Wages and job vacancy durations: Evidence from a spatial discontinuity

Charles Carter[1], Judith Delaney[1], and Kerry L. Papps[2]

[1]Department of Economics, University of Bath

[2] School of Management, University of Bradford and IZA, Bonn.

Abstract

We estimate the causal effect of posted wages on vacancy durations utilising online NHS job vacancy data from *findajob.co.uk* and a sharp border discontinuity resulting from the existence of High Cost Payment Areas in the Greater London area. We find a large significant negative relationship between posted wages and vacancy durations which is consistent with the predictions of a directed search model incorporating heterogeneous job seekers and exogenous wage setting.

1 Introduction

The wages advertised by firms for unfilled vacancies underpin workers' search behaviour. In a typical utility maximising microeconomic framework, wages posted by firms, which We refer to as 'posted wages', are used to incentivise job seekers to apply to unfilled vacancies. As such traditional search modelling often assumes that there exists a trade-off between posted wages and the search costs associated with an open vacancy. This assumption however, despite being intuitive, has little backing in the current empirical literature. This is partially due to the difficulty in characterising all possible jobs available to a particular job seeker and controlling for job and firm fixed effects. The estimation of the relationship between posted wages and vacancy durations is particularly interesting as it relates directly to key issues in the field of labour market search dynamics. The relationship between posted wages and vacancy durations is of key importance to firms, who use wages as a mechanism to fill job vacancies quickly and mitigate costs associated with the job search. Moreover, Ihsaan Bassier and

Petrongolo (2023) argue that firm-level wage elasticity of employment, of which the wage-duration elasticity is a component, is a fundamental measure of a firm's monopsony power.

This theoretical discussion can be directly extended to the critical discussion of staff shortages in Western healthcare systems. Across Western countries, employment in health services has lagged behind rising demand resulting from ageing populations and increasing expectations for healthcare from wealthy and more educated populations Shields (2004). The United Kingdom (UK) National Health Service (NHS) has not been exempt from such issues. The NHS is a publicly funded care system with the aim of providing healthcare service that is free at the point of use for all UK citizens and is, at the time of writing, the seventh biggest employer in the world Statista (2022). However, alongside the prior mentioned systems, evidence suggests that the NHS is critically understaffed, see for example; The Health Foundation (2020) , Rivett (n.d.) and Morgan (2022). As of September 2021, the NHS had 99,460 vacant positions. This number of shortages corresponded to a job vacancy rate of 12% for nurses and 8% for doctors Health and Social Care Committee (2022). Such labour shortages of key workers weigh on the ability of the NHS to deliver effective health care which manifests into poorer health outcomes and longer waiting list times for patients [1]. At the time of writing the NHS has a waiting list of almost 7 million people (Health and Social Care Committee, 2022). Moreover, evidence suggests that labour shortages may increase physical and psychological pressures on the existing workforce (Buchan J, 2019) which could further reduce the ability of the NHS to retain staff, exacerbating the aforementioned challenges over time.

The NHS pay review body, a quasi-independent board which advises the government on issues relating to NHS worker pay, recognises that increasing wages could act as a mechanism to improve labour shortages. They acknowledge that increasing wages could act further to improve staff retention rates and improve wellbeing (NHS Pay Review Body, 2022). Therefore, the NHS pay review body consistently recommends pay increases on the basis that they act as a mechanism to reduce vacancy durations and labour shortages, at least in the short run. While this approach seems intuitive, there exists counteracting economic forces that theoretically inform the relationship between posted wages and vacancy durations. Increasing posted wages improves the consumption possibilities of workers who are successful in their application which, according to traditional microeconomic predictions, leads to

[1] See; Kelly, Stoye and Warner (2022) , McHugh et al. (2011), Shanafelt et al. (2019), Larrabee et al. (2003), Aiken et al. (2008), Francis (2013)

higher utility. However work by Faberman and Menzio (2018) suggests that, in a setting of agents with heterogeneous skill levels, those who are less productive may perceive jobs with higher posted wages as those with higher levels of competition. Given less productive applicants are less likely to be selected, they are less likely to apply. Building on theoretical work by Faberman and Menzio (2018) and Shimer (2005), We produce a directed search model of heterogeneous agents, which investigates the consequences of these counteracting incentive effects. We find that, in the context of an exogenous wage differential with homogeneous firms and heterogeneous agents, there exists a positive relationship between posted wages and the number of job applications a vacancy receives. Symmetrically, this theoretically corresponds to a negative relationship between posted wages and expected job vacancy durations.

We verify this implication using NHS job vacancy data, extracted from the job vacancy website *find-ajob.gov.uk* between April 2021 and March 2023. It is government policy that NHS workers are paid an additional 20%, 15% or 5% annual wage premium if they work in inner, outer or fringe London respectively. This is to reflect the higher cost of living in these areas. As such, in conjunction with the available job vacancy data, this geographical setup provides a strong natural experiment to mitigate potential sources of endogeneity. We use a sharp regression discontinuity specification to estimate the causal effect of posted wages on vacancy durations. In doing so We contribute to the literature by testing search theories predictions of the relationship between wages and vacancy durations while simultaneously providing a policy prescription for the NHS. We estimate an elasticity between job vacancy durations and posted wages of approximately -5.5, controlling for job characteristics, time, local labour market conditions and distance to the border. This estimate is consistent with new evidence from Ihsaan Bassier and Petrongolo (2023), but is larger in magnitude than previous studies in this field. The estimate is robust to changes in the bandwidth, a critical assumption in regression discontinuity approaches. This works builds on the research We completed for the Research Apprenticeship Project (RAP), Carter (2023)

In section 2 We critically explore the current literature which analyses the relationship between the wages firms post, henceforth known as posted wages, the job filling rates, vacancy rates and other hiring outcomes. We describe the literature that tackles this relationship both theoretically and econometrically and discuss where my work builds on previous studies. In section 3 We discuss the institutional background of the NHS in London. In section 4 We modify a directed search model with heterogeneous job seekers and firms by Shimer (2005). In doing so We outline directed search theories predictions of the relationship between posted wages and vacancy durations in the context of a large organisation a with rigorous pay progression criteria. In sections 5 and 6 We describe the data set and the empirical methodology before proceeding to outline the empirical results in sections 7 and 8. We present limitations in section 9 and conclude in section 10.

2 Literature Review

My research is closely connected to prior studies that have explored the correlation between wages and several job-related factors, including the number of job applications, the length of time a position remains vacant and aggregated vacancy rates. The literature extensively employs labour market search models of both random (Burdett and Mortensen, 1998) and directed (Moen (1997), Eeckhout and Kircher (2010), Shimer (2005)) search to investigate the connection between wages and vacancy durations.

In both directed and random search models, firms face a trade-off between the wages they offer and the costs associated with conducting job searches. In directed search job seekers adopt symmetric mixed strategies when making application decisions. The wage paid to an individual serves as a constraint on feasible consumption combinations and ,as a result, higher wages directly increase the expected utility of workers applying to a particular job. This translates into a positive relationship between posted wages and the number of applications. Consequently firms typically set wages endogenously to balance the trade off between wages search costs. In random search models, exemplified by McCall (1970) and others, higher wages increase the likelihood that the offered wage, drawn from a certain probability distribution, surpasses a job seekers reservation wage. Consequently, in both random and directed search frameworks, conventional theory would predict an inverse correlation between initial wages and the duration of a job vacancy.

There exists a small but growing literature that looks to test the negative relationship between wages and vacancy durations, posited by search theory. Faberman and Menzio (2018), utilising data from the Earnings and Opportunity Pilot Project Survey (EOPP) spanning the years 1980 and 1982, explore the association between posted wages, weekly application rates, interview frequencies, and the duration of job vacancies. Counter-intuitively they find that starting wages are negatively correlated with the number of applications a vacancy receives. This relationship naturally corresponds to a positive association between starting wages and posting lengths. Acknowledging the incongruence between these findings and conventional search theory, the researchers adopt a model of heterogeneous firms and jobs, initially introduced by Shimer (2005), to elucidate how unaccounted variations in firm and worker characteristics could underpin the emergence of this counter-intuitive relationship. The authors concede, with the use of their theoretical underpinnings, that the use of more robust firm and job title heterogeneity controls would produce a negative relationship between wages and vacancy durations, aligning more closely with the predictions of labour search theory. We present a more thorough explanation of their directed search model in section 4.

The underlying dynamics of this seemingly paradoxical relationship is further explained by Marinescu and Wolthoff (2020), who examine the significance of wage information displayed on job listing platforms on the number of applicants a posting receives. The authors source data from *CareerBuilder.com*, an online job vacancy platform for postings within the United States which contains detailed information on vacancies such as job titles, wages, and the number of applications for each vacancy. They find that controlling for a vacancies Standard Occupational Classification (SOC) code produces a negative relationship between a jobs wage and the number of applicants it attracts. Intuitively this would imply a positive relationship between posted wages and vacancy durations. After controlling for the more specific job title, however, they find that higher wages attract more applicants and correspondingly result in lower vacancy durations.

Building on the work of Faberman and Menzio (2018), Mueller et al. (2018), merge job advertisements data from the Austrian public market administration (AMS) with data sets encompassing worker-firm dynamics in Austria. As such they have access to a more sophisticated set of controls for worker and job title heterogeneity than Faberman and Menzio (2018). They also include firm fixed effects controls to mitigate potential endogeneity resulting from the relationship between posted wages and establishment characteristics. As is consistent with the theoretical predictions of Faberman and Menzio (2018), they find that there is a negative relationship between the duration of job vacancies and the entry wage for new hires, when using adequate job title and establishment heterogeneity controls. Furthermore, they identify a negative relationship between the vacancy duration and firm level wage premiums. Notably, as is consistent with the work of Marinescu and Wolthoff (2020), they find that the positive relationship between starting wages and vacancy durations they observe in the raw data turns negative with sufficient controls for job and firm heterogeneity.

These findings are also consistent with the results of Belot, Kircher and Muller (2018) who undertake an experimental 'audit study' involving unemployed job seekers conducting job searches under controlled laboratory conditions. They design a job search website using vacancies from Universal Job Match and present job seekers with pairs of similar job openings featuring different, randomly assigned wages. Their experimental results indicate that higher wage vacancies tend to attract a greater number of 'saves' from job seekers, suggesting that a 1% increase in wage corresponds to a 0.7% to 0.9% increase in the number of saves. Belot, Kircher and Muller (2018) also observe that 42% of individuals who save low-wage vacancies do not extend the same interest to high-wage vacancies. This aligns with the concept of a trade-off between elevated wages and the perceived competition for a job, which can shed light on the positive relationship between wages and the observed duration of job vacancies, as observed by Faberman and Menzio (2018) and Banfi and Villena-Roldan (2019). In addition to these empirical observations, Belot et al. (2018) formulate a directed search model that incorporates multiple applications and on-the-job search. Through calibration using UK data, this model effectively replicates the negative relationship between wages and job vacancy durations that their empirical investigations uncover. These results resonate with the findings of Holzer (1990), who finds that firms offering the minimum wage receive a higher number of applicants compared to firms that offer the minimum.

The preceding studies collectively contribute to the understanding of the relationship between wages and job vacancies within the broader context of labour market dynamics. The findings strongly emphasise the necessity of considering firm and job title heterogeneity in econometric specifications. In my estimation We consider a single firm with a stringent pay progression system so as to navigate these concerns. Moreover, the preceding literature reflects the importance of complementing empirical analysis with appropriate theoretical frameworks, to gain a more comprehensive understanding of the underlying job market dynamics.

This literature also builds on the use of regression discontinuity (RD) designs in the empirical literature for questions in both applied and health economics. The method provides a quasi-experimental approach to examining causality, exploiting situations where the probability of a given individual being treated is derived according to the value of a given assignment variable. RD approaches have therefore, been utilised across the social sciences to construct experimental designs without access to fully randomised data. RD methodologies have been extensively analysed from a theoretical perspective in the literature. Hahn, Todd and Van der Klaauw (2001), discuss, in formal detail, the assumptions that are necessary to adequately derive causal estimates in the RD framework. They also provide discussion on the use of non-parametric methods in the use of RD estimation. Imbens and Lemieux (2008) and Lee and Lemieux (2010) survey the existing literature on sharp and fuzzy methodologies. Furthermore, Calonico, Cattaneo and Titiunik (2014) derive data-driven methods for optimal bandwidth selection, a critical assumption in RD methodology and Frölich and Huber (2023) discuss the use of covariates in RD designs.

RD methods have also been extensively employed in practice. In Thistlethwaite and Campbell (1960) seminal paper, they use the RD method to study the relationship between student scholarships and career aspirations. Other early uses of RD designs include Van der Klaauw et al. (1997) and Angrist and Lavy (1999) who, utilising sharp discontinuities, estimate relationships between financial aid and student college attendance decisions on educational outcomes. We contribute to the existing body of literature by employing a RD approach in the presence of a sharp border wage discontinuity, in order to analyse labour market dynamics in the NHS.

This paper contributes to the preceding literature in two ways. Firstly We adapt a directed search model by Shimer (2005) to better understand labour market dynamics in the context of a single firm. This is interesting in the context of large monopsonies, where job and pay progression are strictly defined. The model also introduces implications surrounding discrimination and geographical immobility. In doing so We build on the growing use of directed search models with heterogeneous agents to explain various phenomena relating to job market dynamics [2]. Secondly, We utilise an RD approach to test the implications of the directed search model. The use of a single firm with specialised jobs, allows me to bypass some of the issues surrounding job and firm heterogeneity which plagues the analysis in this field. We therefore, provide a policy prescription by presenting theoretical and econometric evidence which gives credence to the use of increased NHS worker wages to reduce vacancy durations, and consequently improve health outcomes, in the short run.

[2] Refer to Shi (2002),Montgomery (1991),Menzio, Telyukova and Visschers (2016) and Wolthoff (2018).

3 Institutional Background

In this section, We discuss the institutional background of the research, including the geographical structure of the border discontinuity and the pay and progression criteria for NHS workers. We also describe the categorisation of job titles in the NHS.

3.1 High Cost Area Payments

It is government policy that those that work for the NHS, in and around London, are entitled to an additional High-Cost Area Payment (HCAP) as a percentage of their base salary. This is to account for the higher cost of living in and around the Greater London Area. NHS staff working in fringe, outer or inner London are paid an additional 5%,15% and 20% on top of their base salary respectively. The geographical boundaries of London HCAP areas are defined by the 2005 Primary Care Trust (PCT) zones , included in the NHS Terms and Conditions NHS (2023). The areas included in each zone are listed in Table 1.

