

Graduate Labor Economics

Notes to Accompany Lecture 7: Market Concentration and Monopsony Power

Brendan M. Price*
Federal Reserve Board

Up to now, we've been working almost exclusively within a neoclassical, frictionless framework in which workers and firms adjust seamlessly to shocks, the labor market clears, the law of one price holds, and government interventions can be hard to motivate on efficiency grounds (if not on equity grounds).

In this lecture, we'll take a first look at monopsony power stemming from market concentration. Once a somewhat peripheral topic, monopsony has recently resurfaced as an extremely active area of labor (and macro) research, spurred by a growing body of evidence that appears to be consistent with substantial and rising levels of employer market power in the US and other labor markets (though this interpretation is debated). [Furman and Orszag \(2018\)](#) offer a nice summary of the key facts, which include rising firm-level profit margins, the declining labor share of employment, declining rates of job creation and destruction, and growing pay dispersion between firms.

1 Frictions: the broad view

- Why might the labor market differ from the frictionless benchmark? We'll discuss one market failure—monopsony—in detail below, but there are many others (this list is not exhaustive):
 - Search frictions: it takes time for workers to find suitable jobs (and for employers to find suitable workers). Search is a leading paradigm for thinking about the labor market, and we'll discuss it in great detail later in the course.
 - Moral hazard: productivity is hard to observe, raising issues of asymmetric information between workers and firms. Employers deal with this through costly monitoring of workers and by paying “efficiency wages”, above-market wages that incentivize workers to exert more effort by raising the stakes of job loss ([Lazear, 2000](#)). These above-market wages in turn lead to equilibrium unemployment ([Shapiro and Stiglitz, 1984](#)).
 - Holdup: in a well-functioning labor market, workers must invest in skills and firms must invest in capital and technology. Due to search and other frictions, however, the other party usually captures a portion of these investments. The potential for ex post holdup problems may lead to inefficiently low levels of human capital investment and capital acquisition ([Acemoglu, 2001](#)).
 - Fairness norms: workers may balk at nominal wage cuts (even if they tacitly accept real wage cuts all the time) and resent pay differentials relative to similar workers ([Card](#)

*Copyright ©2020 by Brendan M. Price. All rights reserved. Email: brendan.m.price@frb.gov. These notes are adapted from UC Davis course ECN 250A, which I taught in Winter 2018 and Spring 2019. Any views or opinions expressed here are my own and do not necessarily represent the views or policies of the Board of Governors of the Federal Reserve System or its staff. This version: 2020/03/21.

et al., 2012; Dube et al., 2019). These behavioral factors limit firms' ability to cut wages during downturns or to set individual wages equal to individual marginal products.

- Institutions: occupational licensing, non-compete agreements, and other contractual institutions may impede worker mobility and encourage rent-seeking (Kleiner, 2000; Krueger and Posner, 2018). Taxes and transfers can also introduce distortions. Dismissal protections may induce shirking, and employers may respond to them by being inefficiently selective in hiring decisions.
- What happens when we introduce frictions?
 - The law of one price for (efficiency units of) labor breaks down. Identical workers may receive different wages simply because one is lucky enough to land a “good job”.
 - Worker-firm matching becomes really important. If high-skill workers also get matched to high-productivity firms (as some papers contend is increasingly occurring, e.g., Card et al., 2013), inequality will rise relative to a world of random worker-firm matching.
 - Job loss matters. Manning (2003) puts it well: “people go to the pub to celebrate when they get a job rather than greeting the news with the shrug of the shoulders that we might expect if labour markets were frictionless. And people go to the pub to drown their sorrows when they lose their job rather than picking up another one straight away.”
 - Dynamics get more interesting: the market doesn't seamlessly adjust from one steady state to another; there may be large frictional costs along the transition path.
 - Institutions like unions, minimum wages, and mandated benefits potentially raise welfare. Many labor market regulations are motivated on efficiency as well as equity grounds.

