wages by approximately 25%. Early work by Brown and Medoff (1989) and Brown et al. (1990) examines and refutes the idea that the firm size wage premium is driven by compensating differentials. If anything, larger firms seem to have better amenities. Corroborating this view, Holzer et al. (1991) demonstrate that larger firms tend to receive more applications per vacancy and exhibit lower quit rates. More recently, Caldwell et al. (2024b) provide survey evidence that workers believe firms differ in the wages they offer different workers and that jobs at higher wage firms are more desirable.

In this chapter, the tradeoff between firm size and labor costs will be modeled as stemming from the intersection of two fundamental forces: worker heterogeneity and imperfect information. Workers inevitably differ in their outside options and the ability of employers to observe those options is typically limited. The equilibrium search model of Burdett and Mortensen (1998) elegantly captures both of these forces: outside option heterogeneity results from differences in employment status and worker positions on the job ladder, while information constraints feature in the assumption that employers commit to posted wages before meeting workers. Of course, heterogeneity in outside options can also derive from sources besides search frictions, including commuting costs, preferences over workplace amenities, and different valuations of leisure. Another potential font of heterogeneity is worker (mis-)perceptions of outside wage opportunities (Jäger et al., 2024), which can give rise to differences in reservation wages.

Equilibrium search models can be (and often are) embellished with ex-ante heterogeneity in primitives that captures these other forces (Albrecht and Axell, 1984; Van den Berg and Ridder, 1998; Bontemps et al., 1999; Postel-Vinay and Robin, 2002a; Manning and Petrongolo, 2017). While equilibrium models provide a useful guide for thinking through market-wide counterfactuals, the focus here will be on firm-level comparative statics involving wages and employment. Consequently, outside option heterogeneity will, at least initially, be treated as an exogenous constraint that the firm must reckon with in wage setting. Mild non-parametric shape restrictions will be placed on the outside option distribution to ensure the firm's problem has an interior solution. We will then study a

simple class of non-sequential search models where the outside option distribution emerges in equilibrium and verify that these shape restrictions are satisfied.

The analysis here will be limited to environments where strategic considerations can be ignored. Establishing the empirical importance of strategic interactions in wage setting remains an important research frontier. Clear evidence of such interactions has been documented in a few highly concentrated labor markets (Staiger et al., 2010; Prager and Schmitt, 2021; Arnold, 2021) and in settings where employer associations facilitate collusion (Delabastita and Rubens, 2022; Sharma, 2024). However, recent studies evaluating less concentrated contemporary labor markets have found little evidence of strategic interactions in wage setting (Roussille and Scuderi, 2023; Derenoncourt and Weil, 2024). Berger et al. (2022) and Jarosch et al. (2024) describe oligopsonistic models generating effects of labor market concentration on wages, while Chan et al. (2024) characterize the equilibrium of a family of oligopsonistic models featuring multidimensional worker heterogeneity and study comparative statics involving changes to firm productivity and amenities. Azar and Marinescu (2024) provide a review of empirical and theoretical work on the effects of labor market concentration on wages and employment.

2.2 The firm's problem and optimal wages

We begin with some preliminary assumptions that will be maintained throughout this section. There is a unit continuum of workers capable of working at the firm. These workers differ in their outside options b, which are distributed on the interval $\left[\underline{b}, \bar{b}\right]$ according to the twice differentiable distribution function F, which we assume is strictly increasing. Each worker will join the firm if and only if offered a wage that exceeds their outside option, which may reflect the value of leisure, or outside job opportunities inclusive of their non-wage amenities. The firm cannot observe any worker's outside option but knows that these options are distributed according to F, which will be treated as exogenous. Hence, the firm correctly believes it will be able to employ $F(w)$ workers if the wage w is offered. Finally, we make the simplifying assumption that all workers, when employed, exhibit common marginal revenue product $p \in \left(\underline{b}, \bar{b}\right)$.

The firm's problem is to post a wage w that maximizes the profit function $\Pi(w) = F(w)(p - w)$. The first-order necessary condition for optimality is

$$f(w)(p - w) = F(w), \tag{1}$$

where f is the density of worker outside options. In words, the firm seeks to equate the profit made on the marginal worker with the cost of raising the wages of inframarginal workers.

Eq. (1) can be rearranged as

$$f(w)/F(w) = (p - w)^{-1}.$$

The right hand side of this equation is increasing in w. Therefore, because $p \in \left(\underline{b}, \bar{b}\right)$, a sufficient condition for a unique wage w^* to solve this equation is that the ratio $f(w)/F(w)$ be a weakly decreasing function of w. One way to ensure this condition holds is to assume that the distribution function F is log-concave, a property shared by many commonly used parametric distributions (Bagnoli and Bergstrom, 2006). Note that when F is log-concave, $\ln \Pi(w)$ is the sum of a concave function and a strictly concave function, guaranteeing that the firm's objective is strictly concave and has a unique maximum.

Log concavity is known to play an important role in ensuring existence of equilibria in many models of labor market search and imperfect product market competition (Caplin and Nalebuff, 1991; Bontemps et al., 1999). However, it turns out that a weaker shape restriction on F than log-concavity guarantees uniqueness of the monopsony wage. To understand why, consider the inverse wage function

$$\varrho(w) = w + F(w)/f(w),$$

which gives the productivity required for wage level w to solve (1). An optimal wage w^* is one for which $\varrho(w^*) = p$. So long as

$$\varrho'(w) = 2 - F(w) f'(w)/f(w)^2 = \frac{F(w)^3}{f(w)^2} \frac{d^2}{dw^2}\left(\frac{1}{F(w)}\right) > 0$$

for all $w \in \left(\underline{b}, \bar{b}\right)$, any optimal wage that exists must be unique because $\varrho(w)$ crosses p once from below. Since both $F(w)$ and $f(w)$ are positive over $\left(\underline{b}, \bar{b}\right)$ this requirement is satisfied whenever $d^2(1/F(w))/dw^2 > 0$ for all $w \in \left(\underline{b}, \bar{b}\right)$, i.e., whenever $1/F(w)$ is strictly convex. This property is known as *strict* -1-*concavity* because it amounts to the assumption that $-F(w)^{-1}$ is strictly concave. While any log-concave function is -1-concave, the converse is not true (Caplin and Nalebuff, 1991). The following lemma highlights another sense in which log-concavity is stronger than -1-concavity.

Lemma 1. *If* $F : \left[\underline{b}, \bar{b}\right] \to [0, 1]$ *is twice differentiable, strictly increasing, and log-concave then it is strictly* -1*-concave.*

Proof. A function F is log-concave when $\ln F$ is concave and strictly -1-concave when $-1/F$ is strictly concave. Hence, log-concavity implies $d^2 \ln F(w)/dw^2 \leq 0 \Rightarrow f'(w) F(w) \leq f(w)^2$ and strict -1-concavity implies $d^2(-1/F(w))/dw^2 < 0 \Rightarrow f'(w) F(w) < 2f(w)^2$. Since F is strictly increasing, the density $f(w)$ is always positive, implying $f(w)^2 < 2f(w)^2$. Thus, $d^2 \ln F(w)/dw^2 \leq 0 \Rightarrow d^2(-1/F(w))/dw^2 < 0$. □

A historically important example of a distribution that does not satisfy either log-concavity or strict -1-concavity comes from Burdett and Mortensen (1998).

Example (Burdett and Mortensen, 1998). The steady state of the Burdett and Mortensen (1998) model involves an outside option distribution taking the form $F(w) \propto \left(\frac{1}{1+\beta}\right)^2 \left(\frac{p-\underline{b}}{p-w}\right)$, where $\underline{b} > 0$ is the reservation wage and $\beta > 0$ gives the ratio of the on the job arrival rate of offers to the separation rate. The ratio $f(w)/F(w) = (p-w)^{-1} > 0$ is monotonically increasing, revealing that F is log-convex. Moreover, $d^2(1/F(w))/dw^2 = 0$, implying F is inverse-linear.

For this choice of F, uniqueness fails rather dramatically: *any* $w \in [0, p)$ solves (1), revealing that the firm is indifferent between all wage levels below p. This indifference is central to the equilibrium notion in Burdett and Mortensen (1998), which views wage dispersion as arising from a mixed strategy among identical firms. In contrast, part of the appeal of workhorse monopsonistic models is that they yield a deterministic mapping $(F, p) \mapsto w^*$ from primitives to wages that can be scrutinized empirically.

2.2.1 Exploitation, markdowns, and profits

When F is strictly -1-concave and $p \in (\underline{b}, \bar{b})$, a unique interior solution to (1) is assured. Evaluating this equation at the optimal wage w^* and rearranging yields the familiar monopsony wage expression

$$w^* = \frac{\phi(w^*)}{1+\phi(w^*)} p \equiv e(w^*) p, \tag{2}$$

where the function $\phi(w) = d \ln F(w)/d \ln w$ gives the labor supply elasticity to the firm at wage w. Robinson (1933) termed the quantity $e(w^*)$ the *exploitation index*, as it measures the extent to which workers are underpaid relative to their productivity.

Throughout this chapter, we will define the *wage markdown* as $1 - e(w^*)$. The markdown is often a central object of interest in empirical studies of monopsony. A recent meta-analysis by Sokolova and Sorensen (2021) finds an average estimated value of $\phi(w^*)$ among studies published in elite economics journals of 4.5. A monopsonist facing such an elasticity will exhibit an exploitation index of $e(w^*) = 4.5/5.5 \approx 0.82$, implying a wage markdown of roughly 18%. An observation to which we will return many times in this chapter is that labor supply elasticities are likely to depend on wage levels. If $\phi(w)$ is decreasing in the wage, then the markdown will be increasing in the wage.

Plugging (2) into the formula for profits yields

$$\Pi(w^*) = F(w^*)(1 - e(w^*))\, p. \tag{3}$$

In our simple model with constant productivity, the only source of firm profits is the wage markdown. As a result, the profit margin $\Pi(w^*)/(pF(w^*))$ directly identifies both the wage markdown and the labor supply elasticity $\phi(w^*)$. This equivalence is obviously quite fragile. We will consider how the mapping from profit margins to markdowns varies when labor productivity is not constant in Section 6.

2.2.2 Shortages

Since the monopsonist makes a profit of $(1 - e(w^*))\, p$ on each worker, it would like to hire as many workers at wage w^* as possible. The fact that only $F(w^*)$ workers are willing to work at this wage is one explanation for labor "shortages." Shortages can, in principle, be solved by raising wages above w^*; however, doing so would be unprofitable.

Many of the occupational labor markets traditionally cited as exemplars of monopsonistic behavior, particularly the markets for nurses and teachers, have a long history of perceived staffing shortages (Yett, 1970; Landon and Baird, 1971; Sullivan, 1989; Ingersoll, 2003; Staiger et al., 2010). While the literature still awaits a comprehensive empirical account of the relationship between firms' shortage perceptions and empirical estimates of markdowns, Friedrich and Zator (2024) provide quasi-experimental evidence from Germany that raising wages alleviates reported shortages.

Many countries operate guest worker programs designed to address labor shortages in specific occupations. A prominent example is the United States H-1B visa program, which is intended to ease shortages in high-skilled occupations. While guest worker programs may alleviate shortages in some circumstances, they can also amplify them, as restrictions on the job mobility of immigrant labor potentially create the opportunity for greater wage markdowns (Naidu et al., 2016; Doran et al., 2022; Townsend and Allan, 2024).

2.3 Some introductory comparative statics

Comparative statics involving changes to F and p capture the key causal relationships in the basic monopsony model. Recall that an optimal wage solves $\varrho(w^*) = p$ and that strict -1-concavity of F guarantees $\varrho'(w^*) > 0$. It follows that $dw^*/dp = 1/\varrho'(w^*) > 0$. This insight motivates much of the recent empirical literature on productivity passthrough to wages, which studies how firm-specific productivity changes propagate into wages. We discuss this literature in Section 3.1, returning to it again in Section 6.1 where productivity is itself treated as endogenous.

From Eq. (2), wages are uniquely determined by productivity p and the local elasticity $\phi(w^*)$. Hence, a perturbation to F only affects wages insofar as it alters the elasticity $\phi(w^*)$. Specifically, the model predicts that an increase in the elasticity will raise wages by attenuating wage markdowns. Evidence on the validity of this prediction is sparse. In Section 5, we discuss models where the outside option distribution can influence final wages through factors other than the elasticity.

Firm size $F(w^*)$ is, by assumption, a monotone function of wages. It follows immediately that $dF(w^*)/dp = f(w^*)/\varrho'(w^*) > 0$. Likewise, a perturbation to F that raises the elasticity $\phi(w^*)$ will increase the firm's optimal size. In Section 6.4 we discuss models where firm size can depend on recruiting expenditures in addition to wages.

The firm's profits $\Pi(w^*) = F(w^*)(p - w^*)$ serve as a central quantity of interest in assessing the incidence on firm owners of changes to the economic environment. Applying the envelope theorem yields $d\Pi(w^*)/dp = F(w^*)$. That is, small increases in labor productivity are entirely captured by the firm in the form of profits. We return to this observation in Section 2.5, which discusses efficiency in the monopsony model. An additional implication of this envelope result is that profits scale more than proportionately with productivity as $d\ln\Pi(w^*)/d\ln p = p/(p - w) > 1$.

The effects of a small change to the shape of F on profits are more subtle. Once again applying the envelope theorem, a perturbation to F at the point w^* that yields a small increase in the labor supply elasticity $\phi(w^*)$ will reduce profits. Conversely, a small increase in the labor supply level $F(w^*)$ that preserves $e(w^*)$ will serve to increase profits. However, comparative statics involving parameterizations of F typically vary both the elasticity and level of labor supply in ways that conflate these effects.

To illustrate this point, suppose that outside options follow a power function distribution $F(w) = \left(w/\bar{b}\right)^{\phi}$. The parameter $\phi > 0$ gives the elasticity of labor supply, while the parameter $\bar{b}$ governs the labor supply level. However, an increase in ϕ, for any $w < \bar{b}$, will also reduce the level of labor supplied to the firm, which mechanically reduces firm size. It can therefore be useful to compute a *compensated* change in the elasticity that offsets this level effect by reducing $\bar{b}$ in order to hold firm size constant.

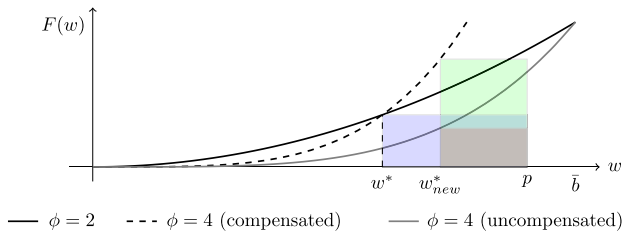

FIGURE 1 An increase in the labor supply elasticity.

Fig. 1 depicts the effects of a discrete jump in the labor supply elasticity ϕ from 2 (solid line) to 4 (gray line). The profits under $\phi = 2$ are given by the purple rectangle, while the profits generated under $\phi = 4$ are depicted by the smaller red rectangle. The compensated labor supply function (dashed line) adjusts the value of $\bar{b}$ governing the uncompensated function (gray line) so as to intersect the point $(w^*, F(w^*))$. In this example, both the compensated and uncompensated changes lead wages to rise from w^* to w^*_{new}, which implies profits per worker fall from $p/3$ to $p/5$. However, the compensated change leads to a greater increase in employment, which yields additional profits, depicted by the green rectangle.

It can be analytically convenient to study infinitessimal, rather than discrete, changes. Countering the mechanical effect of a small change $d\phi$ in the elasticity

on the location of the firm's labor supply curve requires an offsetting change $d\bar{b} = \bar{b} \ln\left(w^*/\bar{b}\right) d\ln\phi < 0$ in the support of outside options, which ensures $dF(w^*) = 0$. The first-order effect of such a compensated elasticity change on profits is $-\left(w^*/\bar{b}\right)^{\phi} (1+\phi)^{-2} p < 0$. Evidently, the larger an elasticity the firm already faces, the smaller is the profit loss from a compensated increase in the labor supply elasticity. Conversely, the higher the wage initially offered, the greater is the loss in profits associated with an increase in elasticity.

2.4 Modeling outside options

The outside option distribution F plays a central role in any monopsony model, serving as the microfoundation of the labor supply curve to the firm. In equilibrium models these distributions are shaped by optimizing behavior, search frictions, and heterogeneity in worker and firm primitives. From the perspective of a single firm, however, this distribution is exogenous.

We turn now to studying some convenient parameterizations of F and develop some results concerning the properties of mixtures of these parametric families. These results are then used to study monopsonistic wage setting in a non-sequential search model. Turning to an equilibrium variant of this model, we find that isoelastic parameterizations of outside options can be microfounded with a careful choice of the productivity distribution across firms and assumptions on the search technology. We then show that introducing taste heterogeneity into the model yields a "logit-like" outside option distribution.

2.4.1 A cookbook of log-concave CDFs

Table 1 lists some benchmark distributions and their properties, which we discuss here.

TABLE 1 Distribution of workers' outside options.

Name	Distribution	CDF	Elasticity	Markdown	Super-elasticity
Power	$b/\bar{b} \sim \text{Beta}(\phi, 1)$	$F(w) = \left(w/\bar{b}\right)^{\phi}$	ϕ	$\frac{1}{1+\phi}$	0
Shifted Power	$(b-\underline{b})/(\bar{b}-\underline{b}) \sim \text{Beta}(\beta, 1)$	$F(w) \propto (w-\underline{b})^{\beta}$	$\beta \frac{w}{w-\underline{b}}$	$\frac{1-\underline{b}/p}{1+\beta}$	$-\left(w/\underline{b}-1\right)^{-1}$
Logit	$\ln b \sim \text{Logistic}(\sigma, \beta)$	$F(w) = \left(1+(w/\sigma)^{-\beta}\right)^{-1}$	$\beta\left(1+(w/\sigma)^{\beta}\right)^{-1}$	$\frac{1+(w^*/\sigma)^{\beta}}{1+(w^*/\sigma)^{\beta}+\beta}$	$-\beta \frac{(w/\sigma)^{\beta}}{1+(w/\sigma)^{\beta}}$
Fréchet	$b/\bar{b} \sim \text{Fréchet}(\beta, 1, 0)$ (truncated to $[0,1]$)	$F(w) = \exp\left(1-(w/\bar{b})^{-\beta}\right)$	$\beta\left(w/\bar{b}\right)^{-\beta}$	$\frac{\left(w^*/\bar{b}\right)^{\beta}}{\left(w^*/\bar{b}\right)^{\beta}+\beta}$	$-\beta$

Log concavity was cited earlier as a property guaranteeing both existence and uniqueness of the profit maximizing wage. A particularly convenient log-concave family that underlies isoelastic characterizations of firm level labor supply comes from a scaled version of the Beta distribution obeying a simple power law.

Example (Power function). If $b/\bar{b}$ follows a Beta$(\phi, 1)$ distribution, then $F(w) = \left(w/\bar{b}\right)^{\phi}$, where $\phi > 0$ gives the elasticity of labor supply to the firm.

A set of power function CDFs was already depicted in Fig. 1 with varying choices of ϕ and $\bar{b}$. While isoelastic parameterizations of labor supply are a staple of the monopsony literature, there are good reasons to be skeptical of such formulations. One is that this distribution assumes that a positive density of workers is willing to work for any wage above zero, which seems unlikely to be true for most firms. Another problem is that the labor supply elasticity should decrease at high wage levels. To see why, observe that if a firm has hired nearly all of the available workers, it cannot expect equivalent employment gains by further hiking the wage. Even in the power function example, the elasticity is only constant up to the point $w = \bar{b}$, above which it becomes zero as there are no more workers available to be hired. It would seem more plausible that the elasticity decreases smoothly before falling to zero discontinuously. Three examples of log-concave distributions exhibiting non-constant labor supply elasticities are given below.

Example (Shifted power function). Card et al. (2018) and Kline et al. (2019) consider the case where $F(w) \propto \left(w - \underline{b}\right)^{\beta}$ for $\beta > 0$, which amounts to assuming $\left(b - \underline{b}\right) / \left(\bar{b} - \underline{b}\right) \sim \text{Beta}(\beta, 1)$. This distribution yields labor supply elasticity $\phi(w) = \beta \frac{w}{w - \underline{b}}$, which is strictly decreasing in w and asymptotes to β.

A notable feature of the shifted power specification is that plugging its elasticity function into (2) yields a linear wage posting rule

$$w^* = \frac{1}{1+\beta}\underline{b} + \frac{\beta}{1+\beta}p.$$

In some respects, this posting rule mirrors the surplus splitting rule delivered by standard Nash bargaining models, with $\beta/(1+\beta)$ playing the role of the firm's bargaining weight and $\underline{b}$ the role of a worker's outside option. Note, however, that in the present model, $\underline{b}$ specifically refers to the outside option of the worker most eager to work for the firm.

A link to traditional discrete choice models can be developed by assuming outside options are log-logistically distributed. This assumption implies that the labor supply curve to the firm takes the familiar "logit" form when plotted as a function of the log wage.

Example (Logit). The log-logistic distribution

$$F(w) = \left[1 + \exp\left(-\beta \ln w/\sigma\right)\right]^{-1} = \left(1 + (w/\sigma)^{-\beta}\right)^{-1},$$

with scale $\sigma > 0$ and shape $\beta > 0$, is defined on $(0, \infty)$. The labor supply elasticity $\phi(w) = \beta\left(1 + (w/\sigma)^{\beta}\right)^{-1}$ is monotone decreasing in w and asymptotes to zero.

Note that in the logit specification the labor supply elasticity obeys $\phi(w^*) = \beta\left[1 - F(w^*)\right]$. Using this relationship, we can express the posted wage in terms of firm size:

$$w^* = \frac{\beta\left[1 - F(w^*)\right]}{1 + \beta\left[1 - F(w^*)\right]} p.$$

All else equal, larger monopsonists will pay higher wages.

A useful summary of how the labor supply elasticity changes with the wage w comes from the "super-elasticity" $d\ln\phi(w)/d\ln w$: the wage elasticity of the labor supply elasticity. As will become clear in later sections, the super-elasticity of F turns out to play an important role in the study of wage passthrough. The (truncated) Fréchet distribution exhibits a constant negative super-elasticity.

Example (Fréchet). If $b/\bar{b}$ follows a truncated Fréchet distribution with shape parameter $\beta > 0$ then $F(w) = \exp\left(1 - \left(w/\bar{b}\right)^{-\beta}\right)$ for $w \in \left[0, \bar{b}\right]$. The labor supply elasticity is $\phi(w) = \beta\left(w/\bar{b}\right)^{-\beta}$. Hence, the super-elasticity of labor supply is $-\beta$.

The Fréchet labor supply elasticity is proportional to log firm size: $\phi(w^*) = \beta\left[1 - \ln F(w^*)\right]$. Thus, posted wages can again be written $w^* = \frac{\beta\left[1-\ln F(w^*)\right]}{1+\beta[1-\ln F(w^*)]} p$.

2.4.2 Mixtures, concavity, and non-sequential search

The distributions in Table 1 can be used as building blocks for generating more complex mixture distributions that offer additional flexibility as models of labor supply. Consider, for instance, a mixture of power function distributions, where the elasticity parameter ϕ is uniformly distributed on the interval $\left[0, \bar{\phi}\right]$. This choice yields marginal distribution

$$F(w) = \bar{\phi}^{-1} \int_0^{\bar{\phi}} \left(w/\bar{b}\right)^{\phi} d\phi = \frac{(w/\bar{b})^{\bar{\phi}} - 1}{\bar{\phi}\ln\left(w/\bar{b}\right)}$$

for $w \in \left(0, \bar{b}\right)$ and endpoints $F(0) = 0$, $F\left(\bar{b}\right) = 1$. Inspecting this CDF reveals that it exhibits an elasticity function that is increasing in w with non-constant super-elasticity.

While mixtures of log-convex functions are necessarily log-convex, mixtures of log-concave functions need not be log-concave (An, 1997). In the mixture of power function distributions example, log-concavity can be shown to fail when $\bar{\phi} > 1$. We saw earlier that a unique wage is assured when F is strictly -1-concave. It turns out that mixtures of concave distributions are strictly -1-concave. This convenient result, which is formalized in the lemma below, reduces the task of verifying uniqueness of the monopsony wage to the problem of verifying that the second derivatives of the distributions being mixed are not positive.