Inner London HCAP = 20%	Hammersmith & Fulham,Kensington & Chelsea, Westminster, Camden, Islington,City & Hackney,Tower Hamlets,Lambeth, Lewisham,Southwark
Outer London HCAP = 15%	Brent, Ealing, Harrow, Hillingdon, Hounslow, Barnet, Enfield, Haringey,Barking & Dagenham, Havering, Newham, Redbridge, Waltham Forest, Bexley, Bromley, Greenwich, Croydon,Kingston,Richmond & Twickenham, Sutton & Merton
Fringe London HCAP = 5%	Basildon, Dartford, Gravesham & Swanley, Billericay, Brentwood & Wickford, Epping Forest, Harlow, Thurrock, Dacorum, Hertsmere, Royston, Buntingford & Bishop Stortford, Southeast Hertfordshire, St Albans & Harpenden, Watford & Three Rivers, Welwyn & Hatfield, Bracknell Forest, Slough, Windsor & Maidenhead, Wokingham, East Elmbridge & Mid Surrey, East Surrey,Guilford & Waverly, North Surrey, Surrey Heath & Woking

Table 1: London Boroughs included in High Cost Payment Areas (NHS Terms and Conditions of Service Handbook, 2023)

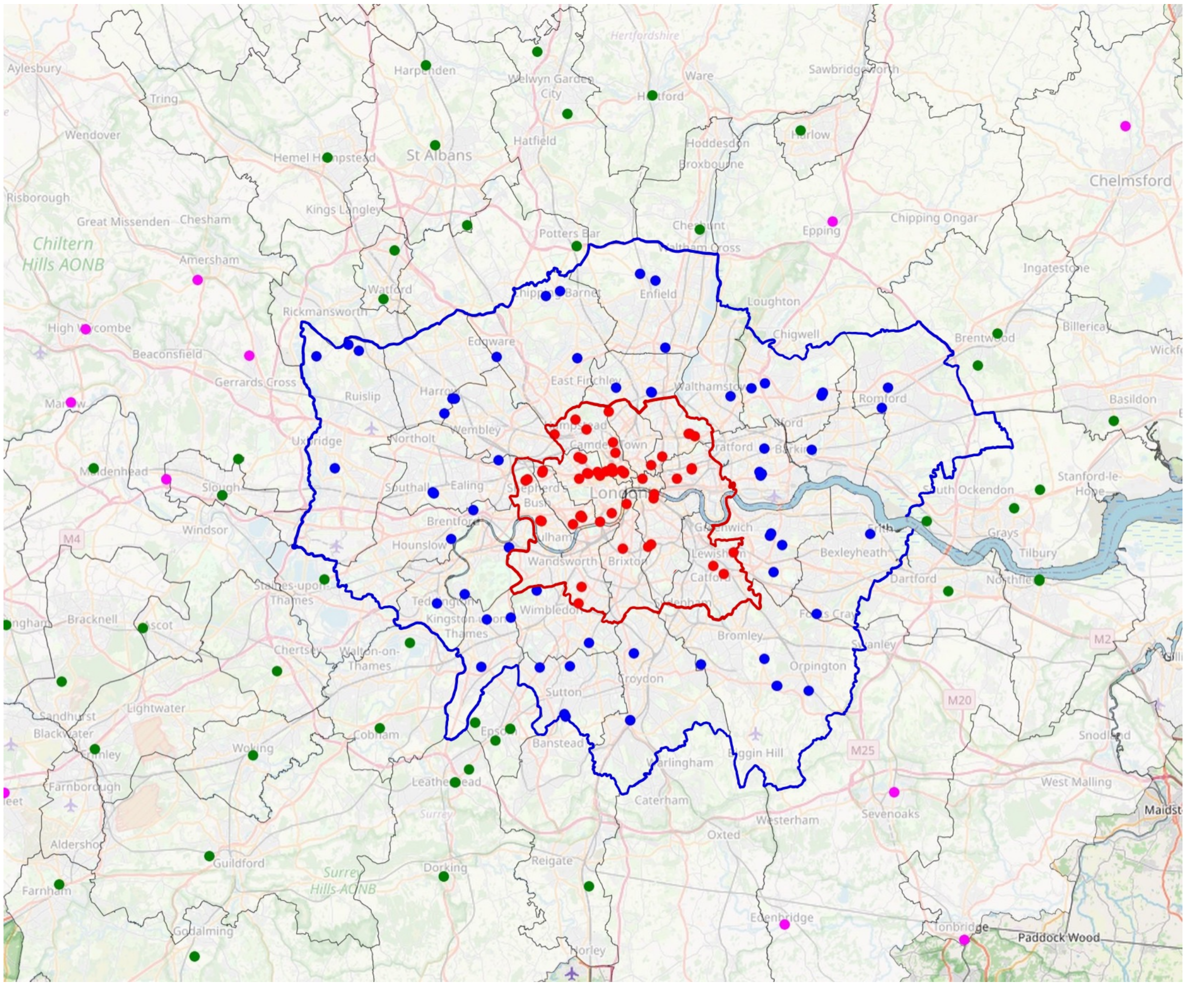


Figure 1: Hospitals in London. Red, Blue, Green and Magenta dots denote hospitals in inner,outer,fringe and non London respectively.

The locations of all NHS hospitals in inner, outer and fringe London and their respective pay zones are shown in figure 1. Each point in figure 1 represents an NHS hospital. All clinics, general practitioners' offices (GPs) and private hospitals in this area are excluded. The hospitals are colour coded according to their respective PCT zone. Notably red, blue, green and magenta dots denote hospitals that fall in the inner, outer, fringe and non-London zones respectively. Figure 1 illustrates the existence of a sharp cut-off point at the inner-outer (IO) and outer-fringe (OF) border, represented by the blue and red lines respectively. At each border wages offered by hospitals, controlling for job title, should drop by the difference of the HCAP. This environment provides a perfect natural experiment to examine the effect of posted wages on vacancy durations at each respective border. This requires that hospitals in zones which necessitate higher HCAP payments operate under similar economic conditions to those that have lower HCAP's. Failure to account for such differences would naturally lead to endogeneity and therefore biased estimates. We discuss this approach in more detail in section 6.

3.2 Job Bands

For the rigour of the empirical methodology, it is important to understand the pay structure and progression system for workers in the NHS. To ensure fairness and non-discriminatory practice, workers in the NHS, excluding Doctors, Dentists and Directors, are paid according to a hierarchical band system.

The system has 8 bands, ranging from Band 2 to Band 9, with Band 2 being the least senior and band 9 being the most senior. Band 1 was closed for new applicants as of December 1st, 2018. Each job in the NHS falls into a specific band, which corresponds to a minimum and maximum salary range. Workers within each band have a predetermined period before they are eligible for a pay rise, based on their experience. The band assigned to a job is determined by the level of responsibilities, required skills, and experience. Band 2 positions are entry-level roles that necessitate minimal or no prior experience or qualifications, such as domestic Support Worker, Housekeeping Assistant, Driver, and Nursery Assistant. Band 3 positions typically require some experience or qualifications, including Emergency Care Assistant, Estates Officer, and Support Workers. Band 4 includes positions requiring greater experience and qualifications. Band 5 encompasses jobs that often require a degree or equivalent qualification, such as Nurses or Paramedics. Band 6 positions involve higher levels of responsibility and expertise, such as specialist Nurses or Physiotherapists. Band 7 comprises managerial or advanced practice roles, such as Nurse Consultants or advanced Physiotherapists. Band 8 is further divided into four sub-bands (8a, 8b, 8c, and 8d) and includes senior managerial roles. Band 9 represents the highest level of responsibility and expertise.

The pay bands are shifted upwards by varying amounts each financial year, which runs from April to March, to accommodate inflation. The pay ranges of each band are defined in the NHS Handbook (2023), which provides clear guidelines for remuneration and progression within the band system.

3.3 Job Titles

The literature in this field, as discussed in section 2, places key importance on the use of job title heterogeneity controls in the estimation of the relationship between posted wages and vacancy durations. The job title controls We propose are informed by the categorisation of jobs the NHS Health Careers website. The jobs are characterised as; Mental Wellbeing Professionals, Domestic Services, Estate Services, Nurses, Midwives, Doctors, Dental Team, Administration, Support Services, Allied Health Professionals, Clinical Support Staff, Healthcare Support Staff, Pharmacists, Research and Science, Corporate Services and Management.

Mental Wellbeing professionals focus on providing mental health and wellbeing support to patients. They include Psychologists, Psychiatrists, Psychotherapists, and Counsellors who work with individuals experiencing mental health issues. Domestic Services categorise staff that are responsible for maintaining cleanliness and hygiene in healthcare facilities. They ensure that hospitals and clinics are kept clean, sanitised, and well-maintained to provide a safe environment for patients and staff. Estate services staff manage and maintain the physical infrastructure of healthcare facilities, such as hospitals and clinics. They handle tasks related to building maintenance, repair, security, and logistics. Nurses are responsible for providing medical care, administering medications, monitoring patients' conditions, and supporting doctors and other healthcare professionals in various settings. Midwives specialise in providing care to expectant mothers before, during, and after childbirth. They assist with prenatal care, childbirth, and postnatal support to ensure the well-being of both the mother and the baby. Doctors, also known as physicians, diagnose and treat medical conditions, prescribe medications, order tests, and perform medical procedures. Each doctor works in a particular speciality, such as general medicine, surgery, paediatrics or cardiology. The dental team includes dentists, dental hygienists, and dental nurses who focus on oral health care. They diagnose and treat dental issues, perform dental cleanings, and educate patients about oral hygiene. Administrators handle tasks related to scheduling appointments, managing patient records, billing, and other administrative duties. Support services staff assist in various areas of patient care and hospital operations, providing help with non-medical tasks like patient transportation, portering, and general assistance. This category also includes Technical Engineers and Information Technology (IT) Assistants. Allied Health Professionals encompasses a diverse group of specialised professionals who support patient care in various ways, such as Physiotherapists, Occupational Therapists, radiographers, and audiologists. Clinical support staff assist Allied Health Professionals in delivering patient care. They may include Medical Assistants, Phlebotomists, and other roles that provide support during medical procedures and tests. Healthcare support staff assist patients. This category may include Healthcare Assistants, Nursing Assistants, and other roles that support patients' daily needs. Pharmacists are responsible for dispensing medications, providing medication-related advice to patients and healthcare professionals, and ensuring the safe and appropriate use of drugs. Research and Science involves professionals engaged in medical and scientific research, including clinical researchers, laboratory technicians, and scientists. They contribute to advancing medical knowledge and developing new treatments. Finally, corporate services and management handle various administrative, financial, and strategic functions within the NHS. They oversee policy development, resource allocation, and overall management of healthcare organisations. Jobs are assigned to particular categories according to their job title and job description. We discuss this in more detail in section 5.4.

4 Theoretical Motivation

In this section, We present the predictions of directed search theory, outlining the theoretical relationship between posted wages, job applications and vacancy durations. We adapt a directed search model, first derived by Shimer (2005), to introduce exogenous wage setting, homogenous firms, a random exclusion parameter and geographical frictions to mimic the structure of the NHS.

4.1 Background

As detailed in section 2, existing literature indicates that, in cases where econometric specifications cannot adequately address firm and job level heterogeneity, a positive correlation between wages and the duration of job vacancies is estimated. Faberman and Menzio (2018) utilise a model, first derived by Shimer (2005), to explain the mechanisms that drive this observed relationship. We summarise their explanation as follows. Imagine a sample of firms, each of which is categorised as either 'sensitive' or 'regular'. Additionally, consider a pool of workers comprising high and low productivity individuals.[3] By definition, sensitive firms derive more substantial benefits from hiring a high-productivity worker compared to a low-productivity counterpart, relative to regular firms. In a model of endogenous wage adjustment, sensitive firms therefore, offer higher wages to attract a superior applicant pool. This

[3]This variation in worker productivity could stem from differences in work ethic, inherent aptitude, interpersonal abilities, and other personality traits Cubel et al. (2016)

also reduces the likelihood that low productivity candidates apply to sensitive firms, given the greater chance of being overlooked in favour of high productivity candidates. If low productivity workers yield lower output in sensitive firms compared to regular firms, Faberman and Menzio (2018) demonstrate that this dynamic leads to an equilibrium where sensitive job positions, characterised by higher wages, receive fewer applications and experience longer vacancy durations.

Empirical studies using vacancy data from diverse firms that do not account for job and firm heterogeneity are thus likely to identify this particular relationship. Nevertheless, Faberman and Menzio (2018) acknowledge that if they were able to control for more detailed job titles, the empirical relationship between wages and vacancy durations would be negative.

We present an adapted version of the framework presented by Faberman and Menzio (2018) to incorporate homogenous firms, exogenous wage setting, a random exclusion parameter and geographical frictions for job seekers. This allows for the analysis of job search dynamics in the context of large, monopolistic firms with a strictly defined pay progression structure.

4.2 The Model

Assume two sets of firms F_1 , F_2 such that $F := F_1 \cup F_2$ and $F_1 \cap F_2 = \emptyset$. Each firm has a single, job vacancy. All firms in F and their corresponding job vacancies, are identical in terms of their relative productivity [4]. Assume further that there exists a set of unemployed workers A which are decomposed into 'high' type $A_h \subseteq A$ and 'low' type workers $A_l \subseteq A$ such that $A := A_h \cup A_l$. Note that $A_h \cap A_l = \emptyset$. We denote $|A_i| := \mu_i \; \forall i \in \{l, h\}$. A firm of type $j \in \{1, 2\}$ matched with a worker of type $i \in \{l, h\}$ will produce output y_{ij}. We assume that agent $a_h \in A_h$ is more productive than $a_l \in A_l$ such that that $y_{lj} < y_{hj}, \forall j \in \{1, 2\}$. Each firm has one open vacancy. The starting wage of the vacancy for firm of type j $\bar{w_j}$ is the same for both high and low type workers. The set of wages $(\bar{w_1}, \bar{w_2})$ are defined exogenously by some, non-utility maximising social planner.[5] We strictly impose that $\bar{w_1} > \bar{w_2}$. Therefore, firms in F_1 and F_2 can be considered to be analogues of hospitals in a high and low HCAP zone respectively.

The model is set in discrete time such that $t \in \{1, 2, 3, ...\}$. The sequence of the directed search is as

[4] Each firm has a single job. For this reason We use the term "firm" and "job" interchangeably.

[5] The concern of this model is not to evaluate aggregated job market efficiency or matching effectiveness. Therefore, this assumption simplifies the analysis without comprising the validity of the outcomes.

follows. Firstly, the exogenous wages are set by the non-utility maximising social planner. Workers then observe the set of posted wages and choose the probability of applying to each firm. This is consistent with other models of directed search that account for coordination frictions (Shimer 2005). An alternative approach would be to solve the model for pure strategy Nash equilibria in which each individual applies to separate firms with deterministic strategy selection. Such assignment models, with perfect communication between job seekers, are explored by Koopmans and Beckmann (1957) and Shapley and Shubik (1971). Despite producing Pareto optimal assignments, Shimer (2005) argues that, in large labour markets, well designed coordination efforts by job seekers are impractical. Therefore the absence of coordination frictions in theoretical frameworks produces results that do not resonate with empirical observations, the most notable being that vacancies and unemployment coexist in modern labour markets. Most models in the literature do not employ asymmetric strategy selection strategies for this reason.

We assume that an individual of type $i \in \{l, h\}$ applies to a firm of type $j \in \{1, 2\}$ with a probability p_{ij}. Job seekers then apply to a given firm after drawing their selection from their probability distribution. Firms observe all job applications and whether they are of a high or low type, with perfect information. If the firm has a collection of high and low type applicants, then it will immediately reject all low type workers and randomly select a high type worker. In the case where they receive only applications from low type workers, they select one randomly. If they receive no applications, then the vacancy remains unfilled. Applicants of type i who are selected for a job j are paid a wage $\bar{w_j}$ less some parameter, g_{ij} which simply denotes the costs associated with travelling to work. These can be thought of as transport expenses, the opportunity cost of leisure, or simply disutility derived from commuting for job seekers of type i. As such this parameter imposes geographical frictions into the model which is a critical aspect in understanding job dynamics in London. We further assume that there is some probability that an application to a firm of type j from a job seeker of type i is randomly rejected, α_{ij}. This parameter captures unobserved, stochastic preferences of the employer, which do not relate to the workers level of output. Such preferences are described in equilibrium search models that explain job market discrimination Black (1995). For the sake of simplicity, We assume that the value of the parameter is common knowledge among all job seekers.

