"Local Monopsony Power”

Nikhil Datta
8th Econ Job Market Best Paper Award

Working Paper Series

n. 161 ■ December 2021

Statement of Purpose

The Working Paper series of the UniCredit Foundation is designed to disseminate and to provide a platform for discussion of either work of the UniCredit economists and researchers or outside contributors (such as the UniCredit Foundation scholars and fellows) on topics which are of special interest to the UniCredit Group. To ensure the high quality of their content, the contributions are subjected to an international refereeing process conducted by the Scientific Committee members of the Foundation.

The opinions are strictly those of the authors and do in no way commit the Foundation and UniCredit.

These Working Papers often represent preliminary work. Citation and use of such a paper should take account of its provisional character.

The Working Papers are also available on our website (http://www.unicreditfoundation.org)

Local Monopsony Power

Nikhil Datta
UCL and CEP *

January 3, 2022

Job Market Paper

[Click here for the latest version]

Abstract

This paper studies the extent of monopsony power in a low pay labour market and explores its determinants. I emphasise the role of the spatial distribution of activity and workers' distaste for commuting in generating imperfect substitutability between jobs, and heterogeneity in monopsony power. Using detailed HR data for a firm with hundreds of establishments across the UK, coupled with two sources of job-establishment level exogenous wage variation, I estimate both the recruitment and separation elasticities as well as the commuting-wage elasticity. Estimates show strong evidence of monopsony power, with wage markdowns of 20-25%, and a sizeable distaste for commuting amongst workers. To formalise the role of commutes in generating monopsony power I develop a job search model where utility depends on wages, commutes and an idiosyncratic component. The model endogenously defines probabilistic spatial labour markets which are point specific and overlapping, and generates labour supply to the firm elasticities which vary across space. Distaste for commuting is shown to increase monopsony power, but does so heterogeneously, increasing monopsony power in rural areas more than in denser urban areas. Spatial heterogeneity in market-power predicted by the model is evidenced in both causal estimates using the HR data and descriptive estimates using a nationally representative dataset. Structurally estimating the model using spatial variation in monopsony power I find that commutes are responsible for approximately 1/3 of the wage markdown.

*I am grateful to Uta Schönberg, Christian Dustmann, Stephen Machin and Alan Manning for their continuous support. I would like to thank Barbara Petrongolo, Abigail Adams-Prassl, Atilla Lindner, John Van Reenen, Fabien Postel-Vinay, Franck Portier, Guy Michaels, Hyejin Ku, David Zentler-Munro, Jan David Bakker, Simon Fuchs, Joshua De-Lyon, Alison Andrews, Pawel Adrjan, and participants at various workshops and seminars for helpful comments and discussions. This research was funded by the Economic and Social Research Council at the Centre for Economic Performance and the UBEL Doctoral Training Partnership. I additionally thank The Living Wage Foundation for institutional information regarding the Living Wage, and The Company for use of the data and institutional information. Contact: nikhil.datta.14@ucl.ac.uk, http://nikhil-datta.com/, Department of Economics, University College London.

1 Introduction

Wages account for 84% of total income for the median working-age household in the UK. The wage-setting behaviour of firms is therefore of first-order economic importance. In a perfectly competitive market where firms are price takers, workers receive their marginal product. But in markets where firms have wage-setting power, workers receive a marked-down percentage of their marginal product, and the size of that markdown depends on the extent of monopsony power their employer exercises.[1] The sources of monopsony power are still not well known and are of key importance in understanding what generates inefficiencies in labour markets, heterogeneity in markdowns and how to to design optimal labour market policy. This paper estimates the extent of monopsony power for a firm with hundreds of establishments across the UK and examines a source of monopsony power, workers' distaste for commuting.

To estimate the extent of monopsony power I am interested in examining two key elasticities; the recruitment-wage elasticity and the separation-wage elasticity, which combined give the labour supply elasticity to the firm. A large labour supply elasticity to the firm suggests a more competitive market, while a low elasticity suggests large markdowns. To estimate workers' distaste for commuting I am interested in estimating the commuting-wage elasticity. To estimate these parameters I utilise a novel dataset for a large UK based firm, which contains rich human resources (HR), vacancy, and applicant data. In addition to the usual information such as wages, tenure, and demographic characteristics, the dataset includes detailed data on specific job roles, entire vacancy text, and worker and applicant home location, so commutes can be accurately calculated.

Estimating the above three elasticities has historically been challenging as they require exogenous wage variation at the establishment-level as a minimum. This is because to isolate the elasticities, we require to see how recruits, separations or willingness-to-commute respond when only the wage in a single establishment changes, and the rest of the market's wages remain unchanged.

To overcome endogeneity concerns I use two novel instruments to establish exogenous variation in the wage at the job-establishment-time level, and the advert-job-establishment-time level respectively. The first a location specific Living Wage floor that only affects firms who are engaged in council procurement contracts and thus affects a fraction of a percentage of firms and workers in the area. The Living Wage is however binding for The Company's establishment in that location, and only affects jobs that were previously paying less than the Living Wage. As the Company's local main competitors all operate purely in the private sector, the location specific Living Wage floor essentially operates as an establishment-job-time wage shifter for those jobs where the Living Wage is binding. The second is an instrument related to the saliency of the advertised wage in a job advert. By law in the UK all jobs must pay 28 days annual leave and some firms decide to pay this as an hourly wage top-up, which works out to

[1] For canonical texts on this see Pigou (1924), Robinson (1933) and Manning (2003).

12.07%. The Company pays this annual leave top-up to all their casual staff, however only some of the advertised positions include the top-up in the posted wage. This is as a result of the member of HR staff who was posting vacancies that day, and thus induced by idiosyncrasies in The Company's HR department. The instrument has the effect of inducing variation in the posted wage randomly for the same jobs, within the same establishment.

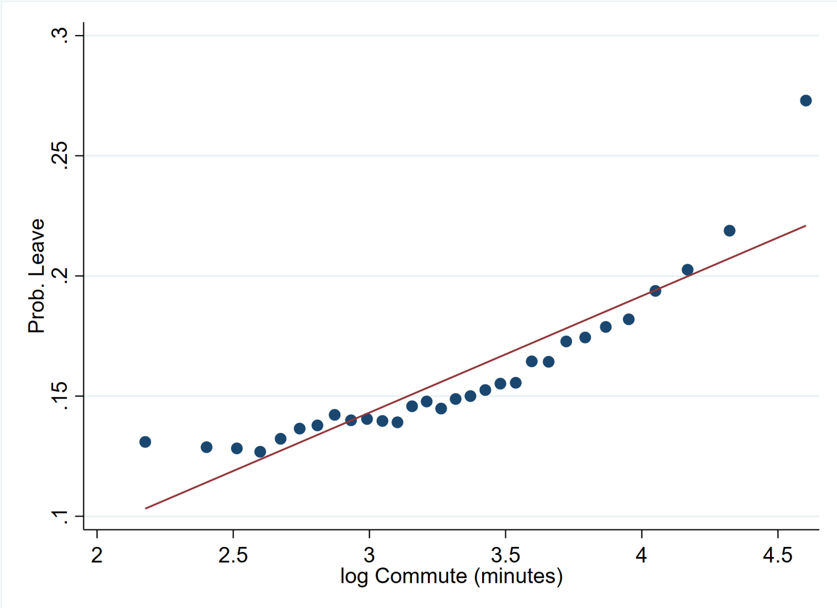
Previous estimates of the labour supply elasticity to the firm have relied on a result from Manning (2003) that the recruitment elasticity and separation elasticity are equal in absolute value under certain restrictive assumptions. This paper is the first to estimate both elasticities separately for the same firm utilising exogenous variation in the wage. In doing so I am able to assess whether there are asymmetries in market power over incumbent workers versus new recruits. I find a separation-wage elasticity of -1.7 and an application-wage elasticity of approximately 3.7, suggesting firms exercise more monopsony power over incumbent workers than in attracting new applicants. This is suggestive of the presence of greater frictions for incumbent workers. However, as the increase in applicants only effects recruitment via increasing the probability of filling a vacancy, the recruitment and separation elasticity are found to be of similar size. The baseline estimates suggest an average markdown in the region of 20-25%. I additionally find workers have a very strong distaste for commuting. In particular I find a commuting-wage elasticity of approximately 1, which implies a minimum wage worker would require an extra £0.47 ($0.55) per hour for an extra 5 minute commute.

The role of commutes in generating monopsony power is straightforward. The length of commute to a place of work is a non-wage factor that can affect the utility from a job, and therefore generate imperfect substitutability between differently located jobs. Figure 1 plots the probability of separation against log commutes for a large representative sample of British workers for the period 2003-2019. As can be seen length of commute is strongly positively correlated with separations, which is suggestive that distance to work is an important factor in worker preferences over jobs.[2] The importance of commutes is determined by the size of the commuting-wage elasticity. If there is evidence of a strong distaste for commuting (a low elasticity), there is scope for this to generate geographical heterogeneity in monopsony power as jobs and workers are not equally spaced across the spatial economy. To formalise this mechanism I develop a job search model of "local monopsony power", where utility depends on the wage, commuting and an idiosyncratic component. The model endogenously defines spatial labour markets which are continuous and overlapping and generates labour supply to the firm elasticities which arise endogenously and vary across space. The model suggests that as distaste for commuting increases, monopsony power increases, but does so heterogeneously across space. Rural areas where job options are more limited are shown to be less competitive than denser more urban areas, as a result of commutes. To validate the model I show that the model predictions in heterogeneity in monopsony power across space are consistent with heterogeneity found in my causal estimates. I additionally show that model predictions for monopsony power across 7,250 Built Up Areas

[2] This correlation is unchanged when including a barrage of controls including wage, age, sex, part time and temporary contracts, and fixed effects fo year, occupation and industry, as shown in figure 6 in the appendix.

in the UK are strongly negatively correlated with wages and worker density for the low pay retail market. I then structurally estimate the model and show that commutes are responsible for approximately 1/3 of the wage markdown.

Figure 1: Separations vs Commutes



Note: *The figure plots the probability of separation within a year against commutes measured in minutes. Commutes are measured using the Open Street map dataset for Great Britain and ArcGIS's networking tool. The sample is based on 1,429,376 worker-year observations from the Annual Survery of Hours and Earnings 2003-2019.*

This paper makes three contributions. Firstly, the empirical section on estimating the labour supply elasticity to the firm, and thus the extent of monopsony power, contributes to the empirical literature on measuring imperfect competition. Till now studies have typically either attempted to estimate the recruitment elasticity, using applicant data,[3] or the separation elasticity[4] and relied on a result from Manning (2003) that states that they are equal in absolute value. There are reasons to believe this result might not hold[5] which would mean estimates of the labour supply elasticity to the firm relying on that result are incorrect. This study is the first to estimate both the recruitment and separation elasticity for the same firm, and assess asymmetries in market power between incumbent and new workers. Furthermore, rather than assuming that the application-wage elasticity is synonymous with the recruitment-wage elasticity this study is also the first to examine how the the number of applications impacts the probability of filling a vacancy. Using these results I offer a range of markdown estimates, where the lowest relates to a firm which employs all its applicants (16%), and the highest relates to a

[3]For examples see Dal Bó et al. (2013); Belot et al. (2018); Pörtner and Hassairi (2018); Azar et al. (2019); Banfi and Villena-Roldan (2019); Dube et al. (2020a); Marinescu and Wolthoff (2020)

[4]Notable recent papers include Ransom and Sims (2010); Dube et al. (2016, 2019) and for a recent survey see Sokolova and Sorensen (2021).

[5]It is derived for a market rather than an individual firm, and relies on assumptions that recruitment from unemployment is invariant to the wage, and that the recruitment and separation elasticity are both constant.

firm with only a single vacancy to fill (25%).

Secondly, it contributes theoretically to the sources of monopsony power. There are two strands of literature on this front, one which attributes monopsony power to search frictions (Burdett and Mortensen, 1998; Manning, 2003), and a newer strand borrowing foundations from Industrial Organization where firms are imperfect substitutes for workers due to idiosyncratic preferences over non-wage aspects of a job (Card et al., 2018).[6] This study belongs to the latter of these two, though develops it by splitting the role of location and commutes out from the single idiosyncratic parameter, which recent research suggests is the most important non-wage aspect of a job for workers (Caldwell and Oehlsen, 2018; Blundell, 2020), and formally models them. In a job search model of labour supply where utility is in part dependent on the commuting distance between the location of the worker and the location of the firm, I show spatial labour markets become point specific and weaken over distances. Monopsony power is increasing in the variance of the idiosyncratic component of utility, and the role of the spatial distribution of activity in generating monopsony power depends on the commuting-wage elasticity. As distaste for commuting increases, so does monopsony power, though heterogeneously across locations. Areas with few local job opportunities, where workers need to commute for work, exercise greater monopsony power than more congested city areas. Firms in more congested city areas are still able to exercise some local monopsony power, as even small commutes within a city generate disutility.

The above result contributes to the literature on the spatial determinants of wages in urban economics. The seminal paper from Glaeser and Mare (2001) noted that urban workers earn on average 33% more than their non-urban counterparts, and this is in part due to level differences in wages (Heuermann et al., 2010; de la Roca and Puga, 2017). Much of the literature has attributed this urban wage premium to productivity gains from agglomeration (Ciccone and Hall, 1996; Glaeser, 1998; Puga, 2010; Moretti, 2011) in a competitive labour market setting. This study adds to new evidence from Hirsch et al. (2020) that part of the premium is likely driven by differences in monopsony power as the thick labour market in urban areas have more jobs in close proximity such that they are closer substitutes, and provides a micro-foundation for this.

Thirdly, it contributes by empirically assessing the sources of monopsony power. The IV estimates of the commuting-wage elasticity are the first I am aware of utilising exogenous wage variation, and suggest a strong distaste for commuting and that a firm's geographic labour market size only mildly responds to wage increases. The estimate suggests that even identical jobs, but differently located, within the same city are imperfect substitutes as workers have very strong travel preferences. By exploiting geographic variation in the location of establishments I additionally show that the model does a good job of spatially predicting more and less competitive areas, giving credibility to the mechanism suggested by the model. A structural estimation exercise matching the model predicted elasticities with the empirically estimated

[6] For evidence of heterogeneity in preferences over non-pecuniary job attributes see Eriksson and Kristensen (2014); Mas and Pallais (2017); Wiswall and Zafar (2018); Datta (2019). For studies using this imperfect substitute framework see Azar et al. (2019); Berger et al. (2019); Lamadon et al. (2019).

elasticities, additionally enables me to estimate the variance parameter related to the distribution of idiosyncratic preferences. A simple transformation into a logit model suggests that in the absence of commutes the firm's application-elasticity would be equal to this parameter, and therefore approximately 80% larger. A short-run partial-equilibrium counterfactual exercise which shrinks all commutes to zero, but holds incumbent wages constant gives a result similar in size. Estimates imply that in the absence of commutes, markdowns would be approximately 10 percentage points lower. Thus, the above mechanism concerning the role of commutes appears to be an important one.

The results in this paper have policy relevance along a number of dimensions. Improved infrastructure and travel links (e.g. new railway lines) are likely to increase competition by expanding the size of local labour markets and therefore have an upward pressure on wages. However, the size of area which is affected may be limited (e.g. small radii around the railway stations) given the strong distaste for commuting. Additionally, the low commuting-wage elasticity suggests that reduced travel-times would also be highly valued by workers. Similarly, changes to work patterns such as increased ability to work from home, would have a similar effect on worker utility, while also have a strong positive effect on reducing monopsony power. The latter of these is driven by the drastic increase in outside options for workers when all commutes essentially shrink to zero.

The analysis also speaks to the recent surge in literature and policy debates concerning employer concentration and market power in the labour market.[7] Until now studies have generally relied on the discretisation of labour markets into non-overlapping, relatively large areas with travel assumed to be costless within this area and infinitely costly at the border. This has recently come under criticism (Berry et al., 2019; Rose, 2019), partly due to the very local nature of labour markets (Manning and Petrongolo, 2017). In the UK this is generally done at the Travel To Work Area (TTWA) and in the US at the Commuting Zone (CZ) level. The mean area for TTWAs is 1,064 km^2 and travel distances within TTWAs can exceed 90 minutes. The results for the commuting-wage elasticity suggest that by using the aforementioned definitions of a spatial labour market, employer concentrations are likely to be underestimated. This is an important development given the increased calls for a stronger regulatory response to market concentration (Krueger and Posner, 2018; Naidu et al., 2018; Marinescu and Hovenkamp, 2019; Marinescu and Posner, 2019). Furthermore, within sizeable discretised areas firms' labour market power are likely to be highly heterogeneous. For example, despite both being within the same TTWA, a firm located in Soho, London is unlikely to have the same labour market power as a firm in the commuter town of Brookmans Park, Hertfordshire. By using a continuous framework with overlapping labour markets a much more accurate and granular measure of labour market power is possible. With that measure in place policy makers will be better suited to choose optimal policies to increase worker power, whether it be through labour market regulation, area

[7]Examples of work in this vein includes Schubert et al. (2021); Naidu and Posner (2021); Azar et al. (2020b); Azkarate-Askasua and Zerecero (2020); Arnold (2020); Azar et al. (2020a); Benmelech et al. (2020); Nimczik (2020); Berger et al. (2019); Jarosch et al. (2019); Qiu and Sojourner (2019); Rinz et al. (2018); Abel et al. (2018); Lipsius (2018); Hershbein et al. (2018).

specific minimum wages or collective bargaining.

The remainder of this paper is organised as follows. Section 2 introduces the data, discusses the identification strategy, and presents estimates of monopsony power and the commuting-wage elasticity. In section 3 I formally develop the model of local monopsony power, discuss its implications, and provide evidence of its validity using both causal and observational data. Section 4 structurally estimates the parameter relating to the distribution of idiosyncratic preferences and quantifies the role of commutes in generating monopsony power. Section 5 concludes.

2 Evidence of Monopsony Power and Distaste for Commuting

This section has two main aims. First, it assesses the extent of monopsony power in a low pay labour market in the UK by estimating the labour supply elasticity to the firm, a key parameter for establishing market power and markdowns. It does this by estimating three separate elasticities, the separation-wage elasticity, the application-wage elasticity and the vacancy fill-application elasticity. The second and third of these elasticities combines to give the recruitment elasticity, which when added to the negative of the separation elasticity gives the labour supply elasticity to the firm[8].

Second, it estimates the commuting-wage elasticity, a parameter for establishing workers' distaste for commuting. This parameter is of interest as commutes have the potential of being a key driver in giving rise to imperfect substitutability between jobs, and therefore generating monopsony power (Caldwell and Danieli, 2020). Furthermore, if there is evidence of a strong distaste for commuting, there is scope for this to generate geographical heterogeneity in monopsony power as jobs and workers are not equally spaced across the spatial economy.

2.1 Data and Identification

2.1.1 Data

I utilise a rich bespoke dataset for a large UK based services firm (The Company) with operations in over 300 establishments across the UK. Establishments are centrally operated by the same company using the same structure of operations and management, but there is establishment level autonomy over employment and workforce composition decisions. While The Company's main competitors are firms operating in the private sector, a large part of the firm's business is government procurement contracts. The dataset includes HR data for the period 2011-2019, and vacancy and applicant data for the period 2016-2019. The HR data covers approximately 31,000 employees and includes information on demographics, job roles, pay, start and leave dates and home location. The vacancy data includes all information that is contained in a job advert including, job role, wage, location and all text within the advert, as well as whether the position was filled or not. The applicant data includes the number of applicants for each

[8] See Manning (2003) for more details.

vacancy, and details on the applicants including home location, gender, ethnicity and whether the applicant was internal or external. The combination of these three datasets allows me to explore separations, applications, vacancy-filling and commutes, and how they respond to wages.

Table 1 presents summary statistics for The Company in March 2019 along with comparative statistics from the Labour Force Survey (LFS). As can be seen 60% of the firm's workforce are female, marginally higher than the national average. The Company's workforce are also younger when compared to the national average. The median worker is 33 years old, while the median worker in the LFS is 42 years old. Almost half of the firm's workers are classified as "Entry-Level". These jobs are typically minimum-wage jobs in the UK, and would be considered unskilled. The Company has a very large number of workers on zero-hours contracts[9], only about 30% of the workforce have permanent contracts, which is much smaller than the national average. These types of contracts are however more prevalent in minimum wage jobs (Datta et al., 2019). The average wage in the firm is £12.88, and this is around 16% lower than the national average. The mean commute for workers in the firm is 24.3 minutes while the median is considerably less at 16.8 minutes. Approximately half of the firm's workforce are based in establishments located in London.

Table 1: Summary Statistics, March 2019

Variable	**The Company**			**LFS**		
	Mean	S.D.	Median	Mean	S.D.	Median
Female	0.60			0.52		
Age	35.9	14.3	33.0	41.9	13.2	42.0
Entry-Level	0.49					
Permanent	0.28			0.94		
Hourly Rate (£)	12.88	5.87	10.20	15.10	10.11	12.03
Commute (minutes)	24.3	17.0	16.8			
London	0.53			0.12		
N						
Workers	18,773			36,125		
Establishments	362					

Note: *The table presents worker-level summary statistics for The Company and the related statistics from the nationally representative Labour Force Survey, as of March 2019.*

Table 10 in the appendix contains summary statistics on the vacancy data which unsurprisingly shows a very similar pattern to those from 1. It additionally shows that the advertised vacancies receive 19 applications on average.

2.1.2 Exogenous Wage Variation

Estimating the recruitment-wage, separation-wage and commuting-wage elasticities relies on being able to isolate exogenous variation in wages at a minimum at the establishment-level.

[9]For a complete description of what these types of contracts entail see Datta et al. (2019).

