THE MEASURE OF MONOPSONY

Monica Langella
Centre for Economic Performance,
London School of Economics and
Political Science, UK

Alan Manning
Centre for Economic Performance,
London School of Economics and
Political Science, UK

Abstract
There has been increasing interest in recent years in monopsony in the labour market. This paper discusses how we can measure monopsony power by combining insights from models based on both frictions and idiosyncrasies. It presents some evidence from the United Kingdom and the United States about how monopsony power varies across the wage distribution within markets, over the business cycle and over time. (JEL: I21, I24)

1. Introduction

This lecture is named after Alfred Marshall, widely regarded as the founder of neo-classical economics and the marginal approach as a coherent body of thought, though he built on and combined the ideas of many others. Neo-classical economics is often characterized as primarily being concerned with the allocation of scarce resources to competing ends through the interplay of demand and supply rather than the more distributional issues discussed by the classical economists. Yet Marshall's *The Principles of Economics* (Marshall 1890) opens with a discussion of inequality (which he thought a thoroughly bad thing, viewing the "degradation" of the poor—his

The editor in charge of this paper was Imran Rasul.

Acknowledgments: We would like to thank the ERC for funding the research under grant 834455 "LPIGMANN". This paper was presented as the Marshall lecture at the EEA Conference in Rotterdam in August 2020. I would like to thank the organizing committee for the invitation to present this lecture.

E-mail: m.langella@lse.ac.uk (Langella); a.manning@lse.ac.uk (Manning)

Journal of the European Economic Association 2021 19(6):2929–2957
https://doi.org/10.1093/jeea/jvab039
© The Author(s) 2021. Published by Oxford University Press on behalf of European Economic Association. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/jeea/article/19/6/2929/6368338 by University Library Utrecht user on 14 January 2022

phrase—more as a consequence than a cause of their poverty[1]) before moving on to a discussion of competition, which he thought, on balance, was good rather than bad and that other well-intended institutions might be more harmful.[2] But he was not starry-eyed; Marshall (1890, p. 12) described how "free competition [. . .] was set loose to run, like a huge untrained monster, its wayward course" and that this had impacts, good and bad: "the abuse of their new power by able but uncultured business men led to evils on every side; it unfitted mothers for their duties, it weighed down children with overwork and disease; and in many places it degraded the race". Competition needed an eye on it and, possibly, some regulation, to channel the wayward, untrained monster to work for good and not for evil.

That agenda of concern about inequality and competition remains as important as ever. We know from basic economics that markets cannot be relied on to produce levels of inequality that are fair and command political legitimacy. Economists do not often regard inequality as a market failure (it is put in a different box), but ordinary people do, and they are right and we are wrong. We also know that markets contain within them the incentives to acquire market power, and we have to be on our guard against the effort put into monopolizing markets. And, because productivity growth at the frontier is based on improvements in knowledge, which is a public good, we know we cannot rely on markets alone to deliver an efficient level of economic growth, which is the basis for improvements in the quality of life.

So, a primary role of economists should be tinkering with the system to make sure that the economy functions well—to train the monster so that its course is not too wayward. This lecture is about one small part of that wider endeavour: labour market competition or monopsony. The idea of employer market power—monopsony or oligopsony—originated with Joan Robinson (1933). For many years, the idea that most employers have market power over their workers had few buyers, but recent years have seen a resurgence in academic interest in monopsony.

There is no doubt that the labour supply curve is not perfectly elastic and that employers do have some market power over their workers. Common sense tells us that if employers cut the wage by a cent, it is simply not true that all the existing workers immediately leave as perfect competition predicts, though they may find it harder to recruit and retain workers. It is not just common sense that supports this conclusion: We have an increasing number of papers with high quality research designs that show that the labour supply curve facing an employer is not infinitely elastic (see, e.g., the experimental studies of Caldwell and Oehlsen 2018; Dube et al 2020).

1. "No doubt their physical, mental, and moral ill-health is partly due to other causes than poverty: but this is the chief cause". "The destruction of the poor is their poverty, and the study of the causes of poverty is the study of the causes of the degradation of a large part of mankind". As a man of his time, he uses the term "man" liberally, but no mention of women anywhere except as mothers.

2. For example, "the kindly meant recklessness of the poor law did even more to lower the moral and physical energy of Englishmen than the hard-hearted recklessness of the manufacturing discipline: for by depriving the people of those qualities which would fit them for the new order of things, it increased the evil and diminished the good caused by the advent of free enterprise".

Downloaded from https://academic.oup.com/jeea/article/19/6/2929/6368338 by University Library Utrecht user on 14 January 2022

The questions of interest are shifting from "Do employers have monopsony power over their workers?" to "How much monopsony power?", "How does monopsony power vary?", "What are the consequences of monopsony power?", and "What should be done about monopsony power?" To answer these questions and for monopsony to be more than an abstract idea, we need to be able to assess the impacts of monopsony and that requires the measurement of monopsony power. A variety of measures exist in the literature, some derived from particular models of monopsony and others drawn from measures of market power used in the literature on product market competition, for example, measures of concentration. But there is no unifying view.

There are two main sources of monopsony power, both identified by Joan Robinson, who was the first to discuss the idea formally (Robinson 1933). First, search frictions mean that it takes time (and perhaps money) for workers to find and change jobs: Burdett and Mortensen (1998) is the classic reference for a model of this type. Examples of measures of monopsony based on this model are the wage elasticity of separations (see the meta-study of Sokolova and Sorensen 2021) and the share of recruits from employment (see, e.g., Hirsch et al. 2020). The second theoretical foundation is the idea that workers care about many dimensions of a job not just the wage so that employers are not perfect substitutes for each other. This is typically modelled by assuming there is an idiosyncratic component to the utility a worker gets from an employer and workers are choosing an employer from a discrete choice model (see, e.g., Card et al. 2018). This approach suggests using the elasticity of the labour supply curve (or applicants as in Azar, Berry, and Marinescu 2019) as a measure of monopsony. On the measurement side, we have some papers using concentration ratios in employment or vacancies (Abel, Terneyro, and Thwaites 2018; Azar, Marinescu, and Steinbaum 2019, 2020; Azar et al. 2020; inter alia). There are also other papers by Caldwell and Danieli (2020) and Jarosch, Nimczik and Sorkin (2019) proposing measures of worker outside options.

All these ideas have some merit, but how they fit together is unclear. One of the purposes of this paper is to present a model that encompasses both frictions and idiosyncrasies and in which one can assess the circumstances in which the different proposed measures of monopsony power are valid. And this more general model is useful because more restrictive models often have specific conclusions that derive more from the particular assumptions made than the underlying economics. The approach is illustrated by an empirical exercise, investigating how the separation elasticity varies across the wage distribution, with the arrival rate of job offers and over time.

The plan of the paper is as follows. The next section presents a brief refresher on monopsony and discusses, in very general terms, how one might measure employer monopsony power. Section 3 presents a simple model of worker effort, Section 4 a simple dynamic model, and Section 5 a model of worker quit behaviour that incorporates both idiosyncrasies and frictions. Section 6 estimates separation elasticities and Section 7 what policies we should be using to address monopsony power. Section 8 concludes.

Downloaded from https://academic.oup.com/jeea/article/19/6/2929/6368338 by University Library Utrecht user on 14 January 2022

2. Measuring Monopsony Power

2.1. The Textbook Model of Monopsony

The static textbook model of monopsony has an employer with a revenue function $F(N)$ facing a labour supply curve $N(w)$. The first-order condition for profit maximization can be written as follows:

$$\frac{F'(N) - w}{w} = \frac{1}{\varepsilon_{Nw}}. \tag{1}$$

The degree of monopsony power is given by the elasticity of the labour supply curve facing the firm. Models of monopsony based on idiosyncrasy have this form (e.g. Card et al. 2018) so this way of looking at the problem leads to the measure used, inter alia, in Azar, Berry, and Marinescu (2019), Lamadon, Mogstad, and Setzler (2019), and Kroft et al. (2020). This approach cannot incorporate frictions in any meaningful sense.

2.2. The Comparison of Wages and Marginal Products

The simplest, most direct way, to assess the extent of monopsony power would be a comparison of the wage with the marginal revenue product, that is, try to estimate the left-hand side of equation (1). Because the marginal revenue product is not directly observable, any such approach is likely to base an estimate on an adjustment to the average revenue product.[3] There is a serious identification problem here: Simply considering the wage relative to the average revenue product (which can, alternatively, be written as the labour share or unit labour costs) requires some identifying assumptions about the technology, the mark-down from monopsony power, and the mark-up from monopoly power. For example, there is a large literature seeking to estimate mark-ups and how they have evolved over time (see Basu 2019, for a recent overview), but these studies typically allow only for monopoly power, and the data could be explained equally well by monopsony power: Stansbury and Summers (2020) go further and argue that much of the variation over time is better explained by declining worker power rather than rising product market power.

One way out of this problem is to try to have a measure of monopsony power such as concentration ratios and see how variables like the labour share (or simply the level of wages) are related to this. A number of recent papers do this (e.g. see Abel, Terneyro, and Thwaites 2018; Benmelech, Bergman, and Kim 2020; Berger, Herkenhoff and Mongey 2019; Caldwell and Danieli 2020, for papers that link concentration and wages; Arnold 2020, for a link between mergers and wages; and the review in Manning 2021). There are some potential difficulties with this approach; concentration ratios by industry could be used to measure both product and labour market power. And the

3. A relevant literature here is the one that seeks to estimate the relationship between price and marginal cost that started with Hall (1988).

Downloaded from https://academic.oup.com/jeea/article/19/6/2929/6368338 by University Library Utrecht user on 14 January 2022

use of concentration ratios to measure market power has fallen out of favour in the industrial organization (IO) literature (see Berry, Gaynor, and Morton 2019; Syverson 2019) because it is an endogenous market outcome and can, in some circumstances, be increased when the market becomes more competitive (this is the case in Burdett and Mortensen 1998).