4.3 Special Case: $|F_j| = 1, \forall j \in \{1,2\}$

We can show the intuitions of this framework in the special case of the model described above without any loss of generality. For this purpose, We assume that there exists one firm in each set F_1 and F_2. To simplify the analysis We also assume that $\alpha_{1i} = \alpha_{2i} = \alpha_i \ \forall \ i \in \{l,h\}$

Proposition 1. *In the case where* $|F_j| = 1, \forall j \in \{1,2\}$*,* $\alpha_{1i} = \alpha_{2i} = \alpha_i \ \forall \ i \in \{l,h\}$ *and* $g_{ij} = 0$ $\forall i \in \{l,h\}, j \in \{1,2\}$*, in a symmetric mixed strategy Nash equilibrium,* $p_{1l} + p_{1h} > p_{2l} + p_{2h}$ *such that* $q_{1l} + q_{1h} > q_{2l} + q_{2h}$*.*

Proof. A job seeker of type i applies to the job j with a probability p_{ji}. Therefore, the number of applications a firm j receives from seeker i follows a Poisson distribution with an average[6]:

$$q_{ji} = p_{ji}\mu_i \tag{1}$$

Therefore the average number of applications for firm j from worker i, that are not randomly rejected with probability α_i, is given by:

$$(1-\alpha_i)q_{ji} = (1-\alpha_i)p_{ji}\mu_i \tag{2}$$

We consider the decentralised mixed strategy Nash equilibrium of this game where the job seeker i chooses $p_{ji} \ \forall j \in \{1,2\}$ to maximise their expected utility. Job seekers within a given type i are identical and therefore, follow symmetrical selection strategies.

Expected Utility

Both $i \in \{l,h\}$ are considered to be risk neutral such that their utility is linearly increasing in their wage. We firstly consider job seekers of type h. Following Shimer (2005), We note that the probability that there are $z \in \{0,1,2,...\}$ other h type workers applying for job j, who are not randomly excluded, is given by a Poisson distribution probability mass function:

$$f\left(z,(1-\alpha_i)q_{ji}\right) = \frac{1}{z!}\left((q_{jh}(1-\alpha_h))^z e^{(\alpha_h-1)q_{ih}}\right) \tag{3}$$

[6] In this analysis workers anticipate the number of other workers that will apply to a given vacancy. Therefore each respective worker assumes an average queue length of applications, other than themselves, of $p_{ij}(\mu_i - 1)$. To simplify notation We assume that μ_h is large meaning that We can denote $p_{ij}(\mu_i - 1) \approx p_{ij}\mu_i$

The probability of a h type worker being offered job j, conditional on applying and given that there are z other h type applicants, is $1/(z+1)$. Therefore, the probability of being hired conditional on applying, for firm j is given by [7]:

$$\sum_{z=0}^{\infty} \frac{1}{(z+1)} \frac{1}{z!} \left((q_{jh}(1-\alpha_h))^z e^{(\alpha_h - 1)q_{ih}} \right) = \sum_{z=0}^{\infty} \frac{1}{(z+1)!} \left((q_{jh}(1-\alpha_h))^z e^{(\alpha_h - 1)q_{ih}} \right) \quad (4)$$

Note that the probability that firm j receives no applications from, or randomly rejects all, workers of type h is given by $f(z=0,(q_{jh}(1-\alpha_h)) = e^{(\alpha_h-1)q_{ih}}$. Therefore equation (4) can be simplified as follows[8]:

$$\sum_{z=1}^{\infty} \frac{1}{z!} \left((q_{jh}(1-\alpha_h))^{z-1} e^{(\alpha_h - 1)q_{ih}} \right) = \frac{1}{(1-\alpha_h)q_{jh}} \sum_{z=1}^{\infty} \frac{1}{z!} \left((q_{jh}(1-\alpha_h))^z e^{(\alpha_h - 1)q_{ih}} \right)$$

$$= \frac{1 - f(z=0, e^{(1-\alpha_i)q_{ji}})}{(1-\alpha_h)q_{jh}} \quad (5)$$

$$= \frac{(1 - e^{(\alpha_h - 1)q_{jh}})}{(1-\alpha_h)q_{jh}}$$

If the workers application is accepted, they achieve a utility of $\bar{w}_j - g_{jh}$, and zero otherwise. Therefore, the expected utility of a high type worker $E[u_h]$ can be expressed as:

$$E[u_h] = \sum_{j \in \{1,2\}} \frac{(1 - e^{(\alpha_h - 1)q_{jh}})}{(1-\alpha_h)q_{jh}} (1-\alpha_h)(\bar{w}_j - g_{jh})p_{jh} = \sum_{j \in \{1,2\}} \frac{(1 - e^{(\alpha_h - 1)q_{jh}})}{\mu_h} (\bar{w}_j - g_{jh}) \quad (6)$$

Now consider workers of type l. The probability of being offered a job depends on the decisions of high type workers, since they are strictly preferred to the low type by the firm. The probability that a worker of type l faces no competition from high type workers is:

$$f(z=0, e^{(1-\alpha_i)q_{ji}}) = 1 - \sum_{z=1}^{\infty} \frac{1}{z!} \left((q_{jh}(1-\alpha_h))^z e^{(\alpha_h - 1)q_{ih}} \right) = e^{(\alpha_h - 1)q_{jh}} \quad (7)$$

[7]To show this consider: $\frac{1}{z+1}\frac{1}{z!} = \frac{1}{z+1}\frac{1}{z(z-1)(z-2)\dots 1} = \frac{1}{(z+1)!}$

[8]The proof is as follows: $\sum_{z=0}^{\infty} \frac{1}{(z+1)!} \left((q_{jh}(1-\alpha_h))^z e^{(\alpha_h-1)q_{ih}} \right) =$
$e^{(\alpha_h-1)q_{ih}} + \frac{1}{2!}(q_{jh}(1-\alpha_h))e^{(\alpha_h-1)q_{ih}} + \frac{1}{3!}(q_{jh}(1-\alpha_h))^2 e^{(\alpha_h-1)q_{ih}} + \dots = \sum_{z=1}^{\infty} \left(\frac{1}{z!}((1-\alpha_h)q_{ih})^{z-1} e^{(\alpha_h-1)q_{ih}} \right)$

The probability that there are $\theta \in (0, 1, 2...)$ other l type workers, who are not randomly excluded, applying for job j is given by a Poisson distribution probability mass function:

$$f\left(\theta,(1-\alpha_i)q_{ji}\right) = \frac{1}{\theta!}\left((q_{jl}(1-\alpha_l))^{\theta}e^{(\alpha_l-1)q_{jl}}\right) \tag{8}$$

In the absence of any h type applicants, the probability of selection in the case of θ other l type workers applying is given by $\frac{1}{1+\theta}$. Therefore, using equations (7) and (8), the probability of the l type worker being hired by firm j can be expressed as:

$$= e^{(\alpha_h-1)q_{jh}}\sum_{\theta=0}^{\infty}\frac{1}{(\theta+1)}\frac{1}{\theta!}\left((q_{jl}(1-\alpha_l))^{\theta}e^{(\alpha_l-1)q_{jl}}\right) \tag{9}$$

$$= e^{(\alpha_h-1)q_{jh}}\sum_{\theta=1}^{\infty}\frac{1}{\theta!}\left((q_{jl}(1-\alpha_l))^{(\theta-1)}e^{(\alpha_l-1)q_{jl}}\right) \tag{10}$$

$$= e^{(\alpha_h-1)q_{jh}}\frac{1}{q_{jl}(1-\alpha_l)}\sum_{\theta=1}^{\infty}\frac{1}{\theta!}\left((q_{jl}(1-\alpha_l))^{\theta}e^{(\alpha_l-1)q_{jl}}\right) \tag{11}$$

$$= \frac{e^{(\alpha_h-1)q_{jh}}}{(1-\alpha_l)q_{il}}(1-e^{-(1-\alpha_l)q_{jl}}) \tag{12}$$

Therefore, it follows that the expected utility of a worker of type l, $E[u_l]$ is given by:

$$E[u_l] = \frac{1}{\mu_l}\sum_{j\in\{1,2\}} e^{(\alpha_h-1)q_{jh}}(1-e^{(\alpha_l-1)q_{jl}})(\bar{w}_j - g_{jl}) \tag{13}$$

The job seeker i aims to maximise their expected utility subject to [9]:

$$\sum_{j\in\{1,2\}} q_{ji} = \mu_i \Rightarrow \sum_{j\in\{1,2\}} p_{ji} = 1 \tag{14}$$

A job seeker of type i's expected utility is increasing in p_{i_j} such that their preferences are monotonic. Therefore, there cannot exist any interior solutions to this optimisation problem meaning that finding the first order conditions of $E[u_i]$ with respect to the respective probabilities is sufficient to solve for optimality. A job seeker of type i chooses their mixed strategies (p_{1i}, p_{2i}) to maximise their expected utility, subject to the condition that the sum of these respective probabilities cannot exceed 1.

[9] we can express the budget constraint as $q_{j1} + q_{j2} = p_{j1}\mu j + p_{j2}\mu_j = \mu_j \Rightarrow p_{j1} + p_{j2} = 1$

High Type

First consider job seekers of type h. Setting the Lagrange multiplier $L_i \ \forall \ i \in \{l, h\}$ equal to:

$$L_h = E[u_h] + \lambda_h(1 - p_{1_h} - p_{2_h}) \tag{15}$$

Taking the first order conditions of L_h with respect to (p_{1h}, p_{2h}) and setting equal to zero yields:

$$\begin{aligned} \frac{\partial L_h}{\partial p_{1h}} &= \frac{\partial u_h}{\partial p_{1h}} - \lambda_h = 0 \\ \frac{\partial L_h}{\partial p_{2h}} &= \frac{\partial u_h}{\partial p_{2h}} - \lambda_h = 0 \end{aligned} \tag{16}$$

This naturally implies that:

$$\lambda_h = \frac{\partial u_h}{\partial p_{1h}} = \frac{\partial u_h}{\partial p_{2h}} \tag{17}$$

Solving:

$$\frac{\partial u_h}{\partial p_{1h}} = \frac{\partial}{\partial p_{1h}} \left[\frac{1 - e^{(\alpha_h - 1)p_{1h}\mu_h}}{\mu_h} (\bar{w}_1 - g_{1h}) \right] = (1 - \alpha_h)(\bar{w}_1 - g_{1h})e^{-(1-\alpha)p_{1h}\mu_h} \tag{18}$$

Using symmetry, it follows that:

$$\frac{\partial u_h}{\partial p_{2h}} = (1 - \alpha_h)(\bar{w}_2 - g_{2h})e^{-(1-\alpha)p_{2h}\mu_h} \tag{19}$$

Therefore, using equation (17) and simple algebra:

$$(1 - \alpha_h)(\bar{w}_1 - g_{1h})e^{-(1-\alpha_h)p_{1h}\mu_h} = (1 - \alpha_h)(\bar{w}_2 - g_{2h})e^{-(1-\alpha_h)p_{2h}\mu_h} \tag{20}$$

$$(\bar{w}_1 - g_{1h})e^{-(1-\alpha)p_{1h}\mu_h} = (\bar{w}_2 - g_{2h})e^{-(1-\alpha)p_{2h}\mu_h} \tag{21}$$

$$\frac{e^{(1-\alpha_h)p_{1h}\mu_h}}{e^{(1-\alpha_h)\mu_h p_{2h}}} = \frac{(\bar{w}_1 - g_{1h})}{(\bar{w}_2 - g_{2h})} \tag{22}$$

$$\ln \frac{(\bar{w}_1 - g_{1h})}{(\bar{w}_2 - g_{2h})} = (1-\alpha_h)p_{1h}\mu_h - (1-\alpha_h)p_{2h}\mu_h \tag{23}$$

$$p_{1h} - p_{2h} = \frac{1}{(1-\alpha_h)\mu_h} \ln \frac{(\bar{w}_1 - g_{1h})}{(\bar{w}_2 - g_{2h})} \tag{24}$$

Rearranging expression (24) and substituting (14) yields:

$$p_{1h} = \frac{1}{2} + \frac{1}{2\mu_h(1-\alpha_h)} \ln (\frac{\bar{w}_1 - g_{1h}}{\bar{w}_2 - g_{2h}}) \tag{25}$$

$$p_{2h} = \frac{1}{2} - \frac{1}{2\mu_h(1-\alpha_h)} \ln (\frac{\bar{w}_1 - g_{1h}}{\bar{w}_2 - g_{2h}}) \tag{26}$$

Low Type

Now consider workers of type l. Analogous to workers of type h they maximise their expected utility (6) subject to the probability constraint (14). Setting up the Lagrangian L_l:

$$L_l = u_l + \lambda_l(1 - p_{1_l} - p_{2_l}) \tag{27}$$

Taking the first order conditions of L_l with respect to (p_{1h}, p_{2h}) and setting equal to zero yields:

$$\frac{\partial L_l}{\partial p_{1l}} = \frac{\partial u_l}{\partial p_{1l}} - \lambda_l = 0 \tag{28}$$

$$\frac{\partial L_l}{\partial p_{2l}} = \frac{\partial u_l}{\partial p_{2l}} - \lambda_l = 0 \tag{29}$$

This naturally implies that:

$$\lambda_l = \frac{\partial u_l}{\partial p_{1l}} = \frac{\partial u_l}{\partial p_{2l}} \tag{30}$$

Solving:

$$\frac{\partial u_l}{\partial p_{il}} = \frac{e^{(\alpha_h - 1)q_{jh}}}{\mu_l}(\bar{w}_j - g_{jl})\frac{\partial u_h}{\partial p_{jl}}(1 - e^{(\alpha_l - 1)\mu_l p_{jl}}) \tag{31}$$

$$= e^{(\alpha_h - 1)q_{jh}}(\bar{w}_j - g_{jl})(1-\alpha_l)e^{(\alpha - 1)\mu_l p_{jl}} \tag{32}$$

$$= e^{(\alpha_h-1)q_{jh}}(1-\alpha_l)(\bar{w}_1 - g_{jl})\mu_l e^{(\alpha_l-1)\mu_l p_{jl}} \tag{33}$$

Therefore, using the equation (33) in the general form and substituting $j \in \{l, h\}$:

$$\begin{aligned} \frac{\partial u_l}{\partial p_{1l}} &= e^{(\alpha_h-1)q_{1h}}(1-\alpha_l)(\bar{w}_1 - g_{1l})e^{(\alpha_l-1)\mu_l p_{1l}} \\ \frac{\partial u_l}{\partial p_{2l}} &= e^{(\alpha_h-1)q_{2h}}(1-\alpha_l)(\bar{w}_2 - g_{2l})e^{(\alpha_l-1)\mu_l p_{2l}} \end{aligned} \tag{34}$$

Rearranging yields:

$$\frac{e^{(1-\alpha_l)\mu_l p_{1l}}}{e^{(1-\alpha_l)\mu_l p_{2l}}} = \frac{(\bar{w}_1 - g_{1l})e^{(\alpha_h-1)q_{1h}}}{(\bar{w}_2 - g_{2l})e^{(\alpha_h-1)q_{2h}}} \tag{35}$$