2 Sources of monopsony power

- Strictly speaking, “monopsony” means one buyer (of labor), perhaps within a particular occupation.¹ The classic example of this is the “company town”: places like Pullman, Chicago, whose residents made the Pullman Company's railroad sleeper cars. Other examples include coal-mining towns, where a single mine operator may employ much of the town's workforce (see the song “Sixteen Tons”), and hospitals, which in many places are the major employer of nurses and other healthcare occupations.
- With a few exceptions, the literal one-buyer case is usually unrealistic, but (somewhat imprecisely) we use the term monopsony to describe any situation in which firms have *some* market power. A degree of monopsony power exists whenever firms face upward-sloping labor supply curves, meaning that a firm won't instantly lose all of its workers if it lowers wages by a cent. Casual introspection suggests that most firms are not the pure wage-takers assumed in models of perfect competition.

¹The etymology is much celebrated and a pure delight: “[T]he stem of monopsony, *opsonein* (*οψωνειν*), has a meaning that is more idiosyncratic than just ‘to go marketing,’ the translation given by Joan Robinson. In classical Greek, the word *opson* (*οψου*) means ‘fish’ (or ‘cooked meat’). And the verb *opsonein* (*οψωνειν*) that Robinson refers to actually means ‘to buy fish,’ making the literal meaning of monopsony a market situation with ‘a single buyer of fish’” (Thornton, 2004).

- Monopsony power stems from many sources (see [Manning, 2011](#), for a nice discussion):
 - Concentration: there may only be a small number of firms in a given market. Under some models of imperfect competition (e.g., Cournot), this yields a degree of market power in wage-setting.
 - Commuting time: even if there are many potential employers within a local labor market, nearby employers will be able to attract workers at lower wages than distant employers.
 - Product differentiation: aside from physical location, jobs differ along a host of other dimensions, making them imperfect substitutes in the eyes of workers.
 - Search costs: even if there are two identical employers for a worker to apply to, she might not immediately know about both of them, in which case they will have some market power.
- Directly testing for the existence of monopsony power is quite difficult because it requires instruments for either firm-level employment or firm-level wages. [Matsudaira \(2014\)](#) exploits a 1999 California law establishing minimum nurse staffing ratios for long-term care facilities and finds that non-compliant firms were able to expand employment without having to increase wages. [Falch \(2010\)](#) exploits centralized wage-setting for Norwegian teachers and finds evidence for upward-sloping labor supply.
- Much of the evidence is more indirect. As one example, [Azar et al. \(2018\)](#) use geolocated data on US job postings to show that most local occupational labor markets are highly concentrated, a fact that promotes monopsonistic behavior on the part of firms.