For this reason, the IO literature has moved towards seeking to estimate the elasticity of demand curves as a measure of product market power. This paper will take this approach to deriving measures of monopsony power, focusing on the elasticity of the labour supply curve to the firm, that is, trying to estimate the right-hand side of equation (1).

One problem with this approach is that, to use the terms of Bronfenbrenner (1956), this leads to a measure of potential monopsony power, but there is a question of how much it is exercised. There may be laws and institutions like minimum wages and trade unions that mean that not all potential monopsony power is translated into lower wages, or other outcomes we ultimately care about. But, perhaps more important and pervasive than all of these is the less organized resistance of individual workers and small groups of workers to the exercise of monopsony power. The next section sketches a simple model with this idea.

3. A Simple Model of Worker Effort

In a competitive labour market, your employer tells you what to do, and because the wage is equal to the marginal product, you risk losing your job if you deviate from this and are found out.[4] But if the wage is below the marginal product, there is some scope for shirking without risk of job loss even if detected as long as you remain profitable. Because labour contracts are incomplete (or in Marx's terminology, there is a distinction between labour and labour power), there is likely to be a tussle between worker and employer on how much work is done.[5] Workers may manage to reduce the gap between marginal products and wages by reducing their marginal product.

To see how this might limit the exercise of monopsony power, consider a very simple model. Workers have utility $u = w - e$, where w is the wage and e is the effort on the job. Firms have constant returns to scale, but the marginal product depends positively on effort $p(e)$. The firm has some monopsony power, and its labour supply depends on the utility that the firm offers workers $N(u)$; for simplicity, we assume this is iso-elastic with elasticity ε_{Nu}.

4. The "shirking" class of efficiency wage models (Shapiro and Stiglitz, 1984) assume that monitoring is imperfect so that employers have to ensure workers have something to lose to induce them to work. We abstract from these considerations here.

5. Or, as Homer Simpson puts it, "If you don't like your job, you don't strike! You just go in every day, and do it really half assed. That's the American way."

Downloaded from https://academic.oup.com/jeea/article/19/6/2929/6368338 by University Library Utrecht user on 14 January 2022

If firms can offer a package of the wage and effort, then it will choose the wage and effort to maximize

$$[p(e) - w] N(w - e) = [p(e) - w] (w - e)^{\varepsilon_{Nu}}. \tag{2}$$

Maximizing equation (2) with respect to the effort and wage leads to the effort being set at the efficient level $p'(e) = 1$, and the usual monopsony power formula will apply though in terms of utility and the surplus rather than wages and marginal products, that is, we will have

$$\frac{p(e) - e - u}{u} = \frac{1}{\varepsilon_{Nu}}. \tag{3}$$

But we will assume a different set-up: a two-step model in which the firm can unilaterally choose the wage in a first stage, but the effort is determined in a second stage. While workers can decide, based on the utility offered, whether to work for the employer in the first stage, we assume that in the second stage they will stay if the utility from doing so is greater than zero. We model the effort determination stage as a bargain between employer and worker to reflect the tussle within the workplace over the determination of effort. Given the wage, we assume effort is chosen to maximize an asymmetric Nash bargain:

$$[p(e) - w] [w - e]^{\beta}, \tag{4}$$

where $\beta(\geq 0)$ is a measure of worker power in the effort determination stage. The first-order condition can be written as

$$p'(e) [w - e] = \beta [p(e) - w]. \tag{5}$$

The wage set in the first stage leads to a certain effort level in the second stage denoted by the relationship $e = e(w)$. The first-order condition for the employer's choice of the wage in the first stage can now be written as

$$\frac{p(e) - e - u}{u} = \frac{1}{\varepsilon_{Nu}} \left[\frac{1 - p' \frac{\partial e}{\partial w}}{1 - \frac{\partial e}{\partial w}} \right]. \tag{6}$$

Compared to equation (3), there is an extra term in square brackets. If this term is bigger than 1, then the employer has more monopsony power than the standard formula, while it has less if it is less than 1.

If $\beta = \varepsilon_{Nu}$, then effort is at the efficient level and the standard formula in equation (3) is correct. The intuition is that, in this case, the objective function in the wage determination stage, equation (2), coincides with that in the effort determination stage, equation (4) (see Manning 1987, for this type of model in a union bargaining context). By the envelope condition, the outcome must then be the same as in the one-step model in which the employer unilaterally chooses wage and the efficient level of effort. And profits must be highest in this case. The implication is that employers want the workers to have some bargaining power in effort determination as a commitment device against excessive ex post exploitation of workers. If employers cannot make

Downloaded from https://academic.oup.com/jeea/article/19/6/2929/6368338 by University Library Utrecht user on 14 January 2022

this commitment, then profits will be lower because workers rationally expect they will be treated badly so will not come to work for the firm in the first place.

It is natural to consider the case where workers have more power in effort determination than employers would like (i.e. $\beta > \varepsilon_{Nu}$). Appendix A shows that in this case, we have effort (and productivity) inefficiently low and employers have less effective monopsony power than the standard formula in equation (3) implies.

This suggests that what happens within the firm, outside the labour market, may be very important in determining labour market outcomes. For example, concerns about fairness (see Dube, Giuliano, and Leonard 2019 for evidence on the relevance of this) could be seen as evolved responses to limit the ability of employers to exploit the greater monopsony power they have over incumbent workers or to act as discriminating monopsonists. Otherwise, if you turn up for work and your employer has cut your wage, then there is little that can be done about that. Like many evolved traits, it may also have unintended consequences, for example, nominal wage rigidity.

The model presented here suggests that the impact of monopsony power on worker effort (and hence productivity) within the workplace may be a neglected topic. One reason it is ignored is that it is a hard process to observe, though Coviello, Deserranno, and Persico (2021) show increases in productivity in response to a mandated increase in the minimum wage, which is consistent with the mechanism proposed here.

The rest of this paper follows the tradition of ignoring worker effort and reverts to a model in which effort is no longer present. What is different about this model is that it tries to incorporate both dynamics and an idiosyncratic aspect to preferences.

4. A Model of Dynamic Monopsony

Consider a very simple dynamic model (which can be found in Manning 2003). At any date t, the state variable for the firm will be the labour force that it had in the last period, N_{t-1}. We assume that in each period the firm sets the wage to maximize the present discounted value of profits.[6] To keep things simple, we assume constant returns to labour, p. Define a value function $\Pi(N_{t-1})$ to be the maximized discounted value of future profits from date t onwards. The value function will satisfy

$$\Pi(N_{t-1}) = \max_w (p - w)N_t + \beta\Pi(N_t). \tag{7}$$

The firm is assumed to face a dynamic labour supply curve where the quit rate of existing workers depends on the current wage offered $q(w)$, and there is a flow of recruits $R(w)$ that also depends on the wage. This means that employment evolves according to

$$N_t = [1 - q(w)]N_{t-1} + R(w). \tag{8}$$

6. A "period" here should be interpreted as the length of time for which employers can commit to wages.

Downloaded from https://academic.oup.com/jeea/article/19/6/2929/6368338 by University Library Utrecht user on 14 January 2022

This setting is general enough to incorporate both the frictional and idiosyncratic views of the foundations of employer monopsony power. The case where there are no frictions, only idiosyncrasy is the specification $q(w) = 1$ in which case each period is independent of the previous period, so the model reduces to the static one. If there are only frictions and no idiosyncrasies, then the specification of the quit function will be that the worker quits whenever they receive a wage offer above the current wage. Perfect competition can be thought of as the case where there are no frictions and no idiosyncrasies in which case the quit rate is 1 and recruits 0 if the pay is below the maximum wage offered by other firms.

The dynamic labour supply curve (equation (8)) is written such that worker decisions depend only on the wage offered, a specification that deserves some discussion given that there is also a state variable (the inherited level of employment).[7] There are several possible interpretations of this: Workers may be myopic so they take into account only the current wage in making their decisions. Or they may think the current wage will last forever, which will be the case in the steady-state but not when considering deviations from the steady-state. Or it may be that the current wage is the only available relevant information for basing a current decision on; this would be the case if we had a Markov model.

Appendix B shows that in the steady-state, the relationship between the marginal product and the wage can be written as

$$\frac{(p-w)}{w} = \frac{[1-\beta\,(1-q)]}{q(\varepsilon_{Rw}-\varepsilon_{qw})} = \frac{1}{\mu}. \tag{9}$$

Equation (9) shows the centrality of the recruitment and quit elasticities for the market power of employers, something that is reflected in their use in many papers seeking to measure monopsony power (see Sokolova and Sorensen 2021, for a meta-study). There is a relationship between equations (9) and (1) because the long-run labour supply to the firm can be written as $N(w) = R(w)/q(w)$ so that $\varepsilon_{Nw} = \varepsilon_{Rw} - \varepsilon_{qw}$. But, there is a difference between the appropriate measures of monopsony power in static and dynamic models. If employers discount the future, then the degree of monopsony power will be higher than the textbook static model would predict given the elasticity of the labour supply curve to the firm. The reason is that the wage elasticity of labour supply is lower in the short- than in the long-run. The ratio of the right-hand side of equation (1) to the right-hand side of equation (9) is approximately given by the quit rate divided by the sum of the quit rate and interest rate. As the quit rate is perhaps about 20% and the real interest about 5%, equation (9) implies that employers have approximately 20% more monopsony power than the standard static model would imply.

7. See Manning (2012) for a model in which it is assumed that workers also pay attention to the level of employment.

Downloaded from https://academic.oup.com/jeea/article/19/6/2929/6368338 by University Library Utrecht user on 14 January 2022

This simple dynamic model offers a justification for using quit and recruitment elasticities but does not explain where these come from. To understand these better, we need to provide more details about the worker's behaviour.