This is necessary as these parameters identify the responsiveness of recruits, separations and commutes to a change in a single firm's wage, while the rest of the market remains unchanged. I exploit two instruments to achieve this. Firstly, I utilise a location specific wage floor that affects a very small number of workers in an area, but is binding for The Company in that location for jobs which are paid less than the Living Wage. The Living Wage Foundation (LWF) is a charitable organisation in the UK that was established in 2011, that campaigns for employers to pay workers a living wage. Organisations can voluntarily sign up to become Living Wage employers and following appropriate audits by the LWF can achieve accredited status. As of July 2020, the LWF lists 6,562 accredited employers and included in this list are 107 local government units. When public bodies achieve accreditation, they are given a temporary amnesty on existing procurement contracts, but are required to enforce the living wage at the start, renegotiation or renewal of contracts. The Company operates in the service sector and the majority of their business is through procurement contracts with local councils. As the firm operates hundreds of establishments across the UK, different establishments become contractually obliged to pay the LLW and UKLW at different times. This is dependent on whether, and when, the local government unit has voluntarily signed up to the LWF's Living Wage, as well as idiosyncratic timings of contractual renewal or renegotiation. The Company's pay structure is centrally determined, and they have two regional pay scales for the UK (London and rest of the UK). When an establishment is exposed to the Living Wage, it effects only those workers within the establishment whose pay point is below the mandated Living Wage (i.e. entry-level workers), and the remainder of wages in the local labour market remain unchanged.

Between 2012 and 2019 107 local government units gained accreditation. For example, of the 32 London Boroughs, 17 have received accreditation, the earliest (Islington) receiving accreditation in May 2012, and the most recent (Redbridge) receiving accreditation in November 2018[10]. As figure 2 shows, this setting gives a large amount of variation in Living Wage treatment for establishments run by The Company. In particular, over the period for which we have HR data, approximately 140 establishments went from being untreated to treated, while run by The Company. The living wage rates for London (LLW) and the rest of the UK (UKLW) are calculated each year by the LWF and the Resolution Foundation and have typically been considerably higher than the mandatory National Minimum Wage (NMW) and National Living Wage (NLW). The LLW rate has typically been approximately 30-35% higher than the mandatory minimums, while the UKLW has been about 15-20% higher as can be seen by figure 7 in the appendix .

To ensure that this instrument can causally identify the recruitment, separation and commuting elasticities, it's necessary to ensure that the Living Wage adoption within an area only affects a very few number of jobs. If for example, it affected all low pay jobs in the area the relative wage offered by the establishment would remain unchanged, and therefore it would be reasonable to assume this would have little affect on separations, recruitment and commutes. It's important to highlight therefore that when a council signs up to the LWF's living wage it only

[10]Correct as of July 2019.

Figure 2: Living Wage Roll-out

Note: *The figure shows the cumulative establishment-level Living Wage treatment for The Company, 2011 - 2019.*

affects council employees and those who are subcontracted to do work for the council. Council employment makes up approximately 3% of employment and is usually made up of workers more skilled than would be affected by the wage floor. As an example table 11 in the appendix gives estimates of the employment counts and shares for the London Borough of Hackney, and shows council employment accounts only for 3.3% of total employment in the borough. Furthermore, examination of the pay scale documentation for the borough show that the lowest paid point is 8% above the binding LLW for 2019. This is suggestive that the council adoptions of the Living Wage affects only a fraction of a percentage of workers in the area. While a significant portion of The Company's business comes from procurement contracts with the council, their main competitors are private sector firms which would not be affected by the Living Wage adoption by the council. Finally, though The Company employs thousands of workers across the UK, their average establishment only has about 25 entry-level workers. Therefore, when an establishment gets treated, there are not concerns regarding wage spillovers to other firms in the local area, as may be the case with very large employers (Derenoncourt et al., 2021).

Secondly, I utilise an instrument related to saliency of the wage posted. In the UK, whether a job has a permanent or zero-hours contract, the firm is required by law to give the worker 28 days paid annual leave. Due to the nature of non-permanent work, many firms opt to give the statutory annual leave as a top up to the wage, which calculates to a wage supplement of 12.07%. Within the vacancy data, some non-permanent jobs (approximately 20%) are advertised with the annual leave top-up included in the advertised wage, while the text of the adverts stay constant. Discussions with the HR team at The Company concluded that this occurred due to idiosyncrasies in whomever happened to be posting the job that day onto the HR system[11]. This lends itself for use as a seemingly random instrument. This instrument however can only be used for non-permanent jobs, as it is not well defined for permanent jobs. Furthermore, it can only be used for the recruitment and commuting elasticity estimates, as it is not relevant to separations.

2.2 Estimates of Monopsony Power

2.2.1 Empirical Framework

To estimate the labour supply elasticity to the firm I estimate the recruitment and separation elasticities separately and then add the negative of the separation elasticity to the recruitment elasticity.

To estimate the recruitment elasticity I consider two settings, one where a firm hires all applicants and thus the application-wage elasticity reflects the recruitment elasticity, and one where a vacancy reflects a single opening, and thus an increase in applicants increases the probability of filling a vacancy. To assess whether it's reasonable to treat the application elasticity as the recruitment elasticity would depend on both the firm's production function and on the demand for their output product. For example, in online task based markets where all workers willing to work at the given wage rate are taken on to carry out a homogenous task, such as in Dube et al. (2020a,b), the responsiveness of applicants to wages is precisely the recruitment elasticity. On the other hand, for a firm which has a Leontief production function, and only one spare unit of capital (for example, a garden maintenance company with a vacant lawnmower), the recruitment elasticity would be akin to the vacancy fill-wage elasticity, which would be less than the applicant-wage elasticity.[12] For The Company however, the latter of these will be more relevant as their structure of vacancy openings more closely resembles this. That is, a vacancy opening generally relates to a single fillable position, and the increase in applicants enables a greater likelihood of filling that vacancy.

To estimate the recruitment elasticity I estimate:

$$log(Apps_{ajemy}) = \beta_1 log(Wage_{ajemy}) + \gamma_{je} + \lambda_{ey} + \nu_{ym} + \theta_{jy} + \epsilon_{ajemy} \tag{1}$$

and

$$Filled_{ajemy} = \beta_2 log(Apps_{ajemy}) + \gamma_{je} + \lambda_{ey} + \nu_{ym} + \theta_{jy} + \epsilon_{ajemy} \tag{2}$$

where $Apps_{ajemy}$ is the number of applicants applying to advert a advertising for job-role j in establishment e in month-year my, $Filled_{ajemy}$ is a binary variable indicating whether the position was filled, $Wage_{ajemy}$ refers to the advertised hourly wage (in £), γ_{je} are job-role es-

[11] This phenomenon was observed consistently over the entire time period, and there is considerable variation within establishments and across jobs.

[12] This point is discussed analytically in section 3.2. However intuitively it can be seen by the fact that the fill-wage elasticity would be equal to the applicant-wage elasticity multiplied by the fill-applicant elasticity, and the fill-applicant elasticity would be less than 1.

tablishment fixed effects, λ_{ey} and θ_{jy} are time-varying establishment and job-role fixed effects [13], ν_{ym} are year-month fixed effects.

To estimate the separation elasticity I regress a linear probability model (LPM) of the form:

$$Leave_{ijemy} = \beta_3 log(Wage_{ijemy}) + \gamma_{je} + \lambda_{emy} + \theta_{jmy} + \epsilon_{ijemy} \tag{3}$$

where $Leave_{ijemy}$ is an indicator variable equal to 1 if individual i leaves job-role j in establishment e in a particular year-month ym, and equal to 0 otherwise.[14]

Equations (1), (2) and (3) utilise variation within the same job role-establishment combination while controlling for establishment-level time shocks, and job-level time shocks. They are akin to a triple-difference specification. Though these specifications are very flexible, concerns relating to endogeneity still exist. Job-location specific time shocks are still conceivable, which in turn could be correlated with wages, recruitment and separations (see Belot et al. (2018); Marinescu and Wolthoff (2020) for evidence of this). As a result I instrument $log(Wage)$ in equations (1) and (3) and $log(Apps)$ in equation (2) with the two instruments outlined in section 2.1.2. The first LW_{jewm}, the Living Wage instrument which is defined at the establishment-job-time level, and the second, AL_{ajemy}, the annual leave wage saliency instrument which is defined at the advert level.

$$LW_{jemy} = \begin{cases} 1 & \text{if establishment is subject to LW \& LW was binding for job} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

$$AL_{ajemy} = \begin{cases} 1 & \text{if advert included annual leave in hourly rate} \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

Assuming parameters β_1, β_2 and β_3 are identified, and we assume that a vacancy only relates to one fillable position, then

$$\varepsilon_{nw} = \underbrace{\varepsilon_{rw}}_{\frac{\hat{\beta}_2}{E[\widehat{Filled_{ajemy}}]} * \hat{\beta}_1} - \underbrace{\varepsilon_{sw}}_{\frac{\hat{\beta}_3}{E[\widehat{Leave_{ijemy}}]}} \tag{6}$$

however, in a setting where all applicants are employed $\varepsilon_{rw} = \hat{\beta}_1$. One could also use these two estimates of the recruitment elasticity to place upper and lower bounds on the parameter. While the structure of a company's vacancies may lend itself more to the interpretation as in equation (6) it's also worth noting that an increase in wages may lead to an increase in quality of applicants (Dal Bó et al., 2013), and therefore an increase in the effective units of labour which would imply $\frac{\hat{\beta}_2}{E[\widehat{Filled_{ajemy}}]} * \hat{\beta}_1$ underestimates the elasticity.

[13] The time varying job-role fixed effects are only used in some specifications due to saturation concerns.

[14] A recent survey from Sokolova and Sorensen (2021) found that results estimated using the more straightforward LPM were not statistically different from results utilising non-linear estimation techniques.

2.2.2 Results

Table 2: Application - Wage Estimates

	(1) log(Apps)	(2) log(Apps)	(3) log(Apps)	(4) log(Apps)	(5) log(Apps)
log(Wage)	0.560**	4.706**	2.407***	3.679**	2.753***
	(0.238)	(2.154)	(0.662)	(1.414)	(0.622)
First Stage	-	0.030***	0.085***	0.066***	-
		(0.005)	(0.008)	(0.009)	
Job-Establishment FE	Yes	Yes	Yes	Yes	Yes
Establishment-Year FE	Yes	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes	Yes
Job-Year FE	Yes	Yes	No	No	No
N	5487	5487	2376	2376	2376
First Stage F-Stat.	-	61.35	385.8	127.1	288.4
Hansen J Stat.	-	-	-	-	0.714
Instrumented With *AL*	No	No	Yes	No	Yes
Instrumented With *LW*	No	Yes	No	Yes	Yes
Sample	All	All	Non-Perm	Non-Perm	Non-Perm

Note: *The table presents estimates of* $\hat{\beta}_1$ *from equation (1) via OLS or where* $log(Wage)$ *is instrumented with either* LW_{jemy}, AL_{ajemy} *or both instruments. If only one instrument is used the table reports the accompanying implied first stage coefficient. If both instruments are used the table reports the over identifying test statistic. All columns report the first stage Cragg-Donald F statistic. Standard errors are reported in parentheses and are clustered at the establishment-job. Col (1) includes a control for whether the job advert was for a salaried position.* * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2 reports estimates of $\hat{\beta}_1$ from (1), column (1) reports OLS estimates and columns (2-5) where $log(Wage)$ is instrumented using one or both of the two instruments discussed in section 2.1.2. It also reports the relevant first stage coefficients for the specifications where only one of the two instruments are employed. Column (2) reports the specification utilising the entire sample and the living wage instrument, column (3) utilising only the sample of non-permanent adverts and the annual leave instrument, column (4) the non-permanent sample and the living wage instrument, and column (5) the non-permanent sample using both instruments. The specifications using the smaller sample do not include job-year fix effects to reduce saturation concerns.

All specifications report a statistically significant applicant-wage elasticity, with estimates ranging between 2.4 to 4.7, aside from the OLS specification which underestimates the elasticity, a finding similar to other recent studies (e.g. Marinescu and Wolthoff (2020)). A possible explanation of this is that firms adjust wages to match local market conditions, and therefore wage changes within the firm are correlated to outside wage options. The midpoint of the IV estimates implies that a 10% increase in the posted wage would increase applicants by 37%.

All specifications report a sizeable Cragg-Donald F-statistic, demonstrating strength in both instruments. Specification (5) additionally reports the Hansen J statistic exploiting the fact the equation is overidentified. The reported χ^2 statistic confirms that the validity of the instruments cannot be rejected.

Table 3: Filled-Applicants Estimates

	(1) Filled	(2) Filled
log(Apps)	0.248***	0.250**
	(0.0127)	(0.127)
Job-Centre FE	Yes	Yes
Centre-Year FE	Yes	Yes
Job-Year FE	Yes	No
Year-Month FE	Yes	Yes
N	4859	2115
R^2	0.522	-
First Stage F-Stat.	-	16.37
Hansen J Stat.	-	1.415
Instrumented With *AL*	No	Yes
Instrumented With *LW*	No	Yes
Sample	All	Non-Perm
Mean of Dep. Var	0.698	0.663

Note: *The table presents estimates of* $\hat{\beta}_2$ *from equation (2). Column (1) reports the OLS specification and column (2) reports the specification where* $log(Apps)$ *is instrumented with* LW_{jemy} *and* AL_{ajemy}*, and additionally reports the over identifying test statistic and the first stage Cragg-Donald F statistic. Standard errors are reported in parentheses and are clustered at the establishment.* * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3 reports estimates of $\hat{\beta}_2$ from (2). Column (1) reports the OLS estimates while column (2) the specification using both instruments. Given the instruments are impacting $log(Apps)$ through two steps[15] the first stage is considerably weaker, and therefore only the specification with both is reported. The implied fill-applicant elasticity is effectively identical across the two specifications. By calculating $\frac{\hat{\beta}_2}{\widehat{E[Filled_{ajemy}]}}$ one can see the elasticity ranges from $0.36 - 0.38$. This implies a 10% increase in the number of applicants would increase the likelihood of filling the vacancy by 3.7%. Similar to table 2 the reported Hansen J statistic implies we can not reject the validity of the instrument. Taking the midpoint of the range of estimates of the applicant-wage elasticity from table 2, 3.7, and a fill-applicant elasticity of 0.37 suggests the recruitment elasticity, ε_{rw} lies between 1.4 and 3.7. The structure of the vacancies for this firm are such that a job posting typically relates to a single opening, and therefore the lower end of this is more likely to be reasonable.

[15] I.e. they act on wages, which in turn acts on applicants.

Table 4: Separation-Wage estimates

	(1) Leave	(2) Leave	(3) Leave
log(Wage)	-0.0234***	-0.0791**	-0.0854***
	(0.0032)	(0.0317)	(0.0323)
First Stage	-	0.0625***	0.0594***
		(0.003)	(0.003)
Job-Centre FE	Yes	Yes	Yes
Centre-Time FE	Yes	Yes	Yes
Job-Time FE	Yes	Yes	Yes
Controls	No	No	Yes
N	1055521	1055521	1055521
First Stage F-Stat.	-	4404.8	4793.3
Instrumented With *LW*	No	Yes	Yes
Mean of Dep. Var		0.048	

Note: *The table presents estimates of $\hat{\beta}_3$ from equation (3) where log(Wage) is instrumented with LW, the accompanying first stage coefficient and the first stage Cragg-Donald F statistic. Standard errors are reported in parentheses and are clustered at the establishment. Col (2) includes controls include Gender, BAME, Contract, Tenure and Age.* * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4 presents estimates of $\hat{\beta}_3$ from (3) via OLS and using the living wage instrument, as well as $E[\widehat{Leave}_{ijemy}]$. Column (1) reports OLS estimates without controls, Column (2) reports IV estimates without controls, while column (3) reports IV estimates with controls for gender, ethnic minority status, whether permanent or not, tenure and age. The OLS estimate as before underestimates the elasticity. The IV estimate is robust to the inclusion of controls and implies an ε_{sw} of approximately -1.7, implying a 10% increase in the wage reduces the likelihood of separation in any month by 17%.

Combining the estimates that $\varepsilon_{rw} \in [1.4,\ 3.7]$ and $\varepsilon_{sw} \approx -1.7$ this implies $\varepsilon_{nw} \in [3.1,\ 5.4]$. According to the traditional wage markdown (μ) equation[16] the estimates suggests a wage markdown ranging between 16-25%. As discussed, the upper end of this markdown estimate is likely to be more accurate. These results suggest considerable market power in a low wage labour market where frictions such as firm-specific capital are less likely to play a role.

2.2.3 Discussion

The estimate of the labour supply elasticity to the firm above makes two advances on the existing literature. First, it does not rely on the result from Manning (2003) that $\varepsilon_{rw} = -\varepsilon_{sw}$

[16] $\mu = 1 - \frac{\varepsilon_{nw}}{1+\varepsilon_{nw}}$.

which relies on a number of restrictive assumptions, including recruitment from unemployment is invariant to the wage, and that the recruitment and separation elasticity are both constant [17]. It is possible there may be differences in firm's monopsony power over new recruits and separations, and the above results test this. My findings suggest that the application-wage elasticity is approximately twice the size of the separation-wage elasticity, and therefore firms exercise more monopsony power over incumbent workers than in attracting new recruits. This is an interesting finding and the source of this deserves investigation which is beyond the scope of this paper. However, a possible explanation is that incumbent workers are likely to have invested time in their existing job (e.g. through building relationships with colleagues, getting to know specifics of the firm) which they place a value on, which in turn creates a greater friction when responding to wages. That said due to the firm's hiring structure (increases in applicants only effects recruitment via increasing the probability of filling vacancies), the recruitment and separation elasticity are similar in size. Second, utilising the information on both applicants and whether the vacancy is filled I offer a range of markdown estimates, where the lowest relates to a firm which employs all its applicants (16%), and the highest relates to a firm with only a single vacancy to fill (25%)

The aforementioned results bare some similarity to recent estimates of monopsony power from the literature, and may also be instructive as to whether restrictions employed in the literature are reasonable. In terms of application and recruitment elasticities the results are within a similar range to some recent studies utilising data on applications (e.g. Dal Bó et al. (2013); Falch (2017); Azar et al. (2019)) and higher than some others (e.g. Dube et al. (2020a); Belot et al. (2018)). That said the results suggest using an application elasticity as synonymous with the recruitment elasticity as done in Azar et al. (2019) is likely to underestimate the extent of monopsony power. The separation elasticity is of a similar magnitude to recent papers with a clear identification strategy. Bassier et al. (2020) report that low pay sectors generally have lower separation elasticities and range from -1.2 to -1.4, which are very similar to those estimated here. Similarly Dube et al. (2019) estimate a separation elasticity of -2.3. The latter of these studies utilises data on a single firm operating in the retail sector in the US and so is similar in set up to this study.

The results are robust to a number of potential issues which are explored further in section C in the appendix. Using a triple-difference event study design I show parallel pre-trends and no evidence of announcement effects in the separation elasticity estimates (see figure 18). Additionally, I address the concerns recently raised regarding the workings of two-way fixed effect estimators, with staggered treatment timing (Borusyak and Jaravel, 2017; Sun and Abraham, 2020; Callaway and Sant'Anna, 2020; Goodman-Bacon, 2021). Despite the fact this study employs a more flexible estimation strategy (akin to a triple-difference estimator), as a matter of caution I check whether any of the issues raised in the aforementioned studies could be sullying the results when using the Living Wage instrument. I do this by comparing a traditional two-

[17]The latter of these is a particularly strong assumption as it is not obvious that a firm in a minimum wage labour market in the UK would expect the same proportional response in recruits when increasing wages from £10 to £11, as when increasing wages from £20 to £22

way fixed effect event study estimator at the establishment level with a robust estimator (similar to that from Sun and Abraham (2020)), examining the impacts on wages. Figure 20 plots the two sets of estimates and as can be seen there is very little noticeable difference between the two panels. Given these results it is unlikely that the more flexible specifications utilising a triple-difference estimator will suffer from the aforementioned issues.

2.3 Estimates of the Commuting-Wage Elasticity

2.3.1 Empirical Framework

To estimate the commuting-wage elasticity I regress

$$log(Comm_{iajemy}) = \beta_4 log(Wage_{ajemy}) + \delta' X_{imy} + \gamma_{je} + \lambda_{ey} + \nu_{ym} + \theta_{jy} + \epsilon_{iajemy} \quad (7)$$

where $Comm_{iajemy}$ is the commuting time measured in minutes between applicant i's home and address of the establishment. Commutes are measured using the Google Maps API and are calculated for an arrival time of 9am so as to account for traffic. Commuting times utilised in this exercise are therefore considerably more accurate than typical "as the crow flies" distances calculated via GIS software. Given the popularity of google maps for routing[18] it is also a reasonable approximate of expected commuting times by job searchers. Commuting time was calculated for both car and public transport, and it is assumed that workers choose the fastest method of the two options.[19] $log(Wage_{ajemy})$ is, as before, instrumented with LW_{jemy} and AL_{ajemy}, and X_{ia} is a set of controls including individual's gender, ethnicity, and whether they were internal or external applicants, and whether the advert was for a non-salaried job.[20]

2.3.2 Results

Table 5 presents results from estimating equation (7). Aside from column (1) which reports the OLS estimates (and is again underestimated), point estimates across all specifications are consistent and imply a commuting-wage elasticity $\varepsilon_{cw} \approx 1$. Specifications containing the annual leave instrument are more precisely estimated, and the specification with both instruments is statistically significant to a 5% level. All specifications have a sizeable first stage F-stat, and the over-identified specifications in column (4) indicates as before that the validity of the instruments cannot be rejected.

The estimated commuting-wage elasticity suggests a strong distaste for commuting, and that a firm's geographic labour market size is only mildly responsive to wage increases. For an hourly wage of £7.50 ($9.67) (the NMW for the midpoint of the sample period) and average commute of 25 minutes, a commuting-wage elasticity ≈ 1 implies an extra 5 minutes of total commute

[18] The app is ranked top for navigation on the apple and android app stores, and has over one billion active monthly users.

[19] In practice this means that many people working in London are assumed to utilise public transport which is a reasonable assumption.

[20] This control was only included for specifications using the entire sample.