Lemma 2. *Suppose* $F(w) = \sum_i \omega_i F_i(w)$, *where* $\omega_i \geq 0$, $\sum_i \omega_i = 1$, *and each* $F_i : \left[\underline{b}, \bar{b}\right] \to [0, 1]$ *is twice differentiable, strictly increasing, and concave. Then* F *is strictly* -1*-concave.*

Proof. Twice differentiability and monotonicity of each F_i implies F is twice differentiable with $f(w) > 0$. Thus, F is strictly -1-concave iff $d^2/dw^2(-1/F(w)) \propto f'(w)F(w) - 2f^2(w) < 0$. Since $F(w) \geq 0$, it suffices to show that $f'(w) \leq 0$. Concavity of F_i implies $F_i((1-\alpha)w_0 + \alpha w_1) \geq (1-\alpha)F_i(w_0) + \alpha F_i(w_1)$ for all $(w_0, w_1) \in [\underline{b}, \bar{b}]$ and $\alpha \in [0, 1]$. By definition, $F((1-\alpha)w_0 + \alpha w_1) = \sum_i \omega_i F_i((1-\alpha)w_0 + \alpha w_1) \geq \sum_i \omega_i [(1-\alpha)F_i(w_0) + \alpha F_i(w_1)] = (1-\alpha)F(w_0) + \alpha F(w_1)$. Hence, F is concave, which implies $f'(w) \leq 0$ for all $w \in [\underline{b}, \bar{b}]$. □

Returning to the mixture of power functions example, recall that strict -1-concavity amounts to the requirement that $d^2/dw^2\,(-1/F(w)) < 0$. Differentiating reveals that this condition is satisfied whenever $\bar{\phi} \leq 1$. Lemma 1, though technically stated in terms of finite mixtures, provides us with a more direct route to the same conclusion: the power function distributions being mixed are concave if and only if $\bar{\phi} \leq 1$.

Non-sequential search models typically yield labor supply curves involving finite mixtures of integer powers of distribution functions. Consider an idealized labor market that contains a continuum of employers, with wage offerings distributed according to the CDF $G : \left[\underline{w}, \bar{w}\right] \to [0, 1]$. A parsimonious approach to modeling search, pioneered by Butters (1977), is to assume that each worker receives a random number of $i.i.d.$ draws from G and selects the sampled employer offering the highest wage. To simplify the problem, suppose that every worker gets at least two offers, so that firms are certain to face competition for each potential employee. Let $\tilde{q}_k$ denote the probability that a worker receives $1 + k$ draws from G, with $\sum_{k=1}^{\infty} \tilde{q}_k = 1$. Hence, each worker expects $1 + \sum_{k=1}^{\infty} k\tilde{q}_k$ offers, a quantity we assume exists.

Suppose the measure of firms happens to equal $1 + \sum_{k=1}^{\infty} k\tilde{q}_k$, a normalization that ensures each firm expects to meet a single worker. The probability that the highest of k draws from G is lower than w is $G(w)^k$. Therefore, the measure of workers expected to be recruited by a firm posting wage w can be written

$$F(w) = \sum_{k=1}^{\infty} q_k G(w)^k, \tag{4}$$

where $q_k = \frac{(1+k)\tilde{q}_k}{1+\sum_{\ell=1}^{\infty} \ell\tilde{q}_\ell}$ gives the expected share of workers encountered that have k outside offers. Suppose there is some maximal number of outside offers $\bar{k}$ such that $q_k = 0$ for $k > \bar{k}$. Then by Lemma 2, F will be strictly -1-concave if $G^{\bar{k}}$ is concave, implying a unique wage will maximize profits for any choice of probabilities $\{q_k\}_{k=1}^{\bar{k}}$ summing to one.

The elasticity function ϕ of such an F will depend on the wage elasticity of G and the distribution of offers. All else equal, the more offers that workers expect to get, the greater will be the elasticity of F. For example, in the

case where each worker who gets an offer from the reference firm also gets an offer from either one or two randomly selected rivals ($q_1 + q_2 = 1$) we can write $F(w) = q_1 G(w) + (1 - q_1) G(w)^2$. Here, the elasticity takes the form $\phi(w) = \phi_G(w) \cdot \frac{q_1 + 2(1-q_1)G(w)}{q_1 + (1-q_1)G(w)}$, where $\phi_G(w)$ is the wage elasticity of G. As q_1 increases, the number of offers falls and $\phi(w)$ decreases. Note that $\phi(w)$ is bracketed in the interval $[\phi_G(w), 2\phi_G(w)]$, corresponding to the elasticities of the two distributions being mixed. Hence, if one works with a G that exhibits a wage elasticity asymptoting to zero, then the elasticity of F will also asymptote to zero. Conversely, if the elasticity of G diverges to infinity, then $\phi(w)$ will diverge as well.

2.4.3 Equilibrium constraints

In an equilibrium model, the cross-sectional distribution of wages G emerges from optimizing behavior, which restricts the set of possible labor supply functions F that can arise. It is worth demonstrating that log-concave labor supply functions of the sort described in Section 2.4.1 can, in fact, be obtained from such an approach.

Let $H : \left[\underline{p}, \bar{p}\right] \to [0, 1]$ denote the cross-sectional distribution of firm productivity. If all firms set wages according to (2), then

$$\begin{aligned} G(w) &= \Pr\left(w^* < w\right) \\ &= \Pr\left(p < w + F(w)/f(w)\right) \\ &= H\left(w + F(w)/f(w)\right), \end{aligned} \tag{5}$$

where the second equality applies the inverse wage transform $\varrho(w^*) = p$. From (2), the lowest wage $\underline{w}$ solves $e\left(\underline{w}\right)\underline{p} = \underline{w}$, while the highest wage $\bar{w}$ solves $e(\bar{w})\,\bar{p} = \bar{w}$, yielding the endpoint conditions $G\left(\underline{w}\right) = 0$ and $G(\bar{w}) = 1$.

An *equilibrium* is a pair (F, G) of distribution functions obeying (4), (5), and the endpoint conditions. When an equilibrium exists, these conditions define a mapping $\left(H, \{q_k\}_{k=1}^{\infty}\right) \mapsto (F, G)$ from productivity and search frictions to labor supply and wages. The following proposition considers a simple choice of H and $\{q_k\}_{k=1}^{\infty}$ that yields analytical solutions for both F and G.

Proposition 1 (K outside offers, power function productivity). *Suppose* (2), (4), *and* (5) *hold, every worker gets* $1 + K$ *wage offers* ($q_K = 1$), *and firm productivity follows a power function distribution on the unit interval with shape parameter* $\lambda > 0$ $\left(H(p) = p^{\lambda}\right)$. *Then* $F(w) = (w/\bar{w})^{\lambda K}$, $G(w) = (w/\bar{w})^{\lambda}$, *and* $\bar{w} = \lambda K/(1 + \lambda K)$.

Proof. The assumption $q_K = 1$ implies $F = G^K$. Imposing the power function form of productivity on (5) yields $F(w) = [w + F(w)/f(w)]^{\lambda K}$, which can be rearranged as the differential equation $f(w) = F(w)/\left[F(w)^{1/\lambda K} - w\right]$. It is straightforward to verify that this differential equation is solved by $F(w) = w^{\lambda K}(1 + 1/(\lambda K))^{\lambda K}$. The wage distribution is $G(w) = F(w)^{1/K} = w^{\lambda}(1 + 1/(\lambda K))^{\lambda}$. It follows that $G(0) = 0$ and $G(\lambda K/(1 + \lambda K)) = 1$. □

The resulting F and G both take the log-concave power function form. The constant wage elasticity of F yields a constant markdown $1/(1+\lambda K)$, implying wages are an affine transformation of productivity supported on the interval $[0, \lambda K/(1+\lambda K)]$. Hence, either a higher λ—indicating a thicker tail of productivity—or a higher K—indicating greater labor market competition—will yield a smaller markdown and a greater maximal wage. However, the mean wage $\int_0^{\lambda K/(1+\lambda K)} w dG(w) = \lambda/(1+\lambda)(\lambda K/(1+\lambda K))$ increases more rapidly with λ than K. This discrepancy reflects that K only governs markdowns, while λ affects both markdowns and the distribution of productivity being marked down. More complex productivity distributions or choices of $\{q_k\}_{k=1}^{\infty}$ will generally yield non-constant markdowns, leading the shape of G to depart more significantly from the shape of the productivity distribution H.

Thus far we have assumed that workers always work for the employer offering them the highest wage. Following Card et al. (2018), it has become popular to work with random utility models implying some horizontal differentiation among firms offering the same wage. This differentiation, which reflects heterogeneity in worker assessments of employers, weakens the grip of equilibrium restrictions on the wage distribution, effectively injecting "noise" into the map between firm productivity and wages.

To explore this approach, suppose that when a worker encounters a firm, information about its non-pecuniary attributes is revealed via a draw ξ from a Fréchet distribution with shape parameter $\beta > 0$. The worker's indirect utility is multiplicative in this Fréchet draw and the offered wage w.[1] Consequently, when faced with K alternative wages $(w_1, \ldots, w_K)$, the probability that a worker chooses the firm offering wage w is

$$\Pr(w\xi > \max\{w_1\xi_1, \ldots, w_K\xi_K\}) = w^\beta / \left(w^\beta + \sum_{k=1}^{K} w_k^\beta\right).$$

Note that, as β grows large, this probability collapses to an indicator function for whether w is larger than the K alternatives. If the K alternative wages are drawn independently from G, then the relevant outside option distribution is

$$\begin{aligned} F(w) &= \mathbb{E}\left[\frac{w^\beta}{w^\beta + \sum_{k=1}^{K} w_k^\beta}\right] \\ &= \int \cdots \int_0^{\bar{w}} \frac{w^\beta}{w^\beta + \sum_{k=1}^{K} w_k^\beta} dG(w_1) \ldots dG(w_K). \end{aligned} \tag{6}$$

In general, F does not have a closed form. However, replacing $w^\beta / \left(w^\beta + \sum_{k=1}^{K} w_k^\beta\right)$ with its second-order Taylor approximation around the

[1] Equivalently, we could assume, as in Card et al. (2018), that workers have indirect utility functions that are linear in log wages and a type I Extreme Value distributed error. The logarithm of an EV1 distributed random variable is Fréchet distributed.

point $\sum_{k=1}^{K} w_k^\beta = K\mathbb{E}\left[w_k^\beta\right]$ and taking expectations yields the more tractable distribution

$$F^\star(w) = \frac{w^\beta}{w^\beta + \sigma^\beta} \cdot \frac{\left(w^\beta + \sigma^\beta\right)^2 + \kappa^2}{\left(w^\beta + \sigma^\beta\right)^2},$$

where $\sigma^\beta = K\mathbb{E}\left[w_k^\beta\right] = K\int_0^{\bar{w}} x^\beta dG(x)$ and $\kappa^2 = K\mathbb{V}\left[w_k^\beta\right] = K\int_0^{\bar{w}} \left(x^\beta - \sigma^\beta/K\right)^2 dG(x)$. While F depends on all the moments of G, $F^\star$ depends only on two moments of the cross-sectional wage distribution. When wage dispersion is modest, $F^\star$ will tend to provide an accurate approximation to F.

The function $F^\star$ is the product of two terms. The first term amounts to the "logit" specification of Section 2.4.1 with a scale parameter σ that depends on the number of offers K and the cross-sectional wage distribution G. This term gives the labor supply function that firms would face if workers (mistakenly) believed that $\sum_{k=1}^{K} w_k^\beta$ always equals its mean. By Jensen's inequality, replacing $\sum_{k=1}^{K} w_k^\beta$ with its expected value will lead to an underestimate of $F(w)$. The second term can be thought of as a correction that accounts for the variance of the outside wage opportunities around their expected value. This term grows monotonically in κ^2. One can show that $F^\star(w)$ is log-concave whenever $\beta > 2$.

The following proposition describes the cross-sectional wage distribution $G^\star$ that emerges when the outside option distribution is given by $F^\star$.

Proposition 2 ($F = F^\star$, K outside offers, power function H). *Suppose* (5) *holds, every worker gets K outside offers, and* $F(w) = \frac{w^\beta}{w^\beta+\sigma^\beta}\frac{\left(w^\beta+\sigma^\beta\right)^2+\kappa^2}{\left(w^\beta+\sigma^\beta\right)^2}$. *If* $H(p) = p^\lambda$, *then*

$$G(w) = w^\lambda\left(1 + \frac{1}{\beta}\left(w^\beta + \sigma^\beta\right)\frac{\kappa^2 + \left(w^\beta + \sigma^\beta\right)^2}{\sigma^\beta\left(w^\beta + \sigma^\beta\right)^2 + \kappa^2\left(\sigma^\beta - 2w^\beta\right)}\right)^\lambda.$$

Proof. From (5), $G(w) = (w + F(w)/f(w))^\lambda$. Differentiating F and substituting into this expression yields the result. □

In contrast to Proposition 1, Proposition 2 reveals that the shape of the cross-sectional wage distribution $G^\star$ induced by $F^\star$ departs substantially from the shape of the productivity distribution. However, when $\kappa^2 \approx 0$, we have $G^\star(w) \approx w^\lambda\left(\frac{1+\beta}{\beta} + \frac{1}{\beta}(w/\sigma)^\beta\right)^\lambda$, which is the product of a power function with a shifted power function. We therefore expect a distributional shape not dramatically different from the power function form when wage inequality is modest.

It is natural to wonder how well the insights derived from the second-order approximation $F^\star$ and its corresponding cross-sectional distribution $G^\star$ carry

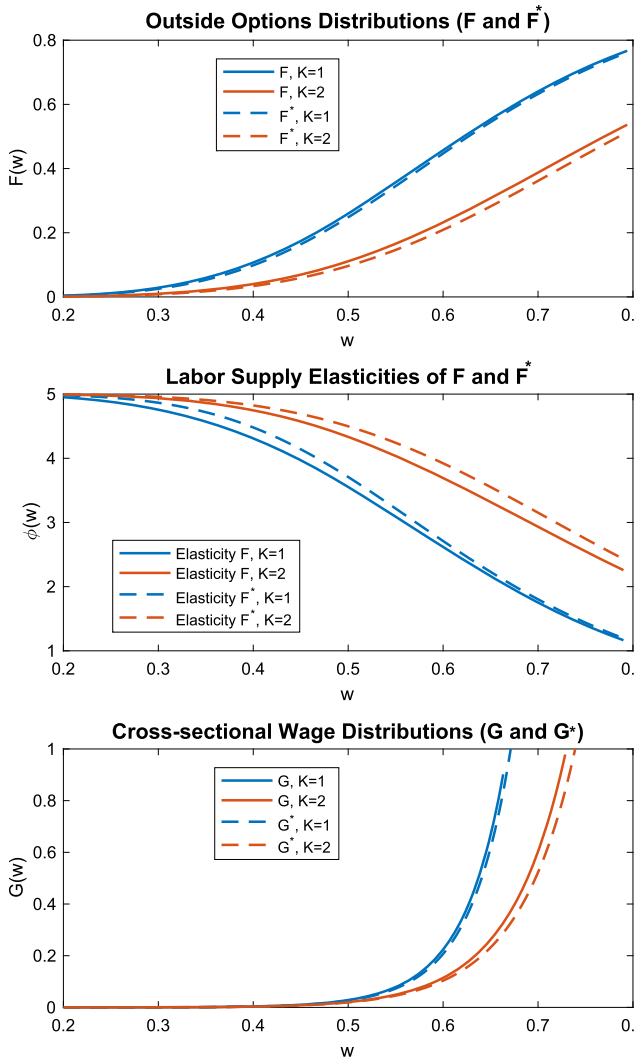

FIGURE 2 Equilibrium F, ϕ, and G with Fréchet taste heterogeneity ($\beta = 5, \lambda = 8$).

over to the exact system described by (5) and (6). Fig. 2 illustrates numerical solutions to the exact system computed via fixed point iteration over spline approximations to (F, G) for two choices of K. For comparison, numerical solutions to the approximate system, which were found by fixed point iteration over the scalars $\left(\kappa^2, \sigma^\beta, \bar{w}\right)$, are also displayed.

The top panel of Fig. 2 shows that the approximation $F^\star$ is nearly indistinguishable from F when $K = 1$ and remains very accurate when $K = 2$. Both $F^\star$ and F turn out to be log-concave numerically. Wage offers below 0.2 have es-

sentially no chance of being accepted. A wage offer of 0.8 has roughly an 80% chance of being chosen when $K = 1$ but only about a 50% chance when $K = 2$. The second panel shows the labor supply elasticities $\phi(w)$ implied by F and $F^\star$. In line with our discussion in Section 2.4.1 of the logit specification, both sets of elasticities converge to $\beta = 5$ at the lowest wage levels and fall gradually as the wage rises. This decline is steepest when $K = 1$. The approximate and exact elasticities converge at the highest wage levels.

The bottom panel plots the cross-sectional wage distribution. As expected, both G and $G^\star$ look very much like a power function and both solutions turn out to be nearly log-concave, with exceptions driven by numerical approximation error. Very few employers offer wages below 0.5. This phenomenon primarily reflects our choice to set $\lambda = 8$, which implies that the share of firms having productivity below $1/2$ is only $2^{-8} \approx 0.004$. As in our earlier example, an increase in K not only raises the mean wage but also boosts dispersion in wages. When $K = 1$, the maximum wage is 0.67, while when $K = 2$, maximum wage rises to 0.78. This increased dispersion leads the second-order approximation $G^\star$ to depart a bit more from the exact solution G. If we had worked with much larger values of K, or much smaller values of λ, a higher order approximation capturing the influence on F of G's skewness and kurtosis would have been required to obtain accurate results.

2.4.4 *Are firm labor supply curves log-concave?*

While log-concave distributions are convenient modeling tools, surprisingly little direct empirical evidence is available on whether and when firm-specific labor supply curves tend to be log-concave. In principle, this question is amenable to testing via the same sorts of research designs used to measure labor supply elasticities. For instance, Dube et al. (2020) estimate elasticities of labor supply to online employers using experimental variation in the wages offered for narrowly defined tasks and find a very low job acceptance elasticity, signifying wide dispersion in outside options. However, they stop short of estimating the full distribution of outside options faced by these online workers.

It is plausible that the shape of outside option distributions varies widely across jobs involving different amenities and skill requirements. Azar et al. (2022) estimate that job postings on an online job board face very different effective labor supply elasticities. However, they rely on a standard nested logit model of preferences that presumes all jobs face log-concave supply curves. Establishing when and whether log-concavity fails is an interesting question for future research. One place to suspect that log-concavity is a reasonable approximation is in jobs subject to minimum wages. If no one is willing to work below the minimum then $f(w)/F(w)$ should be nearly infinite at the minimum wage and much lower at higher wage levels. Even so, there is no guarantee that the ratio $f(w)/F(w)$ will continue to decrease at higher wage levels.

While log-concavity is not required for a unique wage to solve (1), it would nonetheless be quite extraordinary to discover a firm whose labor supply curve

is log-convex. As in the Burdett and Mortensen (1998) model, such a firm could find itself indifferent between multiple profit maximizing wage levels, leading to comparative statics untethered from local supply elasticities. Documenting that such cases actually exist would present an intriguing opportunity to test non-parametric predictions of the theory. In principle, the inverse wage function $\varrho(w)$ can be estimated and the number of potential crossings of p estimated from its shape.

2.5 Match surplus and efficiency

Textbook treatments of monopsony typically conclude that the firm's wage-setting power leads the monopsonist to employ too few workers. The logic of this argument is easy to grasp. The profits of a monopsonist are given by $\Pi(w^*)$, while the rents enjoyed by its workers over their outside options can be written

$$R(w^*) = \int_0^{w^*} (w^* - b)\, dF(b).$$

By the first-order condition for optimization, $\Pi'(w^*) = 0$. In contrast, $R'(w^*) = F(w^*)$. Hence, the total surplus derived from matches with the firm, $\Pi(w^*) + R(w^*)$, can be increased by raising the wage slightly.

Intuitively, while the firm is indifferent about a small wage increase, inframarginal workers value each dollar increase at a full dollar. This insight forms the crux of classic arguments for minimum wages to improve welfare in monopsonistic markets (Robinson, 1933). However, such arguments traditionally presume that workers would be idle if not employed by the monopsonist. When workers' outside options involve productive activities, raising wages can destroy matches that are socially valuable, leading to misallocation.

We now review more carefully the microeconomic forces that can give rise to inefficiency in the monopsony model, starting with the possibility that wage posting leads workers to refuse job offers that would be welfare improving. We then scrutinize the link between wage markdowns and efficiency, arguing that the forces giving rise to these markdowns—taste heterogeneity or search frictions—have very different implications for welfare. This point is then illustrated in a stylized model, where some new conditions are offered for assessing the ex-post efficiency of match formation.

2.5.1 The perils of wage posting

In textbook treatments, the original sin of the monopsonist is its refusal to hire workers with outside options in the range $\left[e(w^*)\, p,\, p\right]$ who would be willing to work for less than their marginal product. This failure to exploit the potential gains from trade can be viewed as arising from an information problem: the firm can't strike a deal with these workers because it doesn't know which of them have outside options in this range. It is plausible that problems of this nature are

widespread. Outside options can be difficult for an employer to verify, consisting in part of non-pecuniary components such as commuting time, job amenities, and the value of leisure.

An equivalent difficulty arises in the Burdett and Mortensen (1998) wage posting model when the value of leisure is private information, which leads some offers to unemployed workers to be rejected even when $\bar{b} < p$. As they note, a judiciously chosen minimum wage can raise employment in this environment by reducing the number of offers to unemployed workers that are rejected. This prediction still awaits careful empirical examination.

One might object that even if the firm doesn't initially know workers' outside options, there are incentives for some sort of deal to be struck. As Manning (2011) notes "economists abhor unexploited surpluses." Yet bargaining can be costly. These costs include the direct time and monetary costs of negotiating wages and also the indirect effects on morale and productivity of paying different wages to workers in roughly equivalent roles (Card et al., 2012; Breza et al., 2018). Weil (2014) notes that "wage discrimination is rarely seen in large firms despite the benefits it could confer," arguing that aversion to within firm wage inequality is a driving force behind outsourcing.

Employers also typically face sharp legal restrictions on wage discrimination. In the United States, the Equal Pay Act of 1963 mandates "equal pay for equal work." Violations of this law can be judged to occur when pay differs between male and female employees performing substantively comparable work, even if their job titles differ. Unwarranted pay disparities involving race, national origin, age, or disability status are respectively prohibited by Title VII of the Civil Rights Act of 1964, the Age Discrimination in Employment Act of 1967, and Title I of the Americans with Disabilities Act of 1990. Amior and Manning (2020) and Amior and Stuhler (2022) argue that constraints on wage discrimination lead firms to apply a common markdown to the wages of immigrant and native workers that potentially gives rise to misallocation.

A final sort of impediment to wage discrimination is that striking individualized deals with workers may be less profitable for the firm than committing to posted wages. We investigate a stylized class of bargaining models in which this phenomenon can arise in Section 5. These models, which involve bargaining under incomplete information, also offer an explanation for how positive surplus matches can sometimes fail to form even in labor markets where negotiation is prevalent.

2.5.2 The tenuous link between markdowns and efficiency

Much of the empirical monopsony literature focuses on estimating wage markdowns $1 - w^*/p$ with the implicit presumption that these parameters provide a gauge of inefficiency. While a large wage markdown can stymie the creation of matches with positive surplus, evaluating the social value of forming a match between a worker and a firm requires assumptions about the forces generating that worker's outside options. Suppose, for example, that all firms mark wages

down by the same proportion. If, as in the non-sequential search environment sketched in (4), all workers get at least one offer and each worker chooses to work for the highest wage firm they encounter, then the mapping of workers to firms will be the same as if each firm had set $w = p$. What ultimately matters for assessing efficiency is not markdowns but allocations: which matches should have formed that didn't?

In the Burdett and Mortensen (1998) model, worker wages are dispersed and always fall below marginal product, even when all workers are identical. Though these markdowns disadvantage workers, match creation is ex-ante efficient in the absence of preference heterogeneity because unemployed workers never turn down job offers and employed workers always accept offers from more productive firms. Likewise, ex-ante efficient allocations arise in the non-sequential consumer search models of Butters (1977), Burdett and Judd (1983), and Menzio (2024) despite the presence of equilibrium gaps between marginal cost and price. In both these search frameworks, the markdowns (or markups in the case of consumer search) serve an allocative role, inducing workers to move as close as possible to their most productive task. When preference heterogeneity is introduced to these models, efficiency tends to break down because firms are unable to tailor wages to latent preference types, leading surplus improving offers to be rejected.