5. The Determination of Quit Rates

In this section, we model the quit rate function of the previous section. To do this, we want a model that incorporates both frictions (so it must be dynamic) and idiosyncratic preferences for jobs. We consider one firm currently paying a wage w. Denote by $V(w)$ the value of being employed in this firm at the start of each period.

In each period, we assume that a worker currently employed in the firm gets an idiosyncratic mean zero draw of utility, η (assumed i.i.d. across time, perhaps more for analytical convenience than realism) so that the utility of staying in this firm after the realization of the idiosyncratic shock can be written as

$$u(w) + \eta + \beta_w V(w'(w)) = X(w) + \eta, \tag{10}$$

where β_w is the discount factor of workers, and we use the notation $w'(w)$ to denote the wage in the next period if the current wage is w. This allows for the possibility that the current wage is expected to be permanent but also that it is more transitory.[8] It will be convenient to use the notation $X(w)$ in what follows: It is the expected discounted utility of remaining in the firm, excluding the current idiosyncratic shock.

The worker does not have to remain with this firm. They always have the option of quitting to unemployment, which is assumed to have the present discounted value of X^u. We do not have to specify an idiosyncratic shock for unemployment because in work that can be defined as relative to unemployment. Define the following function:

$$G(X(w), X^u) = E \max \left[X(w) + \eta, X^u\right], \tag{11}$$

where the expectation is over the distribution of η. Models of this type typically make some specific assumption about the distribution of η (e.g. that it is Frechet or a type 1 extreme value distribution), but it is convenient, simpler, and more general to work with the G function as it contains all the information we need.[9] G will be an increasing function of both its arguments, and an equal increase in (X, X^u) leads to an equal increase in G. McFadden (1978) showed that the probability of choosing each option is the derivative of the G function with respect to that argument: These derivatives sum to 1. So, the quit rate to unemployment if the worker has no other option can be written as $G_{X^u} = 1 - G_X$.

8. A complete model would derive the extent to which current wages persist into the future from the behaviour of firms, but we do not seek a complete model here. See Manning (2012) for how changing the assumption about the ability of firms to commit to future wages changes the outcome of the Burdett–Mortensen model.

9. G will obviously depend on the distribution of η, and one could, for any G, derive the implied distribution of η.

Downloaded from https://academic.oup.com/jeea/article/19/6/2929/6368338 by University Library Utrecht user on 14 January 2022

While the worker always has the option of quitting the job for unemployment, we also assume that they may have an option of quitting to another job if they get a better offer. Assume that the "other" jobs are possibly paying different wages but also have an idiosyncratic component. Assume the worker has an M alternative job offers offering wages $(w_1,..., w_M)$ with idiosyncratic components $(\eta_1,..., \eta_M)$, and with functions $(X_1,..., X_M)$ analogous to equation (10). Define the following function:

$$H(X(w), X^u, X_1, .., X_M) = E \max[X(w) + \eta, X^u, X_1 + \eta_1, .., X_M + \eta_M], \quad (12)$$

where the expectation is now over the joint distribution of $(\eta, \eta_1,..., \eta_M)$. Just as for the G function in equation (11), the derivative with respect to the utility from an option is the probability of choosing that option.

The use of the H function allows the incorporation of a range of possibilities. In search models, it is most common to assume that the worker has at most one other job opportunity in each period; this is a special case of the formulation.[10] Search models often do not have any idiosyncrasy, in which case the H function would just be the max of the X functions. On the other hand, models with idiosyncrasy generally assume the worker has the choice of every firm at every moment; again this is a special case, but one could imagine an intermediate case in which only a sub-set of alternative firms is available at each moment. For the moment, we can be a bit vague because all that matters is the dependence of the H function on the wage offered by this firm. So, with some abuse of notation, simply write $H(X(w))$, which should be interpreted as the expectation of the right-hand side of equation (12) over any stochastic elements in it (the number of other job offers and the wages in those offers).

Assume with probability λ a currently employed worker gets alternative job offers. This means that we can write the value of being in a job paying w, $V(w)$, as

$$V(w) = (1 - \lambda)G[X(w)] + \lambda H[X(w)], \quad (13)$$

where we have suppressed the dependence of the G and H functions on other factors. The separation rate for workers will be given by

$$\begin{aligned} q(w) &= (1 - \lambda)\left[1 - \frac{\partial G\,[X\,(w)]}{\partial X}\right] + \lambda\left[1 - \frac{\partial H\,[X\,(w)]}{\partial X}\right], \\ &= (1 - \lambda)\,[1 - G_X] + \lambda[1 - H_X] \end{aligned} \quad (14)$$

that is, 1 minus the probability of remaining with this firm. In what follows, it is sometimes useful to define the quit rate $q^G(X)$ if the worker only has the option of quitting for non-employment and the quit rate $q^H(X)$ if the worker has the option of quitting for another job, that is,

$$q^G(X) = [1 - G_X], \qquad q^H(X) = [1 - H_X]. \quad (15)$$

10. One might think that one can define a period as sufficiently small such that this is always the case, but a period can also be interpreted as the length of wage commitment, and the length of the period cannot serve both purposes.

Downloaded from https://academic.oup.com/jeea/article/19/6/2929/6368338 by University Library Utrecht user on 14 January 2022

Note that some of the quits when the worker has alternative job offers will be to non-employment so these quit functions are not the same as the quits to non-employment and employment, respectively (though it may be a reasonable empirical approximation that workers very rarely quit to unemployment when they have another job offer.)

We can prove the following result about the quit elasticity.

PROPOSITION: *In a steady-state, the elasticity of the quit function can be written as*

$$\varepsilon_{qw} = \left[\alpha \frac{\partial \ln q^G(X)}{\partial X} + (1-\alpha) \frac{\partial \ln q^H(X)}{\partial X} \right] \frac{w u'(w)}{1 - \beta_w \varphi \left[1 - q(w)\right]}, \tag{16}$$

where

$$\alpha = \frac{(1-\lambda) q^G(X)}{\left[(1-\lambda) q^G(X) + \lambda q^H(X)\right]} \tag{17}$$

is the share of quits to unemployment when the worker has no other job offers and $\varphi = \partial w'/\partial w$*, that is, how much workers expect future wages to change when current wages change.*

Proof: See Appendix C.

This result shows that the quit elasticity has a number of sources. The first is the sensitivity of worker mobility decisions to the present discounted value of jobs as represented by the sensitivity of the functions H and G to X. This, in turn, is likely to be influenced by how close substitutes are jobs with different employers, which is often modelled as the degree of idiosyncrasy.

The final term of the right-hand side of equation (16) tells us that the higher the quit elasticity is, the more sensitive the present discounted value of the job to the wage is. This is influenced by the discount factor, the extent to which current wages are expected to persist into the future, and how long the worker expects to be with the firm (quit rate). If the quit rate is high, then labour supply to the firm is likely to be less elastic; this is the opposite to what is often argued, namely that jobs with high turnover are very competitive.[11]

One can also potentially use this set-up to explain why more concentrated labour markets might be associated with more monopsony power. If greater employer concentration leads to fewer alternative wage offers, then it is plausible that this leads to a lower sensitivity of q^H to the value of the job.

An alternative way of modelling a more concentrated labour market is to assume that the arrival rate of alternative job offers, λ, is lower. From equations (15) and (16), a lower level of λ lowers the quit rate; the impact of this has been discussed earlier. But, it also affects the quit elasticity if q^G and q^H have different sensitivities. In many models, this is an assumption. For example, in Burdett and Mortensen (1998), quits to unemployment are assumed to be completely exogenous, unrelated to the wage, in

11. It is the sensitivity of quits to the wage that is important not their level.

Downloaded from https://academic.oup.com/jeea/article/19/6/2929/6368338 by University Library Utrecht user on 14 January 2022

which case q^G does not depend on X. In this case, a lower arrival rate of job offers increases the monopsony power of employers. This is why the fraction of recruits or quits to employment is sometimes used as a measure of monopsony power. But, we can write down models with the opposite result in which q^G is more sensitive than q^H to the wage so this is ultimately an empirical question.

This discussion has been about the wage elasticity of quits, but equation (9) shows that the elasticity of recruits with respect to the wage is also important. This has received much less attention in the literature largely because it is much harder to estimate. In longitudinal data, one can observe how the current wage affects future separations, but one does not directly observe the wage that a worker would receive if they were recruited to some other firm.

The literature often makes use of a result (discussed in Manning 2003, 2011) that for job-to-job quits there will be a relationship between the recruitment elasticity to this firm and the quit elasticity of the firms from which it recruits. The intuition is that one firm's quit is another firm's recruit. This result is derived under the assumption that when a worker has to choose between two firms, the wages influence choice only through the relative wage and do not depend on which firm the worker is currently employed in. Although many search models have this feature, it is important to be aware of the limitations of this result. First, the link between quit and recruitment elasticities is broken if the worker has a choice of more than two options as in the models with idiosyncrasy. And it is also broken if there are mobility costs so that the choice between two firms depends on which firm the worker is currently employed by. What this means is that it is probably important to have separate estimates of the recruitment elasticity. Some papers (e.g. Azar, Berry, and Marinescu 2019; Dal Bó, Finan, and Rossi 2013) estimate the wage elasticity of applications though applications are not the same as recruits. Hirsch et al. (2021) show how the recruitment elasticity can be estimated using estimated fixed firm effects on wages.

While the recruitment elasticity is an important under-researched area, the rest of this paper discusses some estimates of separation functions.