Table 5: Commuting - Wage Estimates

	(1) log(Comm)	(2) log(Comm)	(3) log(Comm)	(4) log(Comm)	(5) log(Comm)
log(Wage)	0.362***	1.081	0.922*	0.955	0.931**
	0.105	(1.208)	(0.494)	(0.774)	(0.417)
Job-Centre FE	Yes	Yes	Yes	Yes	Yes
Centre-Year FE	Yes	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes	Yes
Job-Year FE	Yes	Yes	No	No	No
N	47313	47313	19585	19585	19585
First Stage F-Stat.	-	673.3	3440.7	1220.7	2562.1
Hansen J Stat.	-	-	-	-	0.00141
Instrumented With *AL*	No	No	Yes	No	Yes
Instrumented With *LW*	No	Yes	No	Yes	Yes
Sample	All	All	Non-Perm	Non-Perm	Non-Perm

Note: *The table presents estimates of* $\hat{\beta}_4$ *from equation (7) where* $log(Wage)$ *is instrumented with either* LW_{jemy}, AL_{ajemy} *or both instruments. If both instruments are used the table reports the over identifying test statistic. All columns report the first stage Cragg-Donald F statistic. Standard errors are reported in parentheses and are clustered at the advert. Regressions are weighted by the inverse number of applicants for each job. Controls include gender, ethnicity, whether applicants were internal and whether the advert was for a permanent job.* * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

would require an extra £3.00 ($4.06) per day[21], or £0.37 ($0.50) per hour assuming an 8 hour work day. The results are suggestive that commuting distances are likely to be a strong factor in generating imperfect substitutability between jobs.[22]

To ascertain robustness of the aforementioned results section C in the appendix explores the assumption of a constant, linear in logs, relationship between wages and commutes as well as the impacts of the higher wage across the distribution of commutes. The results suggest the assumed relationship is reasonable one, and that the impacts across the majority of the distribution are constant, aside from at the tail ends and the CDF under a higher wage would stochastically dominate a CDF under a lower wage, as expected.

[21]The average worker at The Company works 4 hours a day, as noted previously there are many non-permanent and part time staff employed.

[22]The results imply a WTP considerably larger than recent estimates from Le Barbanchon et al. (2021) but similar in size to higher end values of the Value of Travel Time used by the British Deparment of Transport (Batley et al., 2019) and earlier estimates from the literature (Timothy and Wheaton, 2001; Van Ommeren and Fosgerau, 2009).

3 The Role of Commutes in Generating Monopsony Power: Theoretical Framework

The estimates from section 2 suggest strong monopsony power in the labour market and a strong distaste for commuting. It is reasonable to see how the latter of these can cause the former: the length of commute to a place of work is a non-wage factor affecting the utility from a job, and therefore generates imperfect substitutability between differently located jobs. The following section formalises this mechanism in a job search model. I additionally show how this mechanism is not only a factor in generating monopsony power, but that it generates monopsony power heterogeneously across space depending on the spatial distribution of workers and firms. I then validate the model by showing how the model predictions of heterogeneity in monopsony across space are consistent with heterogeneity found in the data used in section 2. I additionally show that model predictions for monopsony power across 7,250 Built Up Areas are strongly negatively correlated with wages, worker density and population as one would expect, for the low pay retail labour market. Given the validity of the model, I then structurally estimate the model in section 4 to quantify how much of monopsony power is due to commutes.

I utilise a job search model where firms gain profits from hiring (or maintaining) workers and worker utility depends on the wage, commuting and an idiosyncratic component. The model endogenously defines spatial labour markets which are continuous and overlapping, generates endogenous labour supply to the firm elasticities which vary across space, and shows the importance of the distribution of spatial activity, and preferences over commuting in generating monopsony power. Sources of monopsony power in the model are generated by idiosyncratic preferences over jobs, preferences over commuting, and the spatial distribution of activity. An extension studying the role of search costs is explored in the appendix in section E. The main take away from this model is that distaste for commuting is a key factor in generating monopsony power but does so heterogeneously. As distaste for commuting increases, rural areas with less outside opportunities experience greater monopsony power than denser urban areas where many firms and workers are located in close proximity; though the latter still sees increases in monopsony power.

3.1 Model Setup

There are many firms j and workers i, spatially located across the economy, l_i, l_j. Firm and worker locations are treated as given and fixed. Firms consider themselves as small and therefore I abstract away from strategic interaction. Furthermore, it's assumed that firms can not wage discriminate workers. In the following subsection I concentrate on the applying (recruitment) worker problem, however results for the maintaining (separation) problem are very similar and can be found in section D in the appendix. Additionally, while the model can flexibly include search costs, I shall predominantly focus on a version where search costs are assumed to be zero. Derivations including search costs can be found in the appendix in section E.

3.2 The Firm Problem

Firm's gain profits from job j according to:

$$\Pi_j = (p_j - w_j)n_j(w_j) \tag{8}$$

where p_j is productivity of job j, w_j is the wage and n_j is the employment. The subscript j on the employment function demonstrates the fact that this will vary across firms.

Therefore the first order condition for the firm can be written as

$$w_j = p_j \frac{\varepsilon_{nw_j}}{1 + \varepsilon_{nw_j}} \tag{9}$$

where ε_{nw_j} is the elasticity of labour supply to the firm. The above is a form of the traditional monopsonistic "rate of exploitation" (Pigou, 1924) where the gap between wages and productivity grows as the labour supply to firm - wage elasticity shrinks. The limiting case where ε_{nw_j} tends to infinite is akin to a perfectly competitive market.

In the steady-state, employment is such that

$$n_j = \frac{r_j(w_j)}{s_j(w_j)} \tag{10}$$

where r_j is recruitment for firm j which is a function of wages, and similarly s_j is separations for firm j which is additionally a function of wages. Therefore taking logs and differentiating by w_j yields the relationship between the labour supply elasticity and the recruitment and separation elasticity.

$$\varepsilon_{nw_j} = \varepsilon_{rw_j} - \varepsilon_{sw_j} \tag{11}$$

Recruits can be further described such that when a firm posts a job advert for job j the number of recruits are

$$r_j(w_j) = \Phi(A_j(w_j)) \tag{12}$$

where Φ describes the relationship between recruitment and the number of job applications A_j, which in turn is a function of wages. Unlike the application-wage function and the separation-wage function, Φ is assumed to be constant across firms. Therefore the recruitment elasticity can be represented by

$$\varepsilon_{rw_j} = \varepsilon_{\Phi A}\varepsilon_{Aw_j} \tag{13}$$

Equation 13 demonstrates what was discussed in section 2. If the firm hired a constant ratio of workers then $\varepsilon_{\Phi A} = 1$ and $\varepsilon_{rw_j} = \varepsilon_{AW_j}$. However, assuming there is only one vacancy to fill as per The Company, then $\varepsilon_{\Phi A} < 1$. I shall assume going forward the latter of these is the case.

Workers are homogenous in productivity, however some worker-firm matches are determined "unsuitable". Therefore there is a non-zero probability of a vacancy being unfilled despite $A_j \geq 1$, as is a feature of the data used in section 2.2.2. In particular I assume the probability of some worker being suitable to be q, which is independent of any other characteristics of the workers and unknown to the worker before applying. Thus

$$\Phi(A_j) = 1 - (1-q)^{A_j} \quad {}^{23} \tag{14}$$

Which implies $\Phi(A)$ is concave and bounded between 0 and 1.

3.3 The Worker Problem

Firms post jobs j, which are advertised with a wage w_j and a location l_j.

Worker i located at l_i, employed in job j gets utility from the job according to:

$$u_{ij} = w_j d_{ij}^{\frac{-1}{\varepsilon_{cw}}} \nu_{ij} \tag{15}$$

where $d_{ij} = |l_i - l_j|$ is a measure of distance between the worker and the job. ν_{ij} is an idiosyncratic utility component known only to the worker[24] and is distributed iid following some distribution $F(v_{ij})$, defined over the support $[0, \infty]$. ν_{ij} could be thought of as j specific characteristics and i's preferences over them, for example, management style, flexibility and firm ethics. The parameter ε_{cw} is the commuting-wage elasticity and reflects the importance of commutes in determining worker utility. The utility function in equation (15) reflects how workers will see jobs as imperfect substitutes as a result of idiosyncratic preferences and the commuting distance and the importance of these two factors will be determined by the size ε_{cw} and the variance of $F(v_{ij})$.

A worker i, who is currently in job j[25] will choose to apply to posted job j' if the expected utility of doing so is greater than not applying:

$$P(A_{j'})u_{ij'} + (1 + P(A_{j'}))u_{ij} \geq u_{ij} \tag{16}$$

where $P(A_{j'})$ is the probability of getting the job, which is a function of the total number of applicants to job j', $A_{j'}$[26]

Rearranged worker i will apply to j' if

[23] At low values of q one can approximate using $\Phi(A_j) = 1 - e^{-qA}$. This would imply $\frac{\partial \Phi}{\partial A} = qe^{-qA} > 0$ and $\frac{\partial^2 \Phi}{\partial A^2} = -q^2 e^{-qA}$

[24] Firms never observe worker locations and ν_{ij}, only the distributions to avoid wage discrimination.

[25] Note the model is flexible enough to allow job j to be unemployment. In such a case w_j would be their unemployment benefit and d_{ij} could be considered the average distance they have to travel to attend job interviews and meetings at their job centre.

[26] This drops out in this setting but is important when search costs are introduced as explored in section E.

$$\nu_{ij'} \geq \underbrace{\frac{w_j d_{ij}^{\frac{-1}{\varepsilon_{cw}}} \bar{\nu}}{w_{j'} d_{ij'}^{\frac{-1}{\varepsilon_{cw}}}}}_{\equiv x_{ijj'}} \tag{17}$$

Therefore the probability of worker i in job j applying to job j' is given by:

$$Pr(Apply_{ij}^{j'}) = 1 - F(x_{ijj'}) \tag{18}$$

Note that $x_{ijj'}$ here is a measure of relative utility between the incumbent job j, and the potential job j'. Additionally note in equation 17 the idiosyncratic component related to the incumbent job is normalised to $\bar{\nu}$, which in practice will be treated as the median draw.[27]

3.4 The Individual's Elasticity

The first key observation here relates to the elasticity of applying for the individual. Under the assumption the distribution $F(x)$ is such that $\frac{\partial h(x)x}{\partial X} \geq 0$, where $h(x)$ is the hazard function related to $F(x)$[28] the individual specific application elasticity is given by

$$\varepsilon_{Aw_{ij}^{j'}} = h(x_{ijj'})x_{ijj'} \tag{19}$$

and is decreasing in potential relative utility. The above assumption on the distribution of ν_{ij} is not a restrictive one and is fulfilled by commonly used distributions including Weibull, exponential, lognormal, loglogistic and Frechet.[29]

The above formulation of the individual's responsiveness of applying with respect to wages has a number of interesting and important features. Firstly, it can be seen that as the posted wage increases, $x_{ijj'}$ decreases and therefore the elasticity of applying to wages decreases. Given that the probability of applying is bounded between 0 and 1, this result is intuitive.[30] Secondly, the smaller the standard deviation of idiosyncratic preferences, the closer the behaviour of the individual to the perfectly competitive benchmark, and vice versa.[31] Both of these points are demonstrated graphically in figure (9) in the appendix.

[27] This normalisation is required for tractability reasons, as distributions under study do not have a neat closed form for the ratio of two random variables. The key patterns from the simulation exercises in sections 3.4 and 3.5 are however unchanged when allowing the incumbent idiosyncratic component to be a random variable.

[28] I.e. $h(x) = \frac{f(x)}{[1-F(x)]}$

[29] Manning and Petrongolo (2017) utilise a pareto distribution which has the feature that $h(x)x$ is equal to a constant parameter of the distribution, and therefore $\varepsilon_{AW_{ij}^{j'}}$ is constant for all worker-firm combinations.

[30] A worker who is being paid £10 an hour in their incumbent job would see a much greater percentage change in their probability of applying if the wage was increased from £8 to £8.80 than from £12 to £13.20 ceteris paribus, as in the former example their initial probability of applying would be closer to 0, and in the latter closer to 1.

[31] It can in fact be shown that in the absence of commuting costs, the shape parameter related to the weibull distribution will equate to the labour supply elasticity facing the firm, which is outlined in greater detail in section 3.5

Thirdly, the lower the elasticity of commuting to wages[32], ε_{cw}, the more important commutes are in determining the elasticity of applying to wages for the individual. For example, assuming $\varepsilon_{cw} < \infty$, the inelastic part of the curve occurs when a firm offers higher wages, or is located relatively nearer to workers than their current job. This would suggest that workers which currently have to commute far, but a job opening comes up close by to them, would have a more inelastic labour supply curve to the posted job. The lower the commuting elasticity, the greater the impact of space on the application elasticity. This is exemplified in figures 10 and 11 in the appendix which present a simulated distribution of two workers and their respective labour supply curves.

3.5 The Elasticity of Labour Supply To The Firm

The labour supply to the firm can be calculated by summing equation (18) over all relevant workers in the economy, which gives

$$A_{j'} = \sum_{i,j} [1 - F(x_{ijj'})] \tag{20}$$

Equation (20) reflects the lack of need to discretise the economy into geographical labour markets. Labour markets are endogenously constructed, point specific, become continuous, and weaken over distances. Ceteris paribus workers close by to the advertised job are more likely to apply and workers farther away from the job are less likely to apply, and the extent of this factor is dependent on their disutility for commuting. Furthermore, it also accounts for workers' incumbent options. For example, if a firm is located far from workers, but those workers are currently commuting a similar distance, the size of the labour market would be larger than a situation where a firm is close to workers and those workers only currently commute a small distance.

It can be shown that the application elasticity to the firm is

$$\varepsilon_{Aw^{j'}} = \frac{\sum_{(i,j)} \varepsilon_{Aw_{ij}^{j'}} [1 - F(x_{ijj'})]}{\sum_{(i,j)} [1 - F(x_{ijj'})]} \tag{21}$$

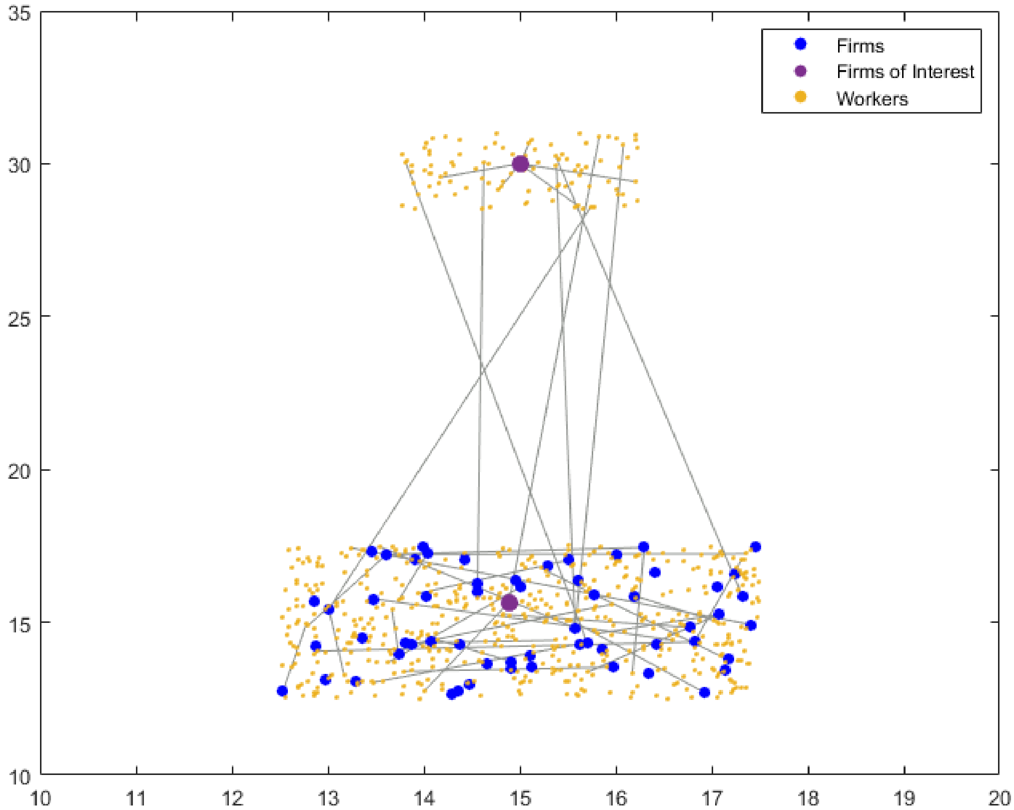
which is simply a weighted average of the individual level elasticity from equation (19), where the weights are the probability of applying to the job. Therefore workers who are more likely to apply to the job (and therefore have a lower individual elasticity) will receive higher weights than those relatively less likely to apply.

As the application elasticity to the firm is a weighted average of the individual level elasticities, many of the features of the individual level elasticities outlined in section 3.4 aggregate up to the firm level. Thus, the larger the spread of idiosyncratic preferences over jobs, the lower the elasticity of labour supply to the firm, and the greater the monopsony power. Additionally, the lower the commuting-wage elasticity, the greater the monopsony power and the more important the spatial distribution of activity in generating heterogeneities in monopsony power. Both of

[32] I.e. The greater the preference for a smaller commute.

these last two points can be demonstrated in a similar simulation exercise as seen in figures 10 and 11.

Figure 3: Spatial Heterogeneity in Labour Supply To the Firm I



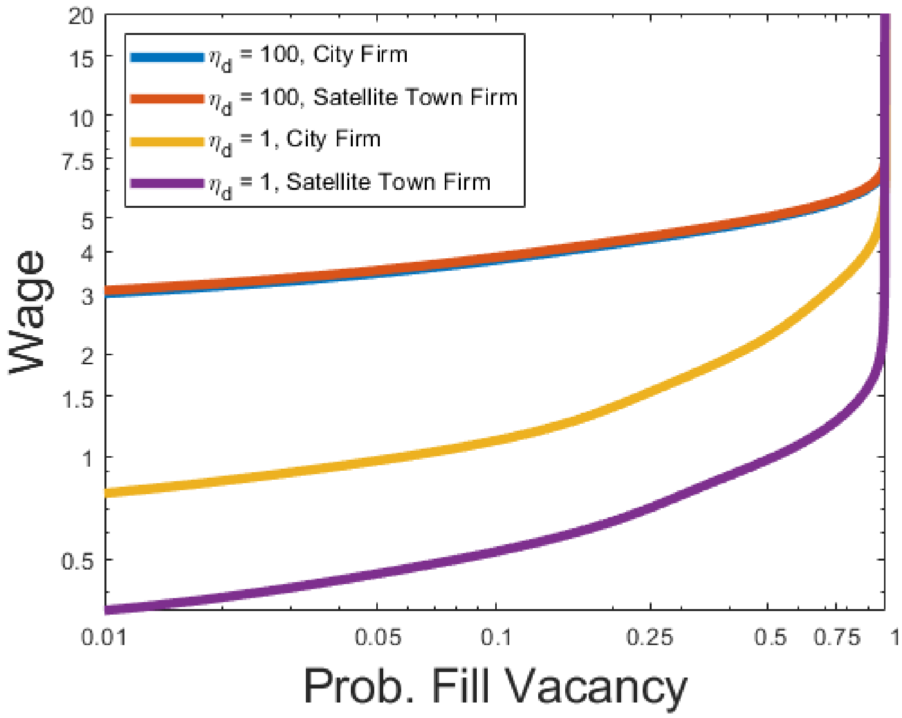
Note: *The figure presents a simulated spatial distribution of 600 workers, and 60 firms. 500 workers and 59 firms are located in the "city" while 100 workers and 1 firm is located in the "satellite town". Lines are drawn to demonstrate some of the existing commutes.*

Figure 3 presents a simulated spatial distribution of 600 workers and 60 firms, where 500 of the workers and 59 firms are located in a "city" while 100 workers and 1 firm is located in a "satellite town". Each firm has a labour force of 10 workers, lines demonstrate existing commutes for some workers and as can be seen some workers living in the satellite town commute into the city for work. As before it's assumed that all workers are currently in jobs with a wage of £10[33]. To explore the impact of commuting on monopsony power I examine the labour supply curve to two firms, one city firm and the satellite town firm, when posting a new vacancy.

As seen in figure 4 in the setting where a worker's commuting elasticity is extremely large ($\varepsilon_{cw} = 100$) there is virtually no difference between the labour supply to the two firms. This is

[33]The above analysis is very much a partial analysis. For such an equilibrium to exist this would require variation in firm level productivity, p_j, such that each maintaining firm's optimal wage offer was £10.

Figure 4: Spatial Heterogeneity in Labour Supply To the Firm II



Note: *The figure plots the labour supply to the firm curves as per figure 3 under different parameterisations of* ε_{cw}*, calculated according to equation (20) and (14), against the advertised wage on log scales. Parameterisation is such that* $c = 0$ *and* $\nu_{ij} = median(\nu_{ij})$*.* F *is assumed to follow a weibull distribution with shape parameter* $k = 5$ *and scale parameter* $\lambda = 1$*, and the exogenous suitability parameter* $q = 0.075$.

generated by the fact that worker's incumbent utilities would be almost identical across space, as would their potential utility from the advertised job. Put another way, when the commuting elasticity is large, the location differences do not generate any heterogeneity in utility. The advertising firms would still have considerable monopsony power under this parameterisation which is generated by the idiosyncratic term.

The parameterisation where $\varepsilon_{cw} = 1$ represents a scenario where workers have a strong distaste for commuting as per the estimates from sectio 2.3. This scenario has two effects, it increases monopsony power for both firms, but does so heterogeneously. At any given wage rate the satellite town firm experiences a more inelastic labour supply curve than the city firm. As many workers living in the satellite town have to commute into the city due to lack of local options, when a job opening appears locally, that job exercises considerable local monopsony power. While the satellite town firm experiences more, both firms experience greater monopsony power (a more inelastic slope) at any given wage rate in comparison to the $\varepsilon_{cw} = 100$ scenario and would therefore be able to markdown wages more than in the high commuting elasticity scenario. This is driven by the fact that distance plays an important role in determining utility from a job, so even the city firm would still enjoy some local monopsony power, as even small commutes within the city generate disutility.

3.6 Model Implications

A key result of this model is that all firms[34] would likely experience some local monopsony power, and therefore some of the monopsony power observed in empirical studies is likely driven by distaste for commuting coupled with the spatial locations of firms and workers. Models which coarsely discretise the spatial economy and treat commuting within areas as costless, and across boundaries as (infinitely) costly are likely to be misspecified, especially given the estimate of $\varepsilon_{cw} \approx 1$ from section 2. For example, many recent models rely on idiosyncratic preferences (typically assumed extreme value distributed) to generate monopsony power and lump distance to work within those idiosyncratic preferences (e.g. Lamadon et al. (2019); Berger et al. (2019); Card et al. (2018); Azar et al. (2019)). However, there is a strong reason to believe that these preferences are not idiosyncratic but rather systematic with clear spatial patterns as outlined above.