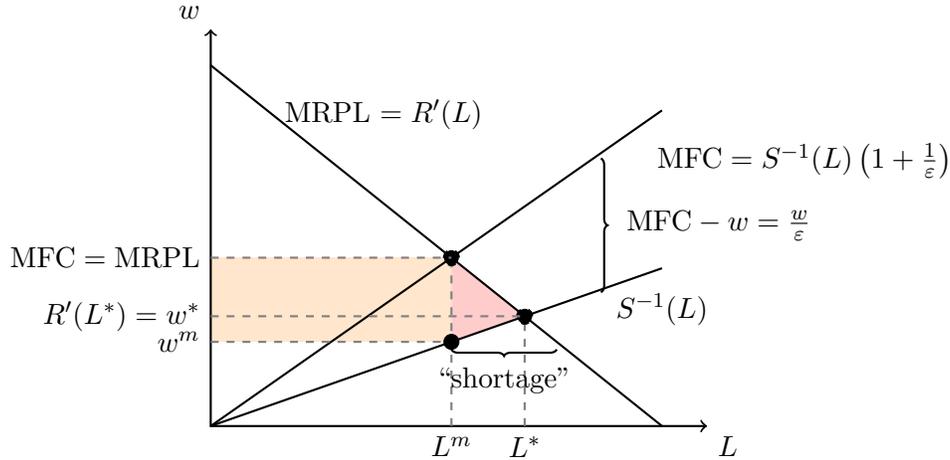
3 Static monopsony

- Let's start with the classic monopsony model, in which a single firm makes a static labor demand decision. Our key assumption is that the monopsonist has to offer everyone the same wage. (We can think of what follows as conditioning on ability: employers can offer higher wages to more productive workers, but we'll assume that they have to pay equally productive workers the same amount.)
 - Why? One answer is imperfect information: the monopsonist may know the distribution of reservation wages it faces (i.e., the labor supply curve) without knowing which workers have high reservation wages and which ones will work for peanuts.
 - Another answer is legal restrictions that constrain the firm's ability to offer dissimilar wages on the basis of certain protected characteristics, such as race ([Hirsch et al., 2013](#)).
 - A third is fairness norms: workers might complain or quit if they think the employer is unfairly favoring some workers over others. There's evidence for this ([Card et al., 2012](#); [Dube et al., 2019](#)).
- The monopsonist earns profits $\pi = \max_L R(L) - w(L)L$, where $R(L)$ denotes revenues net of any other factor costs and $w(L)$ denotes the inverse labor supply curve that the firm faces.² In perfect competition, we have $w(L) = w$, where w is the market wage. But in general we'll think of this as a potentially upward sloping curve.

²We can easily account for product market power, as in $R(L) = p(F(L))F(L)$. It's also easy to accommodate other factors: for example, if the firm uses capital as well as labor (and is a price-taker in the product market), we can write $R(L) = \max_K pF(L, K) - rK$. We simply assume that the firm has already optimized over any other inputs.

- FOC: $R'(L) = w(L) + w'(L)L$. We can rewrite this as $R'(L) = w(1 + \frac{1}{\varepsilon})$, where $\varepsilon \equiv \frac{dL}{dw} \frac{w}{L}$ is the elasticity of labor supply (not necessarily constant). As in the perfectly competitive case, the MRPL of labor is set equal to the marginal factor cost (MFC) of hiring an additional worker. But now, instead of just being w , the MFC has an extra term reflecting the fact that a monopsonist must raise everybody's wages if it wants to hire an extra worker. (See the figure below.)
- We can also rewrite the FOC as $\frac{MRPL - w}{w} = \frac{1}{\varepsilon}$. The lefthand side is sometimes called an index of "exploitation" (Robinson, 1969): the idea is that the firm "exploits" its workers by paying them below their marginal product. (As you might imagine, many economists bristle at this kind of value-laden language.) Elastic workers are hard to exploit in the Robinsonian sense.
- Relative to the efficient benchmark, the monopsonist employs "too few" workers at "too low" a wage. There is deadweight loss because the marginal social benefit of hiring the worker (her output) exceeds the marginal social cost (given by her reservation wage).
- Since monopsonistic employers are reluctant to raise wages for new hires (as doing so may compel them to raise wages across the board), they may look for other margins along which to attract new workers, such as spending resources on recruitment or offering signing bonuses distinct from the base wage (Manning, 2006).
- In the competitive case, a (binding) minimum wage results in higher wages but lower employment, and it creates DWL. Here, a minimum wage can actually *increase* employment and decrease DWL.
 - When the minimum wage is binding, the employer faces a flat supply curve (locally), so that hiring an extra worker doesn't force the firm to raise everybody's wages. This eliminates the incentive to curtail employment to drive down wages.
 - As the minimum wage rises (towards the competitive level) we "walk up the supply curve": more workers want to work at the newly increased wage, and the firm is happy to hire them.
 - A minimum wage set equal to the competitive wage achieves the social optimum. The monopsonist acts just like a price-taker facing the "correct" wage.
 - Marginal increases in the minimum wage above the competitive level reduce employment, just like in the competitive case: now the demand curve is the limiting factor.
- Other observations:
 - If the monopsonist knows every worker's reservation wage, and is able to make individualized offers without legal or political opposition, it holds each worker to his or her outside option. This actually restores the Pareto optimum, though now the firm gets all of the surplus.
 - Suppose instead that workers belong to observable groups known to have different elasticities of supply. (Nota bene: the relevant parameter here is the elasticity of supply *to the firm in question*, not to the market as a whole.) Then the firm makes group-specific wage offers, with lowball offers to inelastic groups. This is third-degree price discrimination.
- Boal and Ransom (1997) also work through a Cournot oligopsony case, in which the Robinson index of exploitation is increasing in market concentration. Intuitively, when there are lots