6. The Estimation of Separation Rates

The previous section has laid out the centrality of the elasticity of the separation rate with respect to the wage for understanding the extent of employer market power, whether that power is rooted in idiosyncrasy or frictions. There is a sizeable literature on estimating these elasticities; see the meta-study of Sokolova and Sorensen (2021). This literature has explored the level of the separation elasticity, and how it varies across demographic groups (e.g. gender or immigrant status); Manning (2021) provides a review of some of these studies.[12]

12. Might wonder about the recruitment elasticity. Appendix A discusses how this is related to the quit elasticity, though it is harder to estimate directly.

Downloaded from https://academic.oup.com/jeea/article/19/6/2929/6368338 by University Library Utrecht user on 14 January 2022

The instantaneous separation rate is modelled as $e^{\beta x}$, where x are suitable controls, including the main variable of interest, the log wage. In most datasets, one only observes whether one is in the same job after some interval of time t. This leads to a model for the probability of separating between the two observed quarters as

$$\Pr(S = 1 \,|x) = e^{e^{\beta x} t}. \tag{18}$$

This is a complementary log–log model for the binary outcome of having a separation, which, compared to linear probability or logit models, has the advantage that it can be rooted in a hazard model and the estimated coefficients can be interpreted directly as the relevant elasticity.

In estimating the wage elasticity of separations, it is obviously important to control adequately for other relevant factors. It is likely that separations depend on the wage in the current job relative to wages in plausible alternative jobs, perhaps proxied by characteristics of the worker that we know are relevant for the average level of wages. A failure to control for the alternative wage is likely to lead to a downward bias in the wage elasticities. On the other hand, separations are likely to be more sensitive to the permanent component of wages than to the part of wages that is a transitory shock or measurement error. In this case, the inclusion of controls correlated with the permanent wage is likely to reduce the estimated wage elasticity. In addition, Manning (2003, chap. 4) shows that unobserved heterogeneity in the separation rate will also cause a bias even if uncorrelated with the wage, because the model is a non-linear one.

Ideally, one would have exogenous variation in the wage and use this to estimate the separation elasticity. This is hard because one needs exogenous variation in wages at the firm level. The interest here is less in the estimate of the level of the separation elasticity but in how it varies. If the biases are constant, then the conclusions about his variation will be valid even if the level is not.

We investigate the separation elasticity using data from the United Kingdom and the United States. For the United Kingdom, the data source is the UK Quarterly Labour Force Survey for 1997–2020. In this survey, addresses are sampled for five consecutive quarters with earnings information being asked in waves 1 and 5. For estimating the separation elasticity, we use the first-quarter data and estimate a model for the probability of leaving the current job by the time of the next quarter. For the United States, we use the Survey of Income and Program Participation (SIPP) for 1996–2016. This is a monthly survey so the model is estimated for the probability of leaving the current job in the next month.

For both data sources, we use the log of the hourly wage as the measure of wages, and residualize it by estimating a standard earnings function using as controls age, gender, ethnicity, marital status and the number of dependent children, region, occupation, industry (both at the two-digit level), and (for the United Kingdom) whether the job is temporary. The residualized wage is the estimated residual from this regression and can be interpreted as an estimate of whether the job is high- or low-paying in a particular labour market (defined by the controls).

Downloaded from https://academic.oup.com/jeea/article/19/6/2929/6368338 by University Library Utrecht user on 14 January 2022

TABLE 1. Descriptive statistics.

	UK QLFS		US SIPP	
	Quarterly frequency		Monthly frequency	
Variable	Mean	Standard deviation	Mean	Standard deviation
Observations	452,618		6,421,560	
Separation	0.097	0.296	0.043	0.202
To employment if separate	0.728	0.445	0.514	0.500
Residualized log wage	0.000	0.401	0.000	0.536
Experience (decades)	2.315	1.258	1.948	1.256
Qual 1: college graduate	0.340	0.474	0.341	0.474
Qual 2: A level/some college	0.253	0.434	0.116	0.320
Qual 3: GCSEs/high school graduate	0.249	0.433	0.440	0.496
Qual 4: no qualification/high school dropout	0.158	0.365	0.104	0.305
Male	0.482	0.500	0.502	0.500
Experience (decades)	2.31	1.26	3.95	1.25
White	0.941	0.235	0.814	0.389
Mixed	0.005	0.074	–	–
South Asian	0.026	0.158	–	–
Black	0.015	0.121	0.119	0.324
Chinese	0.003	0.053	–	–
Asian	–	–	0.032	0.176
Other ethnicity	0.010	0.098	0.035	0.184
Married	0.678	0.467	0.558	0.497
Number of dependent children	0.784	1.019	–	–
Number of kids under 18 in the household	–	–	0.859	1.144
Permanent Job	0.948	0.222	–	–
Job tenure (decades)	0.837	0.862	0.663	0.804
Year	2006.221	6.268	2006.4	6.116

Notes: For the education variable, the first variable name denotes UK category and the second variable name denotes the US category.

Table 1 presents some descriptive statistics on these variables. For the United Kingdom, 9.7% leave the current job in 3 months, and for the United States, 4.3% leave in the next month.

We use this basic specification to investigate a number of issues.

6.1. How the Separation Rate Varies with the Residualized Wage

The first is how the separation rate varies with the residualized wage. This is an interesting question because it tells us about whether firms at the top or the bottom of the labour market facing more or less elastic labour supply. Webber (2015) found that employers have more monopsony power in the lower part of the earnings distribution. To this end, we define a categorical variable that rounds the residualized wage to the

Downloaded from https://academic.oup.com/jeea/article/19/6/2929/6368338 by University Library Utrecht user on 14 January 2022

TABLE 2. The distribution of residualized log wages.

	Percent	
Residualized log wages	United Kingdom	United States
−1	2.39	2.88
−0.9	0.44	0.93
−0.8	0.64	1.38
−0.7	1.08	1.98
−0.6	1.74	2.84
−0.5	2.9	4.03
−0.4	4.72	5.48
−0.3	7.13	6.99
−0.2	10.01	8.36
−0.1	12.38	9.39
0	13.28	9.83
0.1	12.4	9.49
0.2	10.16	8.60
0.3	7.46	7.26
0.4	5.08	5.72
0.5	3.24	4.28
0.6	2.01	3.08
0.7	1.26	2.16
0.8	0.81	1.51
0.9	0.52	1.01
1.0	0.35	2.80

Notes: This is the distribution across categories of the residualized log wage. The category is to the nearest log points.

nearest 10 log points, with all those more than 100 log points from the mean combined in categories. The fractions in each category are reported in Table 2.

Figure 1(a) presents the estimated log quit rate compared to the base category of being within +/–100 log points of 0 together with the 95% confidence interval bands for the United Kingdom, while Figure 1(b) does the same for the United States. The slope of this relationship can be interpreted as the quit elasticity. Both countries have a similar shape: The relationship between the log separation and the wage is strongest in the middle of the distribution and weakest at the extremes. The implication is that employers at the bottom and top of the wage distribution have more monopsony power. This empirical finding is not in line with some of the canonical models used in the existing literature. For example, in the Burdett and Mortensen (1998) model, monopsony power falls as one moves up the wage distribution; in the simplest version of the model with no productivity differences across employers, this has to be the case in equilibrium as the highest wage firms have, by definition, the smallest gap between wages and marginal products. And in the multinomial logit model of idiosyncrasy where individual employers are infinitesimal, the elasticity is constant through the range. And, in models where they are not infinitesimal, the larger the firm is, the typically smaller the labour supply elasticity is (see, e.g., Berger, Herkenhoff, and

Downloaded from https://academic.oup.com/jeea/article/19/6/2929/6368338 by University Library Utrecht user on 14 January 2022

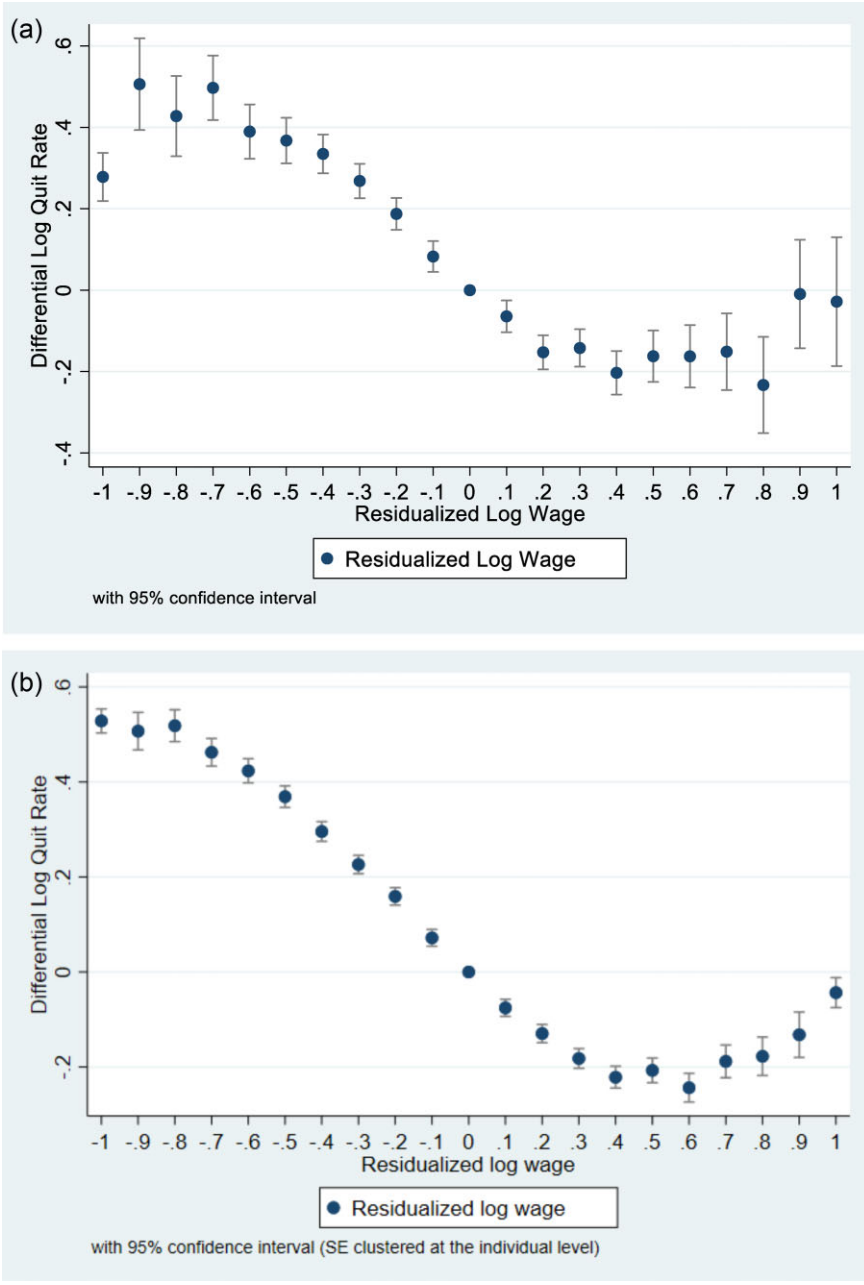

FIGURE 1. (a): Log separations and residualized wage: United Kingdom. (b): Log separations and residualized wage: United States.