The model additionally gives rise to heterogeneity in monopsony power across space, and denser urban areas are likely to see less monopsony power than rural areas with fewer outside options. The model therefore offers a new micro-founded explanation for the urban wage premium. Specifically that markdowns in cities are likely to be smaller in size due to the more competitive nature of the markets. That said, each firm located at a specific point would have its own specific level of monopsony power, and this would likely vary even within a city. Typically labour market analysis has been done at the TTWA level in the UK and the mean area for TTWAs is 1,064 km^2, and the above model suggests there is likely to be considerable heterogeneity within that area.

Figure 12 in the appendix shows the map for the TTWA for London[35], and its surrounding TTWAs, and exemplifies the issue. A study of monopsony power in the UK labour market with discretised areas into TTWAs would suggest the travel time between Rickmansworth (in the North West of London) and Gravesend (in the South East) is costless whereas google-maps estimates a travel time of 90 minutes utilising either car or public transport. Additionally, using a model where distance to work is lumped into the idiosyncratic component, though allowed to vary by area, would result in a firm located on the edge of the TTWA (e.g. Potters Bar) as having the same degree of competitiveness as a firm operating in central London (e.g. Holborn), which seems unlikely.

Another result of this model is that the labour supply elasticity to the firm is no longer a structural parameter, as elasticities are an endogenous result of the spatial distribution of activity.[36]

[34] Except in the extreme case where a firm is relatively located far from all workers.

[35] This TTWA represents almost a sixth of the Great Britain working population.

[36] In models utilising a logit structure where all monopsony power is generated by the idiosyncratic component as those mentioned above, the elasticity of labour supply to the firm is $1/\beta_{\text{area}}$ where β_{area} is the scale parameter related to the extreme value distribution and the subscript reflects that it can vary by area. Thus, in these models the labour supply elasticity to the firm, and therefore the extent of monopsony power, is entirely determined by the area specific scale parameter.

While the labour supply elasticities are still a function of the idiosyncratic component (which could now be assumed to be constant across areas), they are also determined by the spatial distribution of activity, and the wage-commuting elasticity. It's worth noting that if workers did not care about commuting, and $\varepsilon_{cw} = \infty$ then the model would collapse to the typical logit model utilised in other studies, and this is shown in section F.

A final observation from this model is that it suggests that general reductions in travel times should in turn increase market competitiveness, and generate upward pressure on wages. Therefore, one way in which infrastructure investments or changes to working patterns (i.e. increased working from home) could affect wages is through increased competition between firms.

The sources of monopsony explored in the above exposition focus entirely on imperfect substitution between jobs, unlike some of the original literature which focused on dynamic search frictions (see Manning (2003)). The model however can flexibly introduce search frictions and is done so in section E of the appendix. Monopsony power is shown to be weakly increasing in search costs, but more interestingly it is shown that the individual level elasticity can actually be negative in some instances.

3.7 Model Validation

3.7.1 Model Predicted Elasticities vs Empirical Elasticties

To validate the model I first use the data and empirical framework from section 2 and the estimated commuting-wage elasticity from section 2.3.2 to check whether job adverts predicted to be more elastic by the model are. To do this I calculate a version of equation (21) assuming a Weibull distribution, for each job advert, and split the sample of adverts in half at the median based on the model calculated $\varepsilon_{AWj'}$, labelling the top half "high". I then interact this variable in the application-wage elasticity regression from section 2.2.1.

Specifically I calculate

$$\varepsilon_{Aw_{ajemy}} = \frac{\sum_{i,j} k \left(\frac{w_j d_{ij}^{\frac{-1}{\varepsilon_{cw}}} \bar{\nu}}{\lambda w_{j'} d_{ij'}^{\frac{-1}{\varepsilon_{cw}}}} \right)^k exp\left(- \left(\frac{w_j d_{ij}^{\frac{-1}{\varepsilon_{cw}}} \bar{\nu}}{\lambda w_{j'} d_{ij'}^{\frac{-1}{\varepsilon_{cw}}}} \right)^k \right)}{\sum_{i,j} exp\left(- \left(\frac{w_j d_{ij}^{\frac{-1}{\varepsilon_{cw}}} \bar{\nu}}{\lambda w_{j'} d_{ij'}^{\frac{-1}{\varepsilon_{cw}}}} \right)^k \right)} \tag{22}$$

using a value of $\varepsilon_{cw} = 1$ given the estimates from section 2.3. For the incumbent value of ν_{ij}, I assume it is equal to the median value of the distribution $ln(2)^{\frac{1}{k}}$[37], and to ensure robustness I calculate 22 using three different values of $k \in \{3, 5, 7\}$.

I use data from the Annual Survey of Hours and Earnings for 2016-2019 for Great Britain (Office

[37] I use the median rather than the mean ($\lambda\Gamma(1 + \frac{1}{k})$) to avoid the need to utilise the gamma function.

for National Statistics, 2020)[38] for information on w_j, d_{ij} and $d_{ij'}$. d_{ij} and $d_{ij'}$ are calculated at the 6-digit postcode level.[39] I additionally utilise the Open Street map dataset for Great Britain and ArcGIS's networking tool for commute calculation. This allows for a very accurate time of commute between home and work postcodes.[40] For each job advert j', workers with the same occupation code[41] that job j' related to was included in the calculation of (22). Summary statistics for (22) are reported in table 12 in the appendix.

Allocating adverts to the high elasticity group such that

$$HighElast = \begin{cases} 1 \text{ if } \varepsilon_{AW_{ajemy}} \geq med(\varepsilon_{AW_{ajemy}}) \\ 0 \text{ otherwise} \end{cases} \tag{23}$$

I regress

$$\begin{aligned} log(Apps_{ajemy}) =& \beta_5 log(Wage_{ajemy}) + \beta_6 log(Wage_{ajemy}) X HighElast_{ajemy} \\ &+ \beta_7 HighElast_{ajemy} + \gamma_{je} + \lambda_{ey} + \theta_{my} + \epsilon_{ajemy} \end{aligned} \tag{24}$$

utilising the two instruments used previously, LW_{jemy} and AL_{ajemy}. Table 6 presents the parameter estimates for $\hat{\beta}_5$, $\hat{\beta}_6$ and $\hat{\beta}_7$ for the three values of k. The model predicted high group has almost double the elasticity of the low group, suggesting the model does a good job of predicting more and less competitive job adverts. Additionally, the estimate of $\hat{\beta}_7$ is negative as the model would predict.[42] Estimates across all three specifications are consistent, parameter estimates are not statistically significantly different across the various parametrisations of k. This is unsurprising as varying k is only able to alter the model predicted relative difference between the high and low groups' elasticities (as seen in table 12). It would not be able to fundamentally manipulate the composition of the high and low elasticity groups.

3.7.2 Descriptive Correlations for Retail Shop Workers

As a secondary test of validation I look at the relationship between the degree of competitiveness as predicted by the model (i.e. the application-wage elasticity) for different geographic locations across England and Wales, and measures of wages, density and population.

Choice of geographic locations for this exercise is non-trivial. Typically for the UK one would use centroids of TTWAs or Local Authorities (LA), however, the model is particularly granular,

[38] The data does not contain information on Northern Ireland and therefore the sample is marginally smaller than that used in section 2.2.2.

[39] A 6-digit postcode typically relates to either a building (in the case of flats) or a few houses on a street, therefore it is almost the exact location of the individual's residence to their place of (potential) work.

[40] The Open Street map data contains the full road network for Great Britain, speed limits for each road, as well as locations of speed obstructions (roundabouts, traffic lights, crossings etc). The networking tool in ArcGIS allows time obstructions to reflect a delay (measured in seconds) and also allows for delays when crossing or joining new roads.

[41] Typically these are to the 4 digit SOC level, however in some cases multiple 4 digit SOC codes were included.

[42] Those jobs with higher elasticties should be on the flatter part of their labour supply curve and therefore have a lower number of applicants in comparison to the low elasticity group.

Table 6: Model Validation Interaction

	(1) log(Apps)	(2) log(Apps)	(3) log(Apps)
log(Wage)	1.854**	1.864**	2.070***
	(0.788)	(0.844)	(0.701)
log(Wage)X HighElast	1.558**	1.635**	1.549**
	(0.686)	(0.808)	(0.628)
HighElast	-3.752**	-3.925**	-3.550**
	(1.655)	(1.951)	(1.528)
Job-Centre FE	Yes	Yes	Yes
Centre-Year FE	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes
N	2239	2239	2239
First Stage F-Stat.	60.40	35.82	56.48
Hansen J Stat.	4.456	3.009	1.146
Instrumented With *AL*	Yes	Yes	Yes
Instrumented With *LW*	Yes	Yes	Yes
k	3	5	7

Note: *The table presents estimates of $\hat{\beta}_5$,$\hat{\beta}_6$ and $\hat{\beta}_7$ from regression equation (24) utilising both the LW and AL instrument. High and low groups are defined according to the model predicted elasticities $\varepsilon_{AW_{ajemy}}$ from equation 22, with varying values of k. Standard errors are reported in parentheses and are clustered at the establishment.* * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

and therefore using such large areas carries a high risk of measurement error.[43] On the other hand the choice of specific points to calculate competitiveness for any country is effectively infinite, and some structure would be beneficial. I therefore utilise the ONS's geography of Built-Up Areas and their subdivisions (BUAs). BUAs are defined as land which is "irreversibly urban in character" (Harding et al., 2013) and comprise of villages, towns or cities. Approximately 95% of the population of England and Wales live in BUAs and the land area of BUAs only makes up 9.6% of the total land. The smallest BUA has a population of just over 100, while the largest is Greater London with almost 10 million, though for the following analysis however London is split into many subdivisions to improve granularity. Figure 13 in the appendix presents a map of the BUAs and where appropriate, their subdivisions for England and Wales. In total there are 7,625 areas.

For each BUA I take the centroid point and calculate equation 22 using the same parameter values as before ($\varepsilon_{cw} = 1$ and $k \in \{3, 5, 7\}$), for the market of retail shop workers present in ASHE for the year 2019. I choose this occupation as it represents the largest occupation group of minimum wage jobs for the UK, in total there are 10,138 workers located across England and Wales in the dataset. As before distances are measured in minutes utilising the Open Street

[43] By using the centroid of a TTWA or LA it's not clear that specific point would be representative of the rest of the geographic area. For example the centroid of a LA could be agricultural land located between two major towns.

map data and ArcGIS, incumbent wages are taken from ASHE and the elasticities are calculated at the point of the LWF's Living Wages for 2019.[44]

Figure 5 presents binned scatter plots of the worker density against the normalised model predicted application elasticity, and the normalised model predicted elasticity against wages, respectively, for the value of $k = 3$. Worker density is measured by the number of retail workers within a 25 minute drive of the centroid in the sample, while wages is the mean of those workers' wages. Both plots show a strong positive correlation which is reassuring. Denser areas are expected to be more competitive due to the close proximity of outside options, and similarly areas that are more competitive should have higher wages. The slope relating the application elasticity to wages is likely to be slightly muted due to the existence of a national minimum wage policy, however it still shows a strong robust relationship. The strong correlation is invariant to the value of k as shown in figures 14 and 15 in the appendix. The positive relationship between the application elasticity also holds when using measures at the BUA level, such as residential density and population as shown in figures 16 and 17 in the appendix. These results are suggestive that denser more populated areas are likely to experience more competitive labour markets, and thus higher wages, and may go some way in explaining the urban wage premium. They additionally give credence to the mechanism suggested by the model, that distaste for commuting is a key factor in generating monopsony power.

4 Structural Estimation and The Relative Importance of Commutes

In the following section I structurally estimate the model to estimate the value of the Weibull parameter k by matching the model predicted elasticities to the empirical elasticities. By doing so I can comment on the relative importance of commutes in determining monopsony power. Thanks to the relationship between the Weibull distribution and the Extreme Value Type-I (EV) distribution, the Weibull shape parameter $k = \frac{1}{\beta}$ where the variance of the EV distribution is $\frac{\pi^2}{6}\beta^2$. Thus as shown in section F, k has a neat interpretation of being the application elasticity in the absence of commutes. As a secondary step I also use the model and the estimated parameters to examine the effect on competitiveness and markdowns of shrinking all commutes to zero, while holding incumbent wages constant.

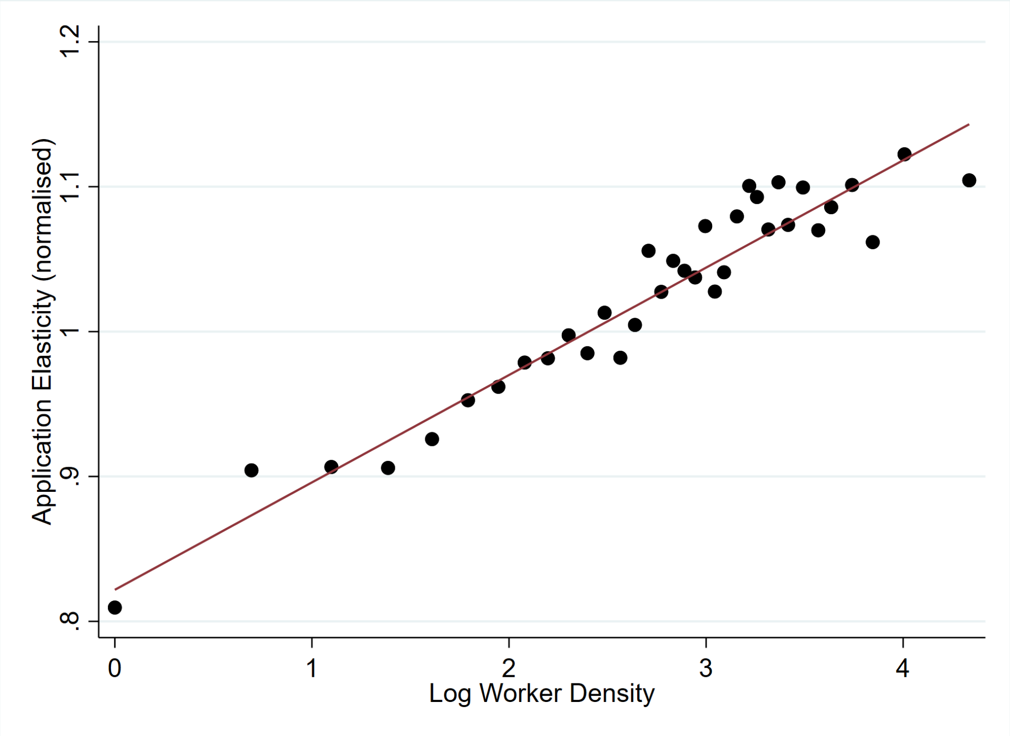
4.1 Method

To elicit the parameter k I use a minimum distance estimator, that minimises the difference between the model predicted elasticities and the empirically estimated elasticities. I additionally introduce a log-scaling parameter α. In practice this parameter ensures that the elasticities match in absolute-value, rather than in just relative value. The parameter, therefore allows for a clearer interpretation of the following results.

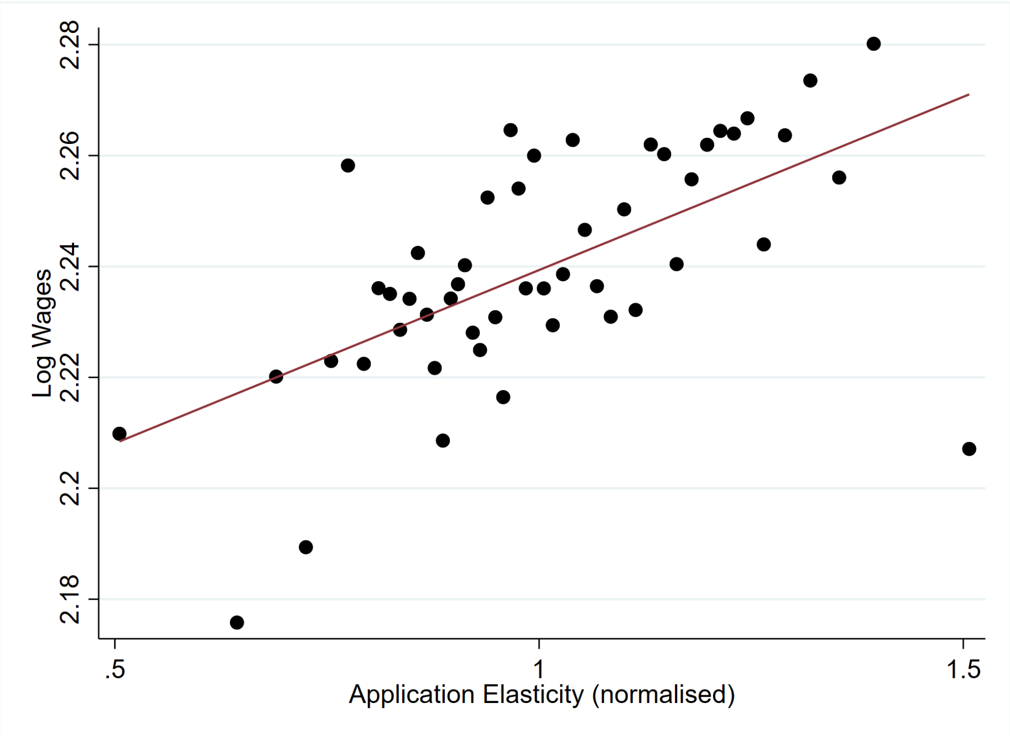
The labour supply function (in logs) then becomes

[44] £9.30 for non-London areas and £10.75 for London.

Figure 5: Model Predicted Elasticities vs Density and Wages



Slope = 0.103, Std. error = 0.003



Slope = 0.062, Std. error = 0.008

Note: *The figure presents binned scatter plots of the application elasticity (normalised by dividing by its mean) as calculated by equation 22 using values of $k = 3$ for centroids of 7,625 BUAs and BUASDs for the market of retail workers against worker density and average log wages. Worker density is measured by those workers within a 25 minute drive of the centroid and wages are calculated by the mean hourly wage of those individuals.*

$$log(A_{j'}) = \alpha \; log\Big(\sum_{ij}[1 - F(x_{ijj'})]\Big) \tag{25}$$

and for each job advert the empirical counterpart becomes

$$\tilde{\varepsilon}_{Aw_{ajemy}} = \alpha * \varepsilon_{Aw_{ajemy}} \tag{26}$$

where $\varepsilon_{Aw_{ajemy}}$ is defined in equation 22. For a given k, I calculate 26 using the same data and method as outlined in section 3.7.1.

The algorithm used to estimate k and α is as follows:

1. Guess parameters k and α.
2. Calculate equation (26) for each job advert using the guessed parameters.
3. Split the sample in half at the median based on the model calculated $\tilde{\varepsilon}_{Aw_{ajemy}}$, labelling the top half "high".
4. Estimate regression equation

$$\begin{aligned} log(Apps_{ajemy}) = & \beta_1 log(Wage_{ajemy}) + \beta_2 log(Wage_{ajemy}) X HighElast._{ajemy} \\ & + \beta_3 HighElast._{ajemy} + \gamma_{je} + \lambda_{ey} + \theta_{my} + \epsilon_{ajemy} \end{aligned}$$

   with the high group as calculated from step 3 and with $log(Wage_{ajemy})$ instrumented with the two instruments discussed in section 2.1.2.
5. Calculate the euclidean distance between the empirical elasticities from step 4 and the means of the model elasticity groups from step 3:

$$Euclidean\ distance = \sqrt{(\hat{\beta}_1 - \bar{\varepsilon}_{Aw,low})^2 + (\hat{\beta}_1 + \hat{\beta}_2 - \bar{\varepsilon}_{Aw,high})^2}. \tag{27}$$

6. Repeat steps 1-5 until equation (27) in step 5 is minimised.

Identification of k comes through the variation the parameter induces in the relative difference between the model predicted elasticities of the high and low group. This is exemplified in table 12 in the appendix which documents the relative difference between the high and low group elasticities. The higher the value of k, the larger the relative difference between the high and low group elasticities. This is consistent with intuition. The lower the value of k the larger the variance of the idiosyncratic term ν_{ij}, and the more important the idiosyncratic term is in determining monopsony power in comparison to location and commutes. Thus, the more similar the levels of monopsony power between the high and low group. Identification of α comes through matching the absolute levels of the elasticities.

4.2 Results

Table 7 presents estimates of the structural parameters using the algorithm outlined in section 4.1. k is estimated to be 5.49, which is reasonable as it would suggest in the absence of commutes the application elasticity would be equal to approximately 5.5 (see section F). Table 8 presents the model implied elasticities along with the empirically estimated counterparts, which are similar in value to the specification with $k = 5$ in the regression table 6.

Table 7: Structural Parameter Estimates

k	α
5.49	2.44

Note: *The table presents the structural estimates for the remaining model parameters, based off the algorithm outlined in section 4.1. The sample includes 2,239 non-permanent job adverts for The Company, and 1,062,022 worker-advert pairs from ASHE.*

Table 8: Model vs Empirical Elasticities

	Elasticity		
	Mean	High	Low
Model	2.93	3.95	1.91
Empirical	3.03	3.71	2.34

Note: *The table presents the model predicted and estimated mean elasticities for the high group, low group and all pooled with the parameter estimates of $k = 5.49$ and $\alpha = 2.44$. The sample includes 2,239 non-permanent job adverts for The Company, and 1,062,022 worker-advert pairs from ASHE.*

The model does a good job of matching the respective empirical elasticities, and gives support to the mechanisms suggested by the model in section 3.5.

4.3 Shrinking Commutes to Zero

The model lends itself to an exercise where all commutes are assumed to shrink to zero[45]. That is, the model can look at the effect on incumbent and potential utilities, and therefore job specific elasticities when commutes are assumed away (or $\varepsilon_{cw} = \infty$), but incumbent wages are held constant. The following analysis does not therefore take into account general equilibrium effects; i.e. the fact that firms would also change wages for their incumbent workers, as their optimal wage would also change. One interpretation of the following exercise is, what would happen to the labour supply elasticity to the firm and therefore markdowns, if all workers were immediately assumed to work from home, and incumbent wages were unchanged in the short run. Another interpretation is that the resulting elasticities will go some way in explaining how much of monopsony power is generated by commutes and the spatial distribution of activity.

[45] Or all workers and firms were to be located at the exact same point.