of smaller firms competing for workers, there will be less of a markdown relative to marginal product. One might suspect that the tech sector is oligopsonistic with respect to computer programmers and web developers, since (the many startups notwithstanding) it's dominated by a handful of major players.



The monopsony problem. Orange: transfer from workers to employer. Red: deadweight loss.

4 Dynamic monopsony

- [Boal and Ransom \(1997\)](#) present a nice dynamic model of monopsony. As in the static case, a firm that “underpays” its workers won’t lose all of them immediately; instead it will face more quits and have a harder time hiring going forward. This is a good moment to bring dynamic considerations into our study of the labor market.
- Suppose that current labor supply depends on both the current wage and on the existing stock of workers: $L_t = L(w_t, L_{t-1})$, with L_t increasing in both arguments: for a given wage, a firm will have more workers today if it had more workers yesterday, because some of the existing workers will stick around even if wages today are low. (We’ll be more explicit about hires/separations shortly.)
- Invert this to get wages as a function of current and past employment: $w_t = w(L_t, L_{t-1})$.
- The firm maximizes the PDV of future profits:

$$\max_{L_0, L_1, \dots} \sum_{t=0}^{\infty} \beta^t [R(L_t) - w(L_t, L_{t-1})L_t]$$

with the initial workforce L_{-1} given.

- This is a recursive problem, and working with the Bellman equation is usually more elegant. We can write the firm’s problem as

$$V(L_0) = \max_{L_N} R(L_N) - w(L_N, L_0)L_N + \beta V(L_N)$$

where L_O is the “old” amount of labor and L_N is the “new” amount of labor. Take the FOC:

$$R'(L_N) = w(L_N, L_O) + \frac{\partial w}{\partial L_N} L_N - \beta V'(L_N)$$

To make further progress, we’ll invoke the envelope theorem, which lets us differentiate the value function w.r.t. the current stock of workers, L_O , and treat L_N as a constant in the process. This gives

$$V'(L_O) = -\frac{\partial w}{\partial L_O} L_N$$

Next, let’s focus on a steady-state solution where $L_{t+1} = L_t$. If the economic environment is stationary, then under some mild assumptions the monopsonist’s employment level converges to this steady-state solution. Setting $L_N = L_O = L$, and plugging the envelope condition into the original FOC, we have

$$R'(L) = w(L, L) + \frac{\partial w}{\partial L_N} L + \beta \frac{\partial w}{\partial L_O} L$$

Rearranging gives

$$\frac{R'(L) - w}{w} = \frac{\partial w}{\partial L_N} \frac{L}{w} + \beta \frac{\partial w}{\partial L_O} \frac{L}{w}$$

Let’s define $\varepsilon_{SR} = \frac{\partial L_N}{\partial w} \frac{w}{L_N} \Big|_{L_O}$ as the short-run elasticity of labor supply. This tells us how changes in wages today affect employment today, holding yesterday’s employment constant. So the inverse short-run elasticity of labor supply is given by $\frac{1}{\varepsilon_{SR}}$. For the long-run elasticity, we need to know how the wage must change to keep employment constant at a marginally higher level forever. It’s easier to compute the inverse elasticity directly. Abusing notation slightly,

$$\frac{1}{\varepsilon_{LR}} = \frac{dw}{dL^*} \frac{L^*}{w} = \left(\frac{\partial w}{\partial L_N} + \frac{\partial w}{\partial L_O} \right) \frac{L^*}{w} = \frac{\partial w}{\partial L_N} \frac{L^*}{w} + \frac{\partial w}{\partial L_O} \frac{L^*}{w}$$