The theoretical possibility that wage markdowns can be efficient presents both challenges and opportunities for the monopsony literature. Are outside options dispersed because of search frictions or because of preference heterogeneity that hinders efficient match formation? Surely, in many markets, the answer involves some mix of these factors. Inefficiencies can also stem from barriers to free entry by firms (see Manning, 2013, chapter 3) or from standard search externalities (Hosios, 1990). While the comparative statics of wages and firm size don't depend on parsing these forces, the welfare consequences do. To formalize these concerns, we conclude this section with a brief example illustrating how the cross-sectional relationship between wages and productivity can be used to assess efficiency in a model exhibiting both preference heterogeneity and search frictions.

2.5.3 *Efficiency with search frictions and taste heterogeneity*

Consider a continuum of workers indexed by $i \in [0, 1]$. Search frictions lead workers to face finite choice sets $\mathcal{C}(i)$ of feasible employment opportunities. Since non-employment is always a feasible option, we assume that $|\mathcal{C}(i)| \geq 1$ but otherwise allow these sets to vary arbitrarily across workers. We will sidestep here the important question of whether the search process could have delivered "better" choice sets, focusing instead on deriving conditions under which match formation subject to these frictions turns out to be constrained efficient.

Suppose that worker i's indirect utility of employment at firm j is given by $V_{ij} = w_j + \varepsilon_{ij}$, where w_j is the wage offered by firm j and ε_{ij} captures

worker i's valuation of the non-pecuniary aspects of the job. Non-employment can be viewed as a firm offering a wage of zero. Worker i will choose the firm in $\mathcal{C}(i)$ offering the highest private value V_{ij}. By contrast, a planner would like for each worker to be paired with the employer that produces the highest match surplus $S_{ij} = p_j + \varepsilon_{ij}$, where p_j is firm j's productivity. Assuming that ties never occur, match formation will be efficient whenever $\arg\max_{j \in \mathcal{C}(i)} V_{ij} = \arg\max_{j \in \mathcal{C}(i)} S_{ij}$ for all workers. When this condition is violated, *misallocation* is present.

Clearly, an efficient allocation will result when $w_j = p_j$ for all firms. However, efficiency is also guaranteed under the weaker requirement that $sign\left(V_{ij} - V_{ik}\right) = sign\left(S_{ij} - S_{ik}\right)$ for all pairs of firms j and k ever found in the same choice set. This condition will be violated when there exists a worker i and firm pair $(j,k) \in \mathcal{C}(i)$ for which

$$w_j - w_k > \varepsilon_{ik} - \varepsilon_{ij} > p_j - p_k \quad \text{or} \quad w_j - w_k < \varepsilon_{ik} - \varepsilon_{ij} < p_j - p_k. \tag{7}$$

That is, when a worker's non-pecuniary preference for firm k over firm j is bracketed by the wage and productivity advantages of firm j over firm k. Pairwise difference conditions of this nature arise often in the literature on matching models with non-transferrable utility, where they are used to characterize the circumstances giving rise to assortative matching (Legros and Newman, 2007).

Inefficient arrangements of the sort described by (7) can be ruled out with the assumption that $w_j - w_k = p_j - p_k$ for all firm pairs. However, this assumption implies that $w_j = p_j - C$ for some constant $C \geq 0$. To rationalize such a wage structure with a monopsony model requires the rather odd assumption that labor supply elasticities are proportional to productivity (i.e., that $\phi_j = p_j/C - 1$).

When the variation in non-pecuniary preferences is restricted, wages can deviate more substantially from productivity without generating inefficiencies. Consider the following margin condition, which stipulates that, in each worker's choice set, wages rise faster with productivity than non-pecuniary valuations fall with it.

Assumption 1 (Slope separation). No two firms have exactly the same productivity and there exists a $B < \infty$ such that $\inf_{i \in [0,1]} \min_{(j,k) \in \mathcal{C}(i)} \left(\frac{w_j - w_k}{p_j - p_k}\right) \geq B$ and $\sup_{i \in [0,1]} \max_{(j,k) \in \mathcal{C}(i)} \left(\frac{\varepsilon_{ik} - \varepsilon_{ij}}{p_j - p_k}\right) \leq B$.

Assumption 1 can be thought of as restricting the dimensionality of the model primitives. When $\inf_{i \in [0,1]} \min_{(j,k) \in \mathcal{C}(i)} \left(\frac{w_j - w_k}{p_j - p_k}\right) > 0$, wages are monotone increasing in firm productivity: i.e., the rank correlation between w_j and p_j is one within workers' choice sets. Likewise, when $\sup_{i \in [0,1]} \max_{(j,k) \in \mathcal{C}(i)} \left(\frac{\varepsilon_{ik} - \varepsilon_{ij}}{p_j - p_k}\right) < 0$, non-pecuniary valuations are monotone increasing in productivity. In such a scenario, the model is one dimensional because wages and valuations are both deterministic functions of productivity.

Non-zero values of B allow for non-deterministic relationships among these quantities. When $|B|$ is a small positive number the relationship between wages, valuations, and productivity will be nearly monotone but there can be "noise" in the relationship involving deviations of particular firms from the central tendency of the economy. A related notion of dependence comes from Theil (1950), who proposed using the median of the slopes fit to all pairs of observations as a robust estimator of the slope coefficient in an error-ridden linear model. Assumption 1 lower bounds the Theil estimate of the linear dependence of w_j on p_j in any choice set $\mathcal{C}(i)$ at B and the corresponding dependence of ε_{ij} on p_j at $-B$. These same bounds can be shown to hold for corresponding least squares regressions within choice sets.[2]

The following proposition describes conditions under which Assumption 1 guarantees that more productive firms exhibit wages high enough to offset any non-pecuniary aversion to working there.

Proposition 3 ($B \leq 1$ ensures efficiency). *If Assumption 1 holds with* $B \leq 1$ *then* $sign\left(V_{ij} - V_{ik}\right) = sign\left(S_{ij} - S_{ik}\right)$.

Proof. Dividing the inequalities in (7) by $p_j - p_k$ reveals that efficiency is violated when either (i) $\left(w_j - w_k\right)/\left(p_j - p_k\right) > \left(\varepsilon_{ik} - \varepsilon_{ij}\right)/\left(p_j - p_k\right) > 1$ or (ii) $\left(w_j - w_k\right)/\left(p_j - p_k\right) < \left(\varepsilon_{ik} - \varepsilon_{ij}\right)/\left(p_j - p_k\right) < 1$. The restriction $\sup_{i\in[0,1]} \max_{(j,k)\in\mathcal{C}(i)} \left(\frac{\varepsilon_{ik}-\varepsilon_{ij}}{p_j-p_k}\right) \leq 1$ rules out (i), while the condition $\inf_{i\in[0,1]} \min_{(j,k)\in\mathcal{C}(i)} \left(\frac{w_j-w_k}{p_j-p_k}\right) \geq \sup_{i\in[0,1]} \max_{(j,k)\in\mathcal{C}(i)} \left(\frac{\varepsilon_{ik}-\varepsilon_{ij}}{p_j-p_k}\right)$ rules out (ii). □

To explore the implications of Proposition 3 it is useful to connect this result to some of the equilibrium models covered in Section 2.4.3. If worker valuations satisfy the bound $\sup_{i\in[0,1]} \max_{(j,k)\in\mathcal{C}(i)} \left(\frac{\varepsilon_{ik}-\varepsilon_{ij}}{p_j-p_k}\right) \leq 0$, the allocation will be efficient whenever the rank correlation between firm wages and productivity is non-negative. Hence, the search equilibrium described in Proposition 1 is efficient because amenities are absent from the model $\left(\varepsilon_{ij} = 0\right)$ and wages are monotone in productivity, implying Assumption 1 is satisfied with $B = 0$. For the same reason, the Burdett and Mortensen (1998) model must be efficient when leisure heterogeneity is absent.

Now consider a model of vertically differentiated amenities where firms have scalar non-pecuniary amenities that are commonly valued ($\varepsilon_{ij} = a_j$). If more productive firms never offer worse amenities or lower wages, then the bound $B = 0$ will be satisfied and efficiency will ensue. Even if more productive firms do sometimes offer lower wages, the allocation will be efficient so long as a unit

[2] Yitzhaki (1996) showed that the least squares slope coefficient in a bivariate regression can be represented as a convex weighted average of the pairwise slopes between observations with adjacent values of the regressor. Since Assumption 1 bounds all pairwise slopes, it also bounds all adjacent pairwise slopes; e.g., slopes of the form $\left(w_j - w_{j+1}\right)/\left(p_j - p_{j+1}\right)$ where firms have been sorted based on their values of productivity.

increase in productivity always yields a sufficiently large increase in amenities. In such a case, Assumption 1 will be satisfied with $B < 0$. Conversely, if more productive firms sometimes have worse amenities, efficiency will ensue if wages are always strongly increasing in productivity, in which case Assumption 1 will be satisfied with $B \in (0, 1]$.

Finally, consider a model where $\varepsilon_{ij} = a_j + \xi_{ij}$, with ξ_{ij} representing an idiosyncratic worker taste inducing horizontal differentiation. The search equilibrium described in Proposition 2 has $a_j = 0$ and assumes ξ_{ij} is drawn from a Fréchet distribution. Because the Fréchet draws are unbounded, Assumption 1 is certain to fail no matter how strong the dependence of wages on productivity. Of course, the popularity of Fréchet and type 1 Extreme Value distributions as modeling devices owes primarily to their analytical convenience rather than their accuracy as a description of preferences. When idiosyncratic worker tastes are bounded, efficiency can ensue if wages and amenities increase sufficiently strongly with productivity that Assumption 1 holds for some $B \in (0, 1]$.

A few studies have used fine-grained data on choice sets to estimate monopsony models featuring unobserved job amenities and idiosyncratic tastes (e.g., Azar et al., 2022; Roussille and Scuderi, 2023). However, those estimates have not typically been used to evaluate the cross-sectional relationship between either wages or amenities and measures of productivity. A first question of interest is whether the relationships between these variables is nearly monotone, which would suggest the "intrinsic dimension" of the data is low. Proposition 3 suggests that even if the empirical relationships are perfectly monotone, the slope of the relationship is important for assessing efficiency. One approach to conducting such an analysis would be to use the Theil (1950) estimator to summarize the strength of the pairwise relationships within choice sets. Sen (1968) proposed a confidence interval for this estimator that can also be used to study the distribution of pairwise slopes.

As mentioned earlier, idiosyncratic tastes are typically modeled as having unbounded support, which will mechanically generate some misallocation. However, the inefficiencies generated by such modeling choices may well be minimal. The empirical literature has made strides in obtaining estimates of marginal labor productivity in the presence of imperfectly competitive labor and product markets (Dobbelaere and Mairesse, 2013; Yeh et al., 2022; Delabastita and Rubens, 2022). Given firm-specific measures of productivity, standard parametric specifications of indirect utility yield identification of the share of workers that are misallocated (i.e., for whom $\arg\max_{j \in \mathcal{C}(i)} V_{ij} \neq \arg\max_{j \in \mathcal{C}(i)} S_{ij}$), as this corresponds to the share of workers that would change employers if $w_j = p_j$ for all firms. One can typically also identify the welfare gap associated with any misallocation, which can be expressed as $\int_0^1 \max_{j \in \mathcal{C}(i)} \left\{ S_{ij} - S_{ij^*(i)} \right\} di$, where $j^*(i) = \arg\max_{j \in \mathcal{C}(i)} V_{ij}$. Producing credible estimates of these quantities based upon granular choice set data would constitute a major contribution to the literature.

3 Empirical implications of the basic model

This section explores in greater depth the predictions of the basic monopsony model regarding the effects of changes to supply and demand conditions at a single firm. A large empirical literature has sprung up testing these comparative statics using idiosyncratic firm shocks. The reduced form effects of these shocks are then typically used to construct estimates of labor supply elasticities and wage markdowns. We begin by studying the model's predictions for the propagation of productivity shocks into wages, which is the area that has received the most attention to date from empiricists (Kline et al., 2019; Lamadon et al., 2022; Garin and Silvério, 2023). This discussion highlights the important role played by the super-elasticity of F. We then proceed to discuss how changes in the outside option distribution can affect wages, focusing on the case where a firm experiences an exogenous change in amenities. The analysis reveals that compensating differentials in the monopsony model are governed by F's curvature, a concept closely related to the super-elasticity.

3.1 Productivity passthrough

An important feature of the monopsony model is that productivity variation "passes through" to workers. While the passthrough of firm-specific productivity shocks to wages signals wage-setting power, it does not directly reveal markdowns or labor supply elasticities. Differentiating (2) yields the passthrough elasticity

$$\begin{aligned}\frac{d\ln w^*}{d\ln p} &= 1+\frac{d\ln e\left(w^*\right)}{d\ln w}\frac{d\ln w^*}{d\ln p}\\ &= 1+\frac{\dot{\phi}\left(w^*\right)}{1+\phi\left(w^*\right)}\frac{d\ln w^*}{d\ln p}\\ &= \frac{1+\phi\left(w^*\right)}{1+\phi\left(w^*\right)-\dot{\phi}\left(w^*\right)} \equiv \rho\left(w^*\right), \end{aligned} \tag{8}$$

where $\dot{\phi}\left(w\right) \equiv \frac{d\ln\phi(w)}{d\ln w}$ denotes the super-elasticity of labor supply. An isoelastic F will exhibit super-elasticity of zero, and therefore, a passthrough elasticity of one. It was argued earlier that the elasticity $\phi\left(w\right)$ should eventually decline with the wage, implying a negative super-elasticity and consequently a passthrough elasticity below one. Intuitively, a positive productivity shock, by lowering the effective elasticity, depresses wages, which serves to mute passthrough.

It will be useful to relate the super-elasticity to the local *curvature* of F, which we measure via the function

$$\chi\left(w\right) = -F\left(w\right)f'\left(w\right)/f\left(w\right)^2.$$

If F is strictly concave then $\chi(w)$ will be positive at all wage levels, while if F is strictly -1-concave then $\chi(w) > -2$.[3] The super-elasticity can be written $\dot{\phi}(w) = 1 - \phi(w)(1 + \chi(w))$. Hence, greater curvature leads to a more negative super-elasticity and, therefore, lower passthrough. An isoelastic F will exhibit super-elasticity of zero and, therefore, a curvature of $\frac{1-\phi}{\phi}$. For $\phi < 1$, F is concave and curvature is positive, while for $\phi = 1$, F is linear (i.e., uniform) and curvature is zero. For $\phi > 1$, F is convex and curvature is negative.

3.1.1 Distribution function redux

Bulow and Pfleiderer (1983) criticized early studies seeking to recover the elasticity of product demand from cost-price passthrough elasticities on the grounds that these exercises were sensitive to functional form assumptions. This concern is reflected in (8), which shows that recovering $\phi(w^*)$ from $\rho(w^*)$ requires prior knowledge of the super-elasticity $\dot{\phi}(w^*)$. The set of curvature-elasticity pairs implied by various product demand systems has now been extensively studied (Mrázová and Neary, 2017; Miravete et al., 2023). In an attempt to catch up with the monopolistic pricing literature, we now compute the passthrough elasticities implied by our example distributions.

Example (Power function). When $F(w) = \left(w/\bar{b}\right)^{\phi}$, we have $\dot{\phi}(w) = 0$. Therefore, for any value of $\phi > 0$, $\rho(w) = 1$.

The unitary passthrough elasticity delivered by the isoelastic labor supply model is highly restrictive and conflicts with the empirical evidence, which typically finds passthrough elasticities far below one (Card et al., 2018). One reason for this discrepancy is that researchers are rarely able to directly measure employer productivity, which may itself respond to the wage level, an issue we study carefully in Section 6. Supposing however that we were able to measure (and directly manipulate) p, it seems unlikely that the elasticity would happen to be one. The outside option distributions introduced in Section 2.4.1 that feature non-constant elasticities are capable of rationalizing departures from this benchmark.

Example (Shifted power function). When $F(w) \propto \left(w - \underline{b}\right)^{\beta}$, we have $\dot{\phi}(w) = 1 - \frac{w}{w-\underline{b}} < 0$. Consequently, $\rho(w) = 1 - \underline{b}/[(1+\beta)w] \in (0, 1)$, which is monotone increasing in w and asymptotes to one.

Example (Logit). When $F(w) = \left(1 + (w/\sigma)^{-\beta}\right)^{-1}$, we have $\dot{\phi}(w) = -\beta \frac{(w/\sigma)^{\beta}}{1+(w/\sigma)^{\beta}} < 0$. Hence, $\rho(w) = 1 - \frac{\beta}{1+\beta} \frac{(w/\sigma)^{\beta}}{1+(w/\sigma)^{\beta}} \in (0, 1)$, which is decreasing in w and asymptotes to $1/(1+\beta)$.

[3] The latter claim follows from noting that the derivative of the inverse wage function can be written $\varrho'(w) = 2 + \chi(w)$.

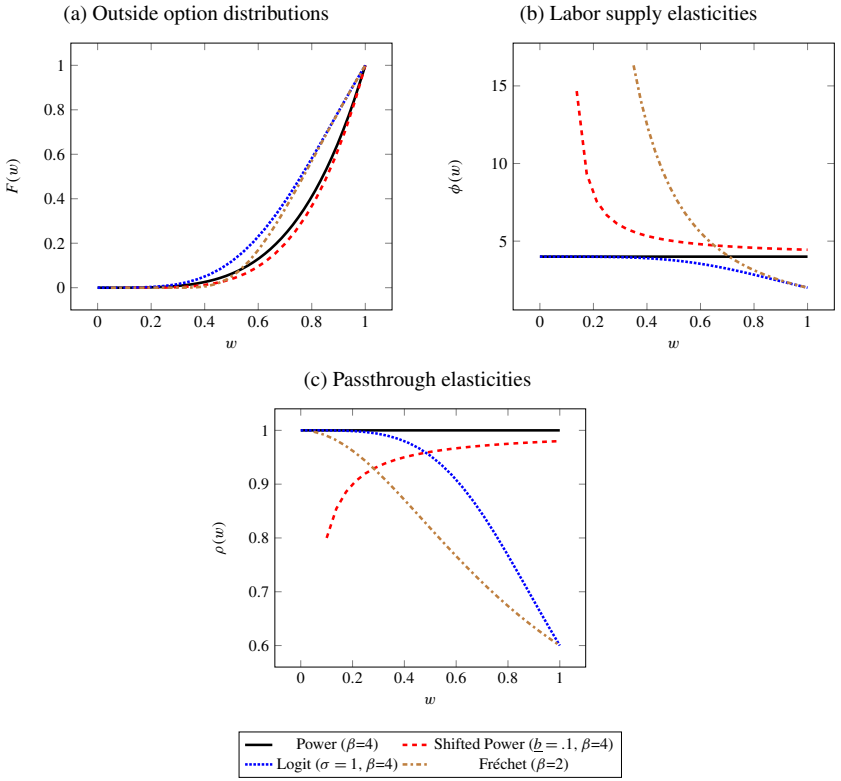

FIGURE 3 Four examples on the unit interval.

Example (Fréchet). When $F(w) = \exp\left(1 - \left(w/\bar{b}\right)^{-\beta}\right)$, we have $\dot{\phi}(w) = -\beta$ and $\rho(w) = \frac{1+\beta\left(w/\bar{b}\right)^{-\beta}}{1+\beta\left(w/\bar{b}\right)^{-\beta}+\beta}$. Passthrough will lie near one at very low wages but fall with the wage, asymptoting to $(1+\beta)^{-1}$.

A version of each of these four distributions is depicted in Fig. 3. Although many of the distribution functions themselves look similar, their passthrough and labor supply elasticities are often quite different. Fig. 4 plots the passthrough elasticity directly against the labor supply elasticity under each of these distributions. While the passthrough elasticity rises with the labor supply elasticity for the logit and Fréchet distributions, it is declining in the labor supply elasticity under the shifted power distribution.

3.1.2 *Can wage-setting power be identified from passthrough alone?*

As these examples make clear, any choice of F implies an elasticity function ϕ and a corresponding passthrough function ρ. It is natural to ask the converse

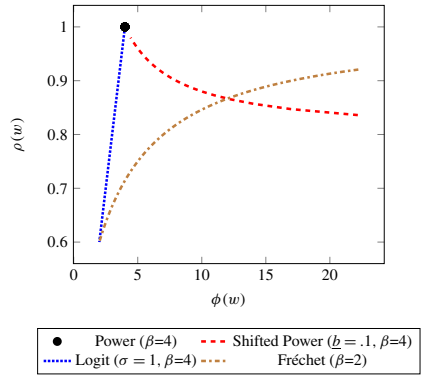

FIGURE 4 Passthrough versus labor supply elasticities.

question: if we have identified ρ can we recover ϕ? Rearranging (8) yields the following differential equation relating the wage elasticity of the exploitation index to passthrough

$$\frac{d\ln e(w)}{d\ln w} = 1 - 1/\rho(w).$$

The general solution to this equation takes the form

$$e(w) = Cw\exp\left(-\int_1^w \frac{1}{x\rho(x)}dx\right).$$

Evidently, the exploitation index is only identified up to an unknown multiplicative constant C. To recover C requires additional information about the value of $\phi(\cdot)$ at some point. In the absence of such prior information, the exploitation index, and consequently the wage markdown, are under-identified. This negative result generalizes Bulow and Pfleiderer (1983)'s pointwise intuition that exercises seeking to infer market power solely from passthrough elasticities are inherently sensitive to functional form.

3.1.3 IV estimation of labor supply elasticities

In settings where it is possible to identify the impact of productivity shocks on wages, it is typically also possible to identify the impact of those shocks on employment. The elasticity of employment with respect to productivity is

$$\frac{d\ln F(w^*)}{d\ln p} = \phi(w^*)\rho(w^*).$$

Hence, we can identify the labor supply elasticity $\phi(w^*)$ from the ratio of elasticities $\frac{d\ln F(w^*)}{d\ln p}/\rho(w^*)$, a result which lends itself naturally to instrumental variables (IV) methods. The recent literature follows variants of this approach,

using firm-specific productivity shocks such as patent grants (Kline et al., 2019), exchange rate fluctuations (Garin and Silvério, 2023), or procurement contracts (Kroft et al., 2020; Carvalho et al., 2023) to instrument wages in employment regressions. One typically finds larger labor supply elasticities over longer horizons, as firms require time to fully adjust to large shifts in productivity. Consequently, elasticities are often estimated over a horizon of 2 to 5 years, by which time adjustment has usually completed.

When F exhibits a non-constant elasticity, these linear IV regressions are subject to misspecification. If the instrument is binary and controls are saturated, IV recovers a weighted average supply elasticity over the range of variation in wages induced by the productivity shifter (Angrist et al., 2000). When the shock is extremely large, this weighted average may not give a good sense of the elasticity relevant for computing markdowns at the current wage. Some instruments yield very small changes in the wage that can plausibly be used to recover labor supply elasticities at particular wages. For instance, Dube et al. (2018) exploit variation in hourly wages stemming from bunching at round numbers, which they argue arises from employer optimization frictions. In principle, this approach can be used to estimate separate elasticities at each wage level where bunching occurs. Likewise, if one has access to a productivity shifter derived from a government policy (e.g. experience-rated taxes) featuring a kinked incentive schedule, the elasticity at a point can be recovered by applying the standard machinery for a "fuzzy" regression kink design (Card et al., 2015). Here, the estimated kink in employment would be divided by the first stage kink in wages to obtain an estimate of the labor supply elasticity to the firm at the going wage level. A disadvantage of regression kink designs is that they typically require very large sample sizes to precisely estimate elasticities of interest, which may explain why this approach has yet to be exploited in the empirical monopsony literature.

In many cases, researchers have access to multiple instruments. If one is willing to commit to a particular functional form for F, then its parameters can typically be estimated by generalized method of moments provided that the number of parameters is less than the number of instruments. When instruments have continuous support, the entire elasticity function $\phi(\cdot)$ can, in principle, be estimated (or bounded) via non-parametric instrumental variables methods (e.g., Newey and Powell, 2003; Blundell et al., 2007; Santos, 2012; Newey, 2013). Shift-share instruments of the sort entertained by Garin and Silvério (2023) and Mertens et al. (2022) are potentially good candidates for such methods because of the sizable range of near-continuous variation they typically capture. The state of the art in non-parametric IV estimation has evolved considerably in recent years and the latest approaches now provide separate guidance for choosing tuning parameters optimally based on whether the labor supply curve, the elasticity schedule, or some particular functional (e.g., the average elasticity over a range of wages) is of interest (Chen et al., 2024).