Given that the budget constraint (14) is binding, this yields:

$$(1-\alpha_l)\mu_l(p_{1l} - p_{2l}) = \ln\frac{(\bar{w_1} - g_{1l})}{(\bar{w_2} - g_{2l})} + \ln\frac{e^{(\alpha_h-1)q_{1h}}}{e^{(\alpha_h-1)q_{2h}}} \tag{36}$$

$$(1-\alpha_l)\mu_l(p_{1l} - p_{2l}) = \ln\frac{(\bar{w_1} - g_{1l})}{(\bar{w_2} - g_{2l})} + \mu_h(\alpha_h - 1)(p_{1h} - p_{2h}) \tag{37}$$

Substituting (24):

$$(1-\alpha_l)\mu_l(p_{1l} - p_{2l}) = \ln\frac{(\bar{w_1} - g_{1l})}{(\bar{w_2} - g_{2l})} - \ln\frac{(\bar{w_1} - g_{1l})}{(\bar{w_2} - g_{2l})} \tag{38}$$

Mixed Strategy Mixed Equilibrium

Therefore, to summarise, the solution to this model can be characterised by the following expressions:

$$p_{1h} - p_{2h} = \frac{1}{\mu_h(1-\alpha_h)}\ln\left(\frac{\bar{w}_1 - g_{1h}}{\bar{w}_2 - g_{2h}}\right) \tag{39}$$

$$p_{1l} - p_{2l} = 0 \tag{40}$$

Firstly assume that $g_{ji} = 0 \ \forall j \in \{1, 2\}, i \in \{l, h\}$. As the wages $(\bar{w}_1, \bar{w}_2)$ are exogenously defined, such that $\bar{w}_1 > \bar{w}_2$, implying that $ln(\frac{w_1}{w_2}) > 0$, one can deduce that for $\bar{w}_j > 0 \; \forall j \in \{1, 2\}$, $p_{1h} > p_{2h}$ and $p_{1l} = p_{2l}$. Therefore, given that the average number of applications for firm j is given by (1) it follows that:

$$q_{1h} + q_{1l} > q_{2l} + q_{2h} \tag{41}$$

This implies that the average queue length for firm 1 is greater than it is for firm 2. The effect of this on the expected vacancy duration for job j, τ_j is shown in proposition 2.

□

Proposition 2. *In the case where* $|F_j| = 1, \forall j \in \{1, 2\}$, $\alpha_{1i} = \alpha_{2i} = \alpha_i \; \forall \; i \in \{l, h\}$ *and* $g_{ij} = 0$ $\forall i \in \{l, h\}, j \in \{1, 2\}$. *In a symmetric mixed strategy Nash equilibrium* $E[\tau_1] < E[\tau_2]$.

Proof. Denote ϕ_i as the probability that a firm of type j receives at least one job application in period t [10]. It must be the case that the average vacancy duration for a firm of type j, $E[\tau_j]$ is given by:

$$E[\tau_j] = {\phi_j}^{-1} \tag{42}$$

Using equations (7) and (8), ϕ_j can be expressed as:

$$\begin{aligned} \phi_j &= 1 - f(z = 0, e^{-(1-\alpha_h)q_{ih}}) \cdot f(\theta = 0, e^{-(1-\alpha_h)q_{il}}) \\ &= 1 - e^{-(1-\alpha_h)q_{ih}} \cdot e^{-(1-\alpha_h)q_{il}} \\ &= 1 - e^{-(1-\alpha_h)(q_{ih}+q_{il})} \end{aligned} \tag{43}$$

Therefore using chain rule:

$$\begin{aligned} \frac{\partial \phi_j^{-1}}{\partial(q_{jh} + q_{jl})} &= -(1 - e^{-(1-\alpha_h)(q_{jh}+q_{jl})})^{-2}((1 - \alpha_h)(e^{-(1-\alpha_h)(q_{jh}+q_{jl})})) \\ &= -(1 - \alpha_h)(e^{-(1-\alpha_h)(q_{jh}+q_{jl})})^{-1} < 0 \end{aligned} \tag{44}$$

Therefore in the case where $|F_j| = 1, \forall j \in \{1, 2\}$, $\alpha_{1i} = \alpha_{2i} = \alpha_i \; \forall \; i \in \{l, h\}$ and $g_{ij} = 0 \; \forall i \in \{l, h\}, j \in \{1, 2\}$ and using proposition 1 it is clear that:

$$E[\tau_1] < E[\tau_2] \tag{45}$$

[10] Note that, in each period, the game is reset such that all workers are separated from their matched job position. Therefore, ϕ_{jt} is constant for all t such that it can be denoted ϕ_j.

□

4.4 Implications

The model presented in the section above, where homogeneous firms and agents with heterogeneous productivity interact, suggests that, in the absence of geographical frictions, firms who offer higher wages attract more applicants. This is driven by increases in the number of applicants from type h applicants. Notably, workers of type l, anticipating that the high wage job will attract more high skilled applicants, assign an equal probability of applying to firms 1 and 2. Given that the average vacancy duration is theoretically inversely proportional to the number of applicants, this would suggest that firms who offer higher wages should experience lower vacancy durations. This is consistent with labour market search models such as Moen (1997), who suggest that firms face a trade-off between higher wages and search costs when choosing the wage to offer. Furthermore, the theoretical results support the work of Faberman and Menzio (2018) who find that posting higher wages can help firms to attract applicants of a higher calibre. This may suggest that the NHS could use higher wages as a mechanism to increase the number of applicants to vacant positions which may ease pressures on labour shortages in the short run.

The model also highlights some issues that arise in the case of analysing vacancy durations and posted wages in the context of geographical immobility. Notably if the costs associated with travelling to firm 1 are sufficiently high, then high type workers assign a higher probability of application to the lower paying firm. For the purpose of illustration, consider workers of type h whose mixed strategies in equilibrium are characterised by expressions (14) and (39). In the case where g_{2h} is sufficiently high such that $g_{2h} < g_{1h}$, it could be such that $w_2 - g_{2h} < w_1 - g_{1h}$ such that $p_{1h} < p_{2h}$. Therefore, firms who are sufficiently costly to access may suffer from higher search costs and fewer applications from high type workers, even if they offer higher wages. The econometric implications of this are clear. When geographical frictions cannot be adequately controlled for, methodologies employing RD approaches should utilise hospitals where there are sufficiently few of the aforementioned frictions. This allows for the effects of wages and geographical immobility on job vacancy durations to be more accurately disentangled.

5 Data

5.1 Overview

In this section We discuss the data set that forms the basis of empirical methodology outlined in the successive sections. This research utilises job vacancy data collected weekly by Dr. Kerry Papps from the online job posting website *findajob.co.uk*. The website is managed by the UK governments Department for Work and Pensions (DWP) and serves as a platform for all UK companies to advertise their vacant positions, while also providing job seekers with free access to search for jobs. The service allows companies advertise job vacancies with accompanying details such as job title, job description, location, and salary. This user-friendly interface facilitates efficient job searching for both employers and job seekers, reducing the costs associated with job search for both. An example job advert from *findajob.co.uk* is provided in appendix C.

The data offers distinct advantages due to the inclusion of a wide range of variables that are not typically available in conventional survey data. Beyond standard variables like wage rate and job role, this vacancy data comprises additional information such as the exact job description, weekly working hours, precise geographical location of the job, and the date of posting. Consequently, this novel data set provides a deeper level of insight into the impacts of wages on the labour market compared to other commonly used datasets like the Labor Force Survey, Annual Survey of Hours and Earnings, or the Apprentice Pay Survey. Given the increasing reliance of job seekers on the internet for job searches, this data is especially relevant, Kuhn and Mansour (2014).

The use of online job vacancy data is especially useful for three reasons. Firstly, it provides information on how long a firm opens a vacancy for, as well as if the job was taken down prior to its predetermined closing date. Secondly, rather than containing the actual wages paid, the adverts contain the wage posted by the firm. While there is no reason to expect significant differences between the two, the use of the latter allows for the behaviour of job seekers in response to the offered wage to be more accurately deduced. Finally, the job descriptions allow for jobs to be assigned to job categories much more easily which bypasses some issues associated with creating sufficiently detailed job title controls Faberman and Menzio (2018).

The data set includes data from 2020, however, We reduce this time period to April 2021 to March 2023 to mitigate the distortionary effects of Covid-19 lockdowns and the resulting social distancing measures during the preceding financial years, 2020-2021 and 2021-22. The data includes all job vacancies posted during the sample period on *findajob.co.uk* in the UK, and is therefore not exclusively comprised of NHS jobs. However, the NHS makes up a significant proportion of job vacancies on the website. From the collected data, We use the job title, posted wage, date the vacancy was posted and closed, name and postcode of the employer, and job description. To bolster the validity of the analysis We only consider hospitals in inner, outer and fringe London, excluding the FN border. We do so due to the poor balance of hospitals on this border. The implications of this concern will become increasingly clear in our discussion of the OF border in successive sections.

5.2 Job Vacancy Duration

The duration of posted vacancy i, denoted as τ_i, is calculated using the opening date of the vacancy τ_o , the date it was scraped from the website τ_s , and its closing date τ_c, which is predetermined by the employer and mentioned in the job advertisement.

In the construction of this variable, We note that some job adverts in the raw data remain unfilled in their original posting and are consequently reopened. This is identifiable as the advertisement number, which is recorded by *findajob.co.uk*, has duplicate observations whose opening and closing dates do not overlap. This implies that the job is reopened, which can be assumed to be because the job remained unfilled at the closing date of the original advert. In the case where there is an advert with a duplicate advert number, job title and job description where the closing and opening date do not overlap, We combine the observations into one by keeping the opening date of the earliest job advertisement as well as the scraping and closing date of the advert with the latest closing date.

Simply using the difference between the opening date and closing date $\tau_o - \tau_c$ to calculate the τ_i would only account for employers, and not job seekers, behaviour and would not be particularly interesting. Therefore We adjust the calculation of τ_i to account for job adverts that were closed early before their predetermined closing date. We do so by utilising the structure of the data scraping process.

The job adverts were scraped weekly. Consequently, it can be inferred that the vacancy was closed earlier than planned if $\tau_c - \tau_s > 7$. τ_i was then defined as follows:

$$\tau_i = \begin{cases} \tau_s - \tau_o + 7, & \text{if } \tau_c - \tau_s > 7 \\ \tau_c - \tau_o, & \text{otherwise} \end{cases} \tag{46}$$

In doing so We account for job vacancies that are taken down early which better reflects job seekers behaviour in response to posted wages. In the absence of this more robust dependent variable, the dependent variable would be more reflective of employers job posting behaviour which is outside the scope of this research.

5.3 Band Controls

Most NHS job vacancies listed on *findajob.co.uk* present the salary as a range, for example £40,000 - £50,000. To simplify the analysis, We calculate the midpoint of the salary range and report it as the posted wage ω_i for each job vacancy. The job band of a particular vacancy was determined by cross-referencing the exogenous minimum and maximum wages associated with each band with the posted wage. While a given vacancies band would ideally be derived from the job title or description, this information was too infrequently included in the data to be reliably used. However, the variable's construction does not pose any issues for the analysis, as workers in the NHS are legally required to be paid according to centrally determined pay scales.

5.4 Job Title Controls

A significant portion of this analysis requires the assignment of each job vacancy i to a particular job category, as defined in section 3.3. Similar to Abi Adams (2020) We take a machine learning approach, as follows.

1. Manually assign job vacancies to the defined job categories in section 3.3 according to key words and phrases included in the job title.
2. Split the manually assigned job vacancies into an 80\20 train\test split.
3. Train a machine learning model on the manually assigned data to classify job vacancies, that cannot be assigned manually, to each job category.
4. Run the model on the job vacancies that were not assigned manually.

Manual Assignment

We label job vacancies manually using key terms in the job title of each vacancy. We use the *strpos* command in STATA to assign job vacancies to the categories defined in section 3.3.

Given the broad scope of job titles in the NHS, manual assignment only accounts for approximately 85% of the observations. Therefore, We employ a machine learning model to assign the job vacancies, that are not matched manually to the predefined job category, using the job description. We divide the manually assigned job vacancies, along with their corresponding job description, into a 80\20 train\test split. We then train a Bayesian classifier machine learning text classification model on 80% of the manually assigned job vacancies.

The Model

The model We employ for the purpose of text classification task is a Multinominal Bayesian classifier (MNBC). We summarise the explanations of MNBC by Rennie et al. (2003) and Kibriya et al. (2005). Let us first assume that are $k \in \{1, ..., K\}$ types of job category C. The MNBC starts by vectorising each job description, $d \in D$, into individual words. Defining n as the number of unique words that are collected following vectorisation, then $\boldsymbol{x} = (x_1, .., x_i, .., x_n)$ is a vector containing the frequency of occurrence for every word in a given job description d. Note that We exclude "stop words" to improve the model's predictive power. This model is interested in determining the probability that a given job vacancy is classified by a particular job category k, C_k, given the collection of words in the job description $\boldsymbol{x}$. Using Bayes theorem, this posterior probability is given in equation (47).

$$P(C_k|\boldsymbol{x}) = \frac{P(C_k)P(\mathbf{x}|C_k)}{P(\mathbf{x})} \tag{47}$$

Each of the job categories are assumed to be equally probable such that the prior probabillity can be expressed as:

$$P(C_k) = \frac{1}{K} \tag{48}$$

There are different approaches to the calculation of $p(\mathbf{x}|C_k)$ however, the most common for the purpose of text classification in the Multinomial Bayesian method [11]. Under a multinomial event model, the

[11]Note that $P(C_k)$ and $P(\boldsymbol{x})$ are constant for each $d \in D$ and $k \in \{1, ..., K\}$. Therefore We only consider the calculation of $p(\boldsymbol{x}|C_k)$ in the estimation of (47).

likelihood of observing a particular histogram $\boldsymbol{x}$ in the job description of a particular vacancy is given by:

$$P(\boldsymbol{x}|C_k) = \frac{(\sum_{i=1}^{n} x_i)!}{\prod_{i=1}^{n} x_i!} \prod_{i=1}^{n} p(x_i|C_k)^{x_i} \tag{49}$$

Estimating the first term of (49) as the prior probability $P(C_K)$ and log linearising yields.

$$P(\mathbf{x}|C_k) \propto \log\left(\frac{(\sum_{i=1}^{n} x_i)!}{\prod_{i=1}^{n} x_i!} \prod_{i=1}^{n} p(x_i|C_k)^{x_i}\right)$$

$$P(\mathbf{x}|C_k) \propto \log\left(\frac{(\sum_{i=1}^{n} x_i!)}{\prod_{i=1}^{n} x_i!}\right) + \sum_{i=1}^{n} x_i \log(p(x_i|C_k)) \tag{50}$$

$$P(\mathbf{x}|C_k) \propto \log\left(P(C_k)\right) + \sum_{i=1}^{n} x_i \log(p(x_i|C_k))$$

The conditional probabilities $p(x_i|C_k)$ can be estimated using the training data. We denote $F_{x_i,k}$ and $F_{x_N,k}$ as the count of word x_i in the training data belonging to class k and the total number of occurrences of x_i in the training data across all classes respectively. $p(x_i|C_k)$ can then be estimated using:

$$P\widehat{(x_i|c)} = \frac{1 + F_{x_N,c}}{n + \sum_{x=1}^{N} F_{x_i,c}} \tag{51}$$

The use of expression (51) is known as Laplace smoothing (Chen and Goodman, 1999) and is used to avoid the zero frequency problem, where, in its absence, words that are not observed with a particular class in the training data would be assigned a posterior probability of zero. This would cancel the whole conditional probability to zero, dramatically reducing the reliability of the estimator.