where I’m using L^* to denote the common level of employment yesterday and today. The first term is the inverse short-run elasticity, and the second term captures the inverse elasticity with respect to last period’s labor choice. So we can rewrite this expression as

$$\frac{1}{\varepsilon_{LR}} = \frac{1}{\varepsilon_{SR}} + \frac{1}{\varepsilon_O},$$

where $\frac{1}{\varepsilon_O} \equiv \frac{\partial w}{\partial L_O} \frac{L}{w}$. Using these new expressions, we can rewrite Robinsonian exploitation as

$$\frac{R'(L) - w}{w} = (1 - \beta) \frac{1}{\varepsilon_{SR}} + \beta \frac{1}{\varepsilon_{LR}}$$

The firm marks down wages according to a weighted average of short- and long-run inverse elasticities. For example, suppose that the firm faces inelastic demand in the short run but elastic demand in the long run. If it is impatient (say because of credit constraints), it will squeeze its workers today by giving them low wages. This raises profits today at the expense of longer-run profits. Conversely, a patient firm will pay its workers more because it wants them to stick around.

- An insight from this formulation is that, if labor supply is increasing in the current stock of workers, firms will treat their workers as an investment. The model also formalizes our intuition that labor supply is probably more elastic in the long run than in the short run.

- Let's dig deeper. Today's labor supply equals yesterday's labor supply plus net hires:

$$L(w_t, L_{t-1}) = L_{t-1} + h(w_t, L_{t-1}) - q(w_t, L_{t-1})$$

Let's assume hires are increasing in wages and quits are decreasing. Quits are probably also increasing in the current stock of workers (the risk set is bigger); hires may be decreasing if it becomes harder to hire when the firm is bigger (the firm starts running out of qualified candidates).

- The short-run elasticity of labor supply tells us how today's employment changes in response to a change in wages, holding yesterday's employment constant:

$$\varepsilon_{SR} = \left. \frac{\partial L_t}{\partial w} \frac{w}{L_t} \right|_{L_{t-1}} = \left(\frac{\partial h}{\partial w} - \frac{\partial q}{\partial w} \right) \frac{w}{L_t}$$

By our assumptions, an increase in wages will increase hires and reduce quits, so both terms here are positive. It can also be shown that the long-run elasticity of labor supply is

$$\varepsilon_{LR} = \varepsilon_{SR} \left(\frac{\partial q}{\partial L} - \frac{\partial h}{\partial L} \right)^{-1}$$

If the two partial derivatives are equal, then long-run labor supply is perfectly elastic: in this case, net outflows of workers are independent of firm size, so big firms don't need to offer higher wages to be big. If the quit derivative is larger, bigger firms must raise wages to offset the rise in quits.

5 What can monopsony explain?

- [Manning \(2011\)](#) argues that monopsony can explain a number of empirical regularities:
 - Inframarginal workers. The competitive model predicts that all workers instantly quit if a firm lowers its wage by a cent. In a monopsonistic world, firms usually have some ability to lower wages without instantaneously losing all of their workers—which certainly seems more realistic.
 - The firm-size premium. In a monopsonistic world, firms have to raise wages if they wish to keep growing, so that in the cross section bigger firms should pay higher wages.
 - Employer complaints of labor shortages. The monopsonist would love to be able to hire additional workers at its offered wage w^m . From its perspective, it faces a “shortage” of workers (even though it could solve its shortage by simply offering higher wages). When employers express frustration about their inability to attract or retain enough workers, it's suggestive of monopsony: a price-taking firm is indifferent about marginal quits because those marginal workers bring in zero profit.
 - Minimum wages may increase employment (or at least have little adverse impact on employment). [Card and Krueger \(1994\)](#) famously found that fast-food employment did not fall, and may have even rose, when New Jersey raised its minimum wage in 1992. The subsequent literature has seldom found evidence that minimum wages actually increase employment, but there is also no consistent evidence of sizable employment losses (at least in the short-run, and in response to minimum wage hikes of the magnitude we've historically seen). Monopsony can help explain this.