Studies utilizing matched employer-employee administrative data typically pool information from many firms. In the simple equilibrium search models of

Section 2.4.3, all firms faced the same outside option distribution F. However, in practice this distribution is likely to vary substantially across firms. Discrete choice models of workplace heterogeneity provide a straightforward way to model such heterogeneity in terms of worker and firm characteristics. For example, it is common to work with nested logit models that allow jobs in the same industry or geographic region to share similar outside option distributions (Lamadon et al., 2022; Azar and Marinescu, 2024). Additional flexibility can be introduced by allowing unobserved worker heterogeneity in the valuation of firm attributes (Roussille and Scuderi, 2023; Volpe, 2024), paralleling the practice in demand estimation of including random coefficients on product characteristics (Berry and Haile, 2021). Endogeneity can then be addressed via an instrumental variables regression of adjusted firm employment shares on wages, mirroring IV methods from industrial organization applied to product markets (Berry et al., 1995). However, standard discrete choice formulations of labor supply do not explicitly account for search frictions, which, as we saw in Section 2.4.3, may interact with worker preferences in complex ways. As richer data become available, a key research direction will be to document heterogeneity in outside option distributions under minimal modeling assumptions. In some cases, it may be possible to leverage changes in the wage policies of large firms (e.g., as in Derenoncourt and Weil, 2024) to estimate firm-specific outside option distributions non-parametrically.

A fundamental requirement of a valid wage instrument is that it should have no direct effect on firm size. A productivity shifter that also affects rival firms in the same market will tend to shift F, violating the exclusion restriction. To allay such concerns, it is common to report diagnostics based upon different market definitions demonstrating that the shock is truly idiosyncratic to the firm. For example, Kline et al. (2019) report that the intraclass correlation of their patent allowance instrument within 5-digit ZIP codes is statistically indistinguishable from zero. Likewise, Garin and Silvério (2023) demonstrate that their preferred shift-share measure of exposure to exchange rate fluctuations fails to predict wages or employment at other firms in the same industry and municipality. Consistent with theory, they find that wages and employment are more responsive to aggregate shocks than the idiosyncratic shock measure they utilize. A potentially fruitful direction for future work is to examine whether labor market definitions based upon worker flows (e.g., Manning and Petrongolo, 2017; Nimczik, 2017; Jarosch et al., 2024) yield similar conclusions about the excludability of popular firm-specific productivity shifters.

In settings where a productivity shifter is known to affect many firms in the same market, identification can often be achieved by modeling market level adjustments to the labor supply function F, an approach pursued by Berger et al. (2022) and Volpe (2024). As in other models of interference between units, accounting for market level adjustments ultimately requires specifying an "exposure mapping" (Manski, 2013; Aronow and Samii, 2017) that details how each firm in a labor market is affected by the aggregate shock. The structure

of this mapping varies across models, depending upon the presumed details of how choice sets are formed, the structure of worker preferences (e.g., whether and when the independence of irrelevant alternatives assumption holds), and assumptions about whether strategic interactions are present in wage setting. Exposure mappings are also necessarily contingent upon labor market definitions. Understanding which market definitions do a better job simultaneously capturing aggregate and firm-specific adjustments is an important task ahead for this literature.

3.2 Shifts in labor supply

Productivity shocks are not the only source of variation useful for identifying markdowns. Recall that our definition of outside options was net of differences in the non-wage amenity level of the firm versus each worker's best outside option. Call the firm's amenity level a. Without loss of generality, the firm's first-order condition can be rewritten:

$$f\left(w^*+a\right)\left(p-w^*\right)=F\left(w^*+a\right),$$

where, so far, we have implicitly normalized a to zero. What is the effect on wages of a small increase in the amenity level? Totally differentiating the first-order condition at the point $a=0$ yields:

$$f'\left(w^*\right)\left(p-w^*\right)\left(dw^*+da\right)-f\left(w^*\right)dw^*=f\left(w^*\right)\left(dw^*+da\right).$$

Rearranging this expression gives the comparative static

$$\frac{dw^*}{da}=-\frac{1+\chi\left(w^*\right)}{2+\chi\left(w^*\right)}. \tag{9}$$

This quantity measures a compensating differential: a dollar increase in the firm's amenities decreases wages by $\frac{1+\chi(w^*)}{2+\chi(w^*)}\times 100$ cents. Note however that if the firm were a price taker in the labor market, as assumed in perfectly competitive models of compensating differentials (Rosen, 1986), the response to a dollar improvement in amenities would necessarily be a dollar drop in the wage. This notion of perfectly equalizing differences has long permeated policy discussion of the incidence of mandated benefits (Summers, 1989; Gruber, 1994; Finkelstein et al., 2023).

With the monopsonist, differences are less than perfectly equalized. Remarkably, the size of the differential depends entirely on the curvature of F. For a uniform F, the compensating differential is 50 cents, while for an isoelastic F, the differential will be $1/\left(1+\phi\right)$ cents. One can, in principle, use a small shock to amenities then to identify the curvature $\chi\left(w^*\right)$, which in conjunction with the passthrough elasticity $\rho\left(w^*\right)$, allows recovery of both the labor supply elasticity $\phi\left(w^*\right)$ and its super-elasticity $\dot{\phi}\left(w^*\right)$.

Eq. (9) can also be used to study a location shift in the distribution of outside options: if all outside options improve by a dollar, it is as if the non-wage amenities of the firm have decreased by a dollar, which should lead to an increase in wages. Jäger et al. (2020) study the effect of an increase in the generosity of the Austrian unemployment insurance system, finding that a dollar increase in UI benefits led to only a 2.6 cent increase in wages after two years. At first glance, this finding might suggest a $\chi(w^*) \approx -1$. However, firms hire both from unemployment and by poaching workers from other firms, which suggests a mixture formulation of outside options

$$F = \iota F_u + (1 - \iota) F_e,$$

where $\iota \in (0, 1)$ is the share of potential workers that are currently unemployed, F_u is the outside option distribution of the unemployed, and F_e is the distribution of the already employed. One would expect the options of the employed to stochastically dominate those of the unemployed. In the case where $F_u(w^*) \approx 1$, nearly all unemployed workers are willing to work at the firm's going wage, suggesting that few such workers are marginal $F_u'(w^*) \approx 0$. If that is the case, we might also expect $F_u''(w^*) \approx 0$, which implies a change in the amenity value of unemployment will have trivial first-order effects on total employment and the wage.[4]

As with productivity shocks, additional identifying power is obtained when we have access to employment. The employment response to a small amenity increase is $f(w^*)\left(1 + \frac{d}{da}w^*\right)$. Hence, we can identify the labor supply elasticity from the restriction $\phi(w^*) = w^*\left[\frac{d}{da}\ln F(w^*)\right] / \left(1 + \frac{d}{da}w^*\right)$. In principle, information on supply shocks can be used in conjunction with productivity shocks to test the monopsony model as both sorts of instruments should identify same the labor supply elasticity. With variable elasticities, a nonparametric test would involve estimating the elasticity schedule $\phi(\cdot)$ separately using supply side and demand side shocks and evaluating whether differences in the estimated functions can be attributed to sampling error.

4 Wage discrimination and sorting

Thus far, we have assumed that every firm offers a unique wage. This section extends the basic monopsony model by allowing the firm to post different wages for different observable worker types. We then discuss the ability of such models to explain worker-firm sorting.

[4] Another potential explanation for the muted wage response documented by Jäger et al. (2020) is that worker productivity may have fallen in response to the increased UI generosity. Lusher et al. (2022) provide evidence from scanner data that increases in UI generosity led to increases in shirking among supermarket cashiers, particularly those with high experience and low productivity. Ahammer et al. (2023) find in Austrian administrative data that increases in UI generosity lead to increases in worker absenteeism.

It is useful to contrast the profits $\Pi(w^*)$ captured by the monopsonist with those that could be achieved by an employer with knowledge of workers' outside options. An employer who can tailor wage offers to match each worker's outside option will hire every worker with $b \leq p$, yielding profits

$$\begin{aligned}\int_{\underline{b}}^{p}(p-w)\,dF(w) &= \int_{\underline{b}}^{w^*}(p-w)\,dF(w) + \int_{w^*}^{p}(p-w)\,dF(w)\\ &= \Pi(w^*) + \underbrace{\int_{\underline{b}}^{w^*}(w^*-w)\,dF(w)}_{\text{extra profits on the inframarginal}}\\ &\quad + \underbrace{\int_{w^*}^{p}(p-w)\,dF(w)}_{\text{profits on extra hires}}.\end{aligned}$$

First-degree wage discrimination yields greater profits than monopsony both by paying lower wages to the workers that the monopsonist would have hired and by making profits on additional workers that the monopsonist would not have hired. Like a perfectly competitive firm, the wage discriminator pays the last worker hired their marginal product. Hence, the perfectly discriminating employer should not perceive itself to be facing labor shortages.

The monopsony model and the first-degree wage discrimination model make polar opposite assumptions about the information presumed available to employers. A similar dichotomy exists in the search literature: Burdett and Mortensen (1998) assume firms cannot tailor wages to workers' outside options, whereas sequential auction models of the sort pioneered by Postel-Vinay and Robin (2002b) allow employers to perfectly wage discriminate. Situated between these extremes are models of third-degree wage discrimination, which assume firms observe some—but not all—aspects of worker outside options. An early contribution in this direction was provided by Van den Berg and Ridder (1998), who considered distinct wage-posting economies differentiated by workers' observable characteristics.

Introducing heterogeneity in both worker productivity and outside options not only enhances realism but also provides an opportunity to examine wage disparities across demographic groups. For example, a substantial empirical literature, spawned by the seminal work of Robinson (1933), investigates the extent to which the gender pay gap can be attributed to differences in the distribution of outside options (Manning and Saidi, 2010; Le Barbanchon et al., 2021; Rong, 2022; Roussille and Scuderi, 2023; Sharma, 2023; Caldwell and Danieli, 2024). Models of third-degree wage discrimination also set the stage for studying worker-firm sorting, a topic on which we will focus below.

4.1 Wage types

Suppose workers are differentiated by a finite number $|\mathcal{T}|$ of observable types. These observable types can be thought of as distinct jobs posted by the firm, each with its own wage and task requirements that serve to attract different sorts of workers. Each type $t \in \mathcal{T}$ may exhibit a different skill level θ_t and distribution F_t of outside options, both of which are known to the firm. The literature has considered many different specifications of how worker and firm types jointly produce output, in some cases involving task assignment problems within the firm (Haanwinckel, 2023). In the interest of conveying the core ideas with minimal overhead, we will confine ourselves to models of production defined directly in terms of worker skills and firm productivity.

Suppose the revenue productivity of a match between a worker of type t and a firm with productivity level p is $p\theta_t$. If production is additive across types, then the firm's profit function can be written

$$\Pi\left(w_1, \ldots, w_{|\mathcal{T}|}\right) = \sum_{t \in \mathcal{T}} F_t\left(w_t\right)\left(p\theta_t - w_t\right).$$

Evidently, the firm's decision problem separates into type-specific sub-problems exhibiting optimums of the form found in (2). In particular, the optimal wage for a worker of type t is $w_t^* = \frac{\phi_t(w_t^*)}{1+\phi_t(w_t^*)} p\theta_t$.

In the special case where each type's outside options follow a power distribution $\phi_t(w) = \phi_t$, we arrive at a log-linear wage equation

$$\ln w_t^* = \ln \frac{\phi_t}{1+\phi_t} + \ln \theta_t + \ln p. \tag{10}$$

This log-additive representation yields a clean separation between the influence of worker features (ϕ_t, θ_t) and firm productivity (p) on the wage. Card et al. (2018) discuss the implications of such a representation for the literature on firm wage effects, connecting variation in statistical firm effects to $\ln p$ and variation in statistical person effects to $\ln \frac{\phi_t}{1+\phi_t} + \ln \theta_t$.

4.2 Three paths to sorting

To study the implications of this wage structure for sorting, consider two worker types: s and t. Sorting can be measured by the employment ratio $F_s\left(w_s\right)/F_t\left(w_t\right)$. From (10), it follows that

$$\begin{aligned}\frac{d \ln\left(F_s\left(w_s^*\right)/F_t\left(w_t^*\right)\right)}{d \ln p} &= \frac{d \ln F_s\left(w_s^*\right)}{d \ln w_s}\frac{d \ln w_s^*}{d \ln p} - \frac{d \ln F_t\left(w_t^*\right)}{d \ln w_t}\frac{d \ln w_t^*}{d \ln p} \\ &= \phi_s - \phi_t.\end{aligned}$$

As the firm's productivity increases, the type with the higher supply elasticity comes to occupy a larger share of employment. If more productive types have

higher elasticities (i.e., if $\phi_s > \phi_t \iff \theta_s > \theta_t$) then more productive firms will tend to have more skilled workers. While this mechanism is plausibly at play in markets where skilled workers have a wide range of job opportunities, it is not obvious that observed skill types and labor supply elasticities should always be positively related. For instance, less skilled workers might exhibit large elasticities because their outside option tends to be a minimum wage job. Conversely, some highly skilled professionals, such as brain surgeons or orchestral conductors, might only have a few possible employers, leading to a low elasticity.

A second way to rationalize sorting is in terms of firm amenities. Suppose that more productive firms have better amenities and higher skilled workers place greater value on those amenities. The logic of this argument is easiest to illustrate with a shifted power specification $F_t(w_t) = F(w_t + v_t a) = (w_t + v_t a)^{\phi}$, where v_t is type t's valuation of the firm's amenity level $a \geq 0$. Under this specification, optimal wages take the form

$$w_t^* = -\frac{1}{1+\phi} v_t a + \frac{\phi}{1+\phi} p\theta_t,$$

which implies that $F_t\left(w_t^*\right) = F\left(\frac{\phi}{1+\phi}\left(v_t a + p\theta_t\right)\right)$. Differentiating reveals that

$$\frac{d}{d\ln a} \ln\left(F_s\left(w_s^*\right)/F_t\left(w_t^*\right)\right) = ap\phi \frac{v_s\theta_t - v_t\theta_s}{\left(v_s a + p\theta_s\right)\left(v_t a + p\theta_t\right)},$$

while

$$\frac{d}{d\ln p} \ln\left(F_s\left(w_s^*\right)/F_t\left(w_t^*\right)\right) = -ap\phi \frac{v_s\theta_t - v_t\theta_s}{\left(v_s a + p\theta_s\right)\left(v_t a + p\theta_t\right)}.$$

If amenity valuations scale greater than proportionately with worker skill type ($v_s/v_t > \theta_s/\theta_t$) then improving the firm's amenity level will improve its skill mix. In contrast, boosting the productivity level of the firm will lead to downskilling. Hence, even if amenities and productivity are positively correlated, the cross-sectional relationship between productivity and skill is ambiguous and potentially non-monotone. However, if the cross-sectional relationship between firm amenities and productivity exhibits an elasticity everywhere above one, then more productive firms will employ more skilled workers.

While many researchers have found that higher wage firms tend to offer better amenities (Sockin, 2022; Lamadon et al., 2022; Caldwell et al., 2024b), it is not entirely clear that skilled workers are willing to pay more for these amenities than their less skilled peers. On one hand, skilled workers command greater earnings, which suggests they should gravitate towards firms offering amenities that are luxuries. On the other hand, skilled workers may value a different set of amenities (e.g., flexibility and growth potential vs air conditioning and lunch breaks) than less skilled workers. Indeed, Roussille and Scuderi (2023) find that unidimensional representations of workplace amenities provide a poor approximation even to the preferences of highly skilled software engineers. It is not yet

clear from this literature how different sorts of amenities scale with firm productivity, which renders sorting explanations predicated on skill differences in willingness to pay for amenities somewhat tentative.

Another way to generate assortative matching is to postulate strong complementarity between worker and firm types. Suppose, for example, that a match between a type-t worker and a firm of productivity p yields revenue productivity p^{θ_t}.[5] This supermodular technology yields a wage equation that is not log-additive:

$$\ln w_t^* = \ln \frac{\phi_t}{1+\phi_t} + \theta_t \ln p.$$

A specification of this form involving an interaction $\theta_t \ln p$ between worker and firm productivity was considered empirically by Lamadon et al. (2022). This wage equation implies

$$\frac{d \ln \left(F_s\left(w_s^*\right)/F_t\left(w_t^*\right)\right)}{d \ln p} = \theta_s \phi_s - \theta_t \phi_t.$$

Hence, even if all types have the same supply elasticity, more productive firms will accrue a larger share of higher skilled workers. While highly skilled jobs exhibiting "superstar effects" (Rosen, 1981) may be characterized by supermodular production technology, it is unclear whether this sort of complementarity is the primary force driving sorting behavior in less skilled sectors. In fact, recent estimates, including those of Lamadon et al. (2022), find that log wages are nearly additive in worker and firm types (Kline, 2024). This observation raises the question of whether plausible economic forces can simultaneously generate sorting and a log-linear wage equation.

4.3 Wages as a screening device

Thus far we have assumed that worker skills vary between but not within observable types. We now relax this assumption by allowing θ to be continuously distributed. A plausible explanation for sorting that has not received much attention in the literature is that worker productivity and outside options covary even within observable types. Suppose that the firm knows each type's joint distribution of worker productivity θ and outside options b but is unable to condition wages on these objects. Let $\hat{\theta}_t(w) = \mathbb{E}_t[\theta | b < w]$ denote the expected productivity of type-t workers hired at wage w and define $\tau_t(w) = \frac{d \ln \hat{\theta}_t(w)}{d \ln w}$ as the wage elasticity of that expectation. It stands to reason that skilled workers will have better outside options, implying $\tau_t(w) > 0$.

[5] Specifications of this form imply any productivity improvements at a firm will be skill-biased. While this assumption is worth entertaining when considering productivity changes stemming from technological innovations, it seems less appropriate for studying changes in revenue productivity driven by shifts in product demand. Lindner et al. (2022) provide evidence that firm-specific technological innovations tend to raise the relative wages of better-educated employees.

With these definitions, the firm's objective is to maximize

$$\Pi\left(w_1,\ldots,w_{|\mathcal{T}|}\right)=\sum_{t\in\mathcal{T}}F_t\left(w_t\right)\left(p\hat{\theta}_t\left(w_t\right)-w_t\right).$$

When an interior solution exists, the optimal wage takes the form

$$w_t^*=\frac{\tau_t\left(w_t^*\right)+\phi_t\left(w_t^*\right)}{1+\phi_t\left(w_t^*\right)}p\hat{\theta}_t\left(w_t^*\right).$$

The presence of $\tau_t\left(w_t^*\right)$ in the numerator of what would ordinarily be the exploitation index reflects the value of wages as a screening device. When $\tau_t\left(w^*\right)>0$, wages are marked down less than in (2) to ensure that the firm obtains the desired level of average worker quality for each observable worker type.

Consider now the isoelastic case where $\phi_t(w)=\phi_t$ and $\tau_t(w)=\tau\in(0,1)$. The latter assumption implies that $\hat{\theta}_t(w)=\tilde{\theta}_t w^\tau$, where $\tilde{\theta}_t$ represents the expected skill level that arises among workers of type t when the offered wage is one. This representation yields a log-additive wage equation:

$$\begin{aligned}\ln w_t^*&=\underbrace{\ln\hat{\theta}_t\left(w_t^*\right)}_{\text{skill}}+\underbrace{\ln\left(\frac{\tau+\phi_t}{1+\phi_t}\right)}_{\text{markdown}}+\underbrace{\ln p}_{\text{productivity}}\\&=\frac{1}{1-\tau}\ln\tilde{\theta}_t+\frac{1}{1-\tau}\ln\left(\frac{\tau+\phi_t}{1+\phi_t}\right)+\frac{1}{1-\tau}\ln p.\end{aligned}$$

As usual, wages are increasing in firm productivity p, implying that (all else equal) firms with higher productivity employ workers with greater outside options. However, because $\tau>0$, workers with better outside options are also more skilled.

The prediction that applicant quality is increasing in the offered wage has been empirically corroborated both experimentally in public sector jobs (Dal Bó et al., 2013) and observationally in online labor markets (Marinescu and Wolthoff, 2020; Escudero et al., 2024). Taking the view that the types correspond to jobs with different task requirements, within type productivity heterogeneity could plausibly be just as important as within type heterogeneity in outside options. Estimates comparing the magnitude of these two dimensions of heterogeneity would be a valuable addition to the literature.

5 Bargaining with incomplete information

There is a long running debate in the empirical literature over the verisimilitude of models involving bargaining versus wage posting (Caldwell and Harmon, 2019; Lachowska et al., 2022; Caldwell et al., 2024a; Townsend and Allan,

2024). How can monopsony provide a suitable model of labor markets if we know that some workers do, in fact, bargain over wages (Hall and Krueger, 2012; Brenzel et al., 2014; Faberman et al., 2022; Caldwell et al., 2024a)?

This section demonstrates that monopsony-like behavior can emerge even when bargaining is present, provided that we retain the assumption that workers' outside options are private information. This point is illustrated first in a simple model where the firm posts a maximum wage and workers decide whether to join the firm taking that maximum as given. We then discuss a richer model in which workers also post a minimum acceptable wage and bargaining yields a set of negotiated wages lying between each worker's minimum and the firm's maximum. When both sides play optimally, an increase in the firm's bargaining strength raises profits, which may provide an explanation for why firms often prefer (when possible) to commit to posted wages. Finally, we discuss some implications of the bargaining paradigm for empirical monopsony research.

5.1 Posting maximum wages

Suppose that, instead of committing to a posted wage, the firm announces a maximum wage $\bar{w}$. If that maximum wage exceeds the worker's private outside option b, the worker and firm negotiate a final wage $w(b) = \omega\bar{w} + (1-\omega)b$ and the worker is hired. Equivalently, the final negotiated wage can be interpreted probabilistically, with wage $\bar{w}$ resulting with probability ω, and wage b with probability $1-\omega$. In either case, this wage-setting protocol presumes that once the firm "meets" the worker, their value of b is revealed and used to determine a final wage. Survey evidence from Caldwell et al. (2024a) corroborates this timing assumption, finding that worker-firm interactions typically begin with the worker providing a salary expectation at the beginning of the bargaining process.

It is customary to refer to the parameter $\omega \in [0,1]$ as the firm's bargaining weight because it governs how closely the final wage aligns with the firm's chosen maximum wage $\bar{w}$. When $\omega = 1$, the firm unilaterally dictates a common wage, as in the basic monopsony model. Conversely, if each worker could freely choose their outside option b, then $\omega = 0$ would correspond to a setting in which workers dictate their final wage to the firm. We postpone discussion of models where workers can set wage demands to Section 5.2, continuing here with our maintained assumption that outside options are exogenous. Under exogenous b, when $\omega = 0$, the firm effectively engages in first-degree wage discrimination, tailoring wages perfectly to worker outside options. Thus, the term $1-\omega$ measures the extent of ex-post wage discrimination by the firm rather than the bargaining strength of workers. Nonetheless, to maintain consistent terminology, we refer to ω as the firm's bargaining weight throughout.

When deciding the maximum wage, the firm doesn't know b, only that $b \sim F$. From the firm's perspective, this setting is equivalent to an auction, where the probability of "winning" the auction is $F(\bar{w})$ and the expected value of the

prize is $p - \omega\bar{w} - (1-\omega)\,\mathbb{E}[b|b < \bar{w}]$. Hence, the firm's expected profits can be written

$$\Pi(\bar{w};\omega) = F(\bar{w})(p - \omega\bar{w}) - (1-\omega)\int_{\underline{b}}^{\bar{w}} b\,dF(b).$$

The necessary first-order condition for the maximum wage is:

$$f(\bar{w})(p - \bar{w}) = \omega F(\bar{w}).$$

When $\omega = 1$ this expression is identical to the first-order condition (1) characterizing the monopsony wage. As before, log-concavity of F assures a unique solution $\bar{w}^*$ to this equation. Rearranging, we obtain an expression analogous to (2) for the optimal maximum wage

$$\bar{w}^* = \frac{\phi(\bar{w}^*)/\omega}{1 + \phi(\bar{w}^*)/\omega} p. \tag{11}$$

This expression for the maximum wage mirrors the wage-setting choice of a monopsonist with perceived labor supply elasticity schedule $\phi(w)/\omega$. As the firm's bargaining weight ω grows large, the maximum wage falls. For any $\omega < 1$, the maximum wage is set higher than the monopsony wage, which raises efficiency by increasing employment. As $\omega \to 0$, hiring becomes fully efficient, with all workers possessing an outside option $b \leq p$ being hired.

