Inequality

The IFS Deaton Review

Monopsony in local labour markets

Alan Manning
Barbara Petrongolo

An IFS initiative funded by the Nuffield Foundation

IFS Institute for Fiscal Studies

Nuffield Foundation

Monopsony in local labour markets

Alan Manning (London School of Economics and Centre for Economic Performance, CEP) and Barbara Petrongolo (University of Oxford and CEP)

Introduction

The decline in the labour share of GDP documented in several countries over recent decades, together with high levels of inequality and employment concentration (see Autor et al., 2020, and references therein), have led to a resurgence of the interest, by researchers and policymakers alike, in the idea of monopsonistic labour markets. The distinctive feature of monopsony is that the labour supply curve to an individual employer is not infinitely elastic with respect to the wage paid, providing employers with market power to set wages below the perfectly competitive rate. Market power may in turn derive from search frictions and/or heterogeneous tastes among workers for non-wage amenities offered by different firms (see Manning, 2021, for a discussion). Boal and Ransom (1997) and Manning (2003) propose comprehensive frameworks to understand the consequences of firms' monopsony power on wage inequality and unemployment, and the role of policy in monopsonistic labour markets. However, it was only more recently that direct estimates of firms' monopsony power started to proliferate, mostly thanks to wider availability of rich administrative data on workers and firms, as well as information on job adverts.

The focus of some of the recent studies is the identification of the elasticity of the labour supply to the firm – via either the quit or hiring margins. The meta-analysis of Sokolova and Sorensen (2021) suggests a wage elasticity of quits between 3 and 5, which implies considerable monopsony power, and recent studies on the wage elasticity of hires, surveyed by Manning (2021), often deliver even lower estimates. Datta (2021) separately identifies recruitment and separation elasticities for the same multi-plant firm and finds evidence of higher market power over incumbent workers than in attracting new recruits.

An even larger body of work has focused on the relationship between wages and employment concentration, as more highly concentrated employment implies reduced outside job opportunities and lower labour supply elasticity to the firm. Based on measures of concentration typically used in the IO literature and antitrust legislation, several studies have established a negative and significant relationship between wages and the Herfindahl–Hirschman index (HHI) of firm-level concentration of employment and online recruitment activity (see, among others, Azar, Marinescu and Steinbaum, 2020a; Benmelech, Bergman and Kim, 2021).

There remain a few unresolved issues in this literature. Firstly, one general issue is about the relationship between the predictions of monopsonistic models of the labour market and those of other non-competitive theories of wage setting – including, for example, search and matching models that predict a role for labour market tightness in wage setting. Secondly, there is a challenge about the identification of appropriate measures of the extent of competition in the labour market. While employment concentration is indeed one important source of monopsony power, the relationship between concentration and wages depends on the source of variation in concentration that is exploited in the empirical analysis. As concentration is endogenous to labour supply shocks, a few papers introduce labour demand shifters as instruments (see, for example, Azar et al., 2020a; Schubert, Stansbury and Taska, 2021). Thirdly, an important challenge is about the correct definition of labour markets, and namely the set of workers who are likely affected by changes in monopsony power in a certain labour market segment. For example, Benmelech et al. (2021) define labour markets in the US by commuting zones and four-digit industry and Azar et al. (2020a) define it by commuting zones and six-digit occupations. In most cases, labour markets are modelled as self-contained segments that are assumed to capture the bulk of outside job opportunities available to workers belonging to those segments – with the exceptions of contemporaneous work by Schubert et al. (2021) and Datta (2021). This practice leads to mismeasured concentration indices whenever segment borders are porous, that is, when workers can consider outside options beyond their labour market segment, according to geographic vicinity and/or industry and occupation similarity.

In this commentary, we propose a monopsony model of the labour market in which market power stems from idiosyncratic worker preferences over non-wage attributes of jobs. In this set-up, the relationship between wages and employment concentration arises from aggregation of firm-level elasticities of labour supply at the market level, with weights given by firms' employment shares. Next, we extend the model to account for worker mobility across labour market segments, leading to spatial labour markets that are continuous and overlapping, as in the framework of Manning and Petrongolo (2017), and we compare the resulting concentration index with the one obtained in a model that assumes self-contained labour markets. Finally, we show a simple application of this model extension on UK data, by measuring model-based and purely local concentration indices in England and Wales and estimating their relationship to local wages. We characterise labour market segments based on 8,848 Census wards, and estimate mobility patterns across wards based on worker commuting flows from the 2011 Census of Population. We find that labour market concentration would be much overstated if one did not take into account worker mobility across wards, and that the concentration index obtained on overlapping local labour markets is negatively and significantly correlated to local wages, leaving very little explanatory power to the purely local index.

In the following section, we derive the relationship between employment concentration and wages in a single, self-contained labour market. We then extend this framework to an economy with worker mobility across labour market segments. In a further section, we describe the administrative data sources on individuals and firm and discuss measurement issues. Finally, we present empirical evidence and provide conclusions.