- Monopsony used to be a bit of a backwater topic, but it’s reemerged in recent years as an active and exciting area of research. This probably reflects new modeling techniques (such as the worker-firm regressions introduced by [Abowd et al. 1999](#)), richer data for testing monopsonistic ideas, and a growing list of credible empirical facts that are difficult to explain in a perfectly competitive model. [Card et al. \(2018\)](#) develop a monopsonistic model that seeks to unify many of the recent theoretical and empirical developments in this area.

References

- Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High Wage Workers and High Wage Firms. *Econometrica*, 67(2):251–333.
- Acemoglu, D. (2001). Good Jobs versus Bad Jobs. *Journal of Labor Economics*, 19(1):1–21.
- Azar, J., Marinescu, I., Steinbaum, M., and Taska, B. (2018). Concentration in US Labor Markets: Evidence from Online Vacancy Data. NBER working paper 24395.
- Boal, W. M. and Ransom, M. R. (1997). Monopsony in the Labor Market. *Journal of Economic Literature*, 35(1):86–112.
- Card, D., Cardoso, A. R., Heining, J., and Kline, P. (2018). Firms and Labor Market Inequality: Evidence and Some Theory. *Journal of Labor Economics*, 36(S1):S13–S70.
- Card, D., Heining, J., and Kline, P. (2013). Workplace Heterogeneity and the Rise of West German Wage Inequality. *Quarterly Journal of Economics*, 128(3):967–1015.
- Card, D. and 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):772–793.
- Card, D., Mas, A., Moretti, E., and Saez, E. (2012). Inequality at Work: The Effect of Peer Salaries on Job Satisfaction. *American Economic Review*, 102(6):2981–3003.
- Dube, A., Giuliano, L., and Leonard, J. (2019). Fairness and Frictions: The Impact of Unequal Raises on Quit Behavior. *American Economic Review*, 109(2):620–663.
- Falch, T. (2010). The Elasticity of Labor Supply at the Establishment Level. *Journal of Labor Economics*, 28(2):237–266.
- Furman, J. and Orszag, P. (2018). A Firm-Level Perspective on the Role of Rents in the Rise in Inequality. In *Toward a Just Society: Joseph Stiglitz and Twenty-First Century Economics*. Columbia University Press, New York.
- Hirsch, J. M., Secunda, P. M., and Bales, R. A. (2013). *Understanding Employment Law: Second Edition*. Carolina Academic Press, Durham, NC.
- Kleiner, M. M. (2000). Occupational Licensing. *Journal of Economic Perspectives*, 14(4):189–202.
- Krueger, A. B. and Posner, E. A. (2018). A Proposal for Protecting Low-Income Workers from Monopsony and Collusion. Hamilton Project.
- Lazear, E. P. (2000). Performance Pay and Productivity. *American Economic Review*, 90(5):1346–1361.

- Manning, A. (2003). *Monopsony in Motion: Imperfect Competition in Labor Markets*. Princeton University Press, Princeton, NJ.
- 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*, volume 4B, pages 976–1041.
- Matsudaira, J. D. (2014). Monopsony in the Low-Wage Labor Market? Evidence from Minimum Nurse Staffing Regulations. *Review of Economics and Statistics*, 96(March):92–102.
- Robinson, J. (1969). *The Economics of Imperfect Competition, 2nd Edition*. Palgrave Macmillan, London.
- Shapiro, C. and Stiglitz, J. E. (1984). Equilibrium Unemployment as a Worker Discipline Device. *American Economic Review*, 74(3):433–444.
- Thornton, R. J. (2004). Retrospectives: How Joan Robinson and B. L. Hallward Named Monopsony. *Journal of Economic Perspectives*, 18(2):257–261.