Although the maximum wage is set in a monopsonistic fashion that trades off size against expected per worker profit, the ex-post wage of each worker depends on their outside option, which generates wage dispersion within the firm. Ex-post wages within the firm are distributed on the interval $\left[\omega\bar{w}^* + (1-\omega)\underline{b}, \bar{w}^*\right]$ with distribution function $w \mapsto F\left(\frac{w - \omega\bar{w}^*}{1-\omega}\right)/F(\bar{w}^*)$. As ω approaches one, within-firm wage dispersion collapses.

The mean wage can be written

$$\mu = \int_{\omega\bar{w}^* + (1-\omega)\underline{b}}^{\bar{w}^*} \frac{w}{1-\omega} f\left(\frac{w - \omega\bar{w}^*}{1-\omega}\right)/F(\bar{w}^*)\,dw.$$

In the special case where F follows a power function distribution, (11) implies that $\bar{w}^* = \frac{\phi}{\phi+\omega} p$, which when plugged into the equation above yields $\mu = \frac{\phi}{1+\phi} p$. Consequently, an increase in ω will reduce the maximum wage but boost the lowest wage $\omega\bar{w}^*$, yielding no effect on the mean wage. As this example illustrates, boosting a firm's bargaining weight ω ought to reduce wage dispersion but the effects on mean wages are ambiguous. To date, surprisingly little evidence exists on how plausibly exogenous changes to firm bargaining power influence within-firm wage dispersion, with most of the literature studying impacts on mean wages. Consistent with the power function example, Jäger

et al. (2021) find no effect on average wages of putting worker representatives on company boards.

The empirical monopsony literature often reports labor supply elasticity estimates derived from instrumenting average wages in a firm size regression (Sokolova and Sorensen, 2021). However, in the present model, only maximum wages are allocative. Somewhat miraculously, in the special case where F takes the power function form, using average instead of maximum wages does not compromise identification because $d\ln\mu = d\ln\bar{w}^* = d\ln p$, which implies that $\phi(\bar{w}^*) = d\ln F(\bar{w}^*)/d\ln\mu$. When F has a variable elasticity, however, instrumenting average wages may under- or over-estimate the elasticity $\phi(\bar{w}^*)$ governing the wage markdown. One approach to circumventing such biases comes from data on job advertisements, which sometimes report wage bands (Batra et al., 2023; Hazell et al., 2023). The upper limits of these bands suggest a wage ceiling, arguably proxying directly for $\bar{w}^*$. A testable implication of this model is that, conditional on $\bar{w}^*$, labor supply (i.e., the flow of job applications) should be insensitive to variation in the ex-post realized wages at a firm.

Returning to the case of a general distribution F, the expected profits of the monopsonist are

$$\Pi\left(\bar{w}^*;\omega\right) = F\left(\bar{w}^*\right)\left(p - \omega\bar{w}^*\right) - (1-\omega)\int_{\underline{b}}^{\bar{w}^*} b\,dF(b).$$

Though increases in ω raise profits per worker, they also reduce the number of workers hired. To compute the net effect on the firm's profits we apply the envelope theorem:

$$\frac{d}{d\omega}\Pi\left(\bar{w}^*;\omega\right) = \int_{\underline{b}}^{\bar{w}^*} b\,dF(b) - F\left(\bar{w}^*\right)\bar{w}^* = -\int_{\underline{b}}^{\bar{w}^*} F(b)\,db < 0,$$

where the second equality follows from integrating by parts. Hence, an increase in the firm's bargaining weight lowers its total expected profits.

It may appear surprising that posting the monopsony wage is less profitable ex-ante than bargaining with $\omega < 1$. Recall, however, that when $\omega = 0$ each worker is paid their outside option. That is, the firm becomes a first-degree wage discriminator, which we saw earlier is more profitable than monopsony wage posting. Thus, in this simplistic model, ω is best thought of as parametrizing ex-post wage setting conduct falling between the first-degree wage discrimination and wage posting benchmarks, a distinction that we argued earlier hinges on the information structure of the market. When $\omega \approx 1$, the firm cannot discriminate between workers with different outside options, while when $\omega \approx 0$ the firm effectively observes worker outside options. This interpretation may explain why wage posting seems to be common in less-skilled jobs (Caldwell and Harmon, 2019; Lachowska et al., 2022; Caldwell et al., 2024a), a setting where firms plausibly have less information (or less incentive to acquire information) about worker outside options.

5.2 The double auction model

The above model treated firms and workers asymmetrically: while firms set maximum wages below productivity taking into account the likely response of workers, workers naively set a minimum acceptable wage equal to their outside option b. However, if workers anticipate that the firm will announce maximum wage $\bar{w}^*$, then they should commit to walking away from any final offer less than $\bar{w}^*$. Better still, if workers know p, they should commit to walking away from any wage below p. In principle, doing so should yield an equilibrium where the firm proposes $\bar{w} = p$ because the workers' wage demands have made the relevant F a step function at p. The empirical relevance of such strategic considerations is unclear. For workers to credibly make such commitments would seem to require some degree of centralized bargaining or institutionalized wage-setting norms. However, in settings where a small firm (e.g., a technology startup) meets a worker with highly specialized skills, commitments on both sides may be viewed as credible, a possibility we now explore.

To formalize this bilateral bargaining scenario, assume now that both the firm's productivity p and the worker's outside option b are private information. Specifically, suppose each firm's productivity p is drawn independently from a distribution G, while each worker's outside option b is drawn independently from a distribution F. Unlike in Section 5.1, both sides strategically commit to wage offers: each worker commits to a minimum acceptable wage $\underline{w}(b)$ and the firm commits to a maximum acceptable wage $\bar{w}(p)$. A match forms whenever $\underline{w}(b) < \bar{w}(p)$, yielding a final wage $w(p, b) = \omega\bar{w}(p) + (1-\omega)\underline{w}(b)$, where $\omega \in [0, 1]$. This problem is formally equivalent to the "double auction" model studied by Chatterjee and Samuelson (1983), where the firm is the buyer of labor and the worker is the seller. In this framework, ω captures the bargaining strength of the firm, while $1-\omega$ captures the bargaining strength of workers. When $\omega = 1$, the firm can dictate wages, whereas when $\omega = 0$, each worker dictates their final wage to the firm.

Restricting to monotone equilibria, in which the offer functions $\bar{w}(p)$ and $\underline{w}(b)$ are increasing, the results of Chatterjee and Samuelson (1983) and Satterthwaite and Williams (1989) establish that the optimal minimum and maximum wages must obey the following equations:

$$p - \bar{w}(p) = \omega \frac{\tilde{F}(\bar{w}(p))}{\tilde{f}(\bar{w}(p))}, \quad \underline{w}(b) - b = (1-\omega)\frac{1-\tilde{G}\left(\underline{w}(b)\right)}{\tilde{g}\left(\underline{w}(b)\right)},$$

where $\tilde{F}(w) = F\left(\underline{w}^{-1}(w)\right)$ and $\tilde{G}(w) = G\left(\bar{w}^{-1}(w)\right)$ give the distribution of offered minimum and maximum wages respectively, and $\bar{w}^{-1}(\cdot)$ and $\underline{w}^{-1}(\cdot)$ denote inverse offer functions. Note that the first condition, which determines the maximal wage $\bar{w}$, takes the same form as in the prior model, yielding a wage rule equivalent to (11) where the relevant labor supply elasticity is a feature of the equilibrium object $\tilde{F}$ rather than F. As the second condition reveals, these objects differ because the worker's bid $\underline{w}(b)$ will now be shaded above b. This

shading leads to a new sort of inefficiency, as workers may walk away from offers with $\bar{w}(p) > b$. This distortion arises because workers act as monopolists, deliberately restricting the probability of a trade to suboptimal levels in order to raise expected wages conditional on being hired.

Cullen and Pakzad-Hurson (2023) show that when F and $1 - G$ are both power function distributions, the offer functions take a simple piecewise linear form that involves workers shading their wage demands up unless b is very high and the firm shading its maximal wage down unless p is very low, where the meaning of very low and very high depends on the elasticity parameters governing the shape of F and G.[6] As the elasticity $d \ln F(w) / d \ln w$ grows large, implying worker outside options are nearly known to the firm, both $\bar{w}$ and $\underline{w}$ approach $\underline{b}$. Conversely, as the elasticity $d \ln (1 - G(w)) / d \ln (1 - w)$ grows large, implying p is nearly known, both $\bar{w}$ and $\underline{w}$ approach the upper support point of G. Consequently, both workers and firms have incentives to conceal or strategically obscure their private valuations (productivity for firms and outside options for workers), because clear revelation would weaken their bargaining positions, forcing concessions on wages or employment.

Contrary to our analysis in the previous subsection of the case where workers naively set $\underline{w} = b$, Cullen and Pakzad-Hurson (2023) show that the firm's expected profits are increasing in ω, implying that monopsonistic wage posting ($\omega = 1$) should be preferred by the firm to bargaining. They point out that when revelations about coworker wages prompt an additional round of bargaining, wage transparency serves to raise the bargaining power ω of the firm. Increased transparency, therefore, raises the firm's profitability by allowing it to more nearly approximate the take-it-or-leave-it monopsony wage. This insight raises the question of why firms are so resistant to measures intended to improve transparency. Cullen and Pakzad-Hurson (2023) argue that firms cannot credibly commit to transparency on their own as they will be tempted to shirk on self-imposed reporting obligations. They provide evidence that legislation mandating transparency suppresses wages, as predicted by their model. The general applicability of the double auction model to the problem of wage determination in modern labor markets is an empirical question, worthy of further investigation.

5.3 Implications of bargaining for empirical work

While the monopsony model implies firms offer high wages only in order to grow large, the bargaining models of this section allow ex-post wages to also reflect non-allocative rent sharing. This distinction has implications both for interpreting estimates of productivity-wage passthrough and IV approaches to estimating labor supply elasticities. To identify a labor supply elasticity, a valid

[6] Formulas for the degree of shading and the thresholds at which shading becomes zero can be found in equations 11 and 14, respectively, of Cullen and Pakzad-Hurson (2023).

instrument should influence firm size only through its effect on wages. However, if a productivity shock triggers renegotiation with workers, it may have direct effects on employment. Suppose, for example, that when a firm lands a lucrative government contract, workers are able to discern that p has increased. In the double auction model, this revelation may lead workers to increase their wage demands $\underline{w}(b)$. In extreme cases, employment could actually fall as workers now reject more offers. When worker shading responses of this nature are present, IV will tend to underestimate the elasticity $d\ln \tilde{F}(\bar{w}^*)/d\ln \bar{w}^*$ because the shock itself alters the effective supply curve $\tilde{F}$ faced by the firm.

Carvalho et al. (2023) provide evidence consistent with such an interpretation. They show that for a sample of large labor intensive firms in Brazil, winning procurement contracts from the government has large effects on wages but no discernible effect on employment. Rather than conclude that the elasticity of supply to the firm is zero, it seems reasonable to infer from these findings that the contract actually shifted the supply curve faced by the firm, violating the exclusion restriction. Quantifying these separate channels nonparametrically is difficult and will tend to require additional instruments and modeling assumptions (Kwon and Roth, 2024). Careful empirical work on the mechanisms mediating the propagation of productivity into wages remains a critically under-explored area of research.

Bargaining models of incomplete information can also potentially rationalize patterns of passthrough heterogeneity across different groups of workers. A common finding in the literature is that productivity-wage passthrough elasticities are higher for higher-wage workers. For instance, Kline et al. (2019) find larger impacts of winning a patent on the wages of workers in the top quartile of earnings and for officers of the company. Garin and Silvério (2023) estimate that exposure to exchange rate fluctuations yields passthrough to high-skilled blue collar workers roughly double that of less-skilled workers. Likewise, Carbonnier et al. (2022) observe that exposure to a corporate tax cut leads wages to rise only for skilled workers, while Lobel (2024) documents that a firm-specific tax reduction leads wages to rise only for workers in occupations in the highest quintile of the earnings distribution. Kennedy et al. (2022) find that firm-specific exposure to tax cuts associated with the U.S. Tax Cut and Jobs Act (TCJA) yields earnings impacts concentrated in the top 10% of the earnings distribution.

In the basic monopsony model, differences in the wage response to a common productivity shock can only be explained by differences in outside option distributions. While it is possible that higher earning and more skilled workers tend to exhibit higher labor supply elasticities, an alternative explanation is that different sorts of workers have different bargaining weights $1-\omega$. A not entirely implausible approximation might be that $\omega = 1$ ("wage posting") for most workers and $\omega \ll 1$ ("bargaining") for managers and executives. Survey evidence from Caldwell et al. (2024a) documents that firms are more likely to bargain with workers when recruiting for management positions. An additional explanation for this finding could be that workers in management occupations have

direct input over the wage setting process. That is, these workers may, to some extent, set their own pay, an idea central to classic "insider-outsider" models (Blanchard and Summers, 1986; Lindbeck and Snower, 1986, 2001) and consistent with evidence that executives often get rewarded for luck (Bertrand and Mullainathan, 2001). Credibly separating the influence of bargaining weights and outside options on wages is a central task for future research in this area (Caldwell and Harmon, 2019).

6 Endogenous productivity

While the basic model with fixed p is useful for developing intuition, it is highly unrealistic. In this section, we extend the basic monopsony model by allowing productivity to depend on wages. We then consider the passthrough of an exogenous shock to productivity on wages. In contrast to the analysis in Section 3, we will see that an isoelastic outside options distribution can deliver a passthrough elasticity below one. We then discuss the tendency of monopsony models to yield implausibly high profit margins and discuss potential reconciliations of this puzzle involving variable labor supply elasticities and adjustment costs. The section concludes with a discussion of some additional difficulties that adjustment costs create for interpreting passthrough evidence.

There are several reasons to expect a firm's labor productivity to vary with wages. First, the firm's production may exhibit increasing or decreasing returns to scale. Second, product markets are unlikely to be perfectly competitive, which suggests that when wages—and consequently employment—rise, output prices should fall. Importantly, this is a long run prediction, as output prices may be sticky. Third, higher wages can exert a direct effect on productivity by boosting morale or preventing shirking on the job, as in classic efficiency wage models (Shapiro and Stiglitz, 1984; Akerlof and Yellen, 1990), or they can attract more skilled workers as discussed earlier in Section 4.3. While the firm does not need to separate these considerations when determining the optimal wage, understanding the relative importance of these factors can help to interpret empirical evidence within the monopsony framework.

Let $p(w)$ denote the value-added (revenue minus the cost of goods sold) per worker achieved when the wage is set to w. Though we will refer to this quantity as productivity, it is important to keep in mind that $p(w)$ is a measure of average labor productivity, which will tend to differ from marginal labor productivity unless production is linear in employment and does not directly depend on wages. Since average labor productivity is usually easier to measure than marginal labor productivity, expressing optimal wages and employment in terms of this quantity can be convenient when studying identification.

The firm's profit function is

$$\Pi(w) = F(w)[p(w) - w].$$

We assume that $p(w)$ is twice continuously differentiable, yielding the necessary first-order condition for optimal wage-setting:

$$\frac{f(w)}{F(w)} = \frac{1 - p'(w)}{p(w) - w}.$$

It is natural here to restrict the function $w \mapsto p(w) - w$ to be strictly log-concave, which guarantees that the right hand side of the first-order condition is increasing. Note that log-concavity does not rule out that $p'(w) > 1$ over some range of w (as could happen if efficiency wage effects are particularly pronounced at certain wage thresholds). Rather, the assumption limits the global convexity of productivity in the wage, requiring that $p''(w) / \left[p'(w) - 1\right]^2 < [p(w) - w]^{-1}$ for all $w \in \left[\underline{b}, \bar{b}\right]$. Let $\pi(w) = wp'(w)/p(w)$ denote the wage elasticity of productivity. In the isoelastic case ($\pi(w) = \pi$), a sufficient condition for strict log-concavity to hold is $\pi \leq 1$, ensuring that wages cannot serve as a money pump for the firm.

Rearranging the first-order condition yields the following expression for the optimal wage w^* as a function of the productivity elasticity $\pi(w^*)$, the elasticity of labor supply $\phi(w^*)$, and average labor productivity $p(w^*)$:

$$w^* = \frac{\pi(w^*) + \phi(w^*)}{1 + \phi(w^*)} p\left(w^*\right). \tag{12}$$

A variant of this expression was encountered earlier in Section 4.3, where a positive productivity elasticity arose due to correlations between worker quality and outside options. Here we are entertaining a wider range of potential productivity shifters, which renders the sign of $\pi(w^*)$ ambiguous a priori. When $\pi(w^*) < 0$, as can arise with inelastic product demand or decreasing returns to scale production, the wage will be marked down further below productivity than would be expected based upon the labor supply elasticity alone. When $\pi(w^*) > 0$ the wage markdown shrinks because it is optimal for the firm to "pay for productivity." Finally, note that while $\pi(w)$ can exceed one at some wage levels, the optimal wage must exhibit $\pi(w^*) \leq 1$, as higher values would yield negative profits.

To think more carefully about the properties of $\pi(w^*)$, it is useful to introduce some additional assumptions. Suppose the firm produces output via a production function

$$Q(w) = zY(F(w), w),$$

where $z > 0$ is total factor productivity (TFP), which we will treat as exogenous. The function $Y(\cdot, \cdot)$ gives the efficiency units of labor produced by a given level of employment $F(w)$ and wage level w. Using subscripts to denote the partial derivatives of Y with respect to these inputs, we expect the marginal product of labor $zY_1(F(w), w) \equiv MPL(w)$ to be positive at all wage levels w. However, if wages can directly influence worker productivity, then the

marginal product of the wage $zY_2(F(w), w) \equiv MPW(w)$ will also tend to be positive. The increase in output achieved by a small increase in the wage level is $Q'(w) = MPL(w) F'(w) + MPW(w)$.

Suppose that output is sold at price $P(Q(w))$. Average productivity can now be written

$$p(w) = \frac{P(Q(w)) Q(w)}{F(w)} = P(zY(F(w), w)) \frac{zY(F(w), w)}{F(w)}. \tag{13}$$

In this formulation $w^*/p(w^*)$ gives the ratio of the firm's wage bill to total revenue, which will serve as "labor's share" in this simple model that neglects capital and recruiting costs. From (12), labor's share depends on both the elasticity of labor supply $\phi(w^*)$ and the productivity elasticity $\pi(w^*)$. If both these elasticities are constant, then labor's share will also be constant.

Differentiating $\ln p(w)$ with respect to $\ln w$ yields

$$\pi(w) = \frac{\varepsilon(w) - 1}{\varepsilon(w)} [\eta_F(w) \phi(w) + \eta_w(w)] - \phi(w), \tag{14}$$

where $\varepsilon(w) \equiv -P(Q(w)) / \left(Q(w) P'(Q(w))\right) > 1$ is the elasticity of demand, $\eta_F(w) \equiv zY_1(F(w), w) F(w) / Q(w)$ gives the returns to scale of production, and $\eta_w(w) \equiv zY_2(F(w), w) w / Q(w)$ measures the direct effect of wages on productivity. Plugging these definitions into (14) and rearranging reveals that $(\varepsilon(w^*) - 1)/\varepsilon(w^*)$ equals the ratio of the marginal cost of production $F(w^*)\left[1 + \phi(w^*)\right]/Q'(w^*)$ to output price $P(Q(w^*))$ as in textbook monopoly pricing models.

While (12) represented monopsony wages in terms of average labor productivity, plugging (14) into (12) yields a representation in terms of marginal revenue productivity:

$$w^* = \frac{F(w^*)}{F(w^*) - [1 - e(w^*)] MRPW(w^*)} e\left(w^*\right) MRPL\left(w^*\right), \tag{15}$$

where $MRPL(w^*) \equiv MPL(w^*) P(Q(w^*)) (\varepsilon(w^*) - 1)/\varepsilon(w^*)$ is the marginal revenue product of labor and $MRPW(w^*) \equiv MPW(w^*) P(Q(w^*)) \times (\varepsilon(w^*) - 1)/\varepsilon(w^*)$ is the marginal revenue product of the wage. When $MRPW(w^*) = 0$ we have the traditional result that the monopsony wage equals the exploitation index times the marginal revenue product of labor.[7] However, when the wage has direct effects on productivity, the usual formula no

[7] In this case, one can also write $w^* = e\left(w^*\right) \cdot \left(\varepsilon\left(w^*\right) - 1\right)/\varepsilon\left(w^*\right) \cdot P\left(Q\left(w^*\right)\right) MPL\left(w^*\right)$. This representation is sometimes used to describe the monopsony wage as governed either by a ratio of price markups to wage markdowns (Dobbelaere and Mairesse, 2013; Delabastita and Rubens, 2022) or in terms of a "double markdown" on the value of the marginal product of labor $P\left(w^*\right) MPL\left(w^*\right)$, with markdowns dictated by the exploitation index $e\left(w^*\right)$ and the ratio $\left(\varepsilon\left(w^*\right) - 1\right)/\varepsilon\left(w^*\right)$ of marginal cost to price (Kroft et al., 2020).

longer applies. In particular, when $MRPW(w^*) > 0$ a markup term premultiplies the usual exploitation index $e(w^*)$, bringing wages closer to the marginal revenue product of labor than would otherwise be the case. Note that wages approach $MRPL(w^*)$ as $e(w^*) \to 1$ regardless of the value of $MRPW(w^*)$, which reflects that the firm becomes a price taker when the elasticity of labor supply is infinite.

It is instructive to rearrange (15) in terms of the firm's total wage bill:

$$w^* F\left(w^*\right) = e\left(w^*\right) MRPL\left(w^*\right) F\left(w^*\right) + \left[1 - e\left(w^*\right)\right] MRPW\left(w^*\right) w^*.$$

Evidently, the wage bill is an exploitation-weighted average of two revenue components: one associated with labor's marginal product, the other with the productivity impact of the wage level. In the special case where outside options follow a power function distribution with parameter ϕ, then $e(w^*) = \phi/(1+\phi)$ and the weights become invariant to the wage level. As $\phi \to \infty$, the wage bill approaches its competitive level $MRPL(w^*) F(w^*)$. However, as $\phi \to 0$ the wage bill approaches $MRPW(w^*) w^*$, which is the level that maximizes profits when the firm's workforce is fixed. Equivalently, if one views the firm's wage level as a factor of production, $MRPW(w^*) w^*$ is the income the wage level would "earn" in a competitive market.

Inspection of (14) reveals that for $\pi(w^*)$ to be positive requires either strongly increasing returns to scale $\left(\eta_F(w^*) > \frac{\varepsilon(w^*)}{\varepsilon(w^*)-1}\right)$ or large direct effects of wages on productivity $\left(\eta_w(w^*) > \phi(w^*) \frac{\varepsilon(w^*)}{\varepsilon(w^*)-1}\right)$. It is common to work with specifications imposing constant returns to scale, in which case (14) simplifies to $[(\varepsilon(w) - 1)\eta_w(w) - \phi(w)]/\varepsilon(w)$. In this scenario, the sign of $\pi(w^*)$ hinges solely on the relative magnitude of the direct wage elasticity $\eta_w(w^*)$ and the labor supply elasticity $\phi(w^*)$. While sizable direct effects of wages on productivity have been documented in certain specialized settings (Cappelli and Chauvin, 1991; Emanuel and Harrington, 2020; Coviello et al., 2022; Ruffini, 2022), most studies assume that $\eta_w(w) = 0$, which implies $\pi(w^*) < 0$. Though a negative $\pi(w^*)$ is plausible, we will see that surprisingly large values of this elasticity are required to rationalize productivity passthrough elasticities of the magnitude typically encountered in the literature.

6.1 Productivity passthrough revisited

Defining passthrough in the present framework is complicated by the potential dependence of productivity on wages. Specifically, one must distinguish between the direct response of wages to exogenous productivity shocks and indirect feedback effects operating through subsequent changes in employment and output prices. To formally capture these direct and indirect effects, we first derive the proportional response $d \ln w^*$ of log wages to a small change $d \ln z$

in TFP. Let $\dot{\pi}(w) = w\pi'(w)/\pi(w)$ denote the super-elasticity of productivity. Totally differentiating (12) yields:

$$\frac{d\ln w^*}{d\ln z} = \frac{1 - 1/\varepsilon(w^*)}{1 - \pi(w^*) - \frac{\pi(w^*)}{\pi(w^*)+\phi(w^*)}\dot{\pi}(w^*) - \left(\frac{\phi(w^*)}{\pi(w^*)+\phi(w^*)} - \frac{\phi(w^*)}{1+\phi(w^*)}\right)\dot{\phi}(w^*)} \equiv \rho_z(w^*).$$

Evidently, a negative super-elasticity of either productivity or labor supply will serve to dampen the passthrough of a TFP shock into wages. In the isoelastic case where $\phi(w) = \phi$, $\varepsilon(w) = \varepsilon$, and $\pi(w) = \pi$, this expression simplifies to $\rho_z(w^*) = \frac{1-1/\varepsilon}{1-\pi}$. As mentioned earlier, it is common to assume π is negative. Thus, unlike in the case of exogenous productivity, where (8) yields a productivity passthrough elasticity of one when the labor supply elasticity is constant, here we expect $\rho_z(w^*) < 1$ in the isoelastic setting.