Equation (47) in combination with (48),(50) and (51) therefore, derives the posterior probability that a particular vacancy will be of class C_k, $\forall k \in \{1, ..., K\}$. The Bayes classifier then requires a decision rule that assigns classes $\hat{y} = C_k$ to a particular job vacancy. A common approach is to simply assign a class to the job vacancy which has the highest probability of being correct, given the histogram of observed vectorised words in a given job description $\mathbf{x}$. This approach is known as a maximum a posteriori (MAP) decision rule. Therefore, job categories are classified according to function (52).

$$\hat{y} = \underset{k \in \{1,...,K\}}{\arg\max} \left(\log(p(C_k)) + \sum_{i=1}^{n} x_i \log(p(x_i|C_k)) \right) \tag{52}$$

There is some debate in the literature as to how reliable a Bayesian classifier is, given the restrictive conditional independence assumptions that the model is based on. However, for text classification purposes, this model is often effective. For a comprehensive discussion of this literature see; Aggarwal and Zhai (2012), Lewis (1992), Zhang and Li (2007), Ying et al. (2021), Kim et al. (2006). We list the precision, recall and F1 test statistics to report on the effectiveness of the text classification model.

Training and Testing

The MNBC model We discuss above is first trained on 119,418 observations. It vectorises each of the job descriptions of the manually assigned vacancies and determines the prior and posterior probabilities, $p(C_k)$ and $p(\boldsymbol{x}|C_k)$, using the multinomial events model. These probabilities will be used in the calculation of (50) for each observation in the data set and $k \in \{1, ..., K\}$.

After being trained on the 80% split of the manually assigned data, the model is then tested on the remaining 20% of the remaining manually assigned data. Following Abi Adams-Prassl (2020), We report three statistics, namely the Precision,Recall and F1 score that is derived from the tested data. Noting that TP, FP and FN denotes the number of true positives, true negatives and false negatives respectively, the equations for each of these metrics is given below:

$$Precision = \frac{TP}{TP + FP} \tag{53}$$

$$Recall = \frac{TP}{TP + FN} \tag{54}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{55}$$

The respective accuracy, precision and F1 score for the test data is 0.87, 0.76, 0.81 respectively. We deem this sufficiently accurate to proceed. Once the model has been trained and tested, the MNBC is used to classify the jobs, that were not assigned manually, according to the MAP rule (52).

5.5 Hospital Location Data

This project also relies on geographical data for each of hospitals included in the analysis as well as the PCT zone boundaries. Each hospital's coordinate was attained using Google Maps. They were then converted from the WGS84 coordinate system to ESPG:3004 so each hospital's distance from each respective PCT zone boundary could be measured in meters and not degrees. [12] Simultaneously each hospitals postcode was recorded. The location of each job location was derived from the postcode included in the job vacancy advert. The hospital distance data was then merged with the job vacancy data to determine each job vacancies distance from each PCT boundary. We plot a histogram demonstrating the distribution of hospitals at the IO and OF borders in figure 2. All hospitals in inner, outer and fringe London are referenced in appendix A.

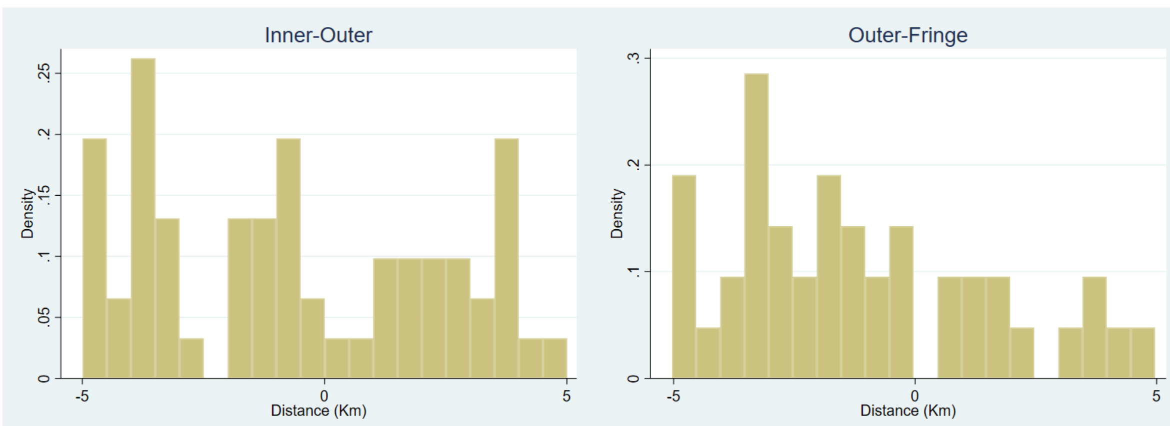


Figure 2: Histogram illustrating the distribution of hospitals at the IO and OF borders.

5.6 Summary Statistics

We outline the summary statistics, including the sample mean of the posted wage and the job posting duration categorised by the respective PCT zone, bands and job titles, in table (2). Across all PCT zones and jobs there are 140,495 observations. The average posted wage and posting length is £44,379.95 and 15.65 days respectivly. The first collection of rows report the number of observations, average wage and vacancy duration for each band within each PCT zone. The number of observations for each band within each PCT boundary follows an approximate normal distribution in which band 5

[12]QGIS software was then used to calculate each hospitals distance to the IO and OF border in kilometres.

and 6 are the most common, with band 2 and band 9 making up much fewer observations. This holds for each PCT zone.

	Inner			Outer			Fringe		
	N	Wage	Duration	N	Wage	Duration	N	Wage	Duration
Band2	3117	25084.2	14.8	3046	24099.2	15.3	2694	21207.4	14.5
Band3	3258	26446.1	14.7	2328	25758	15.3	2157	22689.9	14.2
Band4	4371	28968	14.7	3831	28228.5	15.4	2115	25239.6	14.9
Band5	14995	34577.5	16	9548	33326.5	16.4	4445	30544.2	15.6
Band6	12973	42806.1	15.5	11090	41232.8	15.9	4100	38029.7	15.5
Band7	11139	50392.9	15.6	8180	48623.6	15.9	2980	45307.4	15.2
Band8a	4583	57478.7	15.6	3566	55535.9	16	1733	52284.6	15.5
Band8b	1633	66404.4	15.6	1011	64668.9	15.8	602	61362.2	15.7
Band8c	620	77902.4	15.2	369	76268.9	16.3	274	73117.4	16.1
Band8d	358	91771.5	15.5	261	89514.1	16.2	99	86645.9	15.2
Band9	171	107790.5	15.3	176	105220.6	16.6	80	102936.5	15.1
Senior Manager	38	134793.4	15.1	31	143971.1	20.6	18	139184.2	16.7
Doctor\Dentist\Fellow	8311	67766.5	15.3	6419	75814.3	16.8	3775	73245.1	16.2
Administration	6690	30247.4	14.9	5448	31324.7	15.6	4107	26387.5	14.7
Allied Health	3314	45865.2	15.4	3276	41960.8	16.1	1252	36827.7	15
Clinical Support	2288	43658.9	15.8	1336	40534.2	15.6	442	35670.8	14
Corporate	2026	50596.4	15.1	1354	48506	15	923	44745.8	14.4
Dental	127	49514.8	14.5	5	39017.5	14.8	23	58082.2	14.5
Doctor	8184	68049.7	15.3	6414	75843	16.8	3752	73338	16.2
Estates	1533	37724.1	15.5	1117	37043.2	15.3	965	27308.9	14.9
Healthcare Support	3052	32866.9	15.1	2635	31223.9	15.7	1950	25562.7	15.4
Management	5243	53330.6	14.9	3027	51451.1	15.1	1502	47067.1	14.4
Midwife	1516	46908.7	15.2	1245	44815.9	15.4	717	43523.1	15.8
Nurse	22815	40907.8	16	13757	39644.6	16.2	5383	37331.9	15.5
Pharmacist	2460	44597.5	15.3	2281	43428.5	15.9	888	41199.1	15.1
Public Health	138	40490.6	17.4	62	39499.3	14.9	15	33991.3	14.8
Mental Wellbeing	4853	46996.1	15.6	7194	43397.8	16.4	2719	39948.8	16.2
Research	1328	45465	15.3	705	43108.1	15.6	434	38574.6	13.8
Total	65,567	45,234.4	15.5	49,856	44,915.2	16	25,072	41,064.3	15.3

Table 2: Sample Summary Statistics

Inner London has the highest number of vacancies for each band, while fringe London has the fewest, with some minor exceptions. Due to the existence of the HCAP, the average posted wage is the highest in inner London and the lowest in fringe London for each band. Inner London has lower average posting lengths than outer London for each band. However fringe London has the lowest vacancy duration with a total, full sample average of 15.3 days. We discuss possible reasons for this in section 7.5. Within each PCT zone, the average posted wage of each band is increasing in its seniority, as required by the guidelines set out in the NHS handbook. However less senior bands also have lower average posting durations. Notably Bands 2 and 3 have the lowest average vacancy durations of all bands in each PCT zone. This is consistent with the theoretical predictions of Faberman and Menzio (2018) which We summarise in section 4. Higher wages may act as a signal for more competition for a role, which could reduce application numbers on aggregate. Furthermore jobs in higher bands require more skills and qualifications meaning that finding suitable candidates may be more time consuming for hospitals.

The next set of rows decompose each of the summary statistics by job category. In each PCT zone, Nurse vacancies are posted the most frequently, being observed 41,955 times in total. This is distantly followed by doctors with a total number of 18,350 observations and administrators with 16,245 observations in total. Psychological Wellbeing professionals also comprise a significant proportion of the total number of jobs with 14,766 total observations across all three PCT zones. Job vacancies that are included as part of the Dental team are listed the most infrequently followed closely by public health professionals. Psychological well being professionals have the highest average posting duration on aggregate with an average posting length of 16.02 days while the Dental team has the lowest, with an average posting length of 14.51 days. Vacancies for Doctors also have a relatively high average posting length despite them also being the offered the highest wages, on average, across all PCT zones, with an aggregate average posted salary of £71,885.05 per annum. Note that when, decomposing the relationship by job category, there appears to be no discernible relationship between the average posted wage of particular job and its average posting duration.

6 Empirical Methodology

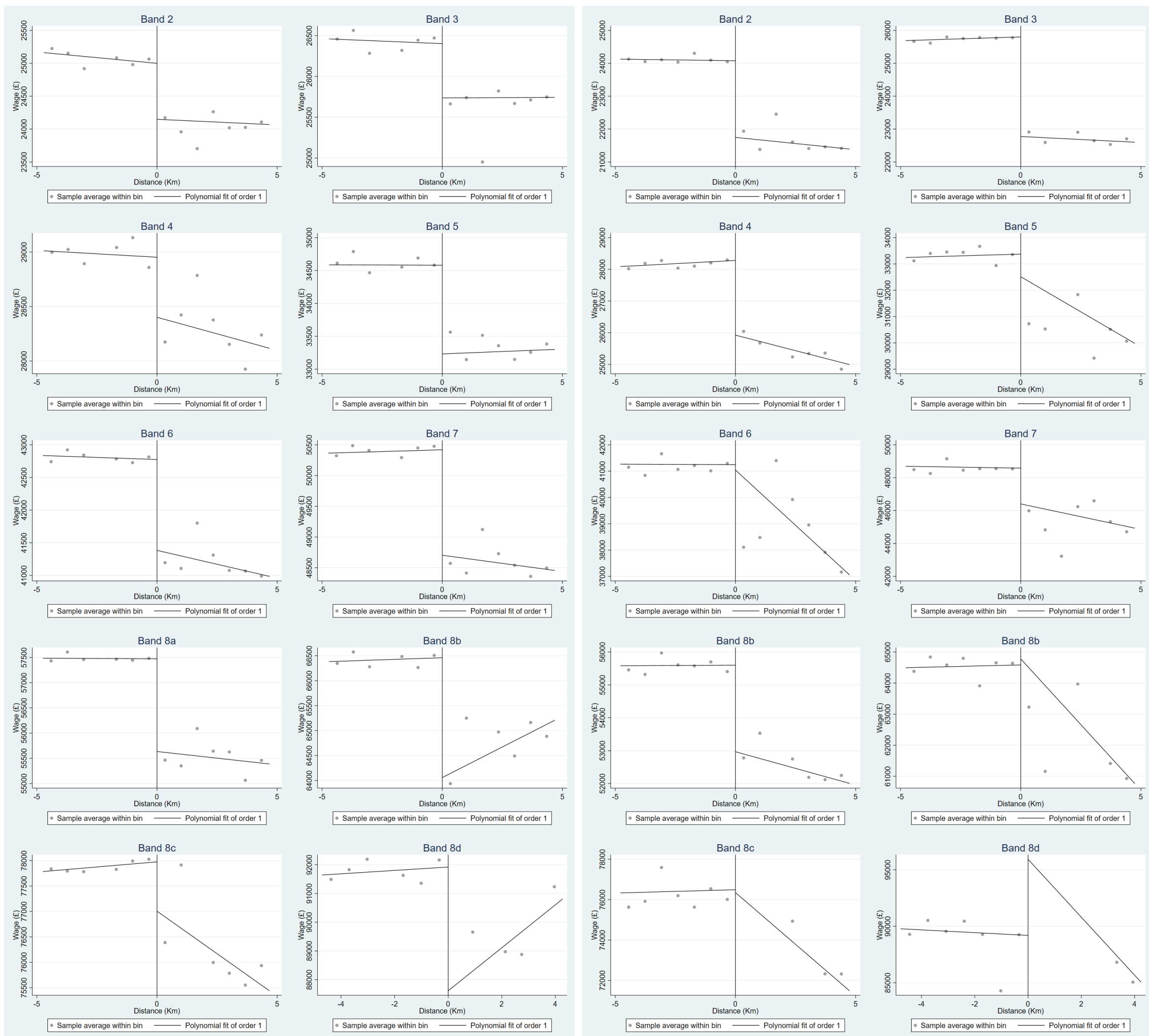
This section describes the empirical methodology We employ to estimate the causal relationship between posted wages and vacancy durations. Consider two hospitals situated on either side of a specific PCT zone boundary. Given their proximity to each other, it is reasonable to assume that these hospitals should operate under similar economic conditions. Factors such as the cost of living, labour market conditions, geographical mobility, and demand for NHS services are likely to be similar between these hospitals. If, despite these similarities, there is a significant variation in the average vacancy duration for hospitals on either side of the border, it may suggest that the exogenous increase in the posted wage, attributed to the difference in wages between London HCAP zones, could influence vacancy durations. This investigation is particularly relevant in London, where excellent transport links mitigate geographical immobility, reducing the likelihood that transport costs play a significant part in influencing job seekers search decisions. Thus, this natural experiment design can help mitigate potential endogeneity, resulting from the relationship between job posting lengths and local labour market conditions. In this analysis We consider both the inner-outer (IO) and the outer-fringe (OF) PCT borders.