Downloaded from https://academic.oup.com/jeea/article/19/6/2929/6368338 by University Library Utrecht user on 14 January 2022

Mongey 2019). But, the estimated relationship does make some sense if there is a wage below which almost no workers are prepared to work but also a natural limit to how many workers can be hired in the short-run whatever the wage paid.

The shape of the relationship between the quit elasticity and the wage also has implications about pass-through from firm productivity to wages. If the labour supply elasticity is constant, then wages are a constant mark-down on productivity so pass-through is 1. If the elasticity falls as the wage rises, then pass-through will be less than 1 as is typically found in empirical studies (see, e.g., Howell and Brown 2020; Kline et al. 2019).

6.2. How the Separation Elasticity Varies with the Job Offer Arrival Rate

As discussed previously, the impact of a changing job offer arrival rate on the monopsony power possessed by employers depends on whether quits to employment are more sensitive to wages than quits to unemployment. We can investigate this by considering how the probability of separations to employment varies with the wage. If the separation rate to employment is $q^E(w)$ and the separation rate to unemployment is $q^U(w)$, then the share of quits to employment can be written as

$$\gamma(w) = \frac{q^E(w)}{q^E(w) + q^U(w)} = \frac{e^{ln[q^E(w)] - ln[q^U(w)]}}{1 + e^{ln[q^E(w)] - ln[q^U(w)]}} \tag{19}$$

so that how the share varies with the wage tells us about whether separations to employment or unemployment are more elastic. We estimate this model with a logit model, again using the categories of residualized wages. The estimated log odds ratio as a function of the residualized wage results is shown in Figure 2(a) for the United Kingdom and in Figure 2(b) for the United States.

For both countries, the shape is the same. The fact that the log odds are increasing in the residualized wage implies that that separations to employment are less sensitive to the wage than quits to unemployment.[13] This is the opposite of what is predicted from the Burdett–Mortensen model (1998). This relationship is strongest in the middle part of the wage distribution and weak at the top. If separations to employment are less sensitive to the wage, then anything that increases the job offer arrival rate is likely to increase monopsony power as it increases the share of separations to employment.

The finding that separations to employment have a lower wage elasticity than separations to non-employment is not common across all studies; for example, Bassier, Dube and Naidu (2020) find the opposite result. And one conceptual problem is that the models used only have voluntary, not involuntary, separations (i.e. firing). If, as is plausible, firing is more common among low-wage workers, then this will make separations to non-employment more sensitive to the wage, but it is not clear whether this should be included in computations of the elasticity of the labour supply curve

13. This is not the same as the recruits from unemployment.

Downloaded from https://academic.oup.com/jeea/article/19/6/2929/6368338 by University Library Utrecht user on 14 January 2022

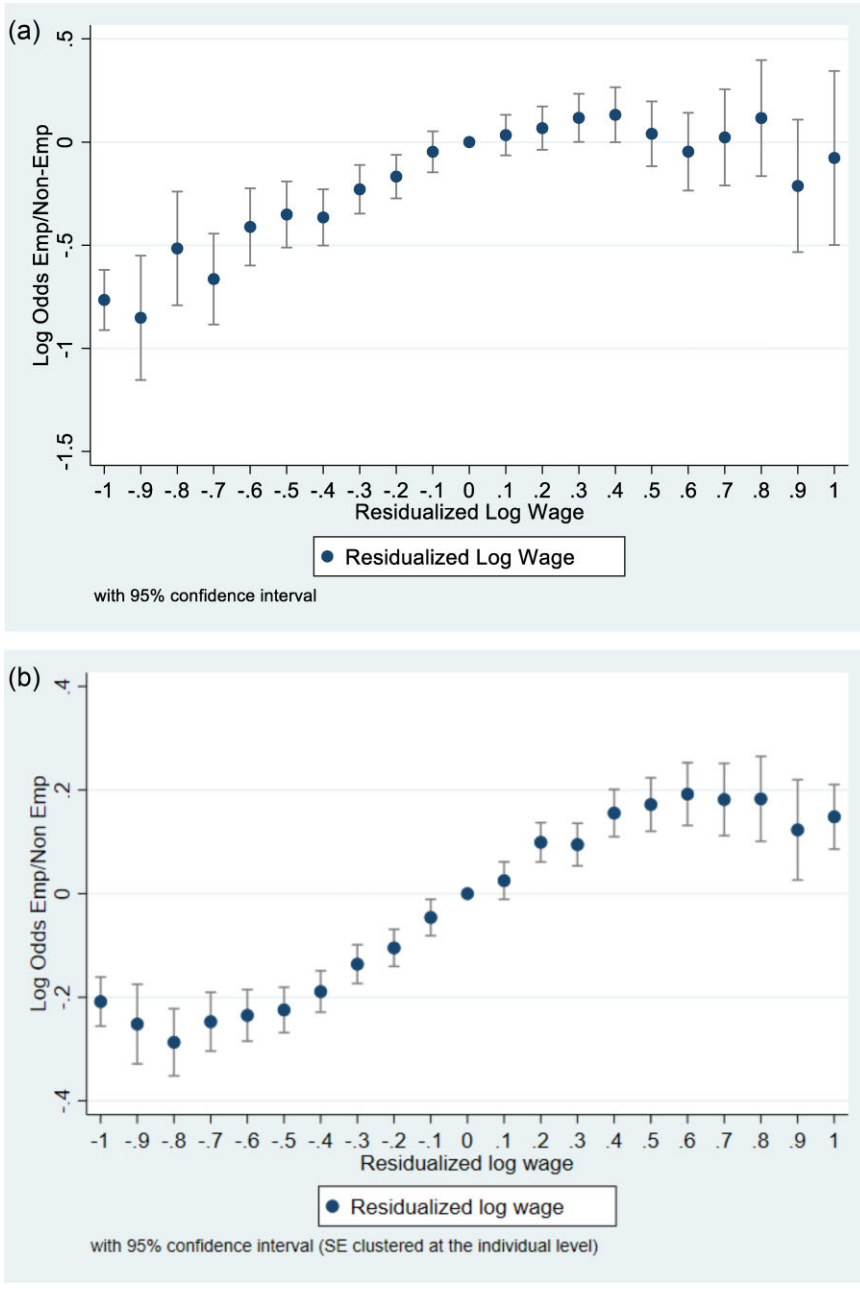

FIGURE 2. (a): Estimated log odds of probability of separation to employment: United Kingdom. (b): Estimated log odds of probability of separation to employment: United States.

Downloaded from https://academic.oup.com/jeea/article/19/6/2929/6368338 by University Library Utrecht user on 14 January 2022

facing a firm. How to think about involuntary separations in these models remains on the "to do list".

Some studies directly consider how the separation elasticity varies over the business cycle with both Hirsch, Jahn, and Schnabel (2018) and Bassier, Dube, and Naidu (2020) reporting evidence that the separation elasticity is higher in booms.

6.3. How the Separation Elasticity Varies over Time

There is a concern, particularly in the United States (see, e.g., Philippon 2019), that markets (primarily product markets) have been becoming less competitive over time and that this can account for, among other things, the decline in the labour share in the United States. Stansbury and Summers (2020) argue that declining worker power is a better explanation of these trends, though this is seen as a result of the decline in institutions that give workers some power to share rents rather than the labour market becoming less competitive. The importance of institutions is also emphasized by Bivens, Mishel, and Schmitt (2018), and Rinz (2018) finds no evidence that concentration ratios in labour markets have been rising.[14] Bell, Bukowski, and Machin (2018) find a lower level of rent-sharing over time in the United Kingdom. None of these studies considers trends in separation elasticities: Hirsch, Jahn and Schnabel (2018) do for Germany and find evidence of cyclicality but no trend. In this section, we consider how the separation elasticity has varied over time in the United Kingdom and the United States.

Because it is hard to keep track of trends in the categories of residualized wages, we simply use the log residualized wage as a regressor and estimate year by year for the period 1997–2018 for the United Kingdom and 1996–2016 for the United States. Figure 3(a) shows the estimated value of the separation elasticity for the United Kingdom and Figure 3(b) for the United States, multiplied by -1 (so it is a positive number). The separation elasticity has been becoming less negative over time in both countries, indicative of rising monopsony power.

A quantitative estimate of this trend is presented in Table 3. The years are pooled together, the same controls are used as for the previous separation models, and an interaction of the residualized log wage with year is included. This coefficient is multiplied by 10 to give an estimate of the change over a decade and centred on the year 2000 so the linear term can be thought of as an estimate in that year.