Methodologically, the exercise is straightforward and is based off equation 26. Using the parameter estimates of $k = 5.49$, $\alpha = 2.44$, and $\varepsilon_{cw} = 1$, I compare the model predicted elasticities with the predicted elasticities where $\varepsilon_{cw} = \infty$. This is equivalent to assuming both d_{ij} and $d_{ij'}$ equal 1.[46]

Table 9 presents the mean advert-level application-elasticities under the true parametrisation of $\varepsilon_{cw} = 1$ and the alternative of $\varepsilon_{cw} = \infty$, where the alternative is equivalent to assuming away all commutes. The average application elasticity is approximately two thirds higher in the counterfactual scenario and suggests that commutes are a key source of monopsony power. Interestingly, under the no-commute counterfactual, the mean application elasticity is close in size to the estimated parameter k, which as discussed previously would be the application elasticity in the absence of commutes in an EV logit model. This gives additional reassurance to the credibility of the exercise.

Table 9: True vs Work form Home Elasticities

	Application Elasticity	
	$\varepsilon_{cw} = 1$	$\varepsilon_{cw} = \infty$
Mean	2.93	4.95

Note: *The table presents the mean of the advert-level elasticities calculated via equation 22 under the true parametrisation of $\varepsilon_{cw} = 1$ and counterfactual of $\varepsilon_{cw} = \infty$ utilising the sample of 2,239 non-permanent job adverts for The Company, 1,062,022 worker-advert pairs from ASHE and the structurally estimated parameter values of $k = 5.49$ and $\alpha = 2.44$.*

Assuming a recruitment-application (i.e. fill-application) elasticity of 0.37 as estimated in section 2.2.2, that $\varepsilon_{rw} = \varepsilon_{ra} * \varepsilon_{aw}$[47] and assuming an equality between the separation and recruitment-elasticity (which is a reasonable assumption given results in section 2.2.2) markdowns can be calculated according to

$$\mu = 1 - \frac{2 * 0.37 * \varepsilon_{AW}}{1 + 2 * 0.37 * \varepsilon_{AW}}. \tag{28}$$

The back of the envelope calculation suggests that markdowns would be approximately 10 percentage points lower in a scenario with no commutes.

These results indicate the spatial distribution of activity and distaste for commuting are a key contributor to monopsony power. The results imply that commutes are responsible for about one third of the markdown of wages. It is reasonable to assume however, that this is a lower bound, as incumbent wages would be expected to rise which would in turn lower markdowns further for the advertising firms, if incumbent wage responses were considered.

[46] As utility is multiplicative assuming they become 0 would not be practical.

[47] Note that this is the lower bound as discussed in section 2.2.2.

5 Conclusion

This paper provides new evidence and theory on the extent, and sources, of monopsony power in the labour market. Utilising two instruments and a rich bespoke dataset that contains HR, vacancy and applicant information for a firm with hundreds of establishments across the UK I evaluate the firm's labour supply elasticity, a key measure of monopsony power, estimating both the recruitment and separation elasticity. Estimates suggest a markdown between 16% - 25%. I additionally estimate the commuting-wage elasticity, and results suggest worker's have a strong distaste for commuting. The estimate for preferences over commuting suggests commutes could be a key factor in generating imperfect substitutability between jobs and monopsony power.

In order to formalise the mechanism of commutes generating monopsony power I develop a search model where a worker's utility from a job is dependent on the wage, an idiosyncratic component and the commute from their home to work. The model avoids discretising labour markets and instead endogenously defines continuous labour markets which are decreasing in distance, and the size depends on worker's preferences over commuting, the commuting-wage elasticity. The model additionally generates endogenous labour supply to the firm elasticities which vary across space. The model suggests that the spatial distribution of activity coupled with a distaste for commuting can play a key role in generating imperfect substitution between jobs. As a result firms in areas with fewer local job opportunities exercise greater monopsony power than firms in urban areas.

I validate the model by showing how the model predictions of heterogeneity in monopsony across space are consistent with heterogeneity found in the causal estimates. I additionally show model consistent measures of competitiveness for Built Up Areas in England and Wales are shown to be strongly correlated with worker and residentail density, residential population, and wages. I then structurally estimate the model by matching the model predicted elasticities with the empirically estimated elasticities. The results of the estimation and an exercise which shrinks all commutes to zero suggests that commutes are responsible for 1/3 of the wage markdown.

The results from this paper go some way in furthering our understanding of the sources of monopsony power, while also contributing to our understanding on the causes of the urban wage premium. By directly modelling in commutes and taking a more granular approach to the spatial economy discussions concerning market concentration can adopt a much more precise measure, as the results from this study suggest that TTWAs and CZs are likely to be overestimating geographic labour market size. Furthermore, the results also speak to the role that transportation infrastructure spending can have on reducing monopsony power. Finally, there is hope that if there are structural shifts in the way we work in response to the COVID-19 pandemic, in particular, a greater shift to working from home, there is scope for increased competition in the labour market as the implied length of commutes would drop.

This paper has placed a strong focus on the definition of a spatial labour market, and its role in generating imperfect substitution between jobs. A next step would be to combine this flexibility in spatial labour markets with the flexibility developed in Schubert et al. (2021) concerning occupational labour market definition in order to construct an even finer measure of market definition for the individual.

References

Abel, Will, Silvana Tenreyro, and Gregory Thwaites (2018), "Monopsony in the uk."

Arnold, David (2020), "Mergers and acquisitions, local labor market concentration, and worker outcomes." *Job Market 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." *Available at SSRN 3456277.*

Azar, José, Ioana Marinescu, and Marshall Steinbaum (2020a), "Labor market concentration." *Journal of Human Resources*, 1218–9914R1.

Azar, José, Ioana Marinescu, Marshall Steinbaum, and Bledi Taska (2020b), "Concentration in us labor markets: Evidence from online vacancy data." *Labour Economics*, 66, 101886.

Azkarate-Askasua, Miren and Miguel Zerecero (2020), "The aggregate effects of labor market concentration." *Unpublished Working Paper.*

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

Bassier, Ihsaan, Arindrajit Dube, and Suresh Naidu (2020), "Monopsony in movers: The elasticity of labor supply to firm wage policies." Technical report, National Bureau of Economic Research.

Batley, Richard, John Bates, Michiel Bliemer, Maria Börjesson, Jeremy Bourdon, Manuel Ojeda Cabral, Phani Kumar Chintakayala, Charisma Choudhury, Andrew Daly, Thijs Dekker, et al. (2019), "New appraisal values of travel time saving and reliability in great britain." *Transportation*, 46, 583–621.

Belot, Michele, Philipp Kircher, and Paul Muller (2018), "How wage announcements affect job search-a field experiment."

Benmelech, Efraim, Nittai K Bergman, and Hyunseob Kim (2020), "Strong employers and weak employees: How does employer concentration affect wages?" *Journal of Human Resources*, 0119–10007R1.

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

Berry, Steven, Martin Gaynor, and Fiona Scott Morton (2019), "Do increasing markups matter? lessons from empirical industrial organization." *Journal of Economic Perspectives*, 33, 44–68.

Blundell, Jack (2020), "Wage responses to gender pay gap reporting requirements." *Available at SSRN 3584259.*

Borusyak, Kirill and Xavier Jaravel (2017), "Revisiting event study designs." *Available at SSRN 2826228*.

Burdett, Kenneth and Dale T Mortensen (1998), "Wage differentials, employer size, and unemployment." *International Economic Review*, 257–273.

Caldwell, Sydnee and Oren Danieli (2020), "Outside options in the labor market."

Caldwell, Sydnee and Emily Oehlsen (2018), "Monopsony and the gender wage gap: Experimental evidence from the gig economy." *Massachusetts Institute of Technology Working Paper*.

Callaway, Brantly and Pedro HC Sant'Anna (2020), "Difference-in-differences with multiple time periods." *Journal of Econometrics*.

Card, David, Ana Rute Cardoso, Joerg Heining, and Patrick Kline (2018), "Firms and labor market inequality: Evidence and some theory." *Journal of Labor Economics*, 36, S13–S70.

Ciccone, Antonio and Robert E Hall (1996), "Productivity and the density of economic activity." *The American Economic Review*, 86, 54–70.

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.

Datta, Nikhil (2019), "Willing to pay for security: a discrete choice experiment to analyse labour supply preferences." *CEP Discussion Paper*.

Datta, Nikhil, Giulia Giupponi, and Stephen Machin (2019), "Zero-hours contracts and labour market policy." *Economic Policy*, 34, 369–427.

Datta, Nikhil and Stepehn Machin (2021), "Consequences of living wages > minimum wages." *LSE Mimeo*.

de la Roca, Jorge and Diego Puga (2017), "Learning by working in big cities." *The Review of Economic Studies*, 84, 106–142.

Derenoncourt, Ellora, Clemens Noelke, and David Weil (2021), "Spillover effects from voluntary employer minimum wages." *Available at SSRN 3793677*.

Dube, Arindrajit, Laura Giuliano, and Jonathan Leonard (2019), "Fairness and frictions: The impact of unequal raises on quit behavior." *American Economic Review*, 109, 620–63.

Dube, Arindrajit, Jeff Jacobs, Suresh Naidu, and Siddharth Suri (2020a), "Monopsony in online labor markets." *American Economic Review: Insights*, 2, 33–46.

Dube, Arindrajit, T William Lester, and Michael Reich (2016), "Minimum wage shocks, employment flows, and labor market frictions." *Journal of Labor Economics*, 34, 663–704.

Dube, Arindrajit, Alan Manning, and Suresh Naidu (2020b), "Monopsony, misoptimization, and round number bunching in the wage distribution." *NBER Working Paper w24991*.

Eriksson, Tor and Nicolai Kristensen (2014), "Wages or fringes? some evidence on trade-offs and sorting." *Journal of Labor Economics*, 32, 899–928.

Falch, Torberg (2017), "Wages and recruitment: Evidence from external wage changes." *ILR Review*, 70, 483–518.

Glaeser, Edward L (1998), "Are cities dying?" *Journal of economic perspectives*, 12, 139–160.

Glaeser, Edward L and David C Mare (2001), "Cities and skills." *Journal of labor economics*, 19, 316–342.

Goodman-Bacon, Andrew (2021), "Difference-in-differences with variation in treatment timing." *Journal of Econometrics*.

Harding, Jenny, Bill South, Mark Freeman, and Babington A Zhou Sh (2013), "Identifying built-up areas for 2011 census outputs." In *Proceedings of the 26th International Cartographic Conference, Dresden, Germany*.

Hershbein, Brad, Claudia Macaluso, and Chen Yeh (2018), "Concentration in us local labor markets: evidence from vacancy and employment data." Technical report, Working paper.

Heuermann, Daniel, Benedikt Halfdanarson, and Jens Suedekum (2010), "Human capital externalities and the urban wage premium: Two literatures and their interrelations." *Urban Studies*, 47, 749–767.

Hirsch, Boris, Elke J Jahn, Alan Manning, and Michael Oberfichtner (2020), "The urban wage premium in imperfect labor markets." *Journal of Human Resources*, 0119–9960R1.

Jarosch, Gregor, Jan Sebastian Nimczik, and Isaac Sorkin (2019), "Granular search, market structure, and wages." Technical report, National Bureau of Economic Research.

Krueger, Alan B and Eric A Posner (2018), "A proposal for protecting low-income workers from monopsony and collusion." *The Hamilton project policy proposal*, 5.

Lamadon, Thibaut, Magne Mogstad, and Bradley Setzler (2019), "Imperfect competition, compensating differentials and rent sharing in the us labor market." Technical report, National Bureau of Economic Research.

Le Barbanchon, Thomas, Roland Rathelot, and Alexandra Roulet (2021), "Gender differences in job search: Trading off commute against wage." *The Quarterly Journal of Economics*, 136, 381–426.

Lipsius, Ben (2018), "Labor market concentration does not explain the falling labor share." *Available at SSRN 3279007*.

Manning, Alan (2003), *Monopsony in motion: Imperfect competition in labor markets*. Princeton University Press.

Manning, Alan and Barbara Petrongolo (2017), "How local are labor markets? evidence from a spatial job search model." *American Economic Review*, 107, 2877–2907.

Marinescu, Ioana and Herbert Hovenkamp (2019), "Anticompetitive mergers in labor markets." *Ind. LJ*, 94, 1031.

Marinescu, Ioana and Eric A Posner (2019), "Why has antitrust law failed workers?" *Cornell L. Rev.*, 105, 1343.

Marinescu, Ioana and Ronald Wolthoff (2020), "Opening the black box of the matching function: The power of words." *Journal of Labor Economics*, 38, 535–568.

Mas, Alexandre and Amanda Pallais (2017), "Valuing alternative work arrangements." *American Economic Review*, 107, 3722–59.

Moretti, Enrico (2011), "Local labor markets, volume 4 of handbook of labor economics, chapter 14."

Naidu, Suresh and Eric A Posner (2021), "Labor monopsony and the limits of the law." *Journal of Human Resources*, 0219–10030R1.

Naidu, Suresh, Eric A Posner, and Glen Weyl (2018), "Antitrust remedies for labor market power." *Harv. L. Rev.*, 132, 536.

Nimczik, Jan Sebastian (2020), "Job mobility networks and endogenous labor markets."

Office for National Statistics (2020), "Annual survey of hours and earnings, 1997-2020: Secure access. [data collection]." *17th Edition. UK Data Service. SN: 6689. http://doi.org/10.5255/UKDA-SN-6689-16*.

Pigou, Arthur Cecil (1924), *The Economics of Welfare*. Macmillan.

Pörtner, Claus C and Nail Hassairi (2018), "What labor supply elasticities do employers face? evidence from field experiments." *Evidence From Field Experiments (October 2018)*.

Puga, Diego (2010), "The magnitude and causes of agglomeration economies." *Journal of regional science*, 50, 203–219.

Qiu, Yue and Aaron Sojourner (2019), "Labor-market concentration and labor compensation." *Available at SSRN 3312197*.

Ransom, Michael R and David P Sims (2010), "Estimating the firm's labor supply curve in a "new monopsony" framework: Schoolteachers in missouri." *Journal of Labor Economics*, 28, 331–355.

Rinz, Kevin et al. (2018), "Labor market concentration, earnings inequality, and earnings mobility." *Center for Administrative Records Research and Applications Working Paper*, 10.

Robinson, Joan (1933), *The economics of imperfect competition*. Macmillan and Co., ltd.

Rose, Nancy L (2019), "Concerns about concentration."

Schubert, Gregor, Anna Stansbury, and Bledi Taska (2021), "Employer concentration and outside options." *Available at SSRN 3599454*.

Sokolova, Anna and Todd Sorensen (2021), "Monopsony in labor markets: A meta-analysis." *ILR Review*, 74, 27–55.

Sun, Liyang and Sarah Abraham (2020), "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of Econometrics*.

Timothy, Darren and William C Wheaton (2001), "Intra-urban wage variation, employment location, and commuting times." *Journal of urban Economics*, 50, 338–366.

Van Ommeren, Jos and Mogens Fosgerau (2009), "Workers' marginal costs of commuting." *Journal of Urban Economics*, 65, 38–47.

Wiswall, Matthew and Basit Zafar (2018), "Preference for the workplace, investment in human capital, and gender." *The Quarterly Journal of Economics*, 133, 457–507.

A Additional Tables

Table 10: Summary Statistics, Adverts

Variable	Mean	S.D.	Median
Hourly Rate (£)	11.07	11.07	10.20
Entry Level	0.48		
No Applicants	18.5	28.9	10
London	0.55		
N	5487		

Note: *The table presents summary statistics for the job adverts for The Company for the period of 2016-2019.*

Table 11: London Borough of Hackney, Employment

London Borough of Hackney (estim)		
Sector	Employment	%
All	133,000	100
Private	115,100	86
Public		
NHS	5,549	4.3
Council	4,390	3.3
Civil Service	1,790	1.4
Education (LEA)	2,148	1.6
Education (Acad.)	2,864	2.1
Other	1159	1.3

Note: *The table presents employment shares by sector for the London Borough of Hackney for the year 2019.*

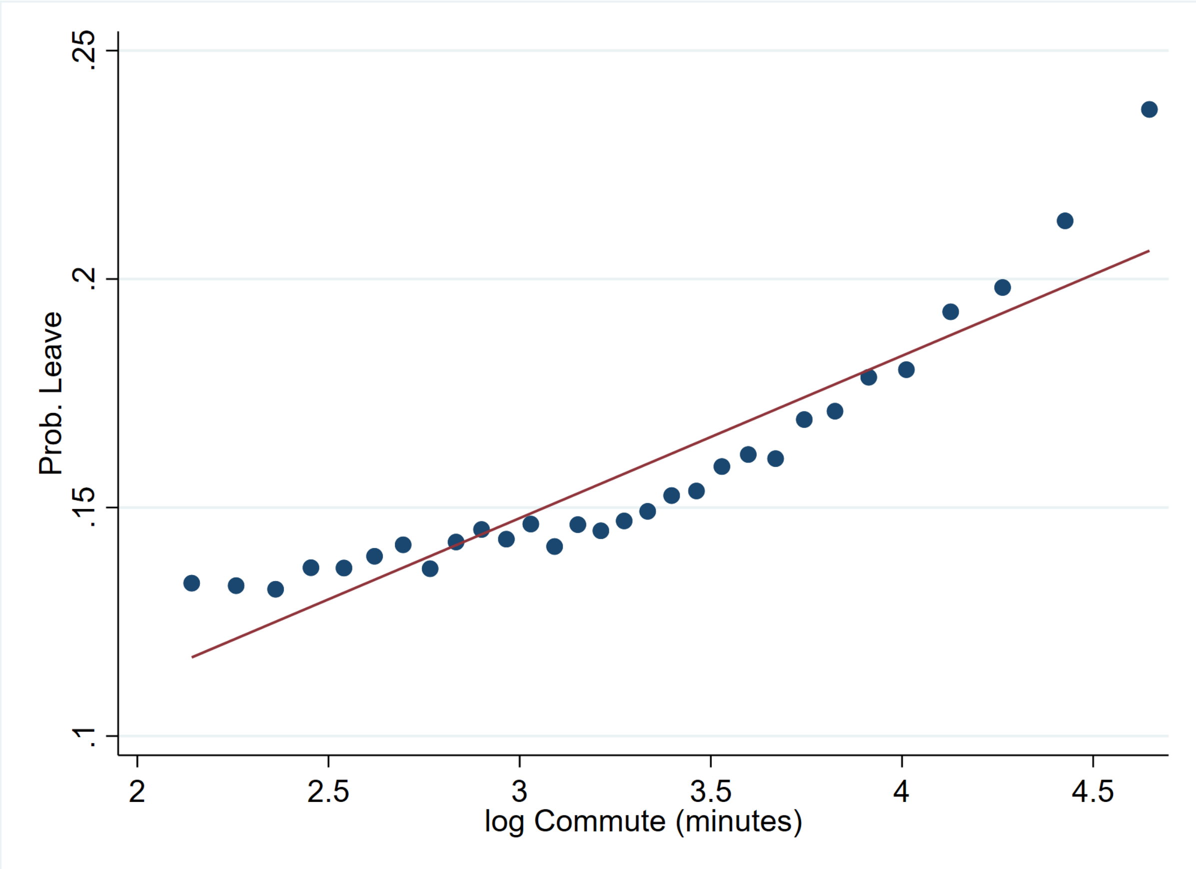
Table 12: Model Elasticity Estimates

k	Pooled Mean	High Elast. Mean	Low Elast. Mean	High/Low Elast. Mean
1	0.97	1.21	0.73	1.66
3	1.21	1.58	0.84	1.88
5	1.21	1.64	0.79	2.08
7	1.23	1.73	0.75	2.31
9	1.24	1.76	0.72	2.44

Note: *The table presents descriptive statistics for the pooled, high group and low group for estimates of* $\varepsilon_{AWj'}$ *based on equation 22 under different values of* k*, for a sample of 2,239 non-permanent job adverts.*