This strategy closely corresponds to the technique employed by Greaves and Sibieta (2019) who utilise HCAP zones in London to estimate the impact of teacher salaries on students' educational standards, as gauged by KS2 Mathematics and English scores. They find that variations in centrally determined salary structures do translate into discernible differences in actual teacher wages.

6.1 Graphical Analysis

For the analysis to be valid, it is vital to demonstrate that the differences in centrally determined pay scales do indeed translate into differences in posted wages for NHS workers. To investigate this, We create a continuous variable, $Distance_{ih}$, that tracks the distance in kilometers from each identified hospital to the PCT border of interest, $h \in \{IO, OF\}$. $Distance_{ih}$ takes a negative value when the hospital is located in the HCAP closer to the center of London and a positive value when located in the HCAP zone further out from the centre of London. $Distance_{ih}$ takes a value of zero at cutoff value $c = 0$ at the border h.

In Figure (3) panel (a), We plot the posted wages against the distance from the IO border, $Distance_{iIO}$, for each band. For the purpose of visual clarity, We bin observations. In each bin all observations are averaged and represented by a single point. The plot clearly illustrates the sharp discontinuity in posted wages, particularly for bands 2 to 8b, when crossing the IO border. Posted wages drop sharply from those advertised at hospitals in inner London to those in outer London. This indicates that the differences in the HCAP between inner and outer London do translate into actual wage differences for NHS staff in each specified band.



(a) (b)

Figure 3: Posted wages (£) against distance to the (a) IO and (b) OF London Borders (Km)

The findings of panel (a) provide credibility to the sharp border discontinuity approach We employ in section 7 of the analysis. Confirming that the proposed HCAP discontinuity is reflected in posted wages supports the validity of using the IO border as a natural experiment to investigate the influence of wages on job vacancy durations in the NHS.

Figure (3) panel (b) repeats the analysis but uses the OF border, plotting posted wages against $Distance_{iOF}$. Here the wage discontinuity is less apparent. The poor balance of hospitals at the OF border, as shown in figure (2),and the correspondingly low sample size for certain bands close to the border on the fringe side, is likely to be the culprit for such an absence. The lack of an observable discontinuity largely invalidates the use of this border, however We proceed with the estimation analysis to illustrate the effects of using a border with a poor balance.

6.2 Variable Construction

Before We proceed to discussion the estimation methodology, We first opt to discuss the construction of the control variables that will later be used to enhance the robustness of the analysis.

We define a dummy variable $Treatment_{ih}$, which takes a value of one when a vacancy for a hospital is located in a PCT zone with the higher mandated HCAP for $h \in \{IO, OF\}$. For example, when considering the regression for the IO border, $Treatment_{iIO}$ takes a value of 1 if vacancy i is located at a hospital in inner London and zero if the vacancy is located in outer London. Similarly if we are considering the estimation of causality in the case of the OF border then $Treatment_{iOF}$ would take a value of 1 if vacancy i was located at a hospital in outer London and zero otherwise. Therefore this is the treatment variable which is assigned according to the distance to border $h \in \{IO, OF\}$ and the cutoff variable $c = 0$ such that:

$$Treatment_{ih} = \mathbb{1}(Distance_{ih} < c = 0) \tag{56}$$

To ensure that We am comparing hospitals that are geographically close to one another, We define geographical areas which We call "segments". Each hospital in inner, outer, and fringe London is assigned to a specific segment according to its location. In doing so, We ensure that hospitals in the same area are geographically close to each other and likely share similar economic conditions. This is key to the robustness of the RD approach We employ, as discussed in the section above.The exact boundaries are defined by Strategic Health Areas (SHA's), as outlined in the NHS Terms and Conditions (2022), which We list in table 3. The areas are called Northwest, North Central, North East, South East, and South West London.

We illustrate this segmentation approach in figure (4) by plotting each hospital in inner, outer and fringe London, on a map. The hospitals are colour-coded to represent the different SHA's: blue for North West London, red for North Central London, orange for North East London, green for South East London, and pink for South West London. We define a vector of dummy variables $\boldsymbol{s_i} = (NW_i, ..., SW_i)$, each of which takes a value of one if vacancy i is located in that particular segment and zero otherwise.

North West London SHA	Hammersmith & Fulham,Kensington & Chelsea, Westminster & Brent, Ealing, Harrow, Hillingdon, Hounslow
North Central London SHA	Camden,Islington, Barnet Enfield,Haringey Brent, Ealing, Harrow, Hillingdon, Hounslow
North East London SHA	City & Hackeny,Tower Hamlets,Barking & Dagenham, Havering, Newham, Redbridge, Waltham Forest
South West London SHA	Wandsworth,Croydon,Kingston,Richmond & Twickenham, Sutton & Merton
South East London SHA	Lambeth,Lewisham,Southwark,Bexley,Bromley, Greenwich

Table 3: PCT's included in each SHA in Inner and Outer London (NHS Terms and Conditions of Service Handbook, 2023)

We control for seasonality trends in the labour market by including a set of time dummies. These dummy variables each represent a separate month across the full length of the sample period, from April 2021 to March 2023. Therefore, the set of month dummies employed in the analysis is defined as $\boldsymbol{m_i} = (April21_i, May21_i, ..., February23_i, March23_i)$. Each of the month dummies take a value of one if the vacancy i is posted in that particular month and zero otherwise.

We also create a vector of dummy variables to control for the band of particular vacancy i, as explained in more detail in section 3.2, $\boldsymbol{b_i} = (Band_{2i}, , ..., Band_9)$. To utilise the job title control for a particular vacancy, We also define a vector of job category controls, $\boldsymbol{j_i} = (Admin_i, Nurse_i, .., Doctors_i, Estates_i)$, informed by the job categories discussed in section 3.3 and 5.4.

Finally We create a vector which controls for the job type of a particular vacancy. Notably each

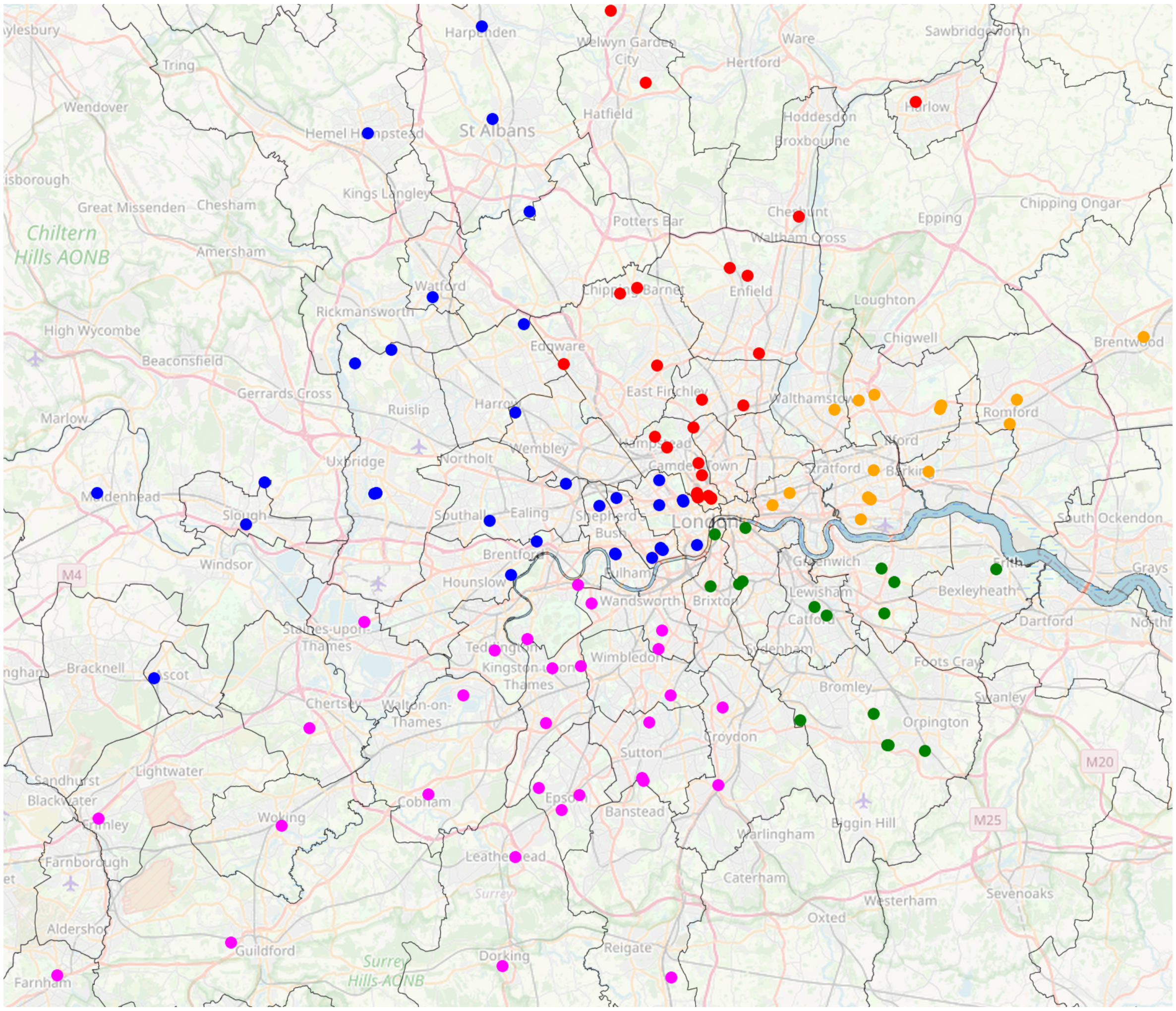


Figure 4: Hospitals in the North West (Blue),North East (Yellow),North Central (Red) ,South West (Magenta),South East (Green) SHA's

vacancy can be part-time or full time and can vary in the length of the contract. A job can be permanent, temporary or an apprenticeship. We therefore, create a vector of job type dummies $\boldsymbol{t_i} = (part_i, temporary_i, apprentiship_i)$ to control for these effects.

6.3 Estimation Methodology

In estimating the causal parameters of interest, We undergo an iterative process of model improvement, starting with a baseline sharp border discontinuity approach with minimal controls[13]. We then progressively include more controls and observe the sensitivity of the purported estimates which serves

[13]We include a more formal description of the canonical sharp border discontinuity approach in the appendix.

as a basic robustness test. We compare these estimates with that produced by ordinary least squares (OLS), using the starting wage as an independent variable.

We start by first discussing the linear OLS approach. We pool observations of the natural logarithm of the vacancy duration of job posting i, from sample size n, $ln(\tau_i)$. The dependent variable is then regressed against the midpoint of the posted salary range ω_i. In the baseline specification We also include the set of segment, $\boldsymbol{s_i}$, month $\boldsymbol{m_i}$, job type $\boldsymbol{t_i}$ and job title $\boldsymbol{j_i}$ alongside sets of corrosponding slope coefficients, $\boldsymbol{\lambda} = (\lambda_1 ..., \lambda_5)'$, $\boldsymbol{\phi} = (\phi_1, ..., \phi_{23})'$, $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \gamma_3)'$, $\boldsymbol{\psi} = (\psi_1, ..., \psi_{12})'$ and $\boldsymbol{\zeta} = (\zeta_1, ..., \zeta_{14})'$, as well as a constant parameter θ_0. The error term is denoted as ϵ_i. This yields the baseline specification (57).

$$ln(\tau_i) = \theta_0 + \theta_1\omega_i + \boldsymbol{s_i\lambda} + \boldsymbol{t_i\gamma} + \boldsymbol{m_i\phi} + \boldsymbol{j_i\zeta} + \epsilon_i \tag{57}$$

$\hat{\theta}_i$ represents the estimated average causal impact of posted wages on the duration of job vacancies. However, the reliability of specification (57) depends on certain stringent assumptions that form the foundation of linear estimation models (Greene, 2000). To begin with, the methodology requires that the independent variable of interest, denoted as ω_i, is exogenous, such that the zero conditional mean assumption, $E[\epsilon_i|\omega_i] = E[\epsilon_i] = 0$ holds. This essentially requires that there exist no factors, encompassed within the error term ϵ_i, that exhibit correlation with ω_i and τ_i. If such factors were present, the estimations of ω_i could be biased.

Several unaccounted factors lie outside the scope of this specification, including elements that influence the demand for workers in specific hospitals. One such a factor could be the size of a hospital's workforce, which is not currently taken into consideration. It is plausible that hospitals located closer to the city center, such as Guys, St.Thomas, and Chelsea & Westminster Hospital, boast larger workforces. This correlation is likely reflected in longer average vacancy durations. Consequently, ω_i could be correlated with a term present in the error component that also relates to τ_i. This scenario would lead to an upward bias in the coefficient $\hat{\theta_1}$.

In recognition of this drawback, We also employ a sharp RD approach which is described as follows. The natural logarithm of the job vacancy duration for observation i, $ln(\tau_i)$, is regressed on a vector of predictor variables $\boldsymbol{x_i}$ that includes the treatment variable $Treatment_{ih}$, the assignment variable $Distance_{ih}$ and an interaction between the two such that $\boldsymbol{x_i} = (Distance_{ih}, Treatment_{ih}, Distance_{ih} \times$

$Treatment_{ih},$). Note that by including an interaction term between the treatment and assignment variables, We allow for the effect of distance from the border h on posting lengths to vary either side of the border. We also include a complete set of segment, month, job type and controls. This specification is given below:

$$ln(\tau_i) = \beta_0 + \boldsymbol{x_i\beta_1} + \boldsymbol{s_i\lambda} + \boldsymbol{m_i\phi} + \boldsymbol{t_i\gamma} + \epsilon_i \tag{58}$$

We build on specification (58) by including the set of band controls $\boldsymbol{b_i}$ and the associated coefficients $\boldsymbol{\psi} = (\psi_1, ..., \psi_{12})'$.

$$ln(\tau_i) = \beta_0 + \boldsymbol{x_i\beta_1} + \boldsymbol{s_i\lambda} + \boldsymbol{t_i\gamma} + \boldsymbol{m_i\phi} + \boldsymbol{b_i\psi} + \epsilon_i \tag{59}$$

We proceed by including a the set of job controls $\boldsymbol{j_i}$ and their associated slope parameters $\boldsymbol{\zeta} = (\zeta_1, ..., \zeta_{14})'$. This produces specification (59):

$$ln(\tau_i) = \beta_0 + \boldsymbol{x_i\beta_1} + \boldsymbol{s_i\lambda} + \boldsymbol{t_i\gamma} + \boldsymbol{m_i\phi} + \boldsymbol{b_i\psi} + \boldsymbol{j_i\zeta} + \epsilon_i \tag{60}$$

As a further basis robustness test We manipulate equation (60) to introduce a non-linear polynomial relationship between the assignment variable, $Distance_{ih}$ and the dependent variable $ln(\tau_i)$. In doing so We introduce an additional 1×2 vector $\boldsymbol{p_i} = (Distance^2_{ih}, Treatment_{ih} \times Distance^2_{ih})$ into the specification alongside the slope parameters $\boldsymbol{\beta_2} = (\beta_4, \beta_5)'$.

$$ln(\tau_i) = \beta_0 + \boldsymbol{x_i\beta_1} + \boldsymbol{p_i\beta_2} + \boldsymbol{s_i\lambda} + \boldsymbol{t_i\gamma} + \boldsymbol{m_i\phi} + \boldsymbol{b_i\psi} + \boldsymbol{j_i\zeta} + \epsilon_i \tag{61}$$

We use linear ordinary least squares (OLS) to estimate each of the slope parameters in the specifications listed above. The coefficient β_1 is of particular interest as it represents the difference in the conditional expectation value of $ln(\tau_i)$ when $Treatment_{ih} = 0$ and when $Treatment_{ih} = 1$. Therefore, in the case where the fundamental assumptions of sharp RD hold, $\hat{\beta_1}$ can be interpreted as the estimated average treatment effect on the treated (ATT) at $Distance_{ih} = c = 0$.