Unfortunately, researchers are rarely able to measure TFP directly. Instead, they typically rely on proxies involving firm value added and employment to translate plausibly exogenous firm-specific shocks into TFP equivalent units. A natural proxy for TFP is average productivity $p(w^*)$. However, a shift in TFP will change wages and output prices, which feed back into value added and employment. To capture these feedback effects explicitly, we totally differentiate (13) with respect to TFP, obtaining

$$\frac{d\ln p(w^*)}{d\ln z} = 1 - 1/\varepsilon(w^*) + \pi(w^*)\frac{d\ln w^*}{d\ln z}.$$

Thus, scaling the elasticity of wages with respect to TFP by the elasticity of average productivity with respect to TFP identifies the average productivity passthrough elasticity

$$\frac{d\ln w^*/dz}{d\ln p(w^*)/dz} = \left[1 - \frac{\pi(w^*)}{\pi(w^*)+\phi(w^*)}\dot{\pi}(w^*) - \left(\frac{\phi(w^*)}{\pi(w^*)+\phi(w^*)} - \frac{\phi(w^*)}{1+\phi(w^*)}\right)\dot{\phi}(w^*)\right]^{-1} \equiv \rho_p(w^*). \quad (16)$$

In the isoelastic case, this expression simplifies to $\rho_p(w^*) = 1$, a result that mirrors our earlier observation that labor's share must be constant in this setting. Many empirical studies of rent-sharing report IV estimates that scale the impact of a firm-specific shock on log wages by the impact of these shocks on log average labor productivity. As both Card et al. (2018) and Jäger et al. (2020) note, such studies typically find relatively modest passthrough elasticities $\rho_p(w^*)$ in the range 0.05-0.20, with a focal value being 0.1. To rationalize such findings in the monopsony framework requires either a non-constant labor supply elasticity

or a non-constant productivity elasticity, possibilities that we explore in more detail below.

Another approach that has been used is to treat total value added $V(w^*) = p(w^*) F(w^*)$ as a proxy for TFP. Scaling the elasticity of wages with respect to TFP by the elasticity of value added with respect to TFP identifies

$$\rho_V(w^*) \equiv \frac{d \ln w^* / d \ln z}{d \ln V(w^*) / d \ln z} = \frac{d \ln w^* / d \ln z}{d [\ln p(w^*) + \ln F(w^*)] / d \ln z}.$$

The response of log firm size to a small change in log TFP is

$$\frac{d \ln F(w^*)}{d \ln z} = \phi(w^*) \frac{d \ln w^*}{d \ln z}.$$

Therefore, the value added passthrough elasticity can be written

$$\rho_V(w^*) = \left[1 + \phi(w^*) - \frac{\pi(w^*)}{\pi(w^*) + \phi(w^*)} \dot{\pi}(w^*) - \left(\frac{\phi(w^*)}{\pi(w^*) + \phi(w^*)} - \frac{\phi(w^*)}{1 + \phi(w^*)} \right) \dot{\phi}(w^*) \right]^{-1}.$$

In the isoelastic case, this expression simplifies to $\rho_V(w^*) = \frac{1}{1+\phi}$. Exploiting this relationship, Lamadon et al. (2022) instrument changes in log value added in a log wage change regression, finding $\rho_V(w^*) \approx 0.7$. This estimate implies a labor supply elasticity of approximately 6.5, which is on the high end of those reviewed in Sokolova and Sorensen (2021). Subsequent work by Kroft et al. (2020) finds a labor supply elasticity closer to 4 using random variation in procurement auction winners as an external instrument.

Recall from (12) that labor's share in the isoelastic case equals $(\pi + \phi)/(1 + \phi)$. Suppose that $w^*/p(w^*)$ is 0.6, which is roughly the labor share reported in the national accounts in recent years (Karabarbounis and Neiman, 2014; Autor et al., 2020; Smith et al., 2022). If we choose $\phi = 4$, then $\pi = 3/5 \times 5 - 4 = -1$. Alternately, if we choose $\phi = 6.5$, then we arrive at $\pi = 3/5 \times 7.5 - 6.5 = -2$. As these examples illustrate, in an isoelastic model, the elasticity of average labor productivity with respect to wages will tend to be negative and fairly large if we work with plausible estimates of labor's share. To date, little direct evidence is available corroborating the prediction that wage increases yield declines in average productivity of this magnitude.

6.2 A profitability puzzle

Empirical monopsony models frequently imply implausibly large profit margins, an issue also highlighted by Bloesch et al. (2024). Three factors likely contribute to this tendency. One is that many monopsony models impose unrealistic functional form assumptions on the labor supply and productivity elasticity

schedules $\phi(\cdot)$ and $\pi(\cdot)$, often relying on specifications that restrict these elasticities to be constant. We consider the effects of relaxing the isoelastic labor supply assumption in Section 6.3. Second, many studies neglect the role of costly inputs other than labor. The addition of capital, materials, and energy can introduce additional costs that scale with labor and affect profit margins. Third, the literature typically ignores adjustment and recruiting costs. In Section 6.4, we show that introducing these factors introduces additional identification challenges.

To appreciate the restrictions placed on firm profitability by the present model, note from (12) that the firm's profit margin can be written

$$\frac{\Pi(w^*)}{p(w^*)F(w^*)} = 1 - \frac{w^*}{p(w^*)} = \frac{1-\pi(w^*)}{1+\phi(w^*)}. \tag{17}$$

The first equality reveals that the profit margin is simply one minus labor's share. Hence, in our example above, where we assumed a labor share of 0.6, the profit margin must be 40%. Notably, this 40% estimate dramatically exceeds aggregate measures of "pure profits" that account for the user cost of capital, which have been estimated to hover around 8% in recent years (Barkai, 2020). While the mechanical connection between labor's share and profitability found in (17) is a logical implication of the premise that labor is the only factor of production, it does not imply that all profits derive from labor market power. Illustrating this point, the second equality in (17) expresses the profit margin in terms of behavioral elasticities. From (14), a large price to cost ratio will yield a small $\pi(w^*)$, which serves to boost the profit margin.

It is tempting to exploit (17) for identification by treating labor's share as a moment to be matched using firm-level data on wages and value added. Unfortunately, estimating labor's share at the firm level is fraught with difficult measurement problems. One problem is that many forms of worker compensation are typically missing from administrative records, including the value of health insurance and other employer provided benefits, self-employment earnings, and labor earnings that are reclassified as business income for tax purposes (Karabarbounis and Neiman, 2014; Autor et al., 2020; Smith et al., 2022). Another problem is that firm value added is often overstated because the cost of goods sold is not reported. Moreover, in addition to ignoring the costs of capital, value added measures typically neglect the costs of recruiting workers, which may be sizable (Bloesch et al., 2024).

In practice, firm-level measures of labor's share often imply dramatically lower aggregate share estimates than those based on the national accounts. For example, Autor et al. (2020) find a labor share of only 25% in the 2012 Census of Manufacturing microdata when comparing total payroll to a relatively detailed measure of value added that accounts for the costs of goods sold. It seems unlikely that economic profits constitute 75% of value added in the manufacturing sector as a whole. Jumping to such a conclusion could lead to a dramatic overstatement of the rents captured by firm owners.

Some business accounting datasets report firm-wide profit margins that could, in principle, be used to avoid some of these difficulties. However, economists have long been wary of equating accounting profits with economic rents (Knight, 1921). In addition to failing to net out relevant opportunity costs, accounting measures often provide a poor measure of expected rents enjoyed by growing firms. For instance, firms sometimes report negative profits for many consecutive years, reflecting temporary losses or intervals without revenue. Moreover, an influential recent literature finds that a non-negligible share of corporate profits in the US and EU is hidden in tax havens (Guvenen et al., 2022; Fuest et al., 2022; Tørsløv et al., 2023).

In cases where the revenue productivity of individual workers can be measured directly (e.g., sales associates paid based on commission), the productivity elasticity $\pi(w^*)$ may be identified without relying on accounting conventions. In such a case, one can use (17) in conjunction with an estimated labor supply elasticity $\phi(w^*)$ to compute profit margins. Alternatively, if one has a credible estimate of the passthrough elasticity $\rho_p(w^*)$ and a separate estimate of the elasticity of labor supply, then it is possible to back out a productivity elasticity using (16). We now illustrate this approach in a setting featuring a non-constant labor supply elasticity.

6.3 A calibration with variable labor supply elasticity

Recall from our discussion of (16) that, in an isoelastic model, the average productivity passthrough elasticity $\rho_p(w^*)$ must equal one, a prediction that is at odds with the findings of many empirical studies (e.g., Kline et al., 2019; Garin and Silvério, 2023). We turn now to investigating whether a simple model with a variable labor supply elasticity can rationalize typical passthrough estimates while generating a plausible productivity elasticity $\pi(w^*)$ and profit margin $1 - w^*/p(w^*)$.

In what follows, we will maintain the assumption that $\dot{\pi}(w^*) = 0$ and follow Card et al. (2018) in considering the case where $\rho_p = 0.1$. Plugging these assumptions into (16) yields the restriction

$$\left(\frac{\phi(w^*)}{\pi + \phi(w^*)} - \frac{\phi(w^*)}{1 + \phi(w^*)}\right)\dot{\phi}(w^*) = -9. \tag{18}$$

For any super-elasticity of labor supply $\dot{\phi}(w^*) \neq 0$ and any choice of labor supply elasticity $\phi(w^*) > 0$, a unique π solves this equation. To build intuition about which sorts of super-elasticities are plausible, it is useful to work with a shifted power specification, which parameterizes the labor supply super-elasticity as $\dot{\phi}(w) = -\left(w/\underline{b} - 1\right)^{-1}$. Hence, any ratio $w^*/\underline{b} > 1$ of the monopsony wage to the outside option of the worker most eager to work at the firm maps to a negative super-elasticity. With this parametrization, it is straightforward to solve numerically for the wage elasticity of worker productivity π

and the firm's profit margin $[1-\pi]/\left[1+\phi\left(w^*\right)\right]$ as functions of $\phi\left(w^*\right)$ and $w^*/\underline{b}$.[8]

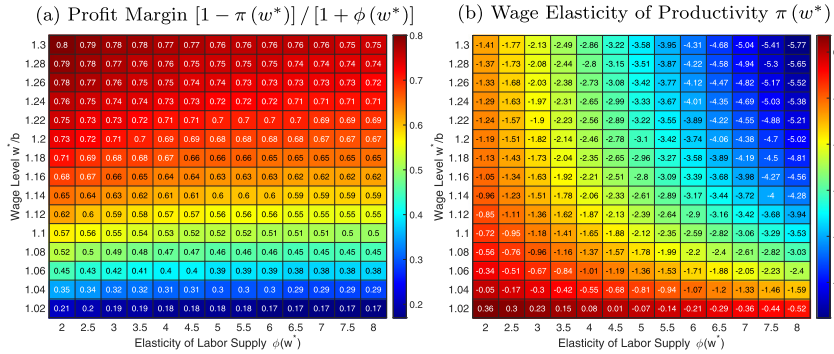

FIGURE 5 Rationalizing $\rho_p\left(w^*\right)=0.1$ with a shifted power distribution of outside options.

The heatmaps in Fig. 5 display the results of such an exercise where the solutions are plotted over a rectangular grid of $\phi\left(w^*\right)$ and $w^*/\underline{b}$ values. As illustrated in Panel (a), the parameters considered yield profit margins ranging from 17% to 80%. For instance, setting $w^*/\underline{b}=1.1$ (a 10% wage rent for the worker with outside option $\underline{b}$) and choosing $\phi\left(w^*\right)=6$ implies a super-elasticity of $\dot{\phi}\left(w^*\right)=-10$ and a profit margin of roughly 51%. These parameter choices yield $\pi=-2.59$, implying that average productivity is highly sensitive to wage levels.

Obtaining smaller profit margins and productivity elasticities requires wage levels very close to $\underline{b}$, which yields very large negative super-elasticities. For example, setting $w^*/\underline{b}=1.04$ and $\phi\left(w^*\right)=6$ yields a profit margin of 30% and a productivity elasticity $\pi=-1.07$. However, the superelasticity of labor supply implied by this configuration of parameters is $\dot{\phi}\left(w^*\right)=-25$.

Panel (b) of Fig. 5 reveals that a positive wage elasticity of worker productivity can emerge when the elasticity of labor supply is low and the monopsony wage is very close to $\underline{b}$. The large negative super-elasticity of labor supply implied by these parameter configurations dissipates passthrough, requiring a countervailing productivity effect to rationalize $\rho_p=0.1$. However, the wage levels generating this behavior are implausibly low, suggesting that workers extract trivial rents from the employment relationship, a finding inconsistent with experimental evidence on employment rents (e.g., Mas and Pallais, 2019).

In sum, although the shifted power specification of outside options allows us to rationalize commonly encountered values of ρ_p, plausible choices of $\phi\left(w^*\right)$ and $w^*/\underline{b}$ yield suspiciously high profit margins and large negative values of $\pi\left(w^*\right)$. Though the shifted power specification is just one of many functional

[8] Recall that in the shifted power distribution $\phi\left(w^*\right)=\beta\frac{w^*/\underline{b}}{w^*/\underline{b}-1}$. Hence, any choice of $\phi\left(w^*\right)$ and $w^*/\underline{b}$ amounts to a choice of $(\beta,\underline{b})$.

forms that can generate variable elasticities of labor supply, these tensions are generic: for any choice of outside option distribution F, extremely large negative super-elasticities of labor supply $\dot{\phi}(w^*)$ will be required to explain profit margins below 30% with plausible choices of $\phi(w^*)$. For example, if one imposes $\phi(w^*) = 6$, then rationalizing a profit margin of 30% via Eq. (18) requires $\dot{\phi}(w^*) = -25$, exactly the same value that was required under the shifted power specification.

A few caveats are in order here. First, this analysis assumed a constant productivity elasticity π. One can show that allowing a positive super-elasticity of productivity $\dot{\pi}(w^*)$ will tend to yield lower profit margins. Unfortunately, the current empirical literature has little to say about the likely sign of $\dot{\pi}(w^*)$, much less its magnitude. We have also ignored capital and other input costs, the introduction of which will tend to diminish profits. Finally, we have ignored adjustment costs associated with bringing workers to the firm, which can further dissipate firm profits. To conclude this section, we next turn to studying the basics of adjustment costs and discuss some additional challenges these costs can pose for identification.

6.4 Adjustment costs

The models discussed so far neglect the non-wage costs associated with bringing workers to the firm. Bloesch et al. (2024) argue that carefully accounting for these costs can produce estimates of profits attributable to labor market power that align more closely with the national accounts. There are at least two such costs that can be economically important. One is the cost of equipping or training a worker when they are first hired. Another is the cost involved in sourcing a candidate (e.g., by posting a vacancy or searching for referrals) and persuading them to join the firm (e.g., via a hiring bonus).

To appreciate the potential implications of accounting for the first sort of cost, suppose that firm profits are given by

$$\Pi(w) = F(w)[p(w) - w] - C(F(w)),$$

where $C(\cdot)$ is a hiring cost function that is increasing but may be concave or convex. In addition to mechanically dissipating profits, the introduction of hiring costs impacts wage setting behavior. The first-order condition for wages becomes

$$\frac{f(w)}{F(w)}\left[p(w) - w - C'(F(w))\right] = 1 - p'(w).$$

With a bit of algebra, the above expression can be rearranged into the following wage equation:

$$w^* = \frac{\pi(w^*) + \phi(w^*)}{1 + \phi(w^*)} p\left(w^*\right) - \frac{\phi(w^*)}{1 + \phi(w^*)} C'\left(F\left(w^*\right)\right)$$

$$= \frac{\pi(w^*)}{1+\phi(w^*)} p(w^*) + e(w^*)\left[p(w^*) - C'(F(w^*))\right].$$

The first line is the monopsony wage in (12) augmented with a term that is decreasing in the hiring cost of the marginal worker times the exploitation index $e(w^*)$. As the elasticity of supply to the firm approaches infinity, $w^* \approx p(w^*) - C'(F(w^*))$, reflecting that the hiring cost effectively lowers the marginal worker's net productivity. The second line provides a reinterpretation of this expression as a markdown of net productivity. The first term in the second line captures the portion of the contribution of the wage change to average productivity captured by the firm in higher profits. The second term marks wages down relative to average productivity net of marginal hiring costs. An upshot of this simple extension is that markdowns may be substantially overstated by comparing wages to average output per worker: the relevant benchmark should be net rather than gross productivity.[9]

Several quasi-experimental estimates of the ratio $C'(F(w^*))/w^*$ exist. Working with an extension of the above model in which firms offer incumbent workers a wage premium to encourage worker retention, Kline et al. (2019, 2021) find that the marginal hiring cost of a new recruit amounts to just over a year's worth of earnings in a sample of innovative small firms. Jäger and Heining (2022) obtain similar estimates when studying firm responses to worker death in a panel of small German firms using a dynamic extension of the model in Kline et al. (2019). Seegmiller (2023) also works with a dynamic extension of the Kline et al. (2019) model and finds that marginal hiring costs as a fraction of entry wages are largest among the least productive firms.

In contrast to these recent estimates based on models where firms offer wages below marginal revenue product, Bloom (2009) finds that rationalizing firm level responses to an aggregate measure of uncertainty shocks in a competitive model yields very low adjustment costs estimates, equivalent to about 2% of a year's worth of annual earnings. Kline (2008) finds similarly small estimates when rationalizing employment and wage responses of the oil and gas field services industry to oil price fluctuations with a competitive model. Dube et al. (2010) study the California Employment Survey and find that replacement costs average about 9% of annual earnings but rise with wage levels. The wide range of estimates provided here suggests there is substantial room to improve on the measurement of direct hiring costs.

Additional empirical difficulties arise when broader notions of recruiting cost are considered. Suppose that at wage w, a recruiting expenditure R attracts $F(w, R)$ workers, with $\partial F(w, R)/\partial R \geq 0$. This specification of F can potentially be microfounded by allowing recruiting effort to change worker consideration sets via the posting and advertising of vacancies or for worker

[9] Manning (2006) considers a dynamic model where the costs of hiring may also depend directly on w. When costs take the form $C(w, F(w))$ in the model above, another term of the form $\frac{w}{\phi(w)f(w)}\frac{\partial}{\partial w}C(w, F(w))$ must be deducted from gross productivity to arrive at net productivity.

reservation wages b to be influenced via recruiting events and signing bonuses. Since recruiting expenses boost output conditional on wages we will write average productivity as $p(w, R)$. If returns to scale are non-increasing and recruiting has no direct effect on worker quality or effort then it is reasonable to assume that $\partial p(w, R)/\partial R < 0$ because increases in output should lower product price. The firm's problem is to choose w and R to maximize

$$\Pi(w, R) = F(w, R)[p(w, R) - w] - R.$$

As with direct hiring costs, the introduction of recruiting effort mechanically dissipates firm profits but has nuanced implications for the proper measurement of wage markdowns. Additive separability of the recruiting cost ensures that optimal wages w^* obey a condition mirroring Eq. (12). Letting $\phi(w, R) = \frac{\partial}{\partial \ln w} \ln F(w, R)$ and $\pi(w, R) = \frac{\partial}{\partial \ln w} \ln p(w, R)$, optimal recruiting expenditures R^* and wages w^* solve the following pair of equations:

$$R^*/F\left(w^*, R^*\right) = \left[p\left(w^*, R^*\right) - w^*\right] \frac{\partial}{\partial \ln R} \ln F\left(w^*, R^*\right) + \frac{\partial}{\partial \ln R} p\left(w^*, R^*\right),$$
$$w^* = \frac{\pi(w^*, R^*) + \phi(w^*, R^*)}{1 + \phi(w^*, R^*)} p\left(w^*, R^*\right).$$

While the semblance of this wage equation to (12) may appear comforting, a closer look reveals that standard IV approaches will fail to identify the elasticities $\pi(w^*, R^*)$ and $\phi(w^*, R^*)$. The fundamental problem is one of excludability: exogenous productivity shifts $d \ln x$ are no longer valid instruments for wages because they also raise optimal recruiting expenditure. Consequently,

$$\frac{d}{d \ln x} \ln F\left(w^*, R^*\right) = \phi\left(w^*, R^*\right) \frac{d \ln w^*}{d \ln x} + \frac{\partial}{\partial \ln R} \ln F\left(w^*, R^*\right) \frac{d \ln R^*}{d \ln x}$$
$$> \phi\left(w^*, R^*\right) \frac{d \ln w^*}{d \ln x}.$$

In words, the ratio of the impact of a productivity increase $d \ln x$ on employment to its impact on wages will tend to overestimate the elasticity relevant for measuring the markdown. Likewise, assuming that $\pi(w^*, R^*) < 0$, standard IV approaches will overstate $|\pi(w^*, R^*)|$ because $\frac{d}{d \ln x} \ln p(w^*, R^*) / \frac{d w^*}{d \ln x} < \pi(w^*, R^*)$. The net result of these two overstatements on estimated markdowns is difficult to express analytically. However, if each elasticity is overstated by the same proportion, the markdown itself will be overstated provided that $\phi(w^*, R^*)$ is at least one, with larger values of that elasticity yielding greater overstatement.[10]

If R were capable of being measured directly, one could resolve these difficulties by instrumenting both wages and recruiting expenditure. Unfortunately,

[10] Suppose that each elasticity is multiplied by a constant $K > 1$. If $\phi\left(w^*, R^*\right) > 1$ then $\frac{K\pi\left(w^*, R^*\right) + K\phi\left(w^*, R^*\right)}{1 + K\phi\left(w^*, R^*\right)} = \frac{K}{1 + K\phi\left(w^*, R^*\right)} \frac{\pi\left(w^*, R^*\right) + \phi\left(w^*, R^*\right)}{1 + \phi\left(w^*, R^*\right)} < \frac{\pi\left(w^*, R^*\right) + \phi\left(w^*, R^*\right)}{1 + \phi\left(w^*, R^*\right)}$.

recruiting costs are notoriously difficult to measure, particularly at the level of individual firms. While data on vacancy posting and filling rates have sometimes been used to develop proxies for recruiting effort (e.g., Davis et al., 2012), Davis et al. (2013) estimate using data from the Job Openings and Labor Turnover Survey that more than one third of hires occur without a vacancy having been posted. Consequently, the dominant approach has been to work with highly structured models of $F(w, R)$, the parameters of which are identified jointly from hiring behavior and wages (e.g., Manning, 2006; Morchio and Moser, 2024; Bloesch et al., 2024). A useful advance for this literature would be the development of improved proxies for R based upon novel data sources, such as accounting measures of recruiting expenses combined with detailed information on employee time allocation.

A key economic implication of large adjustment costs of either sort is that firms have incentives to create long run relationships with workers. As workers stay with the firm their outside options evolve while their value to the firm likely increases as they learn on the job (Stevens, 1994). To support such relationships, the firm may post tenure-dependent wages that exhibit different markdowns and passthrough behavior. Kline et al. (2019) fail to reject in a sample of small innovative firms that the product market rents accompanying a patent grant are shared exclusively with incumbent workers. Likewise, Carbonnier et al. (2022), Garin and Silvério (2023), Seegmiller (2023), and Bıró et al. (2024) find greater passthrough of productivity shocks to incumbent workers than new hires. One interpretation of such patterns is that wage markdowns are smaller for incumbent workers than new hires. Similar predictions arise from agency models, which posit that wages are backloaded in order to stem moral hazard (e.g., Lazear, 1981; Burdett and Coles, 2003), and from models of employer learning, which predict that wages will drift closer to productivity as information about worker types is revealed (e.g., Baker et al., 1994; Kahn and Lange, 2014). A very different interpretation would be that the wage fluctuations of incumbent workers are simply more likely to reflect bargaining behavior, perhaps because the roles within the organization undertaken by more senior workers are substantively different from those of new hires. Understanding when and why firm wage setting for newly hired workers differs from those of more senior workers remains an important frontier in the literature.