The first column of Table 3 shows the estimate without a trend, while the second column shows the inclusion of the trend, which is significantly different from 0. The natural next question is: "Why?" The first possible explanation is demographic change. Although the controls allow for characteristics to affect the level of separations, they may also affect the wage elasticity of separations. To investigate this, column 3 of

14. Even if monopsony power has not changed, it may play a role in explaining economy-wide trends in outcomes. If outcomes $Y = f$(monopsony power, institutions), then the cross-partial is important, and monopsony may increase the explanatory power of changes in institutions. Monopsony power is the catalyst that enables institutional changes have their effect.

Downloaded from https://academic.oup.com/jeea/article/19/6/2929/6368338 by University Library Utrecht user on 14 January 2022

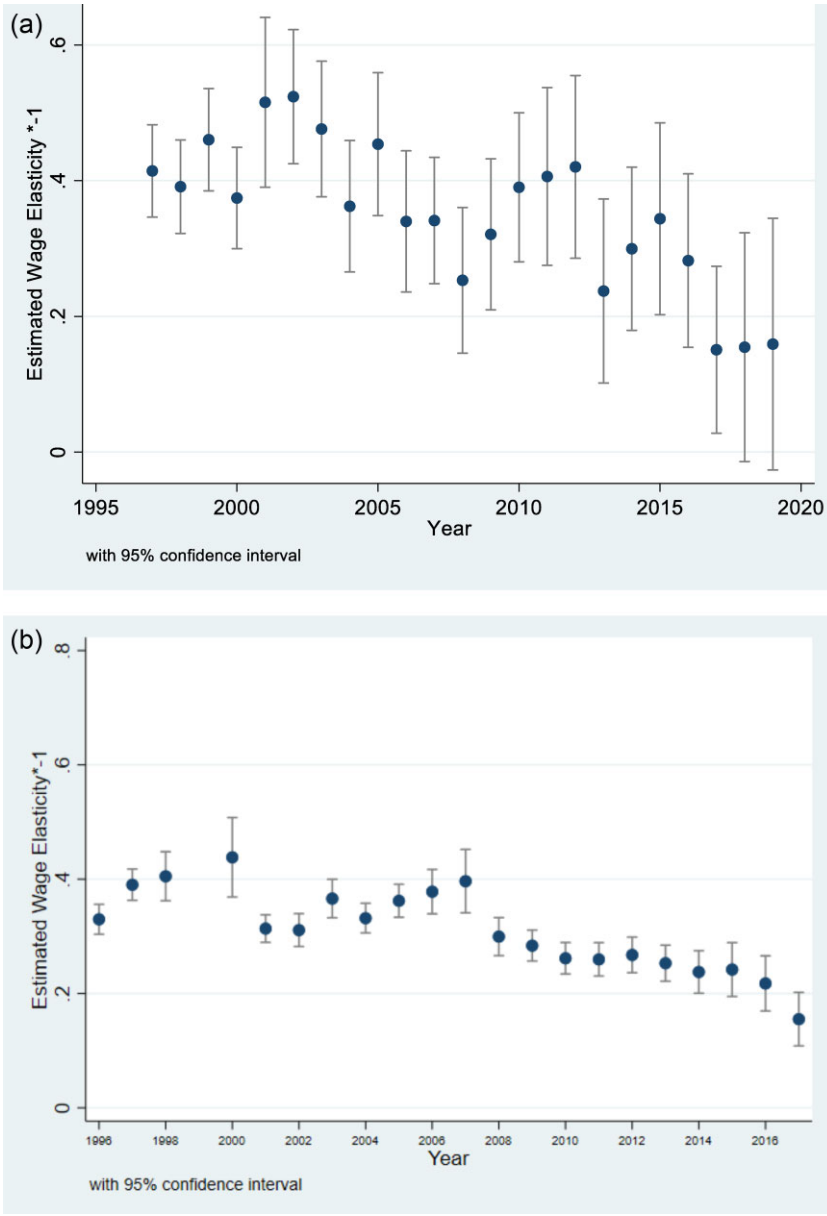

FIGURE 3. (a): Changing separations elasticity over time: United Kingdom. (b): Changing separations elasticity over time: United States.

Table 3 includes interactions with other covariates. Some of these other interactions are significant, implying that the separation elasticity varies with demographics. This is not always in the direction one might expect; for example, men in both the United Kingdom and the United States are estimated to have lower separation elasticities than

Downloaded from https://academic.oup.com/jeea/article/19/6/2929/6368338 by University Library Utrecht user on 14 January 2022

TABLE 3. The variation of the separation elasticity over time.

	UK-QLFS			US-SIPP		
	(1)	(2)	(3)	(4)	(5)	(6)
Residualized wage	−0.370***	−0.414***	−0.351***	−0.308***	−0.344***	−0.383***
	(0.011)	0.013)	(0.028)	(0.004)	(0.005)	(0.010)
Residualized wage* experience			−0.021***			−0.010
			(0.008)			(0.030)
Residualized wage* gender (male)			0.113***			0.032***
			(0.021)			(0.008)
Residualized wage* Qual 2			−0.115***			0.003
			(0.029)			(0.014)
Residualized wage* Qual 3			−0.119***			0.016*
			(0.028)			(0.009)
Residualized wage* Qual 4			−0.101***			0.118***
			(0.034)			(0.013)
Residualized wage* not-White			0.040			0.016*
			(0.044)			(0.010)
Residualized wage* year		0.086***	0.078***		0.064***	0.064***
		(0.016)	(0.017)		(0.006)	(0.006)
Constant	−1.576***	−1.569***	−1.558***	−3.765***	−3.777***	−4.131***
	(0.130)	(0.130)	(0.130)	(0.094)	(0.094)	(0.094)
Observations	452,582	452,582	452,582	6,421,560	6,421,560	6,421,560

Notes: These are estimates from a complementary log–log model in which the other controls whose coefficients are not reported are the same as those used to estimate the residualized log wage. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Downloaded from https://academic.oup.com/jeea/article/19/6/2929/6368338 by University Library Utrecht user on 14 January 2022

women. And there are some differences between the countries; for example, the lowest education group has the highest estimated elasticity in the United Kingdom but the lowest in the United States. However, the significant trend is robust to allowing for differences in separation elasticities across demographic groups.

So, the decline in the separation elasticity does not seem to be the consequence of changing demographics. The explanation for separations becoming less sensitive to the wage is not entirely clear and is not found in all other studies. But, it could be one further aspect of the declining dynamism in labour markets, which has been documented elsewhere (e.g. Molloy et al. 2016).

7. Policies to Address Monopsony Power

If monopsony power is pervasive (and perhaps increasing over time), then we should think about policies to address it.

Monopsony is most often mentioned when discussing the minimum wage. A minimum wage can address monopsony by preventing employers from converting their market power into lower wages. The minimum wage has its place but is inevitably a blunt instrument and can only address issues of monopsony power at the bottom part of the labour market, while the evidence suggests monopsony power is more pervasive. So, some other policy tools are necessary.

As monopsony is about competition in labour markets, it is natural to consider the use of competition policy to address the problems of monopsony, that is, to use the state as a source of counter-vailing power. On paper, the anti-trust laws of many countries treat competition in labour and product markets symmetrically. The practice is different; for example, Shapiro (2019) points out that no mergers in the United States have been blocked because of the impact on competition in the labour market. But, it may be that more thought needs to be given to how to apply competition policy in labour markets; for example, industry may be an appropriate way of thinking about product market competition, but many labour markets cut across product markets (though there are exceptions e.g. in healthcare and airlines), and labour markets are much more local than product markets (Manning and Petrongolo 2017). There are currently many proposals about how to correct the "historic imbalance" (Naidu, Posner, and Weyl 2018) between competition policy in product and labour markets (see, e.g., Krueger and Posner 2018; Marinescu and Hovenkamp 2019; Marinescu and Posner 2019; Naidu, Posner, and Weyl 2018).

Most competition laws prohibit collusive anti-competitive behaviour among employers, but there are serious questions about how much this has been enforced. Adam Smith (1970) had the view that employers "are always and everywhere in a sort of tacit, but constant and uniform combination, not to raise the wages of labour above their actual rate. To violate this combination is everywhere a most unpopular action, and a sort of reproach to a master among his neighbours and equals". And... "We rarely hear, it has been said, of the combination of masters; though frequently of those of workmen. But whoever imagines, upon this account, that masters rarely combine,

Downloaded from https://academic.oup.com/jeea/article/19/6/2929/6368338 by University Library Utrecht user on 14 January 2022

is as ignorant of the world as of the subject." Some recent cases involving high-profile companies[15] have made us aware that these are current and not just historic problems.

More thought also needs to be put in the regulation of employment contracts. The aspect of this that has probably received the most attention is non-competes (sometimes also called restrictive covenants), clauses that limit the ability of workers leaving firms to work for other firms, typically in the same field or area for a limited period (see, e.g., Krueger and Ashenfelter 2018; Marx 2011; Prescott, Bishara, and Starr 2016; Starr, Prescott, and Bishara 2019).

Restrictions on allowable labour contracts can be justified because they reduce the monopsony power of employers. Manning (2003) shows that if one mandates minimum standards that are normal goods, then this has the impact of raising worker utility and efficiency in a monopsonistic labour market because it raises the elasticity of the labour supply curve to the firm reducing monopsony power.

Restrictions on allowable contracts could also be justified by the argument that there are externalities; the possible future employers of a worker may have an interest unrepresented in any negotiations between the worker and their current employer. If a clause in my employment contract limits my ability to move to other employers, then that may make their labour supply less elastic, leading to more monopsony power in the labour market as a whole.