A critical issue in the use of a sharp RD approach is the choice of bandwidth, that is, the observations that are included in the estimation of the parameters specified above, according to some maximum absolute value of the assignment variable. While We would ideally use robust local linear estimation techniques to optimally derive the bandwidth, as shown by Calonico, Cattaneo and Titiunik (2014) and Armstrong and Kolesár (2018), the existence of significant mass points in the data renders these approaches futile. Rather We use a bandwidth of 1.5km, similar to that used by Greaves and Sibieta (2019) in their use of London HCAP zones to estimate the relationship between teacher pay and pupils educational attainment. Therefore, only hospitals within this distance to each border $h \in \{IO, OF\}$ will be included in their respective analysis. The purpose of this approach is to ensure that hospitals being compared across the border are sufficiently close to one another, such that the probability that they operate under similar unobserved conditions is maximised. We discuss the effects of adjusting this bandwidth on the estimated parameters in section 8.

We cluster the standard errors at the postcode level in both regressions. The reason for doing so is that $Distance_{ih}$ is treated as random however, once the hospitals location is fixed, the HCAP premium does not vary. This suggests that the use of clustered standard errors at the postcode level is preferred to their robust counterparts.

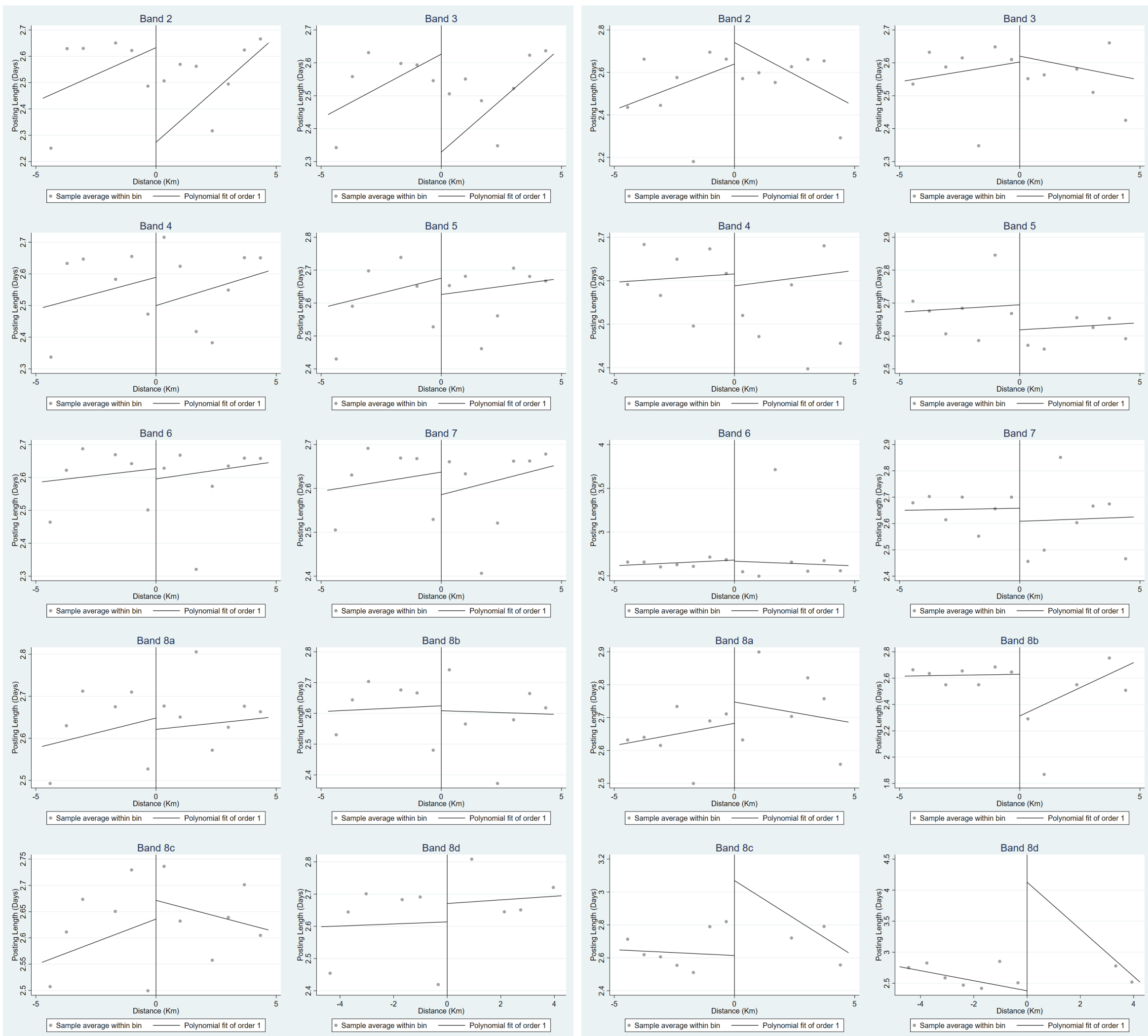
7 Results

In this section We present the empirical results of the estimation methodology described in section 6. We discuss the results of a graphical and regression analysis before proceeding to discuss my results in the context of the theoretical predictions outlined in section 4.

7.1 Graphical Analysis

Before We present the regression results that We detail above, We first employ a graphical analysis. The purpose of this approach to provide a first look into whether there is any immediate effect of crossing the PCT zone border on vacancy durations that is drastic enough to be visually obvious. This is common practice in papers employing sharp RD regression designs as discussed in more detail by, Korting et al. (2023) and Calonico, Cattaneo and Titiunik (2015).

We plot each vacancies posting length, the calculation of which was discussed in section 5, against the distance, in kilometers, to both the IO and OF border. Analogous to the collection of graphics in figure 3, We bin observations. The graphs are characterised by band, as in figure 3, so as to better visualise the discontinuity.



(a) (b)

Figure 5: ln(Vacancy Durations) (days), against distance to the (a) IO and (b) OF London Borders (Km)

Notably figure 5 does not show any clear border discontinuity for any band at either the IO or OF border. While not immediately convincing, the lack of a clear discontinuity does not completely confirm the absence of any significant relationship between wages and vacancy durations at either border. The graphics presented in figure 5 simply contain the raw data which lack any of the control variables listed in preceding section. In the absence of such controls, We cannot be certain that the lack of a clear discontinuity is not the result of other factors that have not been adequately accounted for. For example, in the absence of segmentation controls, these graphs could be comparing hospitals that, although both close to the border, may in fact be far away from each other with one being in North and the other being in South London. In this case, factors other than wage, such as local labour market tightness and the level of demand of local NHS services, may affect the appearance of the discontinuity. Therefore, We proceed to a regression analysis to more formally derive parameter estimates in the presence of appropriate controls.

7.2 Regression Analysis

We present the estimated parameters for each specification using the IO and OF border in tables (4) and (5) respectively. In both tables, column (1) presents the OLS results of specification (57), excluding the set of job title controls, while column (2) describes the estimated coefficients for the complete OLS specification. In column (3) We present the sharp RD parameters for specification (58). This is the baseline sharp RD design, excluding job title and band controls. The results of specification (59), including the set of band controls but excluding job titles are included in column (4). The regression results for the complete linear RD specification (60), including bands and job title controls are shown in column (5). Column (6) reports the estimated coefficients of the non-linear specification (61), with a complete set of controls. All regressions include the set of segmentation, month and job type controls, $\boldsymbol{s_i}$,$\boldsymbol{m_i}$,$\boldsymbol{t_i}$.

Inner - Outer (IO) Border

In this section We discuss the estimated coefficients when using the IO border. The total number of observations is 20,167 after adjusting for the 1.5km bandwidth at this border.

We first consider the results using linear OLS regression as outlined in specification (57). When job title controls are excluded the estimated coefficient on the posted wage is 0.019. This result is not however statistically significant. Including the job title controls in the specification, increases the coefficient to

0.064, which is statistically significant at the 1% confidence level, implying that an increase in posted wages by 1% predicts an increase in the conditional expected vacancy duration by 0.064%. Notably estimates using the basic OLS specification produces positive, albeit small, results, contradicting the predictions of job search theory outlined in section 4. This is likely to reflect the existence of significant upward bias in these estimates in the absence of an appropriate RD methodology, as discussed above.

We now consider the regression results using the sharp RD design specified in equations (58),(59),(60) and (61). In the sharp RD model, the estimated coefficient on the variable $Treatment_{ih}$, $\hat{\beta}_1$, is is primary interest as it represents the difference in the estimated conditional mean of $ln(\tau_i)$ for job vacancies across the HCAP border $h \in \{IO, OF\}$. The estimation assumes that the only difference between job postings in hospitals either side of the IO border is the difference in the HCAP which increases from 15% in outer London to 20% in inner London. This corresponds to an increase in the posted wage by 4.34.%[14]. Therefore, the coefficient on $Treatment_{iIO}$ in this case represents the expected percentage change of vacancy durations resulting from a 4.34% increase in the average wage.

Each of the estimated coefficients using sharp RD with a linear assignment variable, are negative and significantly different from zero at the 1% confidence level. In column (3), the coefficient on $Treatment_{iIO}$ is -0.243, implying that a 4.43% increase in the starting wage results in a decrease in vacancy duration's by 24.3%, which corresponds to a large, negative elasticity of approximately -5.60. In columns (4) and (5), where the latter includes both job titles controls and the former includes only band controls, the estimated coefficients are -0.241 and -0.249, respectively implying an elasticity of -5.55 and -5.74. The sign on the estimates are consistent with the predictions of the directed search model We present in section 4, which suggests that, in an organisation who's pay structure is strictly controlled, there should exist a negative relationship between wages and vacancy durations even in the absence of strict job heterogenity controls. This is therefore, consistent with the literature described in section 2.

[14]To show this, consider a base salary x. Workers in inner London earn $1.2x$ while workers in outer London earn $1.15x$ Calculating the percentage difference is $\frac{1.2x-1.15x}{1.15x} \times 100 \approx 4.34\%$

Dependent variable: ln(vacancy duration)

Inner-Outer (IO)	(1)	(2)	(3)	(4)	(5)	(6)
Wage	0.019 (0.0263)	0.064*** (0.0194)	x	x	x	x
Treatment	x	x	-0.243*** (0.083)	-0.241*** (0.082)	-0.249*** (0.079)	-0.133 (0.228)
Distance	x	x	-0.147** (0.084)	-0.148** (0.081)	-0.145** (0.080)	0.314 (0.228)
Interaction	x	x	0.099 (0.109)	0.102 (0.104)	0.087 (0.101)	-0.226 (0.485)
Band Controls	No	No	No	Yes	Yes	Yes
Job Title Controls	No	Yes	No	No	Yes	Yes
Segment Controls	Yes	Yes	Yes	Yes	Yes	Yes
Month Control	Yes	Yes	Yes	Yes	Yes	Yes
Job Type Control	Yes	Yes	Yes	Yes	Yes	Yes
N	20,167	20,167	20,167	20,167	20,167	20,167
R^2	0.160	0.164	0.165	0.167	0.168	0.170

Table 4: Regression Outputs: Standard errors are clustered by postcode and are presented in parentheses. (***) (**) (*) denotes significance at the 1%,5% and 10% confidence levels.

In column (6) We present the non-linear sharp RD with an additional quadratic assignment variable as well as a corresponding interaction term as shown in specification (61). The interpretation of the coefficient $\hat{\beta}_1$ is identical in the non-linear case in the sharp RD approach as We show in the appendix. Column (6) shows that, while negative, the estimated ATT of the HCAP treatment on posting lengths is not significantly different from zero. This suggests that the estimation methodology is not robust to adjustments in the models functional form.

Outer - Fringe (OF) Border

We now re-estimate the coefficients of interest using the OF border, the results of which are shown in table (5). The total number of observations in all regressions within the specified bandwidth of 1.5km is 8,931 which is much smaller than the number for the IO border.

Analogous to the section above We consider the results of the OLS regression in specification (57) which are reported in columns (1) and (2). When job controls are excluded, the estimated effect of a 1% increase in wages on vacancy durations is 0.024%. This result is not statistically significant however. Once job controls are excluded the estimate increases to 0.073% which is significant at the 5% confidence level. However note that, as above, We express some serious scepticism as to the validity of the results of OLS estimation. Such doubt results from the lack of adequate distance variables which may otherwise control for the local labour market conditions exhibited by different PCT zones.

To account for these concerns We discuss the estimates of the, more robust, sharp RD approach, encapsulated by the specifications (58),(59),(60) and (61). Given that the HCAP changes from 5% to 15% between fringe and outer London, $\hat{\beta}_1$, now represents the percentage increase in vacancy durations resulting from an 8.7% average increase in posted wages. Each linear RD model produces non-significant, positive estimates of β_1 with estimated coefficients ranging from 0.134 to 0.159 which corresponds to an estimated elasticity range of 1.54 to 1.83.

When using a non-linear assignment variable as defined by specification (61), We estimate a very large, negative relationship between posted wages and vacancy durations. This estimate is significantly non-zero at the 1% confidence level which may lead to the suggestion that, when controlling for some underlying non-linear relationship between the treatment variable and vacancy durations, We find a statistically significant negative causal relationship between posted wages and vacancy durations. This would be consistent with the evidence presented using the IO border. However the lack of stability in the reported parameters at the OF border leads me to strongly hesitate to report this result as clear evidence for a negative causal relationship between posted wages and vacancy durations. Rather We simply use this vastly different outcome to show that the results are highly dependent on the specified functional form.

Dependent variable: ln(vacancy duration)

Outer-Fringe (OF)	(1)	(2)	(3)	(4)	(5)	(6)
Wage	0.024	0.073**	x	x	x	x
	(0.040)	(0.490)				
Treatment	x	x	0.159	0.145	0.134	-1.349***
			(0.111)	(0.100)	(0.102)	(0.151)
Distance	x	x	0.039	0.025	0.034	-3.724***
			(0.178)	(0.156)	(0.161)	(0.402)
Interaction	x	x	0.123	0.135	0.115	3.679***
			(0.123)	(0.135)	(0.166)	(0.459)
Band Controls	No	No	No	Yes	Yes	Yes
Job Title Controls	No	Yes	No	No	Yes	Yes
Segment Controls	Yes	Yes	Yes	Yes	Yes	Yes
Month Control	Yes	Yes	Yes	Yes	Yes	Yes
Job Type Control	Yes	Yes	Yes	Yes	Yes	Yes
N	8,931	8,931	8,931	8,931	8,931	8,931
R^2	0.133	0.143	0.139	0.146	0.150	0.151

Table 5: Regression Outputs: Standard errors are clustered by postcode and are presented in parentheses. (***) (**) (*) denotes significance at the 1%,5% and 10% confidence levels.