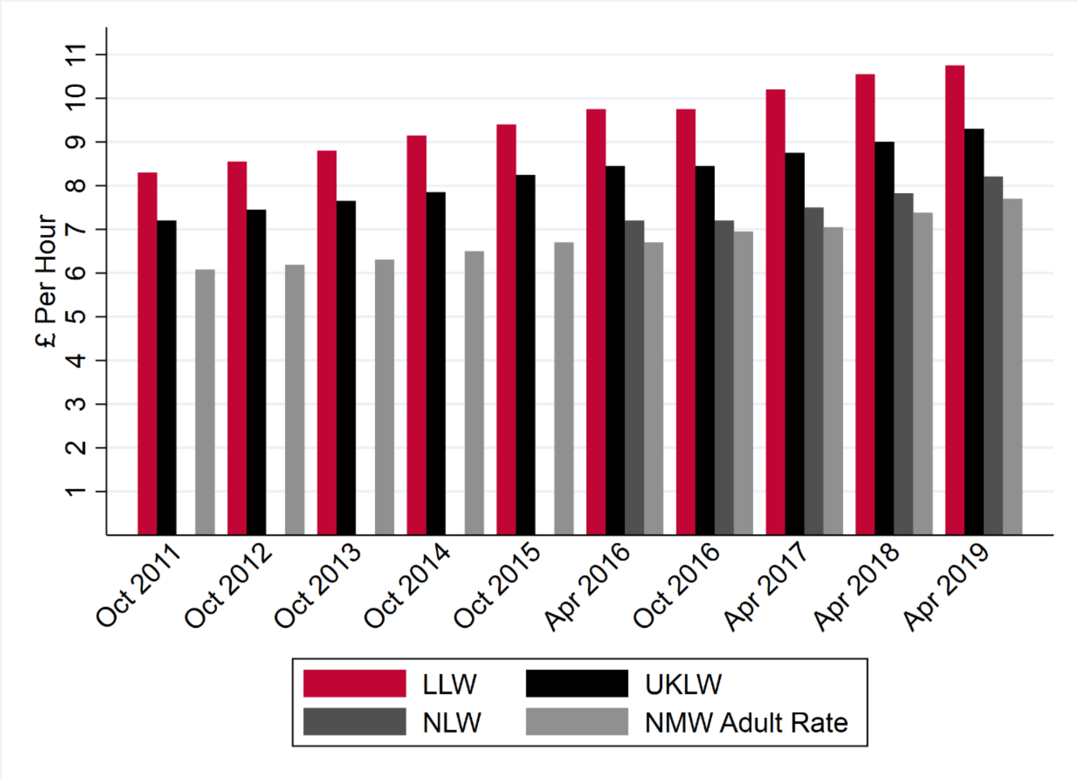
B Additional Figures

Figure 6: Separations vs Commutes with controls



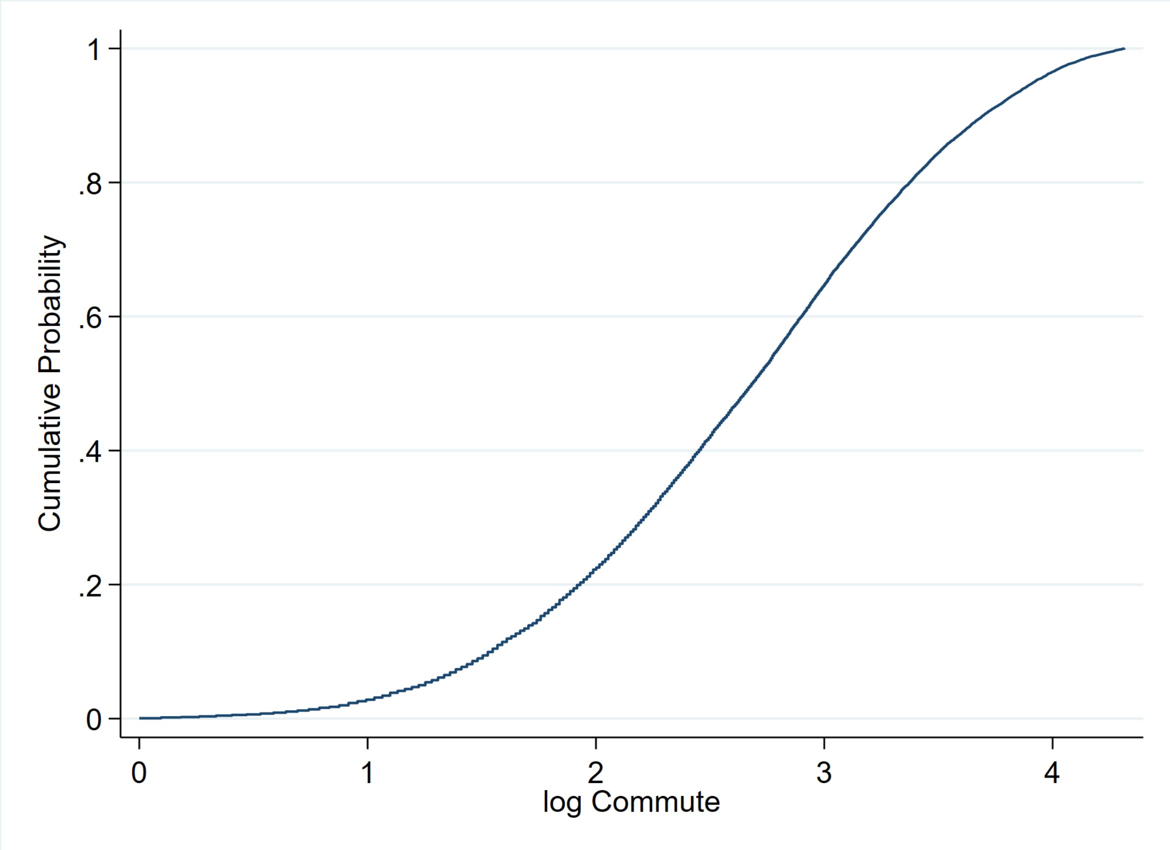
Note: *The figure plots the probability of separation within a year against commutes measured in minutes controlling for wage, age, sex, part time and temporary contracts, year, occupation and industry. Commutes are measured using the Open Street map dataset for Great Britain and ArcGis's networking tool. The sample is based on 1,429,376 worker-year observations from the Annual Survery of Hours and Earnings 2003-2019.*

Figure 7: Living Wage and Minimum Wage Rates



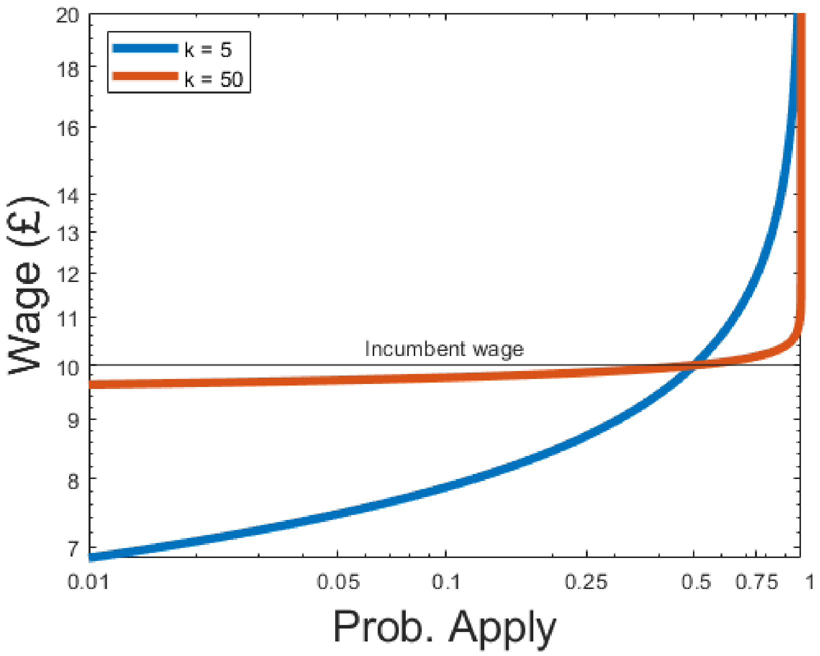
Note: *The figure shows the Living Wage Foundations' London and UK wide rates, as well as the statutory National Living Wage and National Minimum Wage adult rate for 2011 - 2019.*

Figure 8: Log Commute CDF



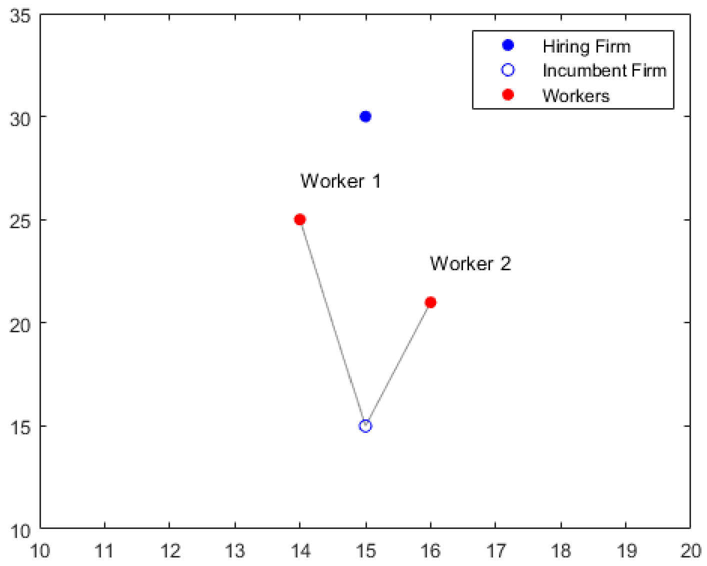
Note: *The figure presents the CDF of log commutes for job applicants for The Company between 2016 and 2019.*

Figure 9: Individual Labour Supply To the Firm



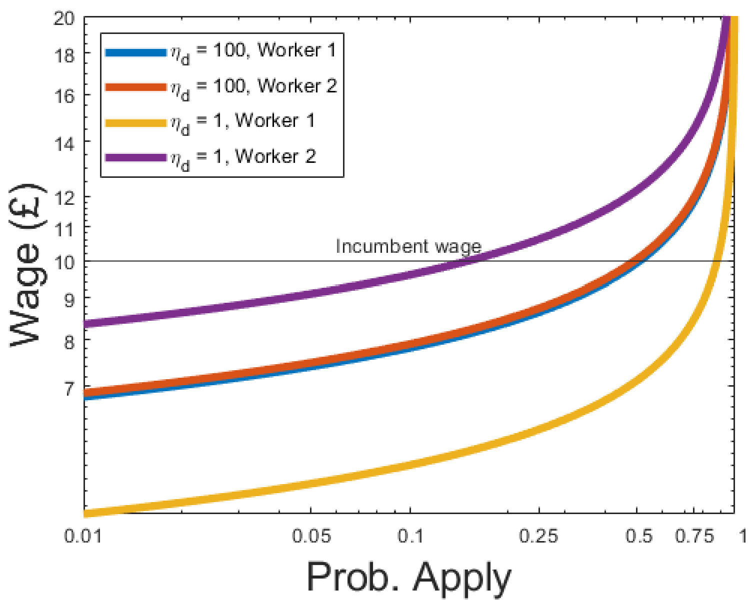
Note: *The figure plots the probability of applying for the worker as per equation (18), against the advertised wage on log scales. Parameterisation is such that* $\varepsilon_{cw} = \infty$ *and* $\nu_{ij} = median(\nu_{ij})$*.* F *is assumed to follow a weibull distribution with scale parameter* $\lambda = 1$ *and the figure plots for both shape parameter values of* $k = 5$ *and* $k = 50$ *.*

Figure 10: Spatial Heterogeneity in Individual Labour Supply To the Firm I



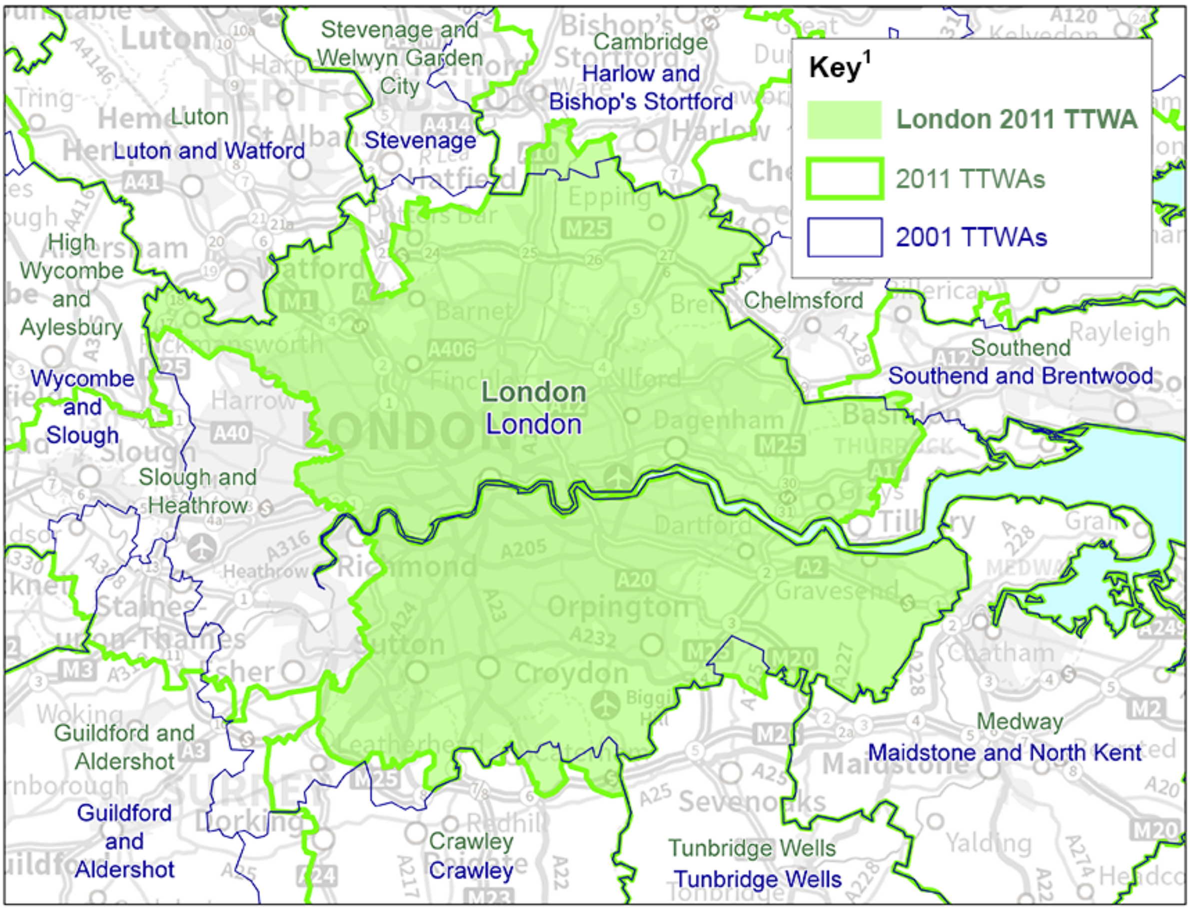
Note: *The figure presents a simulated spatial distribution of two workers with current wage of £10, their incumbent firm and a hiring firm.*

Figure 11: Spatial Heterogeneity in Individual Labour Supply To the Firm II



Note: *The figure plots the probability of applying to the hiring firm for the two workers spatially located as per figure 10 under different parameterisations of ε_{cw}, calculated according to equation (18), against the advertised wage on log scales. Parameterisation is such that $c = 0$ and $\nu_{ij} = median(\nu_{ij})$. F is assumed to follow a weibull distribution with shape parameter $k = 5$ and scale parameter $\lambda = 1$. When $\varepsilon_{cw} = 100$, where distance plays little role in determining utility, the supply curves are close to identical. When $\varepsilon_{cw} = 1$ where worker 2 experiences a lower elasticity of labour supply to the firm for any given wage than worker 1.*

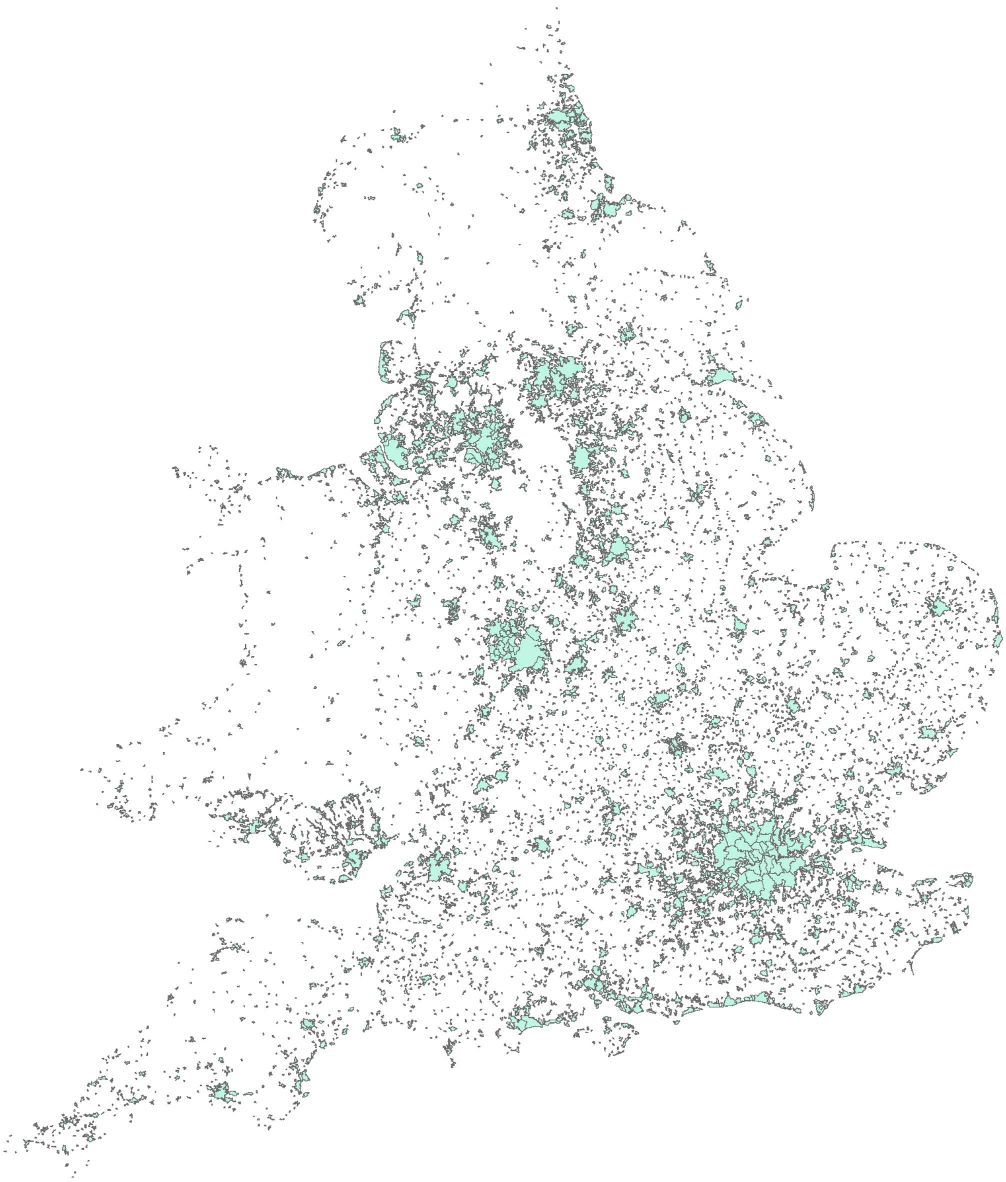
Figure 12: London TTWA



1 Travel to Work Area (TTWA).

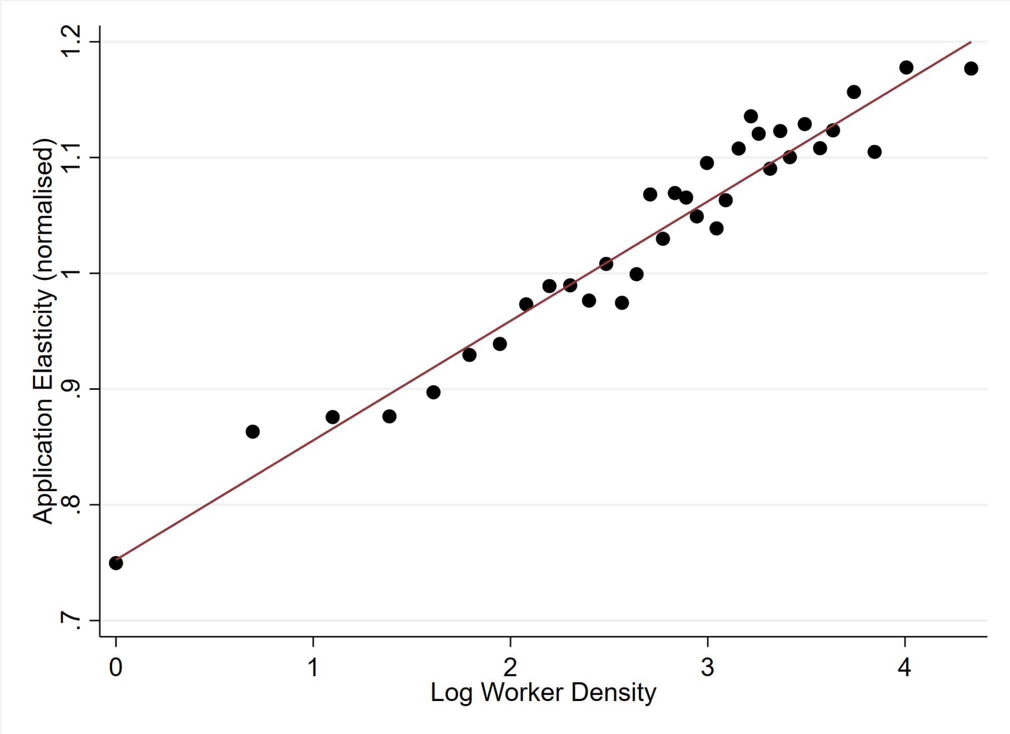
Note: *The figure presents the map of the London TTWA for the years 2001 and 2011.*

Figure 13: Built Up Area Sub Divisions for England and Wales

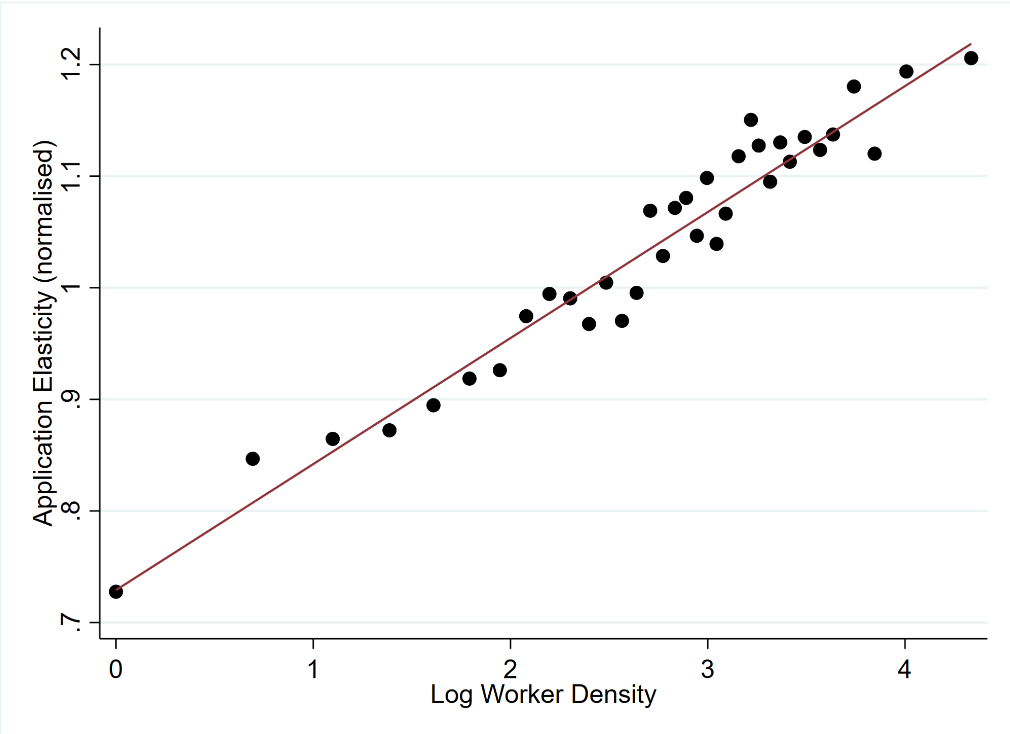


Note: *The figure presents a map of the Built Up Area and Built Up Area Sub Divisions for England and Wales.*