7 Price passthrough of minimum wages

Robinson (1933)'s treatise noted that a carefully chosen minimum wage can increase the employment of a monopsonist by effectively inducing price taking behavior. This classic prediction received renewed interest in the wake of Card and Krueger (1994)'s landmark study of the response of fast food establishments to a hike in New Jersey's minimum wage. Exploiting store specific variation in exposure to the minimum wage, Card and Krueger (1994) found that the hike raised employment at affected establishments, which they suggested was at odds with the predictions of competitive labor market models.

The monopsony interpretation of Card and Krueger (1994)'s findings was almost immediately criticized on the grounds that the study also found evidence that fast food prices rise in response to minimum wages (Brown, 1995; Welch, 1995). The logic of this critique is straightforward: if fast food firms face a stable product demand curve, then greater employment (which presumably generates more fast food output) should lead prices to fall. That is, minimum wage hikes should yield *negative* passthrough to fast food prices. In contrast, the textbook competitive model predicts that minimum wage hikes generate employment losses, which in turn yield positive price passthrough. Reviewing these arguments, Brown (1999) concludes in an earlier volume of this Handbook that "the monopsony model will not replace the competitive diagram in the souls of labor economists."

Though Card and Krueger (1994)'s specifications estimating price passthrough from variation in store specific exposure were statistically insignificant, several modern studies utilizing higher powered research designs confirm that exposure to minimum wage hikes yield substantial increases in product prices (Harasztosi and Lindner, 2019; Renkin et al., 2022; Ashenfelter and Jurajda, 2022). In fact, the latter three studies are unable to reject *full* price passthrough, a common finding in the recent empirical literature (Dube and Lindner, 2024).[11] The strong passthrough of minimum wages to prices continues to be cited as evidence that labor markets are essentially competitive, with some authors even using the price passthrough to estimate employment losses (Aaronson and French, 2007). This challenge to the monopsony framework is sufficiently severe that recent papers studying minimum wages in monopsonistic environments often resort to assuming that product prices are fixed (e.g., Berger et al., 2025).[12]

This section reviews these arguments more carefully using the tools that have been developed so far. Ultimately, the tension with passthrough facts will be seen to lie not with the monopsonistic model of wage setting, but with textbook models of how prices are set. We will show that introducing heterogeneity in demand conditions, variable service quality, or frictions in price setting can generate positive responses of both employment and output prices to a minimum wage hike. Whether the employment and price impacts of minimum wages can be quantitatively rationalized in a monopsonistic framework is an important question for future research.

7.1 Mechanics of minimum wage hikes

In the interest of building up from microeconomic fundamentals, let us return to the problem of a single monopsonist faced with a stable outside option distribution F. Suppose that our monopsonist is subjected to a binding firm-specific

[11] Not all recent studies find full passthrough. Using an expanded version of the McDonald's data studied by Ashenfelter and Jurajda (2022), Wiltshire et al. (2025) estimate that only 55 cents of every dollar of minimum wage induced costs is passed on to consumers.

[12] An exception is Haanwinckel (2023), who specifies and estimates a general equilibrium model containing several other margins of adjustment to minimum wages.

minimum wage $\underline{w} \geq w^*$.[13] Bereft of the power to dictate wages, the firm acts as a (constrained) price taker, seeking a workforce of size $N \leq F\left(\underline{w}\right)$. To simplify the analysis, we shut down efficiency wage effects and decreasing returns to scale by assuming $Y(N, w) = N$. We will additionally assume a constant elasticity of product demand $\varepsilon > 1$, which implies the price of output can be written $P(N) = P_0 N^{-1/\varepsilon}$ for some $P_0 > 0$. Hence, prices and employment are presumed to obey an inverse relationship.

With these assumptions, the firm's profit function can be written $P(N)N - \underline{w}N$. For large enough $\underline{w}$, the optimal employment level N^* will satisfy the first-order condition $(1 - 1/\varepsilon)P(N^*) = \underline{w}$, which equates the marginal revenue product of a worker to the minimum wage. If the N^* that solves this equation is greater than $F\left(\underline{w}\right)$, then the firm hires all $F\left(\underline{w}\right)$ available workers.

Denote by w^{**} the quasi-competitive wage that solves the equation $(1 - 1/\varepsilon)P(F(w^{**})) = w^{**}$. Rearranging (15), the monopsony wage can be expressed as the solution to the equation $(1 - 1/\varepsilon)P(F(w^*))e(w^*) = w^*$. Contrasting these expressions reveals that $w^{**} > w^*$ so long as the labor supply elasticity is finite. At minimum wage levels below w^{**}, the firm will face a labor shortage, seeking more employees than are willing to work for the firm. This shortage is quelled at minimum wage level w^{**}, where the number of workers demanded just equals the number supplied. Above w^{**}, more workers are willing to work for the firm than it wishes to employ.

With these definitions, the optimal employment of the firm can be concisely expressed as the following piecewise function of the minimum wage level:

$$N^*\left(\underline{w}\right) = \begin{cases} F(w^*) & \text{if } \underline{w} < w^* \\ F\left(\underline{w}\right) & \text{if } \underline{w} \in \left[w^*, w^{**}\right] \\ F(w^{**})\left(\underline{w}/w^{**}\right)^{-\varepsilon} & \text{if } \underline{w} > w^{**}. \end{cases}$$

In the first range, the minimum wage does not bind and employment is set at the monopsonistic level. In the second range, the minimum wage binds and further hikes in the minimum raise employment by increasing the number of workers willing to work for the firm. In the third range, supply outstrips demand and employment decreases with wages isoelastically.

Leaving out the points where the elasticity is not defined, we can write the employment and passthrough elasticities to the minimum wage as piecewise functions

$$\frac{d \ln N^*\left(\underline{w}\right)}{d \ln \underline{w}} = \begin{cases} 0 & \text{if } \underline{w} < w^* \\ \phi\left(\underline{w}\right) & \text{if } \underline{w} \in (w^*, w^{**}) \\ -\varepsilon & \text{if } \underline{w} > w^{**}, \end{cases}$$

[13] Derenoncourt and Weil (2024) and Datta and Machin (2024) both study variation in wage floors that is arguably firm-specific.

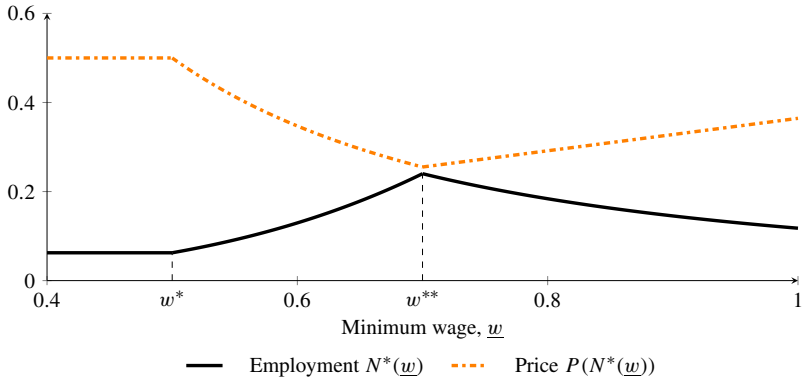

FIGURE 6 Employment and output prices as functions of the minimum wage.

$$\frac{d \ln P\left(N^*\left(\underline{w}\right)\right)}{d \ln \underline{w}} = \begin{cases} 0 & \text{if } \underline{w} < w^* \\ -\phi\left(\underline{w}\right)/\varepsilon & \text{if } \underline{w} \in (w^*, w^{**}) \\ 1 & \text{if } \underline{w} > w^{**}. \end{cases}$$

A striking qualitative implication of the model is that both employment and prices should exhibit a non-monotone relationship with minimum wages, with sign reversals occurring exactly at the quasi-competitive wage. Fig. 6 illustrates this phenomenon, depicting the case where $P_0 = 1/8$, $\varepsilon = 2$, $F(w) = w^4$, $w^* = 0.5$, and $w^{**} = 0.7$.

It is difficult to directly evaluate whether causal relationships of this nature arise empirically, as doing so would seem to require an experiment involving either a particular firm, or a set of firms known to have common thresholds (w^*, w^{**}). Rather, existing studies of minimum wages report evidence on the average behavior of collections of firms that vary in their counterfactual wage levels. In fact, establishment level differences in wages provide a popular source of identifying variation in exposure to minimum wage changes, leveraged by Card and Krueger (1994) and Harasztosi and Lindner (2019) (among others) to infer average effects of the minimum wage on outcomes; see Dube and Lindner (2024) for discussion.

7.2 An aggregation paradox

The non-monotone responses predicted by the monopsony model amplify the formidable challenges involved in inferring microeconomic mechanisms from aggregate responses.[14] To illustrate this point, we will now consider the effects of a small minimum wage hike in a population of firms with different values of

[14] Kline and Tartari (2016) study closely related identification challenges posed by non-monotone labor supply responses to transfer programs exhibiting phase in and phase out regions.

the thresholds (w^*, w^{**}). Each firm faces a stable outside option distribution F, which they believe to be invariant to the level $\underline{w}$ of the minimum wage. This belief, which ensures a non-monotone relationship between employment and minimum wages captured by our previous formulas, might be justified if these firms draw workers primarily from non-employment or uncovered sectors. Despite the negative relationship between employment and prices present at each firm, it will prove possible for the average response to a minimum wage hike of both prices and employment to be positive.

To keep the problem tractable, suppose that each firm is one of two types. Type 1 firms are intrinsically higher wage firms than their type 2 counterparts in the sense that $w_1^* > w_2^*$ and $w_1^{**} > w_2^{**}$. To fix ideas, suppose that this difference is attributable to greater productivity among type 1 firms and let a share $s \in (0, 1)$ of the firms be of type 1. Assume further that the minimum wage is initially set in the range $\underline{w} \in \left(w_1^*, w_1^{**}\right) \cap \left(w_2^{**}, \infty\right)$; that is, the minimum wage is set below the quasi-competitive level for type 1 but not type 2 firms. Finally, suppose the two firm types face a common labor supply elasticity ϕ but the demand elasticity of type 1 firms ε_1 is higher than the elasticity at type 2 firms ε_2.

A clean numerical example results from the choice $\phi = 4$, $\varepsilon_1 = 8$, $\varepsilon_2 = 2$. The average effect of a small increase in the minimum wage on employment and prices is given by the following expressions:

$$\mathbb{E}\left[\frac{d \ln N^*\left(\underline{w}\right)}{d \ln \underline{w}}\right] = s\phi - (1-s)\,\varepsilon_2 = 6s - 2,$$

$$\mathbb{E}\left[\frac{d \ln P\left(N^*\left(\underline{w}\right)\right)}{d \ln \underline{w}}\right] = -s\phi/\varepsilon_1 + (1-s) = 1 - \frac{3}{2}s.$$

For a type 1 share $s \in (1/3, 2/3)$ both the employment and price responses are positive. Evidently, the qualitative pattern of aggregate responses can easily mislead us about the microeconomic structure of product demand. While these parameter values were chosen for analytical convenience, it is clear that this qualitative pattern can be generated for a range of parameter values where $\varepsilon_1 > \phi > \varepsilon_2$.

Do higher wage firms have higher product demand elasticities? There is substantial evidence that firms in tradable sectors tend to exhibit higher wages (Bernard et al., 2007) and to face more difficulty passing cost shocks through to customers (Harasztosi and Lindner, 2019). Likewise, in the fast food sector, one might expect restaurants in urban areas to exhibit higher wages and higher demand elasticities than peer stores in rural areas where competitors tend to be further away. Rationalizing a passthrough elasticity of 1/4, which is what Harasztosi and Lindner (2019) find for manufacturing firms in Hungary, requires $s = 1/2$. This choice yields a net employment elasticity of one, whereas Harasztosi and Lindner (2019) find an empirical employment elasticity in manufacturing of -0.31. There are many combinations of parameters capable of

rationalizing the reduced form finding that $\mathbb{E}\left[d \ln N^*\left(\underline{w}\right)/d \ln \underline{w}\right] = -0.31$ and $\mathbb{E}\left[d \ln P\left(N^*\left(\underline{w}\right)\right)/d \ln \underline{w}\right] = 1/4$. One plausible solution is: $s = 0.38$, $\phi = 4$, $\varepsilon_1 = 4.11$, $\varepsilon_2 = 2.95$.

This back of the envelope calibration is obviously quite crude for at least two reasons. One is that we have ignored factors of production other than labor. When wages account for a smaller share of firm costs then both the employment gains generated by minimum wage hikes at firms with wages in the range (w^*, w^{**}) and the employment losses at firms with wages exceeding w^{**} will tend to be attenuated, though not necessarily by the same amount. Another limitation is that we have assumed the minimum wage affects only two types of firms that account for a small share of employment in a market. A marketwide hike in the minimum wage that already binds at a large share of firms will tend to change the outside option distribution F. One way to model the influence of minimum wages on F would be to use the non-sequential search framework discussed in section 2.4.3, a task that we leave to future research.

7.3 Accounting for quality

It is notable that the studies finding nearly full passthrough of minimum wages to consumers are in sectors involving substantial face to face interaction. An important component of demand in fast food establishments and drug stores is the quality of service. How long must one wait in line to get a prescription filled? How helpful is the person taking the order? These are dimensions of output that, if increased, can raise, rather than lower, prices.

There are two channels through which wages can influence service quality. One is by changing the mix of workers recruited, a channel that was already discussed in Section 4.3. Corroborating this channel, Giuliano (2013) finds that a large retailer adjusted to a minimum wage hike by not only expanding employment of teenagers—behavior consistent with the pre-existing teenager wage falling below the quasi-competitive level—but also by hiring more affluent teenagers. She provides evidence that this compositional shift constitutes a form of quality upgrading, among other reasons because "shrinkage" rates (merchandise lost or damaged) fell at stores where teenage employment increased. Likewise, Horton (2025) provides experimental evidence that forcing employers in an online job market to offer higher wages leads them to hire more skilled workers.

Quality might also respond directly to wages due to traditional efficiency wage effects. While earlier in this section we introduced efficiency wages as a factor boosting output, it seems plausible that the relevant dimension over which efficiency wages operate in these sectors is the quality rather than quantity of output. This channel is most likely to be important in service sectors where it is difficult for supervisors to monitor employee interactions with customers. Ruffini (2022) provides evidence that a minimum wage hike boosted the quality of care in nursing homes, finding that the minimum wage led to a decrease

in the rate of nursing home accidents and deaths. Likewise, Emanuel and Harrington (2020) document sizable productivity improvements among customer service representatives at a Fortune 500 company in the wake of an exogenous pay increase. They find that customer satisfaction with service representatives increased in response to a wage hike. Finally, Brown and Herbst (2023) find that a minimum wage hike affecting childcare centers yields increases in proxies of subjective and objective service quality. It is plausible that these two channels (worker and service quality) together account for a non-trivial share of minimum wage price passthrough in the sectors that have been the focus of this literature.

7.4 Sticky prices

The claim that a firm's prices and employment must respond in opposite directions to a minimum wage change rests fundamentally on the presumption that the firm continuously optimizes the price of its output against a stable demand curve. However, demand is clearly not stable in nominal terms when inflation is present. Moreover, a large body of evidence documents that product prices typically adjust in a lumpy manner: a finding which is almost universally taken to signal the presence of adjustment costs (Nakamura and Steinsson, 2008; Alvarez et al., 2011, 2022).

In an inflationary economy where firms regularly (but infrequently) raise product prices, it is plausible that minimum wage hikes trigger planned price adjustments that would have occurred anyway. If price and wage adjustments share a common fixed cost, then implementing a legally mandated wage increase should lower the cost of additionally adjusting output prices. Contrary to the predictions of our static model, these price adjustments might occur even among firms that do not change their employment in response to the minimum wage. Importantly, these price adjustments will tend to be positive in nominal terms, even if the firm's "frictionless" price target has fallen in real terms.

Suppose that we compare a set of firms for which the minimum wage is initially "just binding" (i.e., $\underline{w} = w^*$) to a set of firms for which the minimum wage does not bind (i.e., $\underline{w} < w^*$). In the short run, the just binding group will hire more workers and hike their nominal prices. By contrast, only a small fraction of the control group of firms unaffected by the minimum wage will update their prices each month, leading to very gradual price adjustment on average. A short run comparison would find that the minimum wage raised the employment of affected fast food restaurants, while also yielding positive passthrough. In the longer run, however, this estimated passthrough would diminish as the control group catches up via regular nominal price adjustments. If the minimum wage hike is permanent and indexed to inflation, it is possible that the long run effect on prices will turn negative, consistent with the negative passthrough prediction of the static monopsony model.

Renkin et al. (2022) find in a panel of US grocery and drug stores that most price adjustments occur within 3 months of the passage (rather than implementation) of a minimum wage law, suggesting that firm pricing decisions

are forward looking. However, they do not test whether the magnitude of price adjustment was affected or if the minimum wage hike simply sped up adjustments that would have occurred anyway. One prediction of the latter hypothesis is that price passthrough should decline over longer horizons. Unfortunately, their main passthrough estimates are limited to 9 months after passage of the law. Indirect evidence that dynamics may be important comes from Benzarti et al. (2020) who study passthrough from value added taxes (VATs) to product prices. They find that VAT hikes yield large price increases almost immediately but the cumulative estimated passthrough falls by nearly half after 20 months. VAT decreases yield much smaller price responses and passthrough is estimated to be negligible after a year.

It is hard to imagine that exposure to a one time nominal minimum wage hike impacts the price of hamburgers a decade later. How many years does it take for price passthrough to diminish? How does passthrough vary with the magnitude of the wage hike and initial wage level of the firm? Do minimum wage decreases, which apparently tend not to generate corresponding wage decreases (Huet-Vaughn and Piqueras, 2023), have symmetric effects on prices? To date, remarkably little evidence on these questions is available.

8 Conclusion

After a long hiatus, the theory of labor market monopsony is back and once again waging war for "the souls of labor economists." A new generation of economists now clamors to measure the scope of firm wage-setting power and use those estimates to inform public policy. An important theme of this chapter has been that closely examining the microeconomic forces governing wage determination—worker outside options and firm motives—is key to understanding both the normative and positive implications of monopsony power. In some respects, this message echoes early lessons from the field of industrial organization, which, decades ago, came to favor models grounded in microeconomic fundamentals over the influential (but ultimately less rigorous) structure-conduct-performance paradigm (Berry et al., 2019). While it is inherently risky to guess what direction a field will go next, several frontiers seem likely to be important in the coming decade.

One avenue for future work is the development of tractable empirical models combining aspects of wage posting and bargaining behavior. A well-developed literature already considers dynamic econometric models of bilateral competition featuring bargaining (Cahuc et al., 2006; Bagger et al., 2014; Bagger and Lentz, 2019) and dynamic contracting (Balke and Lamadon, 2022). However, these models make strong assumptions about the informational environment, often taking the perspective that firms have extraordinary knowledge of worker outside options and the willingness of rival firms to pay for workers. The incomplete information environment reviewed in Section 5 provides a potentially useful weakening of such assumptions that delivers interesting new predictions. Rigorously embedding Chatterjee and Samuelson (1983)'s double auction

model in an equilibrium with realistic search frictions is a non-trivial task that warrants further exploration.

On the empirical front, measuring the frequency with which workers and firms decline to form efficiency-enhancing matches presents a formidable challenge that will likely require rich data on rejected offers, productivity, and outside options. A closely related challenge involves parsing how much of productivity wage passthrough reflects attempts by the firm to grow versus the non-allocative splitting of rents. Separating the strength of these two forces is an exercise in mediation analysis that may require both more sophisticated economic models and new econometric methods. Measurement of how firms grow in response to productivity shocks also remains inadequate. Where do the new hires spawned by a productivity increase come from? If marginal hires are poached from other firms, what was the productivity of the match that was destroyed? Conversely, which workers separate in response to negative shocks, and where do they go? Answering these questions is key to welfare assessments.

Another frontier is more fully describing the shape and structural underpinnings of firm labor supply curves. Are labor supply schedules log-concave? What are the effects of adjusting labor supply schedules for variation in recruiting effort? The answers to these questions likely differ by job type and labor market. A better understanding of the structure of labor supply will strengthen the connection to monopsony theory, which centers fundamentally on the mapping from the distribution of outside options to wages.

Finally, the strong passthrough of minimum wages to product prices remains an important puzzle for monopsony models. We have appealed to explanations invoking the favorite boogeymen of panel data econometricians: unobserved heterogeneity and dynamics. To date, store-level panels on wages and prices have only been available in relatively special settings, usually involving a single company. Better data on product prices, service quality, and wages for a wide range of firms are needed to definitively assess the quantitative importance of the economic explanations offered here.

References

Aaronson, D., French, E., 2007. Product market evidence on the employment effects of the minimum wage. Journal of Labor Economics 25 (1), 167–200.

Acemoglu, D., Autor, D., 2011. Skills, tasks and technologies: implications for employment and earnings. In: Handbook of Labor Economics, vol. 4. Elsevier, pp. 1043–1171.

Ahammer, A., Fahn, M., Stiftinger, F., 2023. Outside options and worker motivation.

Akerlof, G.A., Yellen, J.L., 1990. The fair wage-effort hypothesis and unemployment. The Quarterly Journal of Economics 105 (2), 255–283.

Albrecht, J.W., Axell, B., 1984. An equilibrium model of search unemployment. Journal of Political Economy 92 (5), 824–840.

Alvarez, F., Lippi, F., Oskolkov, A., 2022. The macroeconomics of sticky prices with generalized hazard functions. The Quarterly Journal of Economics 137 (2), 989–1038.

Alvarez, F.E., Lippi, F., Paciello, L., 2011. Optimal price setting with observation and menu costs. The Quarterly Journal of Economics 126 (4), 1909–1960.

Amior, M., Manning, A., 2020. Monopsony and the wage effects of migration.

Amior, M., Stuhler, J., 2022. Immigration and monopsony: evidence across the distribution of firms. Working Paper.
An, M.Y., 1997. Log-concave probability distributions: theory and statistical testing. Duke University Dept of Economics Working Paper 95-03.
Angrist, J.D., Graddy, K., Imbens, G.W., 2000. The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. The Review of Economic Studies 67 (3), 499–527.
Arnold, D., 2021. Mergers and Acquisitions, Local Labor Market Concentration, and Worker Outcomes.
Aronow, P.M., Samii, C., 2017. Estimating average causal effects under general interference, with application to a social network experiment.
Ashenfelter, O., Jurajda, Š., 2022. Minimum wages, wages, and price pass-through: the case of McDonald's restaurants. Journal of Labor Economics 40 (S1), S179–S201.
Autor, D., Dorn, D., Katz, L.F., Patterson, C., Van Reenen, J., 2020. The fall of the labor share and the rise of superstar firms. The Quarterly Journal of Economics 135 (2), 645–709.
Autor, D., Dube, A., McGrew, A., 2023. The unexpected compression: competition at work in the low wage labor market. Tech. Rep. National Bureau of Economic Research.
Azar, J., Marinescu, I., 2024. Monopsony power in the labor market. In: Handbook of Labor Economics.
Azar, J., Berry, S., Marinescu, I., 2022. Estimating labor market power. Tech. Rep. National Bureau of Economic Research.
Bagger, J., Fontaine, F., Postel-Vinay, F., Robin, J.-M., 2014. Tenure, experience, human capital, and wages: a tractable equilibrium search model of wage dynamics. American Economic Review 104 (6), 1551–1596.
Bagger, J., Lentz, R., 2019. An empirical model of wage dispersion with sorting. The Review of Economic Studies 86 (1), 153–190.
Bagnoli, M., Bergstrom, T., 2006. Log-concave probability and its applications. In: Rationality and Equilibrium: A Symposium in Honor of Marcel K. Richter. Springer, pp. 217–241.
Baker, G., Gibbs, M., Holmstrom, B., 1994. The internal economics of the firm: evidence from personnel data. The Quarterly Journal of Economics 109 (4), 881–919.
Baker, M., Halberstam, Y., Kroft, K., Mas, A., Messacar, D., 2023. Pay transparency and the gender gap. American Economic Journal: Applied Economics 15 (2), 157–183.
Balke, N., Lamadon, T., 2022. Productivity shocks, long-term contracts, and earnings dynamics. American Economic Review 112 (7), 2139–2177.
Barkai, S., 2020. Declining labor and capital shares. The Journal of Finance 75 (5), 2421–2463.
Batra, H., Michaud, A., Mongey, S., 2023. Online job posts contain very little wage information. Tech. Rep. National Bureau of Economic Research.
Benzarti, Y., Carloni, D., Harju, J., Kosonen, T., 2020. What goes up may not come down: asymmetric incidence of value-added taxes. Journal of Political Economy 128 (12), 4438–4474.
Berger, D., Herkenhoff, K., Mongey, S., 2022. Labor market power. American Economic Review 112 (4), 1147–1193.
Berger, D., Herkenhoff, K., Mongey, S., 2025. Minimum wages, efficiency, and welfare. Econometrica 93 (1), 265–301.
Bernard, A.B., Jensen, J.B., Redding, S.J., Schott, P.K., 2007. Firms in international trade. The Journal of Economic Perspectives 21 (3), 105–130.
Berry, S., Gaynor, M., Morton, F.S., 2019. Do increasing markups matter? Lessons from empirical industrial organization. The Journal of Economic Perspectives 33 (3), 44–68.
Berry, S., Levinsohn, J., Pakes, A., 1995. Automobile prices in market equilibrium. Econometrica 63 (4), 841–890.
Berry, S.T., Haile, P.A., 2021. Foundations of demand estimation. In: Handbook of Industrial Organization, vol. 4.1. Elsevier, pp. 1–62.
Bertrand, M., Mullainathan, S., 2001. Are CEOs rewarded for luck? The ones without principals are. The Quarterly Journal of Economics 116 (3), 901–932.