7.3 Discussion

The difference in the sign and magnitude of the estimated elasticises produced by the linear OLS and sharp RD models are likely a result of upward bias in the $\hat{\theta}_1$ estimate resulting from endogenity in this parameter, as discussed in section 6.3. The existence of upward bias in $\hat{\theta}_1$ resulting from differences in the size of the workforce is supported by the fact that the estimated coefficient on the $Distance_{iIO}$ variable is negative and significant at the 5% level when using the IO border. When controlling for this effect in the linear RD model, the purported positive estimate of β_1 produced by OLS, turns negative.

The difference between the linearly estimated coefficients using the IO and OF borders sheds some doubt on the robustness of the analysis. However We argue that the lack of significant non-zero estimated treatment coefficients in the use of the latter, better reflects conditions that make this relationship more challenging to derive using this border. Notably, We speculate that the observed, non- significant, positive relationship between vacancy durations and posted wages at the OF border is attributable to two factors; the existence of an additional unobserved discontinuity and the poor balance of hospitals at this border. Firstly We argue that the geographical boundaries of NHS trusts may invalidate the use of a sharp RD methodology in the the context of the OF border. NHS trust borders are also defined by the London borough boundaries that similarly define the OF PCT zone border. Therefore, differences in conditional expectation of job vacancy durations either side of the OF border could also be attributable to differences in NHS trust management. For example, different trusts may be less effective than others, running hospitals with poorer working conditions and longer patient waiting times. If it was such that that trusts in outer London are less desirable to work for, this could negatively influence the supply of applications, thus inducing upwards bias in RD estimates at the OF border. Therefore, the existence of trust boundary lines that perfectly coincide with the OF border, makes the estimation of unbiased estimates significantly more difficult. This would explain the insignificant and positive coefficients shown in table (5).

Furthermore We argue that hospitals, and therefore, job vacancies, are less evenly distributed along the OF border in comparison to the IO border. This can be shown in figure (2) which shows the distribution of hospitals against the distance to the associated PCT border. The lack of hospitals close to the OF border on the fringe side is a cause for concern. We hypothesize that the lack of hospitals increases the probability that unobserved factors influence the relationship between posting durations and wages, resulting in biased estimates at this border. The poor balance of hospitals increases the likelihood that geographical immobility plays a part in determining the relationship between wages and vacancy durations. Notably, geographical mobility is much less effective and expansive across the OF in comparison to IO. This may induce workers to place a higher preference weighting on jobs that are closer to thier place of residence. To explain this in more detail, consider the theoretical directed search model explained in section 4. If it is the case that both workers of type l and h experience higher costs associated with getting to the hospital with higher pay, g_{i2}, then this could outweigh the benefits of the HCAP. This is probable due to the high associated costs with travelling into London from fringe areas. Furthermore, the higher costs of living in Outer London may deter workers who would otherwise move to these areas. These effects could ensure that hospitals in fringe London receive higher applications and thus experience lower posting lengths. Additionally the poor balance of hospitals ensures that the wage discontinuity, for certain bands, is largely absent in the sample, as shown in figures 3 and 5. For these reasons We propose that the IO border is more suitable for the estimation of a causal relationship using the RD approach.

Another feature of interest is the sensitivity of the results in the face of adaptations in the functional form of the assignment variable. Given that the bandwidth is only 1.5km, We hypothesise that the use of non-linear models is not appropriate. We therefore, opt to rely on the estimates of the linear sharp RD, as more robust. The sensitivity of the purported estimates is of interest however. Future research could utilise non-parametric methods to more accurately derive optimal functional forms which so dramatically influence the estimated coefficients.

Finally, it is worth discussing the fact that, in all regressions, estimates are highly stable across the linear RD specifications (58),(59) and (60). This implies that the baseline specification does not suffer from large endogeneity resulting from the exclusion job title and band controls.

7.4 Treatment Heterogeneity

At the time of writing there has been much discussion surrounding the issue of NHS worker pay with an emphasis on nurses and allied health professionals (APH's), with the former grappling with the UK government over the issue through the medium of strike action. Given the recent interest in the nature of these professions job dynamics, We devote another section to discussing the relationship between wages and vacancy durations for APH's and nurses specifically.

In discussing the issue of heterogeneous treatment effects, We limit the sample to exclusively include either AHP's or nurses. This is achieved using the job title controls discussed in sections 5 and 6. We continue to use a bandwidth of 1.5km. Standard errors remain clustered at the postcode level for all the proceeding regressions.

Nurses

In this section We limit the sample to exclusively include Nurses and re-estimate each of the econometric specifications described above using this refined sample. The estimated coefficients for the specification (57),(58),(60) and (61) at the IO and OF border are included in table (6) and (7) respectively. Columns

(1) and (2) reports estimates of the ATT using OLS regression, using the natural logarithm of the wage as the independent variable, with the latter including, and the former excluding, band controls. Columns (3) and (4) report the estimates from the sharp RD specification excluding and including band controls respectively. All specifications include a full set of month, segment and job type controls. Column (5) reports the estimates from the non-linear sharp RD design in specification (61).

Estimates derived using all variants of both the OLS and RD specification at the IO for the relationship between wages and vacancy durations are negative. Estimates produced by linear RD suggest elasticities of -5.67 and -5.85 when band controls are excluded and included respectively. This is consistent with the estimates produced using the full sample. However none of the estimates, using Nurses exclusively, are statistically significant.

When looking at table (7), we see that the linear RD approaches employed for the OF border, as shown in columns (3) and (4) produce large positive estimates that are significant at the 5% confidence level. The estimate turns negative when the model accounts for a non-linear assignment variable, as shown in column (5). This result is statistically significant at the 1% confidence level. Similarly to the full sample, estimates derived using OLS vary significantly from their RD counterparts. Estimates using OLS produce a negative estimate of -0.161 when bands are excluded and 0.432 when bands are included.

Dependent variable: ln(vacancy duration)

Nurses	(1)	(2)	(3)	(4)	(5)
Inner-Outer (IO)					
Wage	-0.038	-0.420	x	x	x
	(0.030)	(0.039)			
Treatment	x	x	-0.246	-0.254	-0.217
			(0.153)	(0.147)	(0.359)
Distance	x	x	-0.122	-0.127	1.242
			(0.129)	(0.124)	(0.767)
Interaction	x	x	0.049	0.055	-1.536*
			(0.160)	(0.154)	(0.833)
Band Controls	No	Yes	No	Yes	Yes
Segment Controls	Yes	Yes	Yes	Yes	Yes
Month Control	Yes	Yes	Yes	Yes	Yes
Job Type Control	Yes	Yes	Yes	Yes	Yes
N	6,584	6,584	6,584	6,584	6,584
R^2	0.144	0.145	0.146	0.149	0.149

Table 6: Regression Outputs: Standard errors are clustered by postcode and are presented in parentheses. (***) (**) (*) denotes significance at the 1%,5% and 10% confidence levels.

Dependent variable: ln(vacancy duration)					
Nurses	(1)	(2)	(3)	(4)	(5)
Outer-Fringe (OF)					
Wage	-0.161 (0.154)	0.423 (0.137)	x	x	x
Treatment	x	x	0.314** (0.134)	0.281** (0.281)	-1.908*** (0.140)
Distance	x	x	0.173 (0.217)	0.137 (0.192)	-5.263*** (0.350)
Interaction	x	x	0.0847 (0.223)	0.108 (0.205)	4.890*** (0.390)
Band Controls	No	Yes	No	Yes	Yes
Segment Controls	Yes	Yes	Yes	Yes	Yes
Month Control	Yes	Yes	Yes	Yes	Yes
Job Type Control	Yes	Yes	Yes	Yes	Yes
N	2,822	2,822	2,822	2,822	2,822
R^2	0.162	0.17	0.173	0.182	0.185

Table 7: Regression Outputs: Standard errors are clustered by postcode and are presented in parentheses. (***) (**) (*) denotes significance at the 1%,5% and 10% confidence levels.

Allied Health Professionals

We now refine the sample to only include APH's, reporting the estimated coefficients for the IO and OF border in tables (8) and (9) respectively. The estimates in table (8) show that the linear RD approaches employed for the IO border, as shown in columns (3) and (4), produce large negative estimates that are significant at the 1% confidence level. This effect is replicated by OLS when band controls are included, although the magnitude varies drastically. Estimates of elasticity between posted wages and vacancy durations using RD are approximately -11 and while OLS produces estimates of -1.61. This further suggests that OLS estimates suffer from upward bias. These negative, significant results using RD, changes positive, and non - significant when a quadratic functional form is introduced, further indicating the sensitivity of slope parameter estimates in the face of functional form adjustments.

Sharp linear RD at the OF border yields non- significant, positive results regardless of whether band controls are included.

Allied Health Professionals	(1)	(2)	(3)	(4)	(5)
Inner-Outer (IO)					
Wage	0.0754	-1.637**	x	x	x
	(0.058)	(0.063)			
Treatment	x	x	-0.479***	-0.490***	0.183
			(0.036)	(0.038)	(0.425)
Distance	x	x	-0.187*	-0.228**	0.461
			(0.093)	(0.091)	(0.799)
Interaction	x	x	-0.004	0.071	1.189**
			(0.218)	(0.216)	(0.519)
Band Controls	No	Yes	No	Yes	Yes
Segment Controls	Yes	Yes	Yes	Yes	Yes
Month Control	Yes	Yes	Yes	Yes	Yes
Job Type Control	Yes	Yes	Yes	Yes	Yes
N	846	846	846	846	846
R^2	0.185	0.187	0.198	0.204	0.209

Table 8: Regression Outputs: Standard errors are clustered by postcode and are presented in parentheses. (***) (**) (*) denotes significance at the 1%,5% and 10% confidence levels.

Allied Health Professionals	(1)	(2)	(3)	(4)	(5)
Outer-Fringe (OF)					
Wage	-0.099 (0.127)	0.141 (0.121)	x	x	x
Treatment	x	x	0.171 (0.103)	0.0245 (0.104)	-0.320** (0.135)
Distance	x	x	0.104 (0.097)	-0.0974 (0.111)	-1.185*** (0.206)
Interaction	x	x	0.0390 (0.039)	0.228 (0.228)	2.016*** (0.368)
Band Controls	No	Yes	No	Yes	Yes
Segment Controls	Yes	Yes	Yes	Yes	Yes
Month Control	Yes	Yes	Yes	Yes	Yes
Job Type Control	Yes	Yes	Yes	Yes	Yes
N	473	473	473	473	473
R^2	0.226	0.260	0.228	0.239	0.242

Table 9: Regression Outputs: Standard errors are clustered by postcode and are presented in parentheses. (***) (**) (*) denotes significance at the 1%,5% and 10% confidence levels.

7.5 Discussion

In summary, We find a large negative relationship between posted wages and vacancy durations for AHP's using the, more reliable, IO border. In contrast, while We do derive negative slope coefficients for the relationship between posted wages and posting lengths for Nurses, We do not find sufficient evidence to reject the possibly that the estimate is zero. From a policy perspective this empirical work may therefore, lend weight to the argument that increases in pay can improve job market matching efficiency for Allied Health Professionals. More work is required to extend this conclusion for Nurses however.

Analogous to the analysis using the entire sample of job advertisements, We find that the estimates using the IO and OF border vary considerably. As We explain above, estimations using data close to the OF border are dubious given the large possibly of endogeneity.

8 Robustness Tests

In this section We present two robustness tests that are used to analyse the stability of the parameters presented in section 7. Firstly We present a placebo test, where each boundaries border, and the corresponding regression sample, is artificially manipulated and the slope parameters are re-estimated. Secondly We adjust the bandwidth used in the estimation of each slope coefficient to analyse the estimates stability.

8.1 Placebo Test

Firstly We present the results of a placebo test to indicate whether the results demonstrated in section 7 are unique to crossing the HCAP boundary where the wage drops or if the results presented are coincidental. We create artificial borders for each original $h \in \{IO, OF\}$ by moving each towards the center of London. This process is designed to create random boundaries that have no relevance in the estimation of the effects of posted wages on vacancy durations. At this newly defined border there is no posted wage discontinuity. Therefore, if the re-estimated parameters under this border are similar to those derived using the real one, then the presented effect on vacancy durations may not result from the sharp change in wages but some other factor that may drive differences in vacancy durations. We present the estimated coefficients of specification (60) for the fictitious IO and OF border in table (10) and table (11).

We shift each border uniformly inwards twice, once at 1km and another at 2km, and run separate regressions each time. We include the estimated coefficient with the real borders to provide a point of reference. At the IO border the coefficient of interest is positive at the 1km border and negative at the 2km border. Neither coefficient however, is statically significant. This provides some evidence to suggest that the difference in vacancy durations at the IO zone border when controlling for other factors is driven by the discontinuity in wages.

Given the concern regarding the estimated effect of posted wages at the OF border, the placebo tests are of minor interest in this analysis. We do however report the results in table (11). The estimates are vastly different for the fictitious borders than the real one. This suggests that the sharp wage discontinuity may drive the observed differences in expected vacancy durations. However, We still hesitate to suggest that there are any casual effects that can be deduced from results at the OF border.

	Baseline (0km)	1km	2km
Inner - Outer (IO)			
Treatment	-0.249***	0.273	-0.039
	(0.079)	(0.037)	(0.072)
Distance	-0.145**	-0.095	0.089**
	(0.080)	(0.079)	(0.042)
Interaction	0.087	0.331	-0.470***
	(0.101)	(0.405)	(0.104)
Band Controls	Yes	Yes	Yes
Job Title Controls	Yes	Yes	Yes
Segment Controls	Yes	Yes	Yes
Month Control	Yes	Yes	Yes
Job Type Controls	Yes	Yes	Yes
N	20,167	11,961	14,518
R^2	0.160	0.1767	0.1712

Table 10: Regression Outputs: Standard errors are clustered by postcode and are presented in parentheses. (***) (**) (*) denotes significance at the 1%,5% and 10% confidence levels.

	Baseline (0km)	1km	2km
Outer -Fringe (OF)			
Treatment	0.134	-0.194	0.215
	(0.102)	(0.061)	(0.236)
Distance	0.034	0.038	-0.021
	(0.161)	(-0.40)	(0.011)
Interaction	0.115	-0.270	0.255
	(0.166)	(0.047)	(0.178)
Band Controls	Yes	Yes	Yes
Job Title Controls	Yes	Yes	Yes
Segment Controls	Yes	Yes	Yes
Month Control	Yes	Yes	Yes
Job Type Controls	Yes	Yes	Yes
N	8,931	5,300	1,530
R^2	0.150	0.153	0.173

Table 11: Regression Outputs: Standard errors are clustered by postcode and are presented in parentheses. (***) (**) (*) denotes significance at the 1%,5% and 10% confidence levels.

8.2 Bandwidth Adjustment

The second robustness test We employ involves repeating the estimation of specification (60) using different bandwidths to determine the sample at each border. In the estimation presented in section 6 and 7, We use a bandwidth of 1.5km. Now We adjust the bandwidth for both the estimation of a causal effect at the IO and OF borders, starting at 1km and increasing in 250 meter increments up to 2.5km.

We report $\hat{\beta}_1$, alongside its corresponding standard error, to indicate the non-zero significance of each respective estimate in table (12). This allows for visual inspection of $\hat{\beta}_1$ in response to adjustments of an assumption that is critical in RD designs. In the first row We report the estimated coefficient