Figure 14: Model Predicted Elasticities vs Worker Density



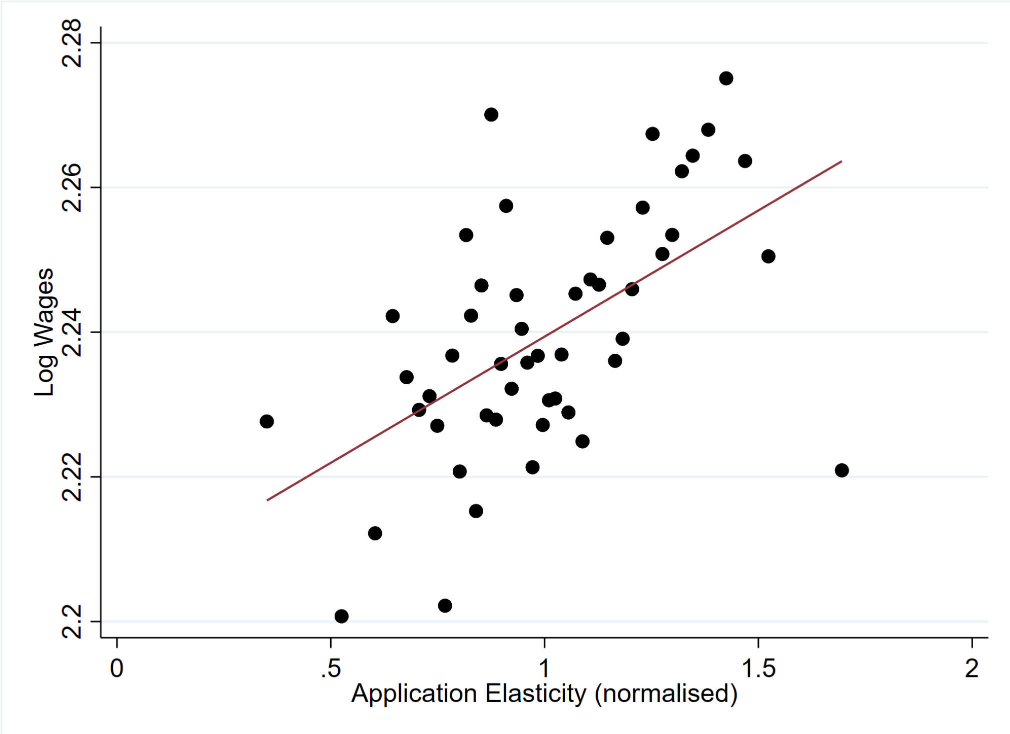
Slope = 0.074, Std. error = 0.002



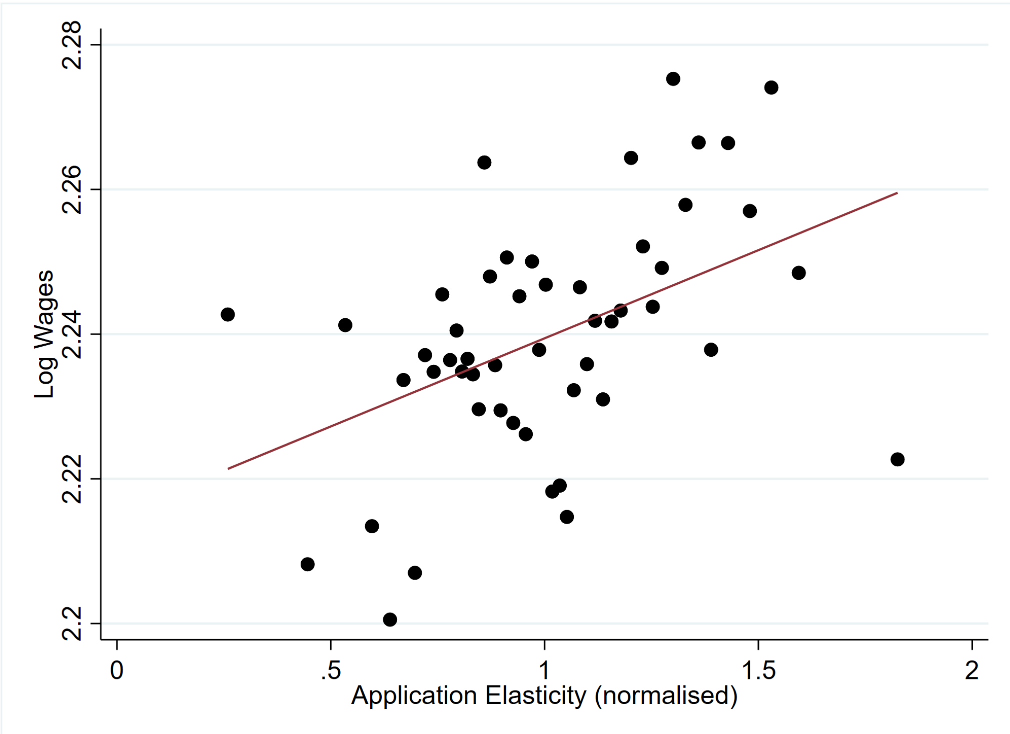
Slope = 0.113, Std. error = 0.003

Note: *The figure presents binned scatter plots of the application elasticity (normalised by dividing by its mean) as calculated by equation 22 using values of* $k = 5$ *and* $k = 7$ *for centroids of 7,625 BUAs and BUASDs for the market of retail workers against log worker density. Worker density is measured by those workers within a 25 minute drive of the centroid.*

Figure 15: Model Predicted Elasticities vs Wages



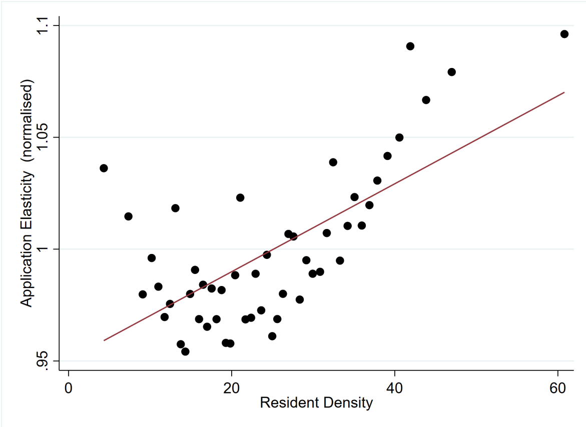
Slope = 0.035, Std. error = 0.006



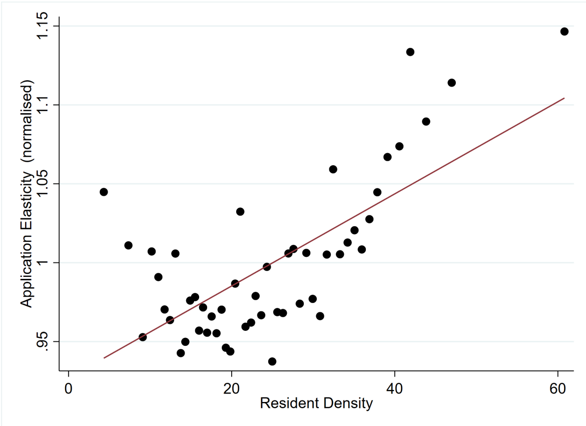
Slope = 0.024, Std. error = 0.005

Note: *The figure presents binned scatter plots of the application elasticity (normalised by dividing by its mean) as calculated by equation 22 using values of $k = 5$ and $k = 7$ for centroids of 7,625 BUAs and BUASDs for the market of retail workers against log average wages. Average wages are measured by the mean wage of workers within a 25 minute drive of the centroid.*

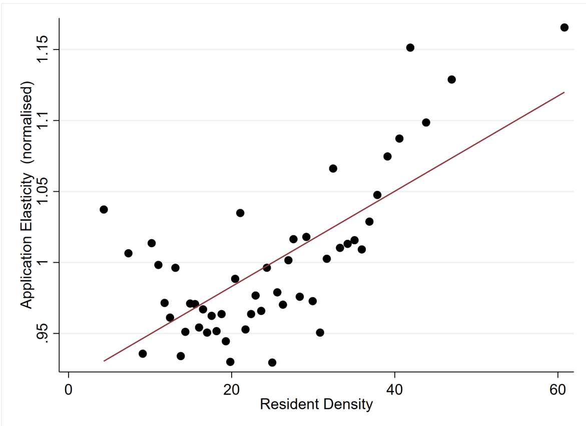
Figure 16: Model Predicted Elasticities vs Residential Density



Slope = 0.0020, Std. error = 0.0002



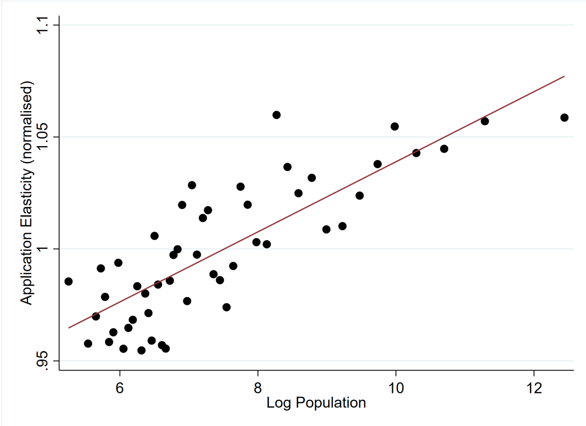
Slope = 0.0029, Std. error = 0.0002



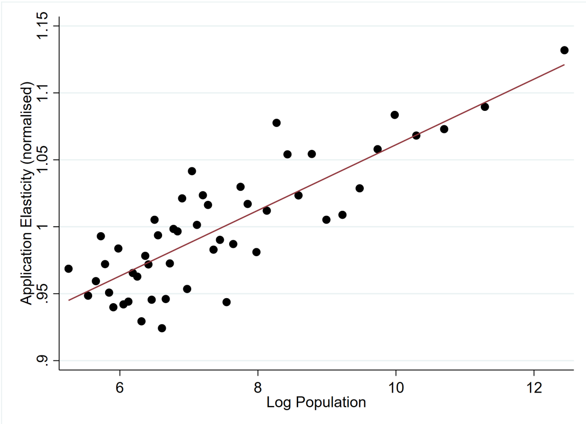
Slope = 0.0030, Std. error = 0.0003

Note: *The figure presents binned scatter plots of the application elasticity (normalised by dividing by its mean) as calculated by equation 22 using values of $k = 3$, $k = 5$, and $k = 7$ for centroids of 7,168 BUAs and BUASDs for the market of retail workers against residential density. Residential density measures from Census 2011 data from NOMIS.*

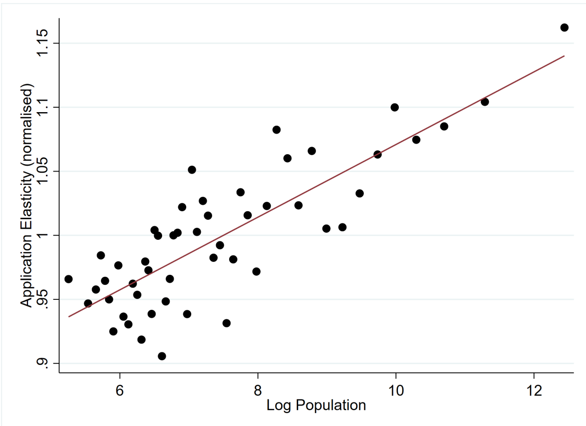
Figure 17: Model Predicted Elasticities vs Population



Slope = 0.0157, Std. error = 0.002



Slope = 0.0245, Std. error = 0.002



Slope = 0.0284, Std. error = 0.002

Note: *The figure presents binned scatter plots of the application elasticity (normalised by dividing by its mean) as calculated by equation 22 using values of $k = 3$, $k = 5$, and $k = 7$ for centroids of 7,168 BUAs and BUASDs for the market of retail workers against log population. Population measures from Census 2011 data from NOMIS.*

C Robustness

C.1 Estimating the Elasticity of Labour Supply to the Firm

To strengthen the credibility of the identification strategy employed and therefore results discussed in section 2.2.2 and 2.3, I perform a number of robustness checks.

C.1.1 Parallel Trends and Anticipation Effects

Firstly, one may have a concern that an announcement effect may cause a bias towards zero in the separation elasticity estimates. Specifically, if workers' employed by The Company in a particular establishment find out many months before the introduction of the Living Wage that they are to receive a substantial pay increase, they may decide to stay on with the firm longer. Discussions with the director of human resources suggests this is unlikely to be a concern treated workers only found out relatively near to the treatment date. However, it is prudent to empirically test for anticipation effects. Figure 18 graphically presents the transformed event study parameter estimates of the reduced-form estimating equation

$$Leave_{ijemy} = \sum_{l \neq -1,-11} \beta_{4,l} LW_{je,my+l} + \gamma_{je} + \lambda_{emy} + \theta_{jmy} + \epsilon_{ijemy} \tag{29}$$

where $l \in \{-12, ..., 12\}$ and the end points are binned such that $LW_{je,my+12} = 1 \; \forall \; \{l \; \geq 12 : LW_{je,my+l} = 1\}$ and $LW_{je,m-12} = 1 \; \forall \; \{l \; \leq -12 : LW_{je,my+l} = 1\}$.[48] Monthly effects are aggregated to the quarter, q such that

$$\hat{\beta}_4^q = \sum_{l \in q} \frac{1}{3} \hat{\beta_{4,l}} \tag{30}$$

and one may note that the monthly parameter effects are normalised to two periods, -1 and -11, as recommended in Borusyak and Jaravel (2017).

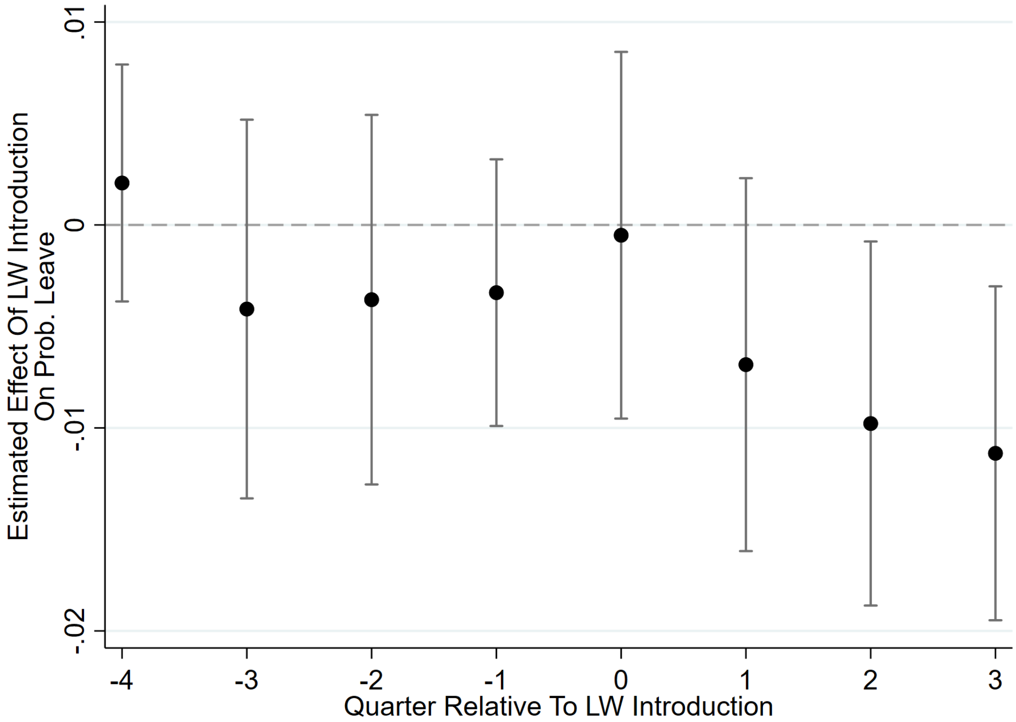
Figure 18 suggests that the restriction of parallel pre-trends can not be rejected, and therefore there is no evidence of anticipatory effects. The figure additionally shows that after one quarter of the introduction of the Living Wage, there is a clear drop in the rate of separations which continues to fall during the second and third quarter proceeding the introduction. For completeness figure 19 reports the similar exercise for the first stage. The estimation is identical to equation 29 and 30 except with the dependent variable changed to $log(Wage)$.

C.1.2 Staggered Treatment Timing

There has been a recent interest in the workings of two-way fixed effect estimators, in particular utilising staggered treatment times (Borusyak and Jaravel, 2017; Sun and Abraham, 2020;

[48] This implies that the first and last parameter estimate in 18 contain longer run pre and post effects.

Figure 18: Separations Event Study

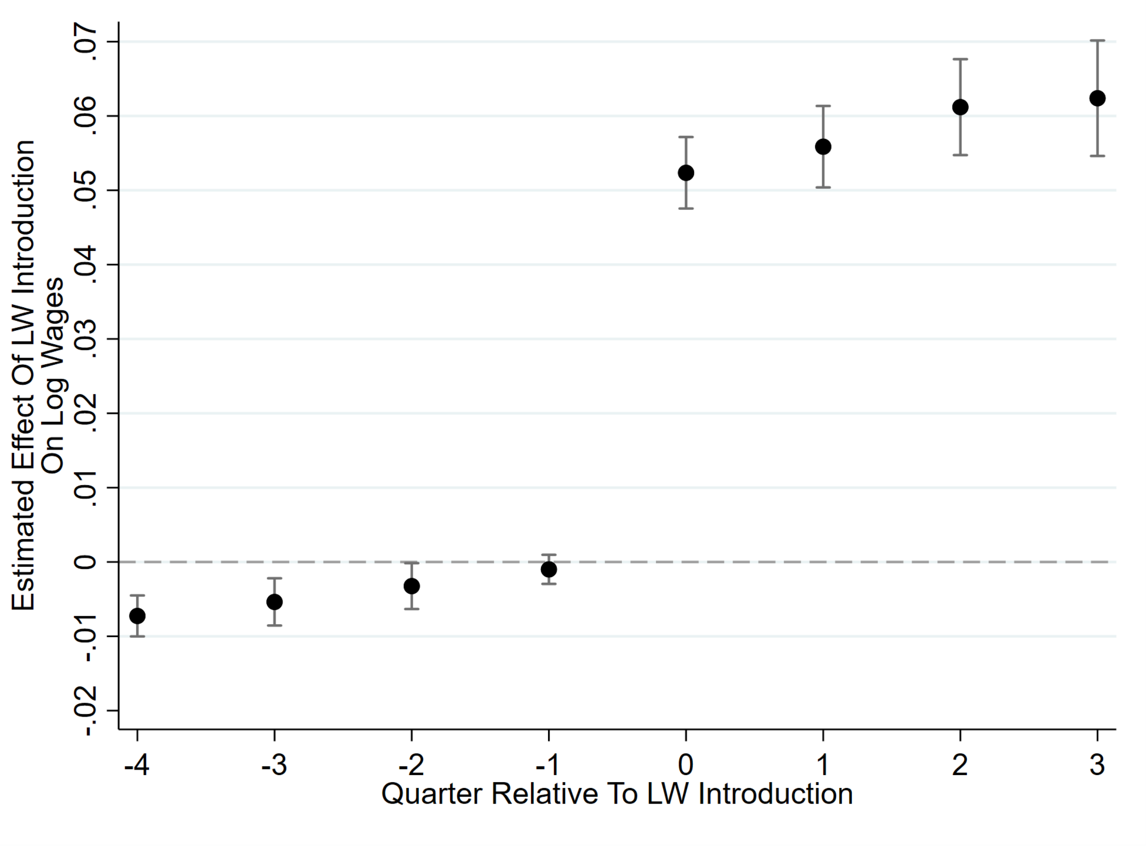


Note: *The figure presents estimates and 95% confidence intervals for parameters* $\hat{\beta}_4^q$ from equation 30.

Callaway and Sant'Anna, 2020; Goodman-Bacon, 2021). Concerns raised include: issues identifying the linear component of the path of pre-trends in traditional event study specifications (Borusyak and Jaravel, 2017), contamination of lead and lag coefficients from other period effects (Sun and Abraham, 2020), biased estimates of treatment effects when the control group contains treated units when dynamic treatment effects are present (Goodman-Bacon, 2021) and the structure of weights assigned across treatment cohorts when estimating dynamic treatment effects (Sun and Abraham, 2020). The approach in this paper is more flexible than a two-way fixed effect estimator, and when utilising just the Living Wage instrument, akin to a triple-difference estimator. Additionally, it is not obvious why some of these issues would be present in the current setting. For example, dynamic treatment effects are unlikely when studying the response of number of applicants in response to wage changes. Despite this fact, as a matter of caution to check whether any of these issues could be sullying the estimated effects when using the Living Wage instrument I implement a two-way fixed effects event study estimator at the establishment level akin to that suggested in Sun and Abraham (2020) while also implementing adjustments as recommended in Borusyak and Jaravel (2017). This estimator is the same implemented in Datta and Machin (2021). I compare these results to a traditional two-way fixed effects event-study estimator at the establishment level to see if there is a fundamental difference between the results.

The robust estimator is as follows. Borrowing notation from Sun and Abraham (2020), let

Figure 19: Wages Event Study



Note: *The figure presents estimates and 95% confidence intervals for parameters of* $\hat{\beta}_4^q$ from equation 30 where the dependent variable in equation 29 is $log(Wage_{ijemy})$.

Y_{et} denote some outcome for unit e at time t with treatment status $D_{et} \in \{0, 1\} : D_{et} = 1$ if e is treated in period t and $D_{et} = 0$ otherwise, where treatment is absorbing, and therefore $D_{es} \leq D_{it}$ for $s < t$. A unit's treatment path can therefore be characterised by $K_e = min\{t : D_{et} = 1\}$, and where we let $K_e = \infty$ if the unit is never treated. Units can therefore be categorized into disjoint cohorts $k \in \{t_{min}, \ldots, t_{max}, \infty\}$, where units in cohort k are first treated at the same time $\{e : K_e = k\}$. Y_{et}^k is the potential outcome in period t when unit e is first treated at time k and Y_{et}^{∞} is the potential outcome at time t if unit e never receives treatment. A cohort-specific average treatment effect on the treated l periods from treatment is thus:

$$CATT_{k,l} = E[Y_{e,k+l} - Y_{e,k+l}^{\infty} | K_e = k] \tag{31}$$

This notation allows treatment effect heterogeneity across cohorts, which in this setting may be important as the bite of the living wage may change over time. I am then interested in some weighted average of 31, for some $l \in g$, to construct a relative period coefficient. As is often the case when firms face a shock to the wage floor, we are interested in the average dynamic effects (which allows an analysis of the pre-trends).

For analysing the average dynamic effects I focus on the weighted average similar to that proposed in Sun and Abraham (2020).

$$v_g = \frac{1}{|g|}\sum_{l \in g}\sum_{k} CATT_{k,l}Pr\{K_e = k|K_e \in l\} \tag{32}$$

which effectively uses weights according to the size of the treated cohort that experiences l periods relative to treatment.

In practice 32 is estimated using the following methodology:

1. For each treatment cohort I estimate an adjusted form of the typical, two-way fixed effect, event study specification, where t is in months and I limit l to 12 months before and after the cohort treatment period.

$$Y_{et} = \alpha_e + \lambda_t + \sum_{l \neq = -1,-12} \delta_{k,l} LW_{i,t+l} + \beta' X_{et} + \epsilon_{et} \tag{33}$$

   Where α_e is the establishment fixed effect, λ_t is a year-month fixed effect, LW_{et} is a dummy variable which represents whether an establishment pays the Living Wage and X_{et} is a set of time varying establishment level controls. For each treatment cohort e, the control group is restricted such that they have not received treatment within the past two years, or will not receive treatment within two years of the relevant treatment cohort treatment date. This is to ensure no overlap of dynamic effects between the treated and control groups. As per the suggestion of Borusyak and Jaravel (2017), I normalise the dynamic effects to two periods, -1 and -12, to deal with the underidentification issues they raise.

2. I estimate the weights $Pr\{K_e = k|K_e \in l\}$ by sample shares of each cohort in the relevant relative period l.

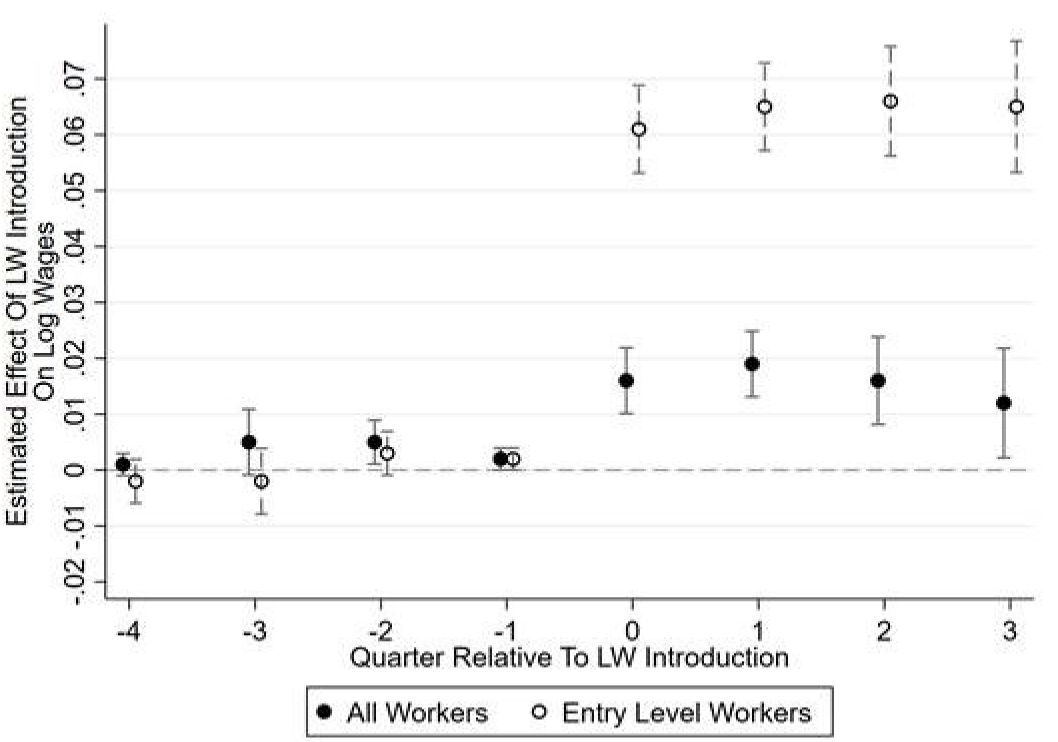
3. I combine steps 1 and 2, and aggregate monthly effects l, to the level of quarters g, for graphical representation by taking a simple equal weighted mean. In particular

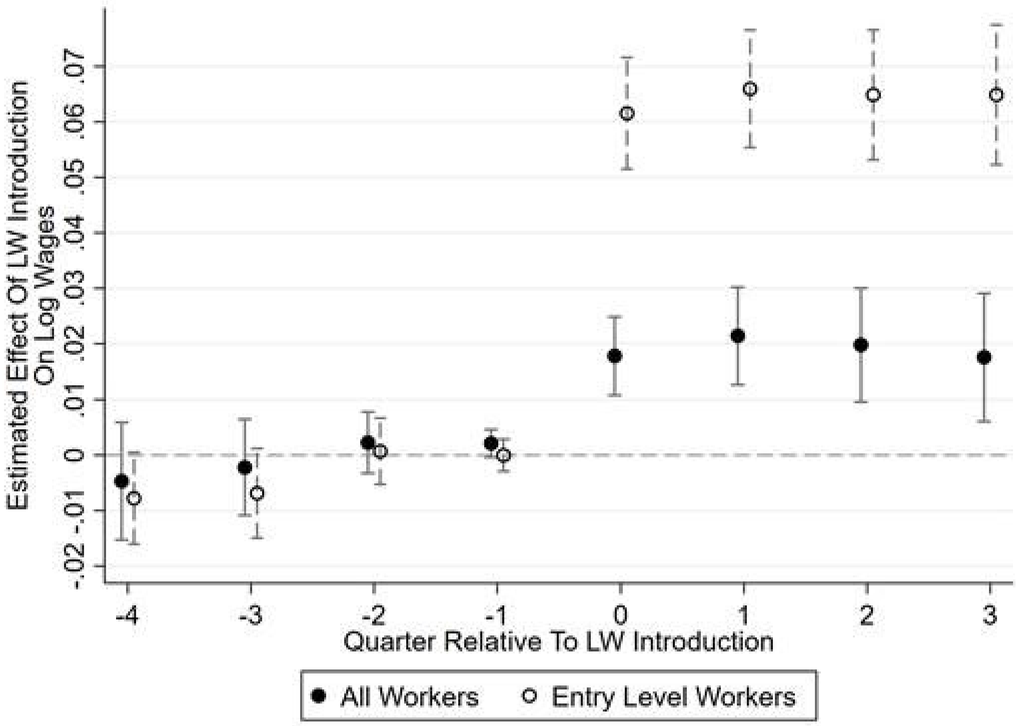
$$\hat{v}_g = \frac{1}{3}\sum_{l \in g}\sum_{k} \hat{\delta}_{k,l}\hat{Pr}\{K_e = k|K_e \in l\} \tag{34}$$

The above methodology comes with a number of benefits. Firstly, it is completely transparent about what weights are being used between treatment cohorts in the estimation of the parameters of interest. These weights are guaranteed to be convex and non-negative, which in the typical event study specification with variation in timing is not necessarily the case Sun and Abraham (2020). Secondly, there is clarity in terms of which groups are being used as treatment and control groups in both the dynamic, and long run treatment effect estimation. Thirdly, it deals with underidentification problems raised previously in the literature.

The top panel of figure 20 presents estimates of $\hat{v}_g$ from equation 34, using the more transparent

Figure 20: Event Study Robustness





Note: *The top panel presents estimates from equation 34, while the bottom panel presents results from equation 36. Both panels use a sample of establishments run by The Company active between January 2011 and April 2019, and are based off 17,879 establishment year-month observations. The vertical bars indicate 95% confidence intervals. For the top panel these are based on 1000 bootstraps.*

methodology outlined above. The bottom panel of figure 20 presents estimates using a standard pooled two-way fixed effects event study estimator,

$$Y_{et} = \alpha_e + \lambda_t + \sum_{l \neq = -1, -12} \delta_l LW_{e,t+l} + \beta' X_{et} + \epsilon_{et} \tag{35}$$

with monthly effects aggregated to the month per

$$\hat{v}_g^{pooled} = \sum_{l \in q} \frac{1}{3} \hat{\delta}_l \tag{36}$$

There is little obvious difference between the two panels for both impacts on all workers' and entry-level workers' wages at the establishment level. Both suggest parallel pre-trends, and stable dynamic treatment effects, with entry level workers experiencing 6% greater wage growth in treated establishments. This result is also consistent with the triple-difference first stage estimate from table 4 and 19. Given these results it is unlikely that the more flexible specifications utilising a triple-difference estimator will suffer from the aforementioned issues.

C.2 Estimating the Commuting-Wage Elasticity

It's reasonable to suspect that assuming a constant, linear in logs, relationship between wages and commutes is a strong assumption. Analysis using observational data however, suggests it is unlikely to be problematic. Figure 21 presents a scatter plot of log commutes against log wages, both orthogonalized against year and industry fixed effects, gender, age, part time and temporary, using data from the nationally representative Annual Survey of Hours and Earnings. There is a clear positive correlation and is approximated well by a linear relationship, except perhaps at the extreme tails.

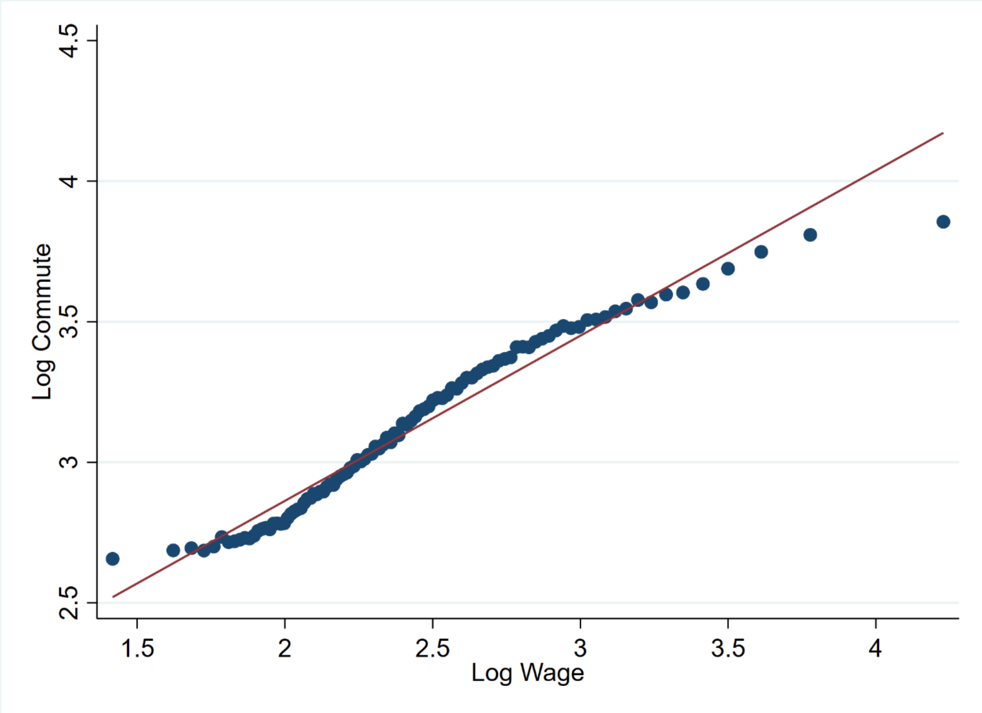
Further to this end figure 8 and table 13 in the appendix presents the CDF of log commutes for the sample used in columns (2) - (4) in table 5 and the results from the instrumented regression[49]

$$P(log(Commute_{iajemy}) \leq x) = \beta_5^x log(Wage_{jemy}) + \delta' X_{ia} + \gamma_{je} + \lambda_{ey} + \nu_{ym} + \epsilon_{iajemy} \tag{37}$$

The estimates from equation (37) show that the CDF under a higher wage would stochastically dominate a CDF under a lower wage, as expected. Furthermore, effects are sizeable across most of the distribution aside from the tail ends. The estimates imply a 10% increase in wages would reduce the probability of being lower than a specific value by between 3-6% between the 10th percentile up to the 75th percentile. Therefore, allowing for a non-parametric relationship between commutes and wages would most likely only marginally improve accuracy, while complicating estimation of the model and losing a degree of analytical tractability.

[49] Ideally one would look at the effects across the distribution using a quantile regression, however at the time of writing this IV quantile regression with high dimensional fixed effects is not possible.

Figure 21: Log Commutes vs Log Wage



Note: *The figure presents a binned scatter plot of log commutes against log wages, with commutes and wages orthogonalised against year and industry fixed effects, age, gender, part time and temporary contract. The sample is based on 2,152,513 observations from the Annual Survey of Hours and Earnings between 2002 and 2019.*

Table 13: Commuting - Wage Distribution Estimates

	(1) x = 1	(2) x = 1.5	(3) x = 2	(4) x = 2.5	(5) x = 3	(6) x = 3.5	(7) x = 4
log(Wage)	-0.05	-0.33**	-0.33*	-0.63**	-0.56**	-0.06	-0.03
	(0.07)	(0.16)	(0.20)	(0.26)	(0.25)	(0.18)	(0.10)
Job-Centre FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Centre-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	19585	19585	19585	19585	19585	19585	19585
Inst. W/ AL	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Inst. W/ LW	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Non-Perm	Non-Perm	Non-Perm	Non-Perm	Non-Perm	Non-Perm	Non-Perm

Note: *The table presents estimates of $\hat{\beta}_5^x$ from equation (37). Standard errors are reported in parentheses and are clustered at the establishment. Regressions are weighted by the inverse number of applicants for each job. Controls include gender and ethnicity.* * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D The Maintaining Worker Problem

For simplicity assume $c = 0$. Worker i, in job j, when job j' is advertised will stay with probability:

$$\varphi_{ijj'}(w_j) = 1 - P(A_{j'})(1 - F(x_{ijj'})) \tag{38}$$

This implies

$$\varepsilon_{\varphi w_{ijj'}} = \frac{P(A_{j'})f(x_{ijj'})}{1 - P(A_{j'})(1 - F(x_{ijj'})}\tilde{x}_{ijj'} \tag{39}$$

Note that

$$\varepsilon_{sw} = -\frac{\varphi}{1-\varphi}\varepsilon_{\varphi w} \tag{40}$$

$$\varepsilon_{sw_{ijj'}} = -h(x_{ijj'})x_{ijj'} \tag{41}$$

It is straight forward to see how the above is the inverse of the hiring problem and particularly obvious when looking at equation 39. As incumbent utility increases, $x_{ijj'}$ increases .
If $h(.)$ is increasing in argument, it follows that $\frac{f(.)}{F(.)}$ is decreasing in argument. Thus

- As w_j increases, $x_{ijj'}$ increases and therefore $\varepsilon_{\varphi w_{ijj'}}$ decreases.
- It is easy to see that the behaviour of $\varepsilon_{\varphi w_{ijj'}}$ follows a similar pattern as $\varepsilon_{Aw^j_{ij'}}$.

Assuming many firms were posting vacancies, aggregation would then follow such that

$$\varphi_{(ij)}(w_j) = \sum_{j'} \varphi_{(ij),j'}(w_j) \tag{42}$$

where as before there is an assumption that the probability of receiving more than one job offer is infinitesimal.

E Introducing Search Costs

The simplest way to introduce search costs into the model is by introducing an application cost into the application decision for the worker. A worker will choose to apply to posted job j' according to:

$$P(A_{j'})u_{ij'} - (1 + P(A_{j'}))u_{ij} - c \geq u_{ij} \tag{43}$$

where c is some fixed application cost. This implies the worker will apply if

$$\nu_{ij'} \geq \underbrace{\frac{w_j d_{ij}^{\frac{-1}{\varepsilon_{cw}}} \nu_{ij} + \frac{c}{P(A_{j'})}}{w_{j'} d_{ij'}^{\frac{-1}{\varepsilon_{cw}}}}}_{\equiv \tilde{x}_{ijj'}} \tag{44}$$

where $\tilde{x}_{ijj'}$ is akin to $x_{ijj'}$ in the model without search costs.

E.1 The Individual's Elasticity

The individual's elasticity of applying with respect to posted wage becomes

$$\tilde{\varepsilon}_{AW_{ij}^{j'}} = h(\tilde{x}_{ijj'})\left(\tilde{x}_{ijj'} + \frac{c}{P(A_{j'})w_{j'} d_{ij'}^{\frac{-1}{\varepsilon_{cw}}}} \varepsilon_{PA}\, \tilde{\varepsilon}_{AW^{j'}}\right) \tag{45}$$

where $\varepsilon_{PA} \leq 0$ is the elasticity of the probability of getting the job with respect to the number of applicants, and $\varepsilon_{AW^{j'}}$ is as before the aggregate elasticity of applications to wages for the firm, but now in the presence of search costs included.

One can see a key change with the addition of search costs that as more workers apply for a job, they have an externality on other worker's choice decision, and this in turn can have a negative impact on the size of the responsiveness of an individual worker applying for a job. Intuitively, if a wage is posted much higher than the going market rate, some workers may not apply as they believe that the probability of getting the job goes down, and this matters when applications are costly. The size of this externality depends on two factors c and ε_{PA}. The role of the first of these is trivial while the latter depends somewhat on the production function of the firm. For example, if firms hire all workers that apply for a vacancy (such as on some online task markets) then $\varepsilon_{PA} = 0$ and the problem collapses to a similar framework as in section 3. In the situation where a firm only has one job vacancy on the other hand, $\varepsilon_{PA} > 0$ and any increase in applicants would reduce the probability of getting an offer. Additionally, without imposing any additional restrictions it is not clear that $\varepsilon_{AW_{ij}^{j'}}$ is positive for all individuals, nor that the elasticity is decreasing in the posted wage.

E.2 The Elasticity of Labour Supply To The Firm

Aggregating to the firm level the application elasticity is the application elasticity absent of search costs divided by one plus a "search wedge", $S^{j'}$. $S^{j'} \geq 0$, and this wedge is increasing in c and ε_{PA}. Furthermore this search wedge is a weighted average of the individual specific search wedges, $S^{j'}_{i,j}$.

$$\tilde{\varepsilon}_{AW^{j'}} = \frac{\varepsilon_{AW^{j'}}}{1 + S^{j'}} \tag{46}$$

such that

$$S^{j'} = |\varepsilon_{PA}| \frac{\sum_{(i,j)} S^{j'}_{i,j}(1 - F(x_{ijj'}))}{\sum_{(i,j)}(1 - F(x_{ijj'}))} \tag{47}$$

and

$$S^{j'}_{i,j} = h(x_{ijj'}) \frac{c}{P(A_{j'}) w_{j'} d_{ij'}^{\frac{-1}{\varepsilon_{cw}}}} \tag{48}$$

As a result the introduction of search costs reduces the application elasticity to the firm, and the extent of this depends on the size of c and ε_{PA}.

F Relationship with Logit Models of Monopsony

A number of recent studies have used logit models of search (Card et al., 2018; Azar et al., 2019; Lamadon et al., 2019) to illustrate monopsony power in the labour market. The utility function set up in this setting with a Weibull distribution speaks directly to these models due to the relationship between the Extreme Value Type 1 distribution and the Weibull distribution.

Given utility is such that $u_{ij} = w_j d_{ij} \nu_{ij}$ where ν_{ij} is Weibull distributed with shape parameter k and scale parameter λ utility can be log transformed such that

$$\hat{u}^*_{ij} = log\ w_j - \frac{1}{\varepsilon_{cw}} log\ d_{ij} + \xi^*_{ij}. \tag{49}$$

Where ξ^*_{ij} is distributed Extreme Value type 1 (i.e. Gumbel) with scale parameter $\beta = \frac{1}{k}$ and location parameter $\mu = log(\lambda)$.

Assuming $\varepsilon_{cw} = \infty$, that is, commuting does not matter, the utility function can be rewritten

$$\hat{u}_{ij} = \frac{1}{\beta} log\ w_j + \xi_{ij} \tag{50}$$

where ξ_{ij} now has variance $\frac{\pi^2}{6}$.

The choice decision facing the searching worker implies the probability they will apply to some job j within choice set $j \in \mathcal{J}$ if

$$P_{ij} = Pr(\xi_{ij'} - \xi_{ij} < w_j - w_{j'}) \forall j' \neq j \in \mathcal{J} \tag{51}$$

Which gives the logit formulation

$$P_{ij} = \frac{exp(\frac{1}{\beta} w_j)}{\sum_{j'} exp(\frac{1}{\beta} w_{j'})} \tag{52}$$

Therefore, if there are L workers in the market the number of applications to the firm is given by

$$A_j = L * \frac{exp(\frac{1}{\beta} w_j)}{\sum_{j'} exp(\frac{1}{\beta} w_{j'})} \tag{53}$$

Taking logs, equation (53) can be re written as

$$a_j = log(L) + \frac{1}{\beta} w_j - log\Big(\sum_{j'} exp(\frac{1}{\beta} w_{j'})\Big). \tag{54}$$

And therefore the elasticity of applications to the firm is given by

$$\varepsilon_{AW} = \frac{1}{\beta}(1 - s_j) \tag{55}$$

where s_j is akin to the share of firm j in the labour market.

Assuming there is a large number of firms, $s_j \approx 0$ and therefore $\varepsilon_{AW} = \frac{1}{\beta} = k$. This therefore demonstrates that in the absence of commuting costs, the elasticity of applications to the firm is equal to the shape parameter of the Weibull model.

UniCredit Foundation

Piazza Gae Aulenti, 3

UniCredit Tower A

20154 Milan

Italy

Giannantonio De Roni – *Secretary General*

e-mail: giannantonio.deroni@unicredit.eu

Annalisa Aleati - *Scientific Director*

e-mail: annalisa.aleati@unicredit.eu

Info at:

www.unicreditfoundation.org

UniCredit Foundation