But, over and above these justifications for restricting allowable labour contracts are questions about whether any contract agreed between two consenting parties should be enforceable. The economists' approach is often that if the agreement is voluntary, then there is a presumption that both parties are better off because they are the best judge of their own welfare. But, there are questions about the circumstances in which this is a reasonable approximation to reality (not just in labour markets—see Schoar 2021). Economists are often happy to assume that people differ in their ability when modelling wage inequality but, with the exception of behavioural economics, much more reluctant to assume that people differ in their ability to maximize their utility. For a discipline that prides itself on its logical consistency, this does seem something of an oversight. And it has important implications about the view of the merits of voluntary trade between agents who may differ in their decision-making abilities. Problems may often occur where we have interactions between what we might call amateurs and professionals. An employer will have many more workers than a typical worker has employers so the employer might be expected to be more skilled (and have more incentives to become more skilled) in manipulating the terms of the employment contract to their advantage and to the disadvantage of workers. One possible example of this is that workers may not always read the "small print" of the contracts they are signing: Lipsitz and Starr (2021) show larger effects of non-competes on labour mobility and reductions in wages when they are presented to workers after accepting the job offer.

The current attitude to the details of labour contracts is very different from the attitude taken in product markets. For example, we would not dream to allow my

15. See, for example, https://www.lieffcabraser.com/antitrust/high-tech-employees/.

Downloaded from https://academic.oup.com/jeea/article/19/6/2929/6368338 by University Library Utrecht user on 14 January 2022

current supermarket to only sell to me on the condition that if I want to do my grocery shopping elsewhere, I can only shop in more distant shops. Yet, in the United Kingdom, it has been ruled acceptable for the employer of a hairdresser to restrict their ability to work for other hairdressers within half a mile. That may not sound like a very restrictive condition, but, according to the Labour Force Survey, 25% of hairdressers have a commute of less than 5 minutes and 50% less than 10 minutes. Labour markets are very local so these restrictions may be considerable. One way of thinking about these restrictive covenants is that employers are seeking to appropriate the human capital of workers (see Lobel 2013, for a wider discussion of these issues).

Another example would be notice periods, which increase the amount of time it takes for workers to change jobs. As these are switching/mobility costs, one would expect them to increase frictions in the labour market, making it less competitive. In product markets, competition policy often seeks to reduce or eliminate switching costs. In the labour market, we accept them as normal; it may be that there are pro-competitive justifications, but the important point is that we do not consider the issue at all.

Using the power of the state to set minimum wages and regulate employment contracts is important but has its limitations as they are inevitably blunt instruments and there is a limit to the effectiveness of "top-down" policies when enforcement may be difficult. There is also a potentially important role for "bottom-up" policies to give counter-vailing power of workers through trade unions or worker board-level representation. There is a recent resurgence of interest in this area (see Jäger, Schoefer, and Heining 2021, for a recent assessment of German co-determination) though more needs to be done.

8. Conclusions

As the idea that employers have considerable market power over their workers comes to be widely accepted, we need to have measures of the extent of that power, how it varies within and across labour markets, how it has changed over time, and what can be done to address the problems it causes. While this paper has offered some views on many of these issues, there is much more that needs to be done. And it does need to be done if economists are to make sure that our economy works for the benefit of all.

Appendix A. The Effort Determination Model

We first show how equation (5) implies that $\partial e/\partial w \geq 0$. Differentiating equation (5) leads to

$$p''(e)\,[w-e]\,\frac{\partial e}{\partial w} + p'(e)\left[1-\frac{\partial e}{\partial w}\right] = \beta\left[p'(e)\frac{\partial e}{\partial w} - 1\right]. \tag{A.1}$$

Downloaded from https://academic.oup.com/jeea/article/19/6/2929/6368338 by University Library Utrecht user on 14 January 2022

Re-arranging equation (A.1) leads to

$$\frac{\partial e}{\partial w} = \frac{p' + \beta}{p' + \beta p' - p''[w - e]}. \tag{A.2}$$

The denominator must be positive from the second-order condition, so this implies $\partial e/\partial w \geq 0$.

Comparing equations (5) and (6), we also have that

$$\frac{p'}{\beta} = \frac{1}{\varepsilon_{Nu}}\left[\frac{1 - p'\frac{\partial e}{\partial w}}{1 - \frac{\partial e}{\partial w}}\right], \tag{A.3}$$

which can be re-arranged to give

$$p' = \frac{\beta}{\beta\frac{\partial e}{\partial w} + \varepsilon_{Nu}\left(1 - \frac{\partial e}{\partial w}\right)}. \tag{A.4}$$

We are interested in the case where $\beta > \varepsilon_{Nu}$ and will show that this implies $p' > 1$. Suppose that not $p' \leq 1$. From equation (A.4), this can only be the case if $\partial e/\partial w > 1$. Equation (A.2) then implies that $p' < 1$, a contradiction. This then implies that $\partial e/\partial w < 1$.

Appendix B. Deriving the Mark-Down in the Simple Dynamic Model

Taking the first-order condition of equation (7) with respect to w and taking account of the dependence of N_t on w leads to the first-order condition

$$[p - w + \beta\Pi'(N_t)]\frac{\partial N_t}{\partial w} = N_t. \tag{B.1}$$

We also have the envelope condition that allows us to derive the derivative of the value function. Differentiating equation (7), we have that

$$\Pi'(N_{t-1}) = \left[p - w + \beta\Pi'(N_t)\right]\frac{\partial N_t}{\partial N_{t-1}} = [p - w + \beta\Pi'(N_t)](1 - q). \tag{B.2}$$

In a steady-state[16] where wages and employment are constant, we can solve equation (B.2) for $\Pi'(N)$, which leads to

$$\Pi'(N) = \frac{(1 - q)(p - w)}{1 - \beta(1 - q)}. \tag{B.3}$$

Substituting equation (B.3) into equation (B.1) and re-arranging leads to

$$\frac{(p - w)}{1 - \beta(1 - q)}\frac{\partial N_t}{\partial w} = N_t. \tag{B.4}$$

16. This might seem innocuous, but Manning (2012) shows it is not by means of presenting a one-period version of the Burdett–Mortensen model where no steady-state equilibrium with constant wages and employment exists.

Downloaded from https://academic.oup.com/jeea/article/19/6/2929/6368338 by University Library Utrecht user on 14 January 2022

By differentiating equation (8), we have that

$$\frac{\partial N_t}{\partial w'} = R'(w') - q'(w')N_{t-1}. \tag{B.5}$$

In a steady-state, equation (B.5) can be written as

$$w\frac{\partial N_t}{\partial w} = R\varepsilon_{Rw} - qN\varepsilon_{qw} = qN(\varepsilon_{Rw} - \varepsilon_{qw}). \tag{B.6}$$

Substituting equation (B.6) into equation (B.4) then leads to equation (9).

Appendix C. Proof of Proposition about the Quit Elasticity

Differentiating equation (14), we have that

$$q'(w) = -[(1-\lambda)\, G_{XX} + \lambda H_{XX}]X'(w). \tag{C.1}$$

From equation (10), we have that

$$X'(w) = u'(w) + \beta_w \frac{\partial w'}{\partial w} V'(w'(w)), \tag{C.2}$$

where $\partial w'/\partial w$ is a measure of how much the future wage is expected to change when the current wage changes. Denote this derivative by φ. And from equation (13), we have

$$V'(w) = [(1-\lambda)G_X + \lambda H_X]X'(w). \tag{C.3}$$

Combining equations (C.2) and (C.3), and using equation (14), we can derive

$$X'(w) = \frac{u'(w)}{1 - \beta_w \varphi\,[1 - q(w)]}. \tag{C.4}$$

This has the intuition that workers discount the wages paid by this firm by their discount factor times the probability they remain with the firm. If the quit rate is very high, then an increase in wages is worth less to the workers. Substituting equation (C.4) into equation (C.1), the elasticity can be expressed as

$$\varepsilon_{qw} = -\frac{[(1-\lambda)G_{XX} + \lambda H_{XX}]}{[(1-\lambda)(1-G_X) + \lambda(1-H_X)]} \frac{wu'(w)}{1 - \beta_w \varphi[1 - q(w)]}. \tag{C.5}$$

Equation (C.5) can be written as

$$\varepsilon_{qw} = -\left[\alpha \frac{G_{XX}}{(1-G_X)} + (1-\alpha)\,\frac{H_{XX}}{(1-H_X)}\right] \frac{wu'(w)}{1 - \beta_w \varphi\,[1 - q(w)]}, \tag{C.6}$$

where

$$\alpha = \frac{(1-\lambda)G_X}{[(1-\lambda)(1-G_X) + \lambda(1-H_X)]}, \tag{C.7}$$

that is, the share of quits who did not have another job offer. Using equation (15), this proves the proposition.

Downloaded from https://academic.oup.com/jeea/article/19/6/2929/6368338 by University Library Utrecht user on 14 January 2022

References

Abel, Will, Silvana Tenreyro, and Gregory Thwaites (2018). "Monopsony in the UK." CFM Discussion Paper No. 1827. http://www.centreformacroeconomics.ac.uk/Discussion-Papers/2018/CFMDP2018-27-Paper.pdf.

Arnold, David (2020). "Mergers and Acquisitions, Local Labor Market Concentration, and Worker Outcomes." Working paper. https://scholar.princeton.edu/sites/default/files/dharnold/files/jmp.pdf.

Azar, José, Steven Berry, and Ioana Elena Marinescu (2019). "Estimating Labor Market Power." SSRN Working Paper No. 3456277.

Azar, José, Ioana Marinescu, and Marshall Steinbaum (2019). "Measuring Labor Market Power Two Ways." *AEA Papers and Proceedings*, 109, 317–321.

Azar, José, Ioana Marinescu, and Marshall Steinbaum (2020). "Labor Market Concentration." *Journal Human Resources*, 56, 1218–9914R1.