Bıró, A., Branyiczki, R., Lindner, A., Márk, L., Prinz, D., 2024. Firm Heterogeneity and the Impact of Payroll Taxes.
Blanchard, O.J., Summers, L.H., 1986. Hysteresis and the European unemployment problem. NBER Macroeconomics Annual 1, 15–78.
Bloesch, J., Larsen, B., Yding, A., 2024. Monopsony with Recruiting. Available at SSRN.
Bloom, N., 2009. The impact of uncertainty shocks. Econometrica 77 (3), 623–685.
Bloom, N., Guvenen, F., Smith, B.S., Song, J., von Wachter, T., 2018. The disappearing large-firm wage premium. AEA Papers and Proceedings 108, 317–322.
Blundell, R., Chen, X., Kristensen, D., 2007. Semi-nonparametric IV estimation of shape-invariant Engel curves. Econometrica 75 (6), 1613–1669.
Boal, W.M., Ransom, M.R., 1997. Monopsony in the labor market. Journal of Economic Literature 35 (1), 86–112.
Bontemps, C., Robin, J.-M., Van den Berg, G.J., 1999. An empirical equilibrium job search model with search on the job and heterogeneous workers and firms. International Economic Review 40 (4), 1039–1074.
Borjas, G.J., 1999. The economic analysis of immigration. In: Handbook of Labor Economics, vol. 3, pp. 1697–1760.
Borjas, G.J., Edo, A., 2023. Monopsony, efficiency, and the regularization of undocumented immigrants. Tech. Rep. National Bureau of Economic Research.
Brenzel, H., Gartner, H., Schnabel, C., 2014. Wage bargaining or wage posting? Evidence from the employers' side. Labour Economics 29, 41–48.
Breza, E., Kaur, S., Shamdasani, Y., 2018. The morale effects of pay inequality. The Quarterly Journal of Economics 133 (2), 611–663.
Brown, C., 1995. Myth and measurement: the new economies of the minimum wage. ILR Review 48 (4), 828–830.
Brown, C., 1999. Minimum wages, employment, and the distribution of income. In: Handbook of Labor Economics, vol. 3, pp. 2101–2163.
Brown, C., Hamilton, J., Medoff, J.L., 1990. Employers Large and Small. Harvard University Press.
Brown, C., Medoff, J., 1989. The employer size-wage effect. Journal of Political Economy 97 (5), 1027–1059.
Brown, J.H., Herbst, C.M., 2023. Minimum Wage, Worker Quality, and Consumer Well-Being: Evidence from the Child Care Market.
Bulow, J.I., Pfleiderer, P., 1983. A note on the effect of cost changes on prices. Journal of Political Economy 91 (1), 182–185.
Burdett, K., Coles, M., 2003. Equilibrium wage-tenure contracts. Econometrica 71 (5), 1377–1404.
Burdett, K., Judd, K.L., 1983. Equilibrium price dispersion. Econometrica: Journal of the Econometric Society, 955–969.
Burdett, K., Mortensen, D.T., 1998. Wage differentials, employer size, and unemployment. International Economic Review, 257–273.
Butters, G.R., 1977. Equilibrium distributions of sales and advertising prices. The Review of Economic Studies 44 (3), 465–491.
Cahuc, P., Postel-Vinay, F., Robin, J.-M., 2006. Wage bargaining with on-the-job search: theory and evidence. Econometrica 74 (2), 323–364.
Caldwell, S., Danieli, O., 2024. Outside options in the labour market. The Review of Economic Studies 91 (6), 3286–3315.
Caldwell, S., Dube, A., Naidu, S., 2023. Monopsony Makes it Big. Tech. Rep.
Caldwell, S., Haegele, I., Heining, J., 2024a. Bargaining in the Labor Market.
Caldwell, S., Haegele, I., Heining, J., 2024b. Firm Pay and Worker Search.
Caldwell, S. Harmon, N., 2019. Outside options, bargaining, and wages: evidence from coworker networks. Unpublished manuscript. Univ. Copenhagen.
Caplin, A., Nalebuff, B., 1991. Aggregation and imperfect competition: on the existence of equilibrium. Econometrica: Journal of the Econometric Society, 25–59.

Cappelli, P., Chauvin, K., 1991. An interplant test of the efficiency wage hypothesis. The Quarterly Journal of Economics 106 (3), 769–787.

Carbonnier, C., Malgouyres, C., Py, L., Urvoy, C., 2022. Who benefits from tax incentives? The heterogeneous wage incidence of a tax credit. Journal of Public Economics 206, 104577.

Card, D., 2022. Who set your wage? American Economic Review 112 (4), 1075–1090.

Card, D., Cardoso, A.R., Heining, J., Kline, P., 2018. Firms and labor market inequality: evidence and some theory. Journal of Labor Economics 36.S1, S13–S70.

Card, D., Krueger, A.B., 1994. Minimum wages and employment: a case study of the fast-food industry in New Jersey and Pennsylvania. American Economic Review 84 (4), 772.

Card, D., Lee, D.S., Pei, Z., Weber, A., 2015. Inference on causal effects in a generalized regression kink design. Econometrica 83 (6), 2453–2483.

Card, D., Mas, A., Moretti, E., Saez, E., 2012. Inequality at work: the effect of peer salaries on job satisfaction. American Economic Review 102 (6), 2981–3003.

Carvalho, M., da Fonseca, J.G., Santarrosa, R., 2023. How are Wages Determined? A Quasi-Experimental Test of Wage Determination Theories. Tech. Rep. Rimini Centre for Economic Analysis.

Chan, M., Kroft, K., Mattana, E., Mourifié, I., 2024. An empirical framework for matching with imperfect competition. Tech. Rep. National Bureau of Economic Research.

Chatterjee, K., Samuelson, W., 1983. Bargaining under incomplete information. Operations Research 31 (5), 835–851.

Chen, X., Christensen, T., Kankanala, S., 2024. Adaptive estimation and uniform confidence bands for nonparametric structural functions and elasticities. The Review of Economic Studies rdae025.

Coviello, D., Deserranno, E., Persico, N., 2022. Minimum wage and individual worker productivity: evidence from a large US retailer. Journal of Political Economy 130 (9), 2315–2360.

Cullen, Z.B., Pakzad-Hurson, B., 2023. Equilibrium effects of pay transparency. Econometrica 91 (3), 765–802.

Dal Bó, E., Finan, F., Rossi, M.A., 2013. Strengthening state capabilities: the role of financial incentives in the call to public service. The Quarterly Journal of Economics 128 (3), 1169–1218.

Datta, N., Machin, S., 2024. Government contracting and living wages > minimum wages. Tech. Rep. IZA Discussion Papers.

Davis, S.J., Faberman, R.J., Haltiwanger, J.C., 2012. Recruiting intensity during and after the Great Recession: national and industry evidence. American Economic Review 102 (3), 584–588.

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

Deb, S., Eeckhout, J., Patel, A., Warren, L., 2024. Walras–Bowley lecture: market power and wage inequality. Econometrica 92 (3), 603–636.

Delabastita, V. Rubens, M., 2022. Colluding against workers. Available at SSRN 4208173.

Derenoncourt, E., Weil, D., 2024. Voluntary minimum wages. Tech. Rep. National Bureau of Economic Research.

Dobbelaere, S., Mairesse, J., 2013. Panel data estimates of the production function and product and labor market imperfections. Journal of Applied Econometrics 28 (1), 1–46.

Doran, K., Gelber, A., Isen, A., 2022. The effects of high-skilled immigration policy on firms: evidence from visa lotteries. Journal of Political Economy 130 (10), 2501–2533.

Dube, A., Freeman, E., Reich, M., 2010. Employee replacement costs.

Dube, A., Jacobs, J., Naidu, S., Suri, S., 2020. Monopsony in online labor markets. American Economic Review: Insights 2 (1), 33–46.

Dube, A., Lindner, A., 2024. Minimum wages in the 21st century. In: Handbook of Labor Economics, vol. 5, pp. 261–383.

Dube, A., Manning, A., Naidu, S., 2018. Monopsony and employer mis-optimization explain why wages bunch at round numbers. Tech. Rep. National Bureau of Economic Research.

Emanuel, N., Harrington, E., 2020. The Payoffs of Higher Pay. Tech. Rep. Working Paper.

Escudero, V., Liepmann, H., Vergara, D., 2024. Directed Search, Wages, and Non-wage Amenities: Evidence from an Online Job Board. Tech. Rep. IZA Discussion Papers.
Faberman, R.J., Mueller, A.I., Şahin, A., Topa, G., 2022. Job search behavior among the employed and non-employed. Econometrica 90 (4), 1743–1779.
Finkelstein, A., McQuillan, C.C., Zidar, O.M., Zwick, E., 2023. The health wedge and labor market inequality. Tech. Rep. National Bureau of Economic Research.
Friedrich, B.U., Zator, M., 2024. Price Discovery in Labor Markets: why do Firms say They Cannot Find Workers?.
Fuest, C., Greil, S., Hugger, F., Neumeier, F., 2022. Global profit shifting of multinational companies: Evidence from cbcr micro data.
Garin, A., Silvério, F., 2023. How responsive are wages to firm-specific changes in labor demand? Evidence from idiosyncratic export demand shocks. The Review of Economic Studies, rdad069.
Giuliano, L., 2013. Minimum wage effects on employment, substitution, and the teenage labor supply: evidence from personnel data. Journal of Labor Economics 31 (1), 155–194.
Gruber, J., 1994. The incidence of mandated maternity benefits. American Economic Review, 622–641.
Guvenen, F., Mataloni Jr, R.J., Rassier, D.G., Ruhl, K.J., 2022. Offshore profit shifting and aggregate measurement: balance of payments, foreign investment, productivity, and the labor share. American Economic Review 112 (6), 1848–1884.
Haanwinckel, D., 2023. Supply, demand, institutions, and firms: a theory of labor market sorting and the wage distribution. Tech. Rep. National Bureau of Economic Research.
Hall, R.E., Krueger, A.B., 2012. Evidence on the incidence of wage posting, wage bargaining, and on-the-job search. American Economic Journal: Macroeconomics 4 (4), 56–67.
Harasztosi, P., Lindner, A., 2019. Who pays for the minimum wage? American Economic Review 109 (8), 2693–2727.
Hazell, J., Patterson, C., Sarsons, H., Taska, B., 2023. National Wage Setting. Tech. Rep. IZA Discussion Papers.
Holzer, Harry J., Katz, Lawrence F., Krueger, Alan B., 1991. Job queues and wages. The Quarterly Journal of Economics 106 (3), 739–768.
Horton, J.J., 2025. Price floors and employer preferences: evidence from a minimum wage experiment. American Economic Review 115 (1), 117–146.
Hosios, A.J., 1990. On the efficiency of matching and related models of search and unemployment. The Review of Economic Studies 57 (2), 279–298.
Huet-Vaughn, E., Piqueras, J., 2023. The Asymmetric Effect of Wage Floors: A Natural Experiment with a Rising and Falling Minimum Wage.
Ingersoll, R.M., 2003. Is there really a teacher shortage? A research report. Center for the Study of Teaching and Policy.
Jäger, S., Heining, J., 2022. How substitutable are workers? Evidence from worker deaths. Tech. Rep. National Bureau of Economic Research.
Jäger, S., Roth, C., Roussille, N., Schoefer, B., 2024. Worker beliefs about outside options. The Quarterly Journal of Economics, qjae001.
Jäger, S., Schoefer, B., Heining, J., 2021. Labor in the boardroom. The Quarterly Journal of Economics 136 (2), 669–725.
Jäger, S., Schoefer, B., Young, S., Zweimüller, J., 2020. Wages and the value of nonemployment. The Quarterly Journal of Economics 135 (4), 1905–1963.
Jarosch, G., Nimczik, J.S., Sorkin, I., 2024. Granular search, market structure, and wages. The Review of Economic Studies 91 (6), 3569–3607.
Kahn, L.B., Lange, F., 2014. Employer learning, productivity, and the earnings distribution: evidence from performance measures. The Review of Economic Studies 81 (4), 1575–1613.
Karabarbounis, L., Neiman, B., 2014. The global decline of the labor share. The Quarterly Journal of Economics 129 (1), 61–103.
Katz, L.F., et al., 1999. Changes in the wage structure and earnings inequality. In: Handbook of Labor Economics, vol. 3. Elsevier, pp. 1463–1555.

Kennedy, P.J., Dobridge, C., Landefeld, P., Mortenson, J., 2022. The efficiency-equity tradeoff of the corporate income tax: Evidence from the Tax Cuts and Jobs Act. Unpublished manuscript.

Kline, P., 2008. Understanding sectoral labor market dynamics: an equilibrium analysis of the oil and gas field services industry.

Kline, P., 2024. Firm wage effects. In: Handbook of Labor Economics, vol. 5, pp. 115–181.

Kline, P., Petkova, N., Williams, H., Zidar, O., 2019. Who profits from patents? Rent-sharing at innovative firms. The Quarterly Journal of Economics 134 (3), 1343–1404.

Kline, P., Petkova, N., Williams, H., Zidar, O., 2021. Corrigendum to 'Who profits from patents?'. Tech. Rep. Working Paper.

Kline, P., Tartari, M., 2016. Bounding the labor supply responses to a randomized welfare experiment: a revealed preference approach. American Economic Review 106 (4), 972–1014.

Knight, F.H., 1921. Risk, Uncertainty and Profit, vol. 31. Houghton Mifflin.

Kroft, K., Luo, Y., Mogstad, M., Setzler, B., 2020. Imperfect competition and rents in labor and product markets: the case of the construction industry. Tech. Rep. National Bureau of Economic Research.

Kwon, S., Roth, J., 2024. Testing mechanisms. arXiv preprint. arXiv:2404.11739.

Lachowska, M., Mas, A., Saggio, R., Woodbury, S.A., 2022. Wage posting or wage bargaining? A test using dual jobholders. Journal of Labor Economics 40.S1, S469–S493.

Lamadon, T., Mogstad, M., Setzler, B., 2022. Imperfect competition, compensating differentials, and rent sharing in the US labor market. American Economic Review 112 (1), 169–212.

Landon, J.H., Baird, R.N., 1971. Monopsony in the market for public school teachers. American Economic Review 61 (5), 966–971.

Lazear, E.P., 1981. Agency, earnings profiles, productivity, and hours restrictions. American Economic Review 71 (4), 606–620.

Le Barbanchon, T., Rathelot, R., Roulet, A., 2021. Gender differences in job search: trading off commute against wage. The Quarterly Journal of Economics 136 (1), 381–426.

Legros, P., Newman, A.F., 2007. Beauty is a beast, frog is a prince: assortative matching with nontransferabilities. Econometrica 75 (4), 1073–1102.

Lindbeck, A., Snower, D.J., 1986. Wage setting, unemployment, and insider-outsider relations. American Economic Review 76 (2), 235–239.

Lindbeck, A., Snower, D.J., 2001. Insiders versus outsiders. The Journal of Economic Perspectives 15 (1), 165–188.

Lindner, A., Muraközy, B., Reizer, B., Schreiner, R., 2022. Firm-level technological change and skill demand.

Lobel, F., 2024. Who Benefits from Payroll Tax Cuts? Market Power, Tax Incidence and Efficiency.

Lusher, L., Schnorr, G.C., Taylor, R.L., 2022. Unemployment insurance as a worker indiscipline device? Evidence from scanner data. American Economic Journal: Applied Economics 14 (2), 285–319.

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

Manning, A., 2011. Imperfect competition in the labor market. In: Handbook of Labor Economics, vol. 4. Elsevier, pp. 973–1041.

Manning, A., 2013. Monopsony in Motion: Imperfect Competition in Labor Markets. Princeton University Press.

Manning, A., 2021. Monopsony in labor markets: a review. ILR Review 74 (1), 3–26.

Manning, A., Petrongolo, B., 2017. How local are labor markets? Evidence from a spatial job search model. American Economic Review 107 (10), 2877–2907.

Manning, A., Saidi, F., 2010. Understanding the gender pay gap: what's competition got to do with it? ILR Review 63 (4), 681–698.

Manski, C.F., 2013. Identification of treatment response with social interactions. Econometrics Journal 16 (1), S1–S23.

Marinescu, I., Wolthoff, R., 2020. Opening the black box of the matching function: the power of words. Journal of Labor Economics 38 (2), 535–568.

Mas, A., 2017. Does transparency lead to pay compression? Journal of Political Economy 125 (5), 1683–1721.
Mas, A., Pallais, A., 2019. Labor supply and the value of non-work time: experimental estimates from the field. American Economic Review: Insights 1 (1), 111–126.
Menzio, G., 2024. Markups: a search-theoretic perspective. Tech. Rep. National Bureau of Economic Research.
Mertens, M., Müller, S., Neuschäffer, G., 2022. Identifying rent-sharing using firms' energy input mix. Tech. Rep. IWH Discussion Papers.
Miravete, E., Seim, K., Thurk, J., 2023. Elasticity and Curvature of Discrete Choice Demand Models. Tech. Rep. Working paper.
Moore, H.L., 1911. Laws of Wages: An Essay in Statistical Economics. Macmillan.
Morchio, I., Moser, C., 2024. The gender pay gap: micro sources and macro consequences. Tech. Rep. National Bureau of Economic Research.
Mrázová, M., Neary, J.P., 2017. Not so demanding: demand structure and firm behavior. American Economic Review 107 (12), 3835–3874.
Naidu, S., Nyarko, Y., Wang, S.-Y., 2016. Monopsony power in migrant labor markets: evidence from the United Arab Emirates. Journal of Political Economy 124 (6), 1735–1792.
Naidu, S., Posner, E.A., Weyl, G., 2018. Antitrust remedies for labor market power. Harvard Law Review 132 (2), 536–601.
Nakamura, E., Steinsson, J., 2008. Five facts about prices: a reevaluation of menu cost models. The Quarterly Journal of Economics 123 (4), 1415–1464.
Newey, W.K., 2013. Nonparametric instrumental variables estimation. American Economic Review 103 (3), 550–556.
Newey, W.K., Powell, J.L., 2003. Instrumental variable estimation of nonparametric models. Econometrica 71 (5), 1565–1578.
Nimczik, J.S., 2017. Job mobility networks and endogenous labor markets.
Oi, W.Y., Idson, T.L., 1999. Firm size and wages. In: Handbook of Labor Economics, vol. 3, pp. 2165–2214.
Planck, M., 1949. Scientific Autobiography and Other Papers. Trans. by F. Gaynor (New York, 1949). pp. 33–34.
Posner, E.A., 2021. How Antitrust Failed Workers. Oxford University Press.
Postel-Vinay, F., Robin, J.-M., 2002a. Equilibrium wage dispersion with worker and employer heterogeneity. Econometrica 70 (6), 2295–2350.
Postel-Vinay, F., Robin, J.-M., 2002b. The distribution of earnings in an equilibrium search model with state-dependent offers and counteroffers. International Economic Review 43 (4), 989–1016.
Prager, E., Schmitt, M., 2021. Employer consolidation and wages: evidence from hospitals. American Economic Review 111 (2), 397–427.
Renkin, T., Montialoux, C., Siegenthaler, M., 2022. The pass-through of minimum wages into US retail prices: evidence from supermarket scanner data. Review of Economics and Statistics 104 (5), 890–908.
Robinson, J., 1933. Economics of imperfect competition.
Rong, M., 2022. Monopsony Power and Worker Mobility: Evidence from Coworking Couples.
Rosen, S., 1981. The economics of superstars. American Economic Review 71 (5), 845–858.
Rosen, S., 1986. The theory of equalizing differences. In: Handbook of Labor Economics, vol. 1, pp. 641–692.
Roussille, N., Scuderi, B., 2023. Bidding for Talent: A Test of Conduct in a High-Wage Labor Market.
Ruffini, K., 2022. Worker earnings, service quality, and firm profitability: evidence from nursing homes and minimum wage reforms. Review of Economics and Statistics, 1–46.
Santos, A., 2012. Inference in nonparametric instrumental variables with partial identification. Econometrica 80 (1), 213–275.

Satterthwaite, M.A., Williams, S.R., 1989. Bilateral trade with the sealed bid k-double auction: existence and efficiency. Journal of Economic Theory 48 (1), 107–133.
Seegmiller, B., 2023. Valuing Labor Market Power: the Role of Productivity Advantages.
Sen, P.K., 1968. Estimates of the regression coefficient based on Kendall's tau. Journal of the American Statistical Association 63 (324), 1379–1389.
Shapiro, C., Stiglitz, J.E., 1984. Equilibrium unemployment as a worker discipline device. American Economic Review 74 (3), 433–444.
Sharma, G., 2023. Monopsony and gender. Working Paper.
Sharma, G., 2024. Collusion Among Employers in India.
Smith, M., Yagan, D., Zidar, O., Zwick, E., 2022. The rise of pass-throughs and the decline of the labor share. American Economic Review: Insights 4 (3), 323–340.
Sockin, J., 2022. Show Me the Amenity: Are Higher-Paying Firms Better All Around?.
Sokolova, A., Sorensen, T., 2021. Monopsony in labor markets: a meta-analysis. ILR Review 74 (1), 27–55.
Staiger, D.O., Spetz, J., Phibbs, C.S., 2010. Is there monopsony in the labor market? Evidence from a natural experiment. Journal of Labor Economics 28 (2), 211–236.
Stevens, M., 1994. A theoretical model of on-the-job training with imperfect competition. Oxford Economic Papers 46 (4), 537–562.
Sullivan, D., 1989. Monopsony power in the market for nurses. The Journal of Law & Economics 32 (2, Part 2), S135–S178.
Summers, L.H., 1989. Some simple economics of mandated benefits. American Economic Review 79 (2), 177–183.
Theil, H., 1950. A rank-invariant method of linear and polynomial regression analysis. Indagationes Mathematicae 12 (85), 173.
Tørsløv, T., Wier, L., Zucman, G., 2023. The missing profits of nations. The Review of Economic Studies 90 (3), 1499–1534.
Townsend, W., Allan, C., 2024. How Restricting Migrants' Job Options Affects Both Migrants and Existing Residents.
Van den Berg, G.J., Ridder, G., 1998. An empirical equilibrium search model of the labor market. Econometrica, 1183–1221.
Van Reenen, J., 2024. A comment on: "Walras–Bowley lecture: market power and wage inequality" by Shubhdeep Deb, Jan Eeckhout, Aseem Patel, and Lawrence Warren. Econometrica 92 (3), 643–646.
Volpe, O., 2024. Job Preferences, Labor Market Power, and Inequality. Tech. Rep. Discussion paper, Working Paper.
Weil, D., 2014. The Fissured Workplace. Harvard University Press.
Welch, F., 1995. Myth and measurement: the new economies of the minimum wage. ILR Review 48 (4), 842–849.
Wiltshire, J., McPherson, C., Reich, M., Sosinskiy, D., 2025. Minimum wage effects and monopsony explanations. Journal of Labor Economics. Forthcoming.
Yeh, C., Macaluso, C., Hershbein, B., 2022. Monopsony in the US labor market. American Economic Review 112 (7), 2099–2138.
Yett, D.E., 1970. The chronic shortage of nurses: a public policy dilemma. In: Empirical Studies in Health Economics, pp. 357–389.
Yitzhaki, S., 1996. On using linear regressions in welfare economics. Journal of Business & Economic Statistics 14 (4), 478–486.