Monopsony in a single labour market

Labour supply

We first consider a simple model in which the labour market is well defined, and can be thought of as a single, self-contained area with several firms. We assume that labour supply has a nested logit structure, in which the upper nest is the decision about whether to work at all and the lower nest is the decision about which firm to work for, conditional on the decision to work. We assume further that the utility from working for an individual firm i is related to its wage W_i, a firm-specific amenity θ_i, and an idiosyncratic component of utility with a type-1 extreme value distribution. Under these assumptions, the labour supply to firm i, N_i, is given by

$$N_i = \frac{e^{\beta \log W_i + \theta_i}}{\sum_j e^{\beta \log W_j + \theta_j}} N = \frac{e^{\beta \log W_i + \theta_i}}{U} N, \qquad (1)$$

where N denotes the total labour supply to the market (which, in the interests of simplicity we assume is completely inelastic) and $U \equiv \sum_j e^{\beta \log W_j + \theta_j}$ denotes the inclusive value, representing the expected utility from working. A similar model of labour supply to individual firms has been used, for example, in Card et al. (2018), and Azar, Berry and Marinescu (2019) also use it to model job applications to the firm.

Measures of market concentration

By differentiating equation (1), the labour supply elasticity to firm *i* can be written as

$$\epsilon_i \equiv \frac{\partial \log N_i}{\partial \log W_i} = \beta - \frac{\partial \log U}{\partial \log W_i}. \tag{2}$$

From the definition of the inclusive value *U*, we can compute

$$\frac{\partial \log U}{\partial \log W_i} = \beta \eta_i, \tag{3}$$

where $\eta_i = N_i/N$ denotes the share of firm *i* in total employment. Substituting equation (3) into equation (2), the elasticity of labour supply to firm *i* can be written as

$$\epsilon_i = \beta(1 - \eta_i). \tag{4}$$

The employment-weighted average of the labour supply elasticity can be written as

$$\epsilon = \sum_i \eta_i \, \epsilon_i = \beta \sum_i \eta_i \, (1 - \eta_i) = \beta - \beta \sum_i \eta_i^2. \tag{5}$$

Result (5) shows how, other things equal, the Herfindahl–Hirschman index ($\mathrm{HHI} \equiv \sum_i \eta_i^2$) is negatively related to the average labour supply elasticity, which is one key measure of the degree of competition in the labour market. Note that this is not the only way to model the relationship between the HHI and competition in the labour market. Schubert et al. (2021) develop a similar relationship based on a wage bargaining model in which the size of firm at which a worker is employed reduces their likelihood to receive an outside wage offer.

There has been a recent resurgence of interest in the role of concentration indices in labour markets. Using data from online job adverts in the US, Azar et al. (2020a, 2020b) defined labour market segments as a six-digit occupation and a commuting zone in a quarter and found that the HHI for vacancies in most labour markets was above the threshold for high concentration according to the Department of Justice/Federal Trade Commission horizontal merger guidelines. A number of other studies (see, for example, Benmelech et al., 2021; Berger et al., 2021; Rinz, 2021) obtain HHI using firm-level employment and defining labour market segments by various interactions between (narrowly defined) industry and geography. They obtain qualitatively similar results on the relationship between the HHI and wages as papers using online job adverts.[1]

All of these studies assume that, when appropriately defined, labour markets are self-contained, as in our simple model above. The model outlined later in the next section will relax this assumption.

Sources of employment concentration

The model of the previous subsection illustrates why one should expect a negative relationship between employment concentration and wages, which is indeed validated in much of the empirical work cited above. However, it is important to understand the source of variation in concentration to correctly interpret any empirical relationship between concentration and wages.

[1] This discussion is far from exhaustive of the body of work in this area. See Manning (2021) for a broader survey of the literature.

There are three possible sources for variation in labour market concentration: the number of firms in the market; the dispersion in productivity and amenities across firms; and the sensitivity of firm-level labour supply to the wage, β.

Variation in the number of firms provides the most natural intuition for variation in market concentration. If the number of active firms in the labour market decreases without altering the distribution of amenities and productivity across firms, market power of individual firms increases, market concentration increases and wages fall.

Suppose next that the distribution of productivity across firms becomes more dispersed. This also generates a rise in market concentration, as workers become disproportionately clustered in the higher productivity firms. But this kind of shock generates an increase, rather than a decrease, in the average level of wages because the employment-weighted average level of productivity is rising. However, mark-downs will also change because the share of employment in different firms will change. The low-productivity firms will see their market share fall so their mark-downs also fall. But the high-productivity firms will see their market share rise so their mark-downs will rise.

Finally, consider the case in which worker utility becomes more sensitive to wage changes, implying an increase in β. This increases competition among firms and leads to both higher wages and higher levels of market concentration, as workers again tend to cluster in high-wage firms. In labour market models with search frictions, an equivalent shock would be an increase in the probability to receive outside job offers, as workers become more likely to be poached by high-wage firms.

In summary, the relationship between market concentration and wages depends on the source of variation in market concentration that is exploited in the empirical analysis. As both concentration and employment levels are endogenous, natural instruments for concentration would be labour demand shifters, such as (the inverse of) the number of posting employers for the same occupation in other areas (see, e.g., Azar et al., 2020a) or the national level growth of large firms with plants in the local market (see, e.g., Schubert et al., 2021). Other papers leverage instead variation from merger and acquisition activity (see, e.g., Currie, Farsi and Macleod, 2005; Arnold, 2021; Prager and Schmitt, 2021).

Defining the labour market

We have assumed so far that there exists a self-contained labour market that could be defined, for example, by geography, industry or occupation. But, in practice, labour market boundaries are porous. For example, workers can consider jobs outside their local area, or outside their current or previous occupation and industry. And the intensity of flows across labour market segments depends on geographic vicinity and/or skill transferability across occupations and industries. Worker mobility is taken into account by Schubert et al. (2021), who segment local labour markets by occupation and model the degree of outward mobility from different occupations based on observed transition patterns.

In our analysis, we focus on the geographic dimension of labour markets and we extend the model of the previous section to allow the labour supply to firm *i* to come from different areas, denoted by *a*. We can think of these areas as the smallest measurable building blocks of local labour markets.

The labour supply from area *a* to firm *i* is a modified version of equation (1):

$$N_i(a) = \frac{e^{\beta \log W_i + \theta_i - d_i(a)}}{U(a)} N(a). \tag{6}$$

Here, $d_i(a)$ is the distance from the centroid of area *a* to firm *i*, capturing the (inverse of the) attractiveness of working for firm *i* for residents of area *a*, $N(a)$ is total labour supply in area *a* and $U(a)$ is the inclusive value of working in area *a*. As the attractiveness of jobs at various firms decays continuously with the distance to those firms, this set-up leads to a framework with overlapping local labour markets, as modelled in Manning and Petrongolo (2017).

From the labour supply schedule in equation (6) and by analogy to equation (4), the elasticity of supply to firm *i* from area *a* can then be written as

$$\epsilon_i(a) = \beta\big(1 - \eta_i(a)\big), \tag{7}$$

where $\eta_i(a) = N_i(a)/N(a)$ denotes the share of firm *i* in the total employment of people living in *a*. For firm *i*, the overall labour supply elasticity will be a weighted average of equation (7) across feeder areas *a*, with weights $s_i(a)$ given by the share of area *a* in the total labour supply to firm *i*:

$$s_i(a) = \frac{\eta_i(a)N(a)}{\sum_{a'}\eta_i(a')N(a')}. \tag{8}$$

Using equation (7), the wage elasticity of labour supply to firm *i* can then be written as

$$\epsilon_i = \sum_a s_i(a)\epsilon_i(a) = \beta\left[1 - \frac{\sum_a N(a)\eta_i(a)^2}{\sum_a N(a)\eta_i(a)}\right]. \tag{9}$$

The aggregate concentration index

The average labour supply elasticity across the whole economy is a weighted average of firm-level elasticities (9), using as weights the firm-level shares in total employment, which we denote by η_i,

$$\eta_i = \frac{\sum_a \eta_i(a)N(a)}{\sum_a N(a)} = \sum_a \eta_i(a)\phi(a), \tag{10}$$

where $\phi(a') \equiv N(a)/\sum_a N(a)$ is the share of area *a* in total labour supply. Combining equations (9) and (10) leads to the following expression for the aggregate elasticity:

$$\epsilon = \sum_i \eta_i\epsilon_i = \beta\left[1 - \frac{\sum_{a,i} N(a)\eta_i(a)^2}{\sum_a N(a)}\right] = \beta\left[1 - \sum_{a,i}\phi(a)\eta_i(a)^2\right] \equiv \beta[1 - H], \tag{11}$$

where $H \equiv \sum_{a,i}\phi(a)\eta_i(a)^2$ is the relevant concentration index for the case of overlapping local labour markets. We refer to *H* as the model-based concentration index.

It is helpful to compare *H* in equation (11) with measures of concentration for a fully integrated labour market as a homogeneous whole, or measures that would treat areas as self-contained labour markets with no linkages across them.

Note first that *H* can be expressed as a weighted average of local indices, defined by workers' areas of residence, $\sum_i \eta_i^2(a)$, with

$$H = \sum_a \phi(a) \sum_i \eta_i^2(a). \tag{12}$$

Consider next the concentration index that one would obtain if the labour market were a single integrated space, in which distance has no relevance. In this case, what matters is each individual employer's share in total employment, as given by equation (10), leading to the following aggregate concentration index:

$$H_{int} = \sum_i \eta_i^2 = \sum_i \left(\sum_a \phi(a)\, \eta_i(a) \right)^2. \tag{13}$$

Intuitively, the index for the integrated market H_{int} can never be larger than the true, model-based, concentration index. To see this, note that using equations (12), (13) and (10) we obtain

$$\begin{aligned} H - H_{int} &= \sum_{a,i} \phi(a)\eta_i^2(a) - \sum_{a,i} \phi(a)\eta_i(a)\eta_i = \sum_{a,i} \phi(a)\eta_i(a)[\eta_i(a) - \eta_i] \\ &= \sum_{a,i} \phi(a)[\eta_i(a) - \eta_i]^2 \geq 0, \end{aligned} \tag{14}$$

where equation (14) holds with equality only in the case in which $\eta_i(a)$ does not vary across areas. The intuition is that, under equal $\eta_i(a)$, areas become unimportant for the size of firms.

At the other extreme, we consider the concentration index that one would obtain under the assumption that each area *a* is a fully isolated labour market (i.e. the attractiveness of jobs at any positive distance from one's local area falls to zero). In this case, the relevant concentration index is simply based on the employment shares of each firm in a given area. Define μ_i as the share of firm *i*'s employment in the area where it is located. The following relationship must hold:

$$\mu_i = \frac{\eta_i}{\sum_{j \in F(i)} \eta_j}, \tag{15}$$

where $j \in F(i)$ denotes the set of firms in the same area as firm *i*. The concentration index in area *a* is then given by

$$H_{work}(a) = \sum_{i \in F(i)} \mu_i^2(a), \tag{16}$$

where the subscript *work* indicates that the index is based on workers' area of work. The corresponding aggregate index is given by a weighted average of equation (16),

$$H_{sep} = \sum_a \phi_{work}(a)\, H_{work}(a), \tag{17}$$

where $\phi_{work}(a)$ is the share of people who work in area *a*, given by

$$\phi_{work}(a) = \sum_{i \in F(i)} \phi(a)\, \eta_i(a). \tag{18}$$

To understand the relationship between these alternative concentration indices, consider a simple example in which there are *A* identical areas, each containing *F* identical firms and *N* workers. Assume that a fraction α of workers work in the local area, choosing one of the local firms at random. The remaining fraction $(1 - \alpha)$ then randomly choose among all firms, whether in the local area or not. Given these assumptions, firms will end up identical in size. The integrated index will be given by $H_{int} = 1/AF$ (i.e., the inverse of the total number of firms in the economy). The index for fully isolated markets will be H_{sep} = 1/*F* (i.e., the inverse of the number of firms in each local area). Finally, the true concentration index will lie between these two extremes. To see this, note that in each residential area there are *F* local firms, who each have a share of local employment equal to

$$\eta_i(a) = \frac{\alpha}{F} + \frac{1-\alpha}{AF}, \tag{19}$$

and each of the *F*(*A* – 1) non-local firms has a share of

$$\eta_i(a) = \frac{1-\alpha}{AF}. \tag{20}$$

Combining these expressions, the true index will be given by

$$H = F\left[\frac{\alpha}{F} + \frac{1-\alpha}{AF}\right]^2 + F(A-1)\left[\frac{1-\alpha}{AF}\right]^2 = \frac{\alpha^2}{F} + \frac{1-\alpha^2}{AF}, \tag{21}$$

which is a weighted average of H_{int} and H_{sep}, with weights α^2 and $1-\alpha^2$.

The intuition is that, ignoring mobility across local labour market segments overstates the degree of market concentration, because it ignores employment opportunities outside one's local area, but ignoring the local nature of mobility would understate it, because it implicitly assumes that workers would equally value outside options at any distance from their residences.

Local concentration indices

The discussion so far has been about the measurement of the average labour supply elasticity across the market as a whole and the resulting aggregate concentration indices. But, for understanding local variation in wages, what matters is the local variation in the elasticity of labour supply, either across areas of residence or areas of work. This measure is relevant to address the question whether firms in some areas have higher market power than firms in other areas or, alternatively, whether residents of a particular area face a more monopsonistic labour market than residents of other areas.

Equation (9) provides an expression for the labour supply elasticity facing each individual firm. To obtain the average elasticity among all the firms in area *a*, which we denote by $\epsilon^f(a)$, we need to take a weighted average of equation (9), using as weights the employment shares given by equation (10). If we denote the set of firms in area *a* by *F*(*a*), this can be written as

$$\epsilon^f(a) = \frac{\sum_{i\in F(a)} \eta_i \epsilon_i}{\sum_{i\in F(a)} \eta_i}. \tag{22}$$

Combining equations (9) and (10) leads to the following expression for $\epsilon^f(a)$:

$$\epsilon^f(a) = \beta\left[1 - \frac{\sum_{a',i\in F(a)} \phi(a')\eta_i(a')^2}{\sum_{a',i\in F(a)} \phi(a')\eta_i(a')}\right]. \tag{23}$$

We can similarly define the average elasticity faced by residents of area *a*, which we denote by $\epsilon^r(a)$. This is a weighted average of the elasticity of each firm in the economy, denoted by equation (9), with weights given by the probability that this is the firm they work for. This can be written as

$$\epsilon^r(a) = \sum_i \eta_i(a)\epsilon_i = \beta\left[1 - \sum_i \eta_i(a)\frac{\sum_{a'}\phi(a')\eta_i(a')^2}{\sum_{a'}\phi(a')\eta_i(a')}\right]. \quad (24)$$

Data and measurement

Data sources

To show evidence on alternative concentration indices, we combine data from three different sources. Information on individuals is drawn from the Annual Survey of Hours and Earnings (ASHE) (ONS, 2019a), an employer-based survey, covering a 1% sample of employee jobs in the UK, randomly selected from the Pay As You Earn records of HM Revenue and Customs. The survey has been carried out in April of each year since 1997. It represents the main administrative data source on UK employees and contains information on personal and work-related variables, as well as geographic identifiers for employees' residences and workplaces at the full postcode level.[2] We use ASHE waves from 2000 onwards, as the postcode of work is misclassified in a very large share of observations before 2000.

Information on establishments is drawn from the Business Statistics Database (BSD; ONS, 2019b), an employer-based annual survey that covers the near universe of business organisations in the UK from 1997 onwards. It combines data collected by HM Revenue and Customs via VAT and Pay As You Earn records and ONS business surveys. Relevant information is recorded at both the firm and establishment level. From this database, we select all active establishments with at least one employee. The smallest geographic identifier in the BSD is the Census Output Area. As of the 2011 Census, Output Areas on average included 129 households.

We define labour market segments according to their location and we measure the extent of local mobility based on information on commuting patterns across Output Areas from the 2011 Census. As of 2011, half of the employed population commuted less than 6.2 km to work (one way), and three-quarters commuted less than 14.3 km.

As our main geographic units, we consider Census Area Statistics (CAS) wards, created for the 2001 Census outputs, which we match to postcodes in the ASHE data and to Output Areas in the BSD and commuting data. There are 10,656 CAS wards in the UK, with an average resident population of 6,000 persons. As the commuting data are only available for England and Wales, our working sample covers England and Wales from 2000 onwards and includes nearly three million individual–year observations, across 8,848 wards.

Empirical implementation

To measure employment concentration, we need to measure $\eta_i(a)$, which is the share of firm *i* in the employment of people who live in *a*. If one had matched employer–employee data for the universe of workers, one could directly measure the number of people who live in an area and work in any firm located in any possible area. But the ASHE data only cover a 1% sample of employees, so the resulting matrix of employment patterns at the ward level would be very noisy.

We use instead information on commuting patterns from the 2011 Census, on which we obtain the fraction of residents of area *a'* who commute to any area *a* – denoted by $\eta(a', a)$. In addition, we obtain the share of employment of each firm *i* in its local area *a*, denoted by μ_i, from the BSD. A natural assumption is that that the fraction of area *a'* residents who work for firm *i*, located in *a*, is

[2] There are currently 1.8 million postcodes in the UK, with an average of 15 households per postcode.

proportional to firm *i* employment share in *a*, by a factor that captures the intensity of commuting from a' to *a*, i.e., $\eta_i(a) = \eta(a', a)\mu_i$. This is equivalent to assuming that commuters from a' to *a* are randomly distributed across firms in *a* according to their respective employment shares. Substituting $\eta_i(a)$ into equation (9) gives the elasticity of labour supply to firm *i*:

$$\epsilon_i = \beta \left[1 - \mu_i \frac{\sum_{a'} \phi(a')\eta(a', a)^2}{\sum_{a'} \phi(a')\eta(a', a)}\right]. \tag{25}$$

The final term in square brackets can be thought of as a measure of the concentration of workers who work in *a* in terms of the areas they come from. So the intuition for equation (25) is that the elasticity is smaller for firms that are large for their area and for areas that have a more concentrated pool of workers. If we average equation (25) across all firms in an area, we obtain

$$\epsilon^f(a) = \beta \left[1 - \sum_{i \in F(a)} \mu_i^2 \frac{\sum_{a'} \phi(a')\eta(a', a)^2}{\sum_{a'} \phi(a')\eta(a', a)}\right] = \beta[1 - H(a)]. \tag{26}$$

Here, *H*(*a*) denotes the model-based local concentration index, and is equal to the product of two terms: the first term, $\sum_{i \in F(a)} \mu_i^2$, is the index one would compute for area *a* if each area was an isolated labour market; the second term,

$$\frac{\sum_{a'} \phi(a')\eta(a', a)^2}{\sum_{a'} \phi(a')\eta(a', a)},$$

is akin to a concentration index across all source areas for employment in area *a*.

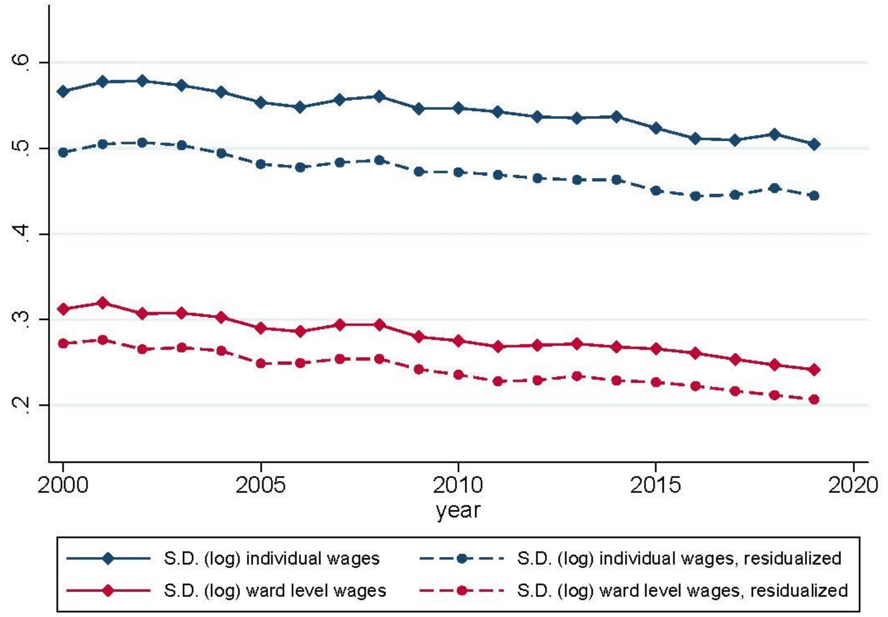
To obtain the average elasticity for residents of an area, one needs to take a weighted average of equation (26) using commuting shares as weights, i.e.

$$\epsilon^r(a) = \sum_{a'} \eta(a', a)\, \epsilon^f(a') = \beta \left[1 - \sum_{a'} \eta(a', a)\, H(a')\right]. \tag{27}$$

Empirical evidence

Before relating wages to the model-based concentration index, we present descriptive evidence on changes in wage inequality during 2000–19 at both the individual and the local level. Figures are based on hourly wages, which were on average growing by 2.4% per year in real terms during our sample period. The top two plot lines in Figure 1 represent the standard deviation in (log) individual wages. The sample size grows from about 137,000 individuals in 2000, to 156,000 in 2019. The solid line is based on raw wages, and shows a slightly declining trend in inequality, with the standard deviation of log wages falling from about 0.57 in the early 2000s, to 0.51 in 2019. The dashed line is based on wages that have been residualised in each year with respect to the impact of a small set of individual characteristics – namely, gender, unrestricted age effects and two-digit industry effects. As one would expect, the standard deviation of wages is slightly lower, but it follows exactly the same trend. The slight decline in wage inequality since the early 2000s is in contrast with trends in inequality in the previous two decades, and has been documented in further detail in Giupponi and Machin (2022).

Figure 1. Trends in wage inequality in England and Wales, 2000–19



Note: The figure shows, from top to bottom, the standard deviation in (i) log individual hourly wages, (ii) wage residuals (obtained in a regression of log hourly wages on fixed effects for gender, age and two-digit industry for each year), (iii) mean log hourly wages at the ward level and (iv) residualised mean log hourly wages at the ward level, obtained as in (ii).

Source: ASHE 2000–19.

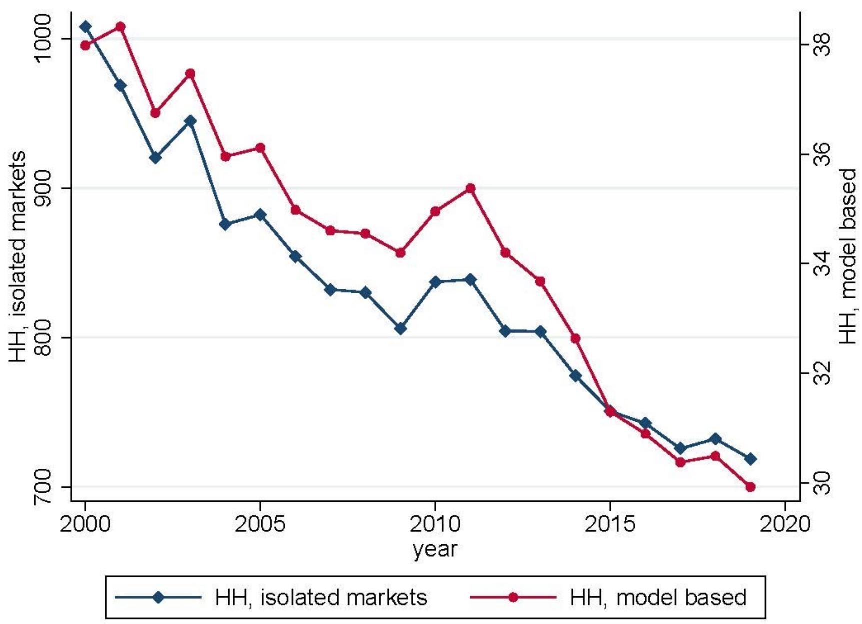
The bottom two plot lines in Figure 1 focus on the local dimension of inequality, by showing the standard deviation of mean log wages at the ward level, using ward-level employment as weights. The number of wards in England and Wales is 8,848, but the sample size is on average 8,346, because not all wards are represented each year in the 1% ASHE data. Ward-level wage inequality follows a very similar trend as overall inequality (whether on raw or residualised wages), and its level is about half as large, from 31% in 2000, to 25% in 2019. The important result here is that local inequality explains (in an accounting sense) both a large portion of overall inequality, and its slight decline. In other words, within-ward wage inequality remained roughly constant throughout the sample period. This motivates our focus on the geographic dimension of concentration.

Another reason for restricting to geographic labour market segments is data driven. Given the 1% sampling in the individual-level data, to obtain reasonably sized cells at the ward level, we need to abstract from industry and occupation divides. One consequence of this is that our concentration indices are much lower than those commonly reported by studies that consider interactions across these dimensions, as we implicitly assume that all existing jobs at a certain commuting distance represent equally valid outside options, regardless of their industry or occupation. Hence, our focus will be more on the trends in concentration, and its correlation with wages, rather than on levels.

We compute the employment concentration index on fully isolated labour markets, H_{sep} (shown in equation (17)), and compare it to the model-based index H (shown in equation (12)). H_{sep} can be directly obtained on BSD data, measuring firms' employment shares in their local wards, while to obtain H we combine firm-level data from the BSD with commuting flows.

Figure 2 plots series for H_{sep} and H. Their respective scales are very different (and hence plotted on different *y*-axes), because cross-ward commutes dilute the extent of local employment concentration. In particular, concentration would be overstated by a factor of about 24 if one did not take into account job opportunities outside workers' residential wards. The important point to note is that, even moderate local mobility – with a median commute of about 6 km and 10% of individuals working in the ward where they live – has quantitatively important effects on the resulting concentration index. Both indicators are declining over time, except for a moderate blip during the Great Recession.[3] Similar trends have been documented for the UK by Abel, Tenreyro and Thwaites (2018), based on a coarser segmentation of the labour market by two-digit industry and region. For the US, indices on industry and commuting zones tend to deliver either stable or slightly declining concentration over the same sample period (Benmelech et al., 2021; Rinz, 2021), depending on specific classifications used.

Figure 2. Trends in employment concentration in England and Wales, 2000–19



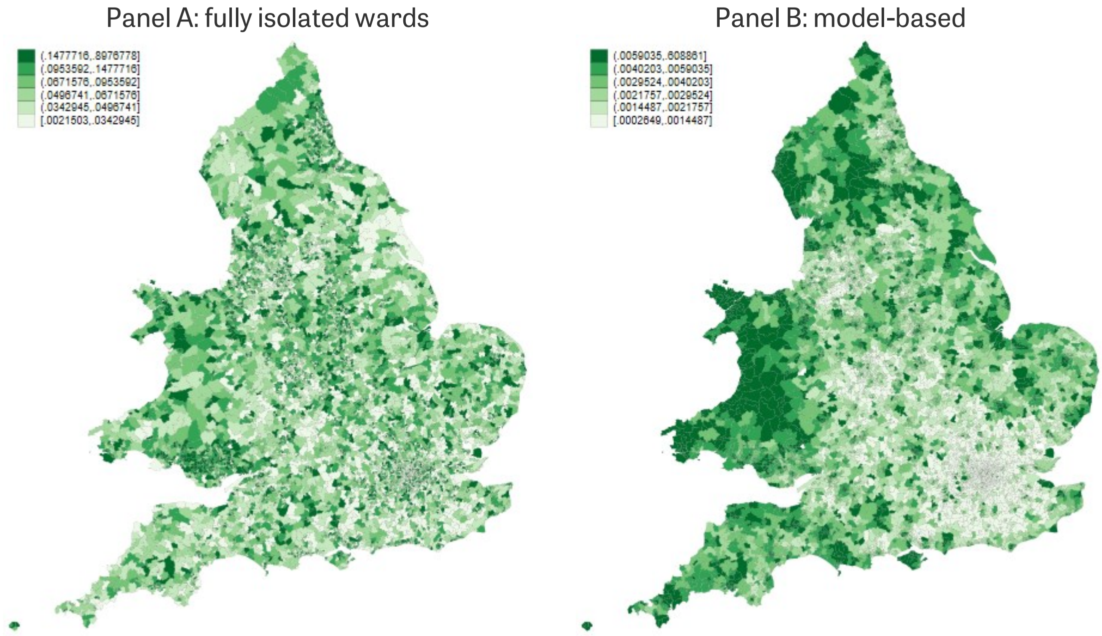
Note: The figure shows trends in employment concentration based on fully isolated labour markets (see equation (17); left-hand axis), and a model with cross-wards commutes (see equation (12); right-hand axis). Underlying employment shares are multiplied by 100.

Source: UK Census 2011 and BSD 2000–19.

Figure 3 shows evidence on the local variation in employment concentration. Panel A plots the time average of the local indices $H_{work}(a)$, based on isolated markets (see equation (16)), and Panel B plots the time average of the model-based $H(a)$ (see equation (26)). Both indices display very wide variation at the local level (compared, for example, with their modest decline over time). As one would expect, the model-based index $H(a)$ smooths variation at the local level, because it is based on local moving averages of $H_{work}(a)$. Large metropolitan areas such as London, Birmingham and Manchester are characterised by lower concentration according to $H(a)$, because workers commute relatively longer distances in urban areas.

[3] The two indices are bound to follow similar trends, because we are using a time-invariant measure of commuting flows.

Figure 3. Spatial variation in employment concentration



Note: Panel A plots the time average of $H_{work}(a)$ (see equation (16)), and Panel B plots the time average of $H(a)$ (see equation (26)). Underlying employment shares are multiplied by 100.

Source: UK Census 2011 and BSD 2000–19.

We finally illustrate the predictive power of our model of labour supply across local labour markets by showing correlations between local wages and employment concentration. The regression results are reported in Table 1. Column 1 regresses the average log wage at the ward level on the model-based index and shows that they are negatively and significantly correlated, as the model would predict. In Column 2, we control for ward and year fixed effects. As one would have expected from the evidence shown in Figure 3, ward fixed effects absorbed much of the variation in the concentration index, but its correlation with local wages is still negative and highly significant. Columns 3 and 4 replicate the regressions on wage residuals and show a very similar picture of correlations with the employment concentration index, as in Columns 1 and 2. Finally, Column 5 additionally controls for the purely local concentration index. If our model of local mobility fully captures the availability of job opportunities within and beyond a worker's ward of residence, we expect that the model-based index would be a sufficient statistics for such opportunities, and the local index should have no additional explanatory power on wages. We find instead that the local index is negatively correlated to local wages, even after controlling for the model-based concentration of jobs. The associated coefficient, however, is extremely small, while the coefficient on the model-based index remains entirely robust to the introduction of the local index. One reason why the local index could retain some explanatory power on local wages is the potential mismeasurement of the model-based index, which implicitly assumes constant commuting flows over time. Quantitatively, the near-unit semi-elasticity of wages to concentration implies that the fall in concentration observed over the past two decades had only a tiny effect on wage growth of about 0.08%.

Table 1. Correlation between local wages and employment concentration

Dependent variable	Log hourly wage		Log wage residuals		
	(1)	(2)	(3)	(4)	(5)
Model-based index	–11.2970***	–0.6243**	–8.8516***	–1.0674***	–0.9213***
	(0.1462)	(0.2450)	(0.1132)	(0.2133)	(0.2253)
Local index					–0.0130**
					(0.0065)
Year and ward fixed effects	No	Yes	No	Yes	Yes
Observations	158,395	158,394	158,395	158,394	158,394
R^2	0.0363	0.7933	0.0372	0.7386	0.7386

Note: The table shows regressions of log hourly wages (Columns 1 and 2) and log wage residuals (Columns 3–5) on the model-based and local concentration indices. Wage residuals are obtained in regressions of log hourly wages on gender, unrestricted age effects and two-digit industry effects for each year. Standard errors are reported in brackets. Significance at 1%, 5% and 10% is denoted by ***, ** and *, respectively.

Source: ASHE 2000–19, BSD 2000–19 and UK Census 2011.

Conclusions

We propose a monopsonistic labour market model in which the attractiveness of workers' outside options decays with distance from their residential location as commuting is costly. Using this framework, we show how to construct concentration indices that capture the elasticity of labour supply to the firm, providing a measure of the degree of employer market power faced by workers at each location.

We compute the model-based concentration indices using a combination of firm-level administrative data and commuting flows from the Census. One important caveat is that we are unable to model on these data the industry or occupational segregation of the labour market, so these measures are bound to understate the level of concentration in labour markets, because they implicitly assume that, conditional on location, jobs in any occupation provide equally plausible outside options for a given worker. However, variation across locations and over time may nonetheless be informative about variation in market power.

Our analysis has shown two main results. First, measurement of concentration is clearly sensitive to the characterisation of workers' local labour markets. In our empirical set-up based on Census wards, employment concentration would be overstated by a factor of about 24 if one did not take into account that ward boundaries are porous for job search purposes. This point is remarkable if one considers that commuting flows are relatively 'local' in our data such that, in the 2011 Census, 10% of the employed population live and work in the same ward, and 50% commute less than 6 km.

Second, we find that employment concentration in local labour markets was slightly falling in England and Wales during 2000–19. The model-based concentration index is negatively correlated with wages and performs better than other, purely local concentration measures. However, in quantitative terms, the observed fall in concentration can predict only a negligible increase in average wages. The evidence shown here suggests that changes in monopsony power may not be an important factor behind changes in wages over the past two decades.

References

Abel, W., Tenreyro, S., and Thwaites, G. (2018), 'Monopsony in the UK', CfM Discussion Paper No. 2018-27.

Arnold, D. (2021), 'Mergers and Acquisitions, Local Labor Market Concentration, and Worker Outcomes', mimeo, Princeton University.

Autor, D., Dorn, D., Katz, L. F., Patterson, C., and Van Reenen, J. (2020), 'The Fall of the Labor Share and the Rise of Superstar Firms', *Quarterly Journal of Economics*, 135, 645–709.

Azar, J., Berry, S., and Marinescu, I. E. (2019), 'Estimating Labor Market Power', available at SSRN, https://ssrn.com/abstract=3456277.

Azar, J., Marinescu, I., and Steinbaum, M. (2020a), 'Labor Market Concentration', *Journal of Human Resources*, published online before print (http://doi.org/10.3368/jhr.monopsony.1218-9914R1).

Azar, J., Marinescu, I., Steinbaum, M., and Taska, B. (2020b), 'Concentration in US Labor Markets: Evidence from Online Vacancy Data', *Labour Economics*, 66, 101886.

Benmelech, E., Bergman, N. K., and Kim, H. (2021), 'Strong Employers and Weak Employees: How Does Employer Concentration Affect Wages?', *Journal of Human Resources*, published online before print (http://doi.org/10.3368/jhr.monopsony.0119-10007R1).

Berger, D., Herkenhoff, K., and Mongey, S. (2021), 'Labor Market Power', BFI Working Paper No. 2021-58.

Boal, W. M., and Ransom, M. (1997), 'Monopsony in the Labor Market', *Journal of Economic Literature*, 35, 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.

Currie, J., Farsi, M., and Macleod, W. B. (2005), 'Cut to the Bone? Hospital Takeovers and Nurse Employment Contracts', *ILR Review*, 58, 471–93.

Datta, N. (2021), 'Local Market Power', mimeo, University College London.

Giupponi, G., and Machin, S. (2022), 'Labour Market Inequality', IFS Deaton Review of Inequalities.

Manning, A. (2003), *Monopsony in Motion: Imperfect Competition in Labor Markets*, Princeton, NJ: Princeton University Press.

Manning, A. (2021), 'Monopsony in Labor Markets: A Review', *ILR Review*, 74, 3–26.

Manning, A., and Petrongolo, B. (2017), 'How Local Are Labor Markets? Evidence from a Spatial Job Search Model', *American Economic Review*, 107 (10), 2877–907.

ONS (2019a), 'Annual Survey of Hours and Earnings, 1997–2019: Secure Access', Technical Report.

ONS (2019b), 'Business Structure Database, 1997–2018: Secure Access', Technical Report.

Prager, E., and Schmitt, M. (2021), 'Employer Consolidation and Wages: Evidence from Hospitals', *American Economic Review*, 111 (2), 397–427.

Rinz, K. (2020), ‘Labor Market Concentration, Earnings, and Inequality’, *Journal of Human Resources*, published online before print (http://doi.org/10.3368/jhr.monopsony.0219-10025R1).

Schubert, G., Stansbury, A., and Taska, B. (2021), ‘Employer Concentration and Outside Options’, mimeo, Harvard University.

Sokolova, A., and Sorensen, T. (2021), ‘Monopsony in Labor Markets: A Meta-Analysis’, *ILR Review*, 74, 27–55.