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

Bassier, Ihsaan, Arin Dube, and Suresh Naidu (2020). "Monopsony in Movers: The Elasticity of Labor Supply to Firm Wage Policies." NBER Working Paper No. 27755.

Basu, Susanto (2019). "Are Price-Cost Markups Rising in the United States? A Discussion of the Evidence." *Journal of Economic Perspectives*, 33(3), 3–22.

Bell, Brian, Pawel Bukowski, and Stephen Machin (2018). "Rent Sharing and Inclusive Growth." CEP Discussion Paper No. DP1584, Centre for Economic Performance, London School of Economics and Political Science.

Benmelech, Efraim, Nittai K. Bergman, and Hyunseob Kim (2020). "Strong Employers and Weak Employees: How Does Employer Concentration Affect Wages?" *Journal Human Resources*, 56, 0119–10007R1.

Berger, David W., Kyle F. Herkenhoff, and Simon Mongey (2019) "Labor Market Power." NBER Working Paper No. 25719.

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

Bivens, Josh, Lawrence Mishel, and John Schmitt (2018). "It's Not Just Monopoly and Monopsony: How Market Power Has Affected American Wages." Working paper. Economic Policy Institute.

Bronfenbrenner, Martin (1956). "Potential Monopsony in Labor Markets." *ILR Review*, 9, 577–588.

Burdett, Kenneth and Dale T. Mortensen (1998) "Wage Differentials, Employer Size, and Unemployment." *International Economic Review*, 39, 257–273.

Caldwell, Sydnee and Oren Danieli (2020). *Outside Options in the Labor Market*. Manuscript, Harvard University.

Caldwell, Sydnee and Emily Oehlsen (2018). "Monopsony and the Gender Wage Gap: Experimental Evidence from the Gig Economy." Working paper, Massachusetts Institute of Technology.

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

Coviello, Decio, Erika Deserranno, and Nicola Persico (2021). "Minimum Wage and Individual Worker Productivity: Evidence from a Large US Retailer." Working paper, Searle Center for Law, Regulation, and Economic Growth, Northwestern University. https://drive.google.com/file/d/0B1g6WDc8t1zLUVpEU0ZSdFUySUE/view.

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

Dube, Arindrajit, Laura Giuliano, and Jonathan Leonard (2019). "Fairness and Frictions: The Impact of Unequal Raises on Quit Behavior." *American Economic Review*, 109(2), 620–663.

Dube, Arindrajit, Jeff Jacobs, Suresh Naidu, and Siddarth Suri (2020). "Monopsony in Online Labor Markets." *American Economic Review: Insights*, 2(1), 33–46.

Downloaded from https://academic.oup.com/jeea/article/19/6/2929/6368338 by University Library Utrecht user on 14 January 2022

Hall, Robert E. (1988). "The Relation between Price and Marginal Cost in US Industry." *Journal of Political Economy*, 96, 921–947.
Hirsch, Boris, Elke J. Jahn, Alan Manning, and Michael Oberfichtner (2020). "The Urban Wage Premium in Imperfect Labor Markets." *Journal of Human Resources*, 56, 0119-9960R1.
Hirsch, Boris, Elke J. Jahn, Alan Manning, and Michael Oberfichtner (2021) "The Wage Elasticity of Recruitment." CEP Discussion paper.
Hirsch, Boris, Elke J. Jahn, and Claus Schnabel (2018) "Do Employers Have More Monopsony Power in Slack Labor Markets?" *ILR Review*, 71, 676–704.
Howell, Sabrina T. and J. David Brown (2020). "Do Cash Windfalls Affect Wages? Evidence from R&D Grants to Small Firms." NBER Working Paper No. 26717.
Jäger, Simon, Benjamin Schoefer, and Jörg Heining (2021). "Labor in the Boardroom." *The Quarterly Journal of Economics*, 136, 669–725.
Jarosch, Gregor, Jan Sebastian Nimczik, and Isaac Sorkin (2019) "Granular Search, Market Structure and Wages." NBER Working Paper No. 26239.
Kline, Patrick, Neviana Petkova, Heidi Williams, and Owen Zidar (2019). "Who Profits from Patents? Rent-Sharing at Innovative Firms." *The Quarterly Journal of Economics*, 134, 1343–1404.
Kroft, Kory, Yao Luo, Magne Mogstad, and Bradley Setzler (2020). "Imperfect Competition and Rents in Labor and Product Markets: The Case of the Construction Industry" NBER Working Paper No. w27325.
Krueger, Alan B. and Orley Ashenfelter (2018). "Theory and Evidence on Employer Collusion in the Franchise Sector." NBER Working Paper No. w24831.
Krueger, Alan B. and Eric A. Posner (2018). "A Proposal for Protecting Low-Income Workers from Monopsony and Collusion." Policy Proposal No. 2018-15, The Brookings Institution.
Lamadon, Thibaut, Magne Mogstad, and Bradley Setzler (2019). "Imperfect Competition, Compensating Differentials and Rent Sharing in the US Labor Market." NBER Working Paper No. w25954.
Lipsitz, Michael and Evan Starr (2021). "Low-Wage Workers and the Enforceability of Non-Compete Agreements." SSRN: https://ssrn.com/abstract=3452240, retrieved 20 October 2020.
Lobel, Orly (2013). *Talent Wants to Be Free: Why We Should Learn to Love Leaks, Raids, and Free Riding*. Yale University Press, New Haven, NJ.
McFadden, Daniel (1978) "Modeling the Choice of Residential Location." *Transportation Research Record*, 673, 72–77.
Manning, Alan (1987). "An Integration of Trade Union Models in a Sequential Bargaining Framework." *The Economic Journal*, 97, 121–139.
Manning, Alan (2003). *Monopsony in Motion: Imperfect Competition in Labor Markets*. Princeton University Press, Princeton, NJ.
Manning, Alan (2011). "Imperfect Competition in Labor Markets." In *Handbook of Labor Economics*, Vol. 4, edited by Orley Ashenfelter and David Card. Elsevier, Amsterdam, pp. 973–1041.
Manning, Alan (2012), "Steady-State Equilibrium in a Model of Short-Term Wage-Posting," CEP Discussion Paper No. 1158.
Manning, Alan (2021) "Monopsony in Labor Markets: A Review." *ILR Review*, 74, 3–26.
Manning, Alan and Barbara Petrongolo (2017) "How Local Are Labor Markets? Evidence from a Spatial Job Search Model." *American Economic Review*, 107(10), 2877–2907.
Marinescu, Ioana E. and Herbert Hovenkamp (2019), "Anticompetitive Mergers in Labor Markets." *Indiana Law Journal*, 94, 1031.
Marinescu, Ioana E. and Eric A. Posner (2019). "Why Has Antitrust Law Failed Workers?" *Cornell Law Review*, 105, 1343.
Marshall, Alfred (1890), *The Principles of Economics*. McMaster University Archive for the History of Economic Thought.
Marx, Matt (2011). "The Firm Strikes Back: Non-Compete Agreements and the Mobility of Technical Professionals." *American Sociological Review*, 76, 695–712.
Molloy, Raven, Riccardo Trezzi, Christopher L. Smith, and Abigail Wozniak (2016). "Understanding Declining Fluidity in the U.S. Labor Market." *Brookings Papers on Economic Activity*, 43, 187–259.

Downloaded from https://academic.oup.com/jeea/article/19/6/2929/6368338 by University Library Utrecht user on 14 January 2022

Naidu, Suresh, Eric A. Posner, and Glen Weyl (2018). "Antitrust Remedies for Labor Market Power." *Harvard Law Review*, 132, 536.
Philippon, Thomas (2019). *The Great Reversal: How America Gave Up on Free Markets*. Harvard University Press, Cambridge, MA.
Prescott, James J., Norman D. Bishara, and Evan Starr (2016). "Understanding Noncompetition Agreements: The 2014 Noncompete Survey Project." *Michigan State Law Review*, 2, 369–464.
Rinz, Kevin (2018). "Labor Market Concentration, Earnings Inequality, and Earnings Mobility." CARRA Working Paper No 2018-10, Center for Economic Studies, U.S. Census Bureau.
Robinson, Joan (1933). *The Economics of Imperfect Competition*. Macmillan, London.
Schoar, Antoinette (2021). "How is Big Data Changing Consumer Financial Markets?" *Journal of the European Economic Association*, forthcoming.
Shapiro, Carl (2019). "Protecting Competition in the American Economy: Merger Control, Tech Titans, Labor Markets." *Journal of Economic Perspectives*, 33(3), 69–93.
Shapiro, Carl and Joseph E. Stiglitz (1984). "Equilibrium Unemployment as a Worker Discipline Device." *The American Economic Review*, 74(3), 433–444.
Smith, Adam (1970). *The Wealth of Nations*. Penguin, London.
Sokolova, Anna and Todd Sorensen (2021) "Monopsony in Labor Markets: A Meta-Analysis." *ILR Review*, 74, 27–55.
Stansbury, Anna and Lawrence H. Summers (2020). "Declining Worker Power and American Economic Performance", *Brookings Papers on Economic Activity*, Spring, 1–96.
Starr, Evan, James J. Prescott, and Norman Bishara (2019) "Noncompetes in the U.S. Labor Force." University of Michigan Law & Economics Research Paper No. 18-013.
Syverson, Chad (2019). "Macroeconomics and Market Power: Facts, Potential Explanations, and Open Questions." *Journal of Economic Perspectives*, 33(3), 23–43.
Webber, Douglas A. (2015). "Firm Market Power and the Earnings Distribution." *Labour Economics*, 35, 123–134.

Supplementary data

Supplementary data are available at *JEEA* online.

Downloaded from https://academic.oup.com/jeea/article/19/6/2929/6368338 by University Library Utrecht user on 14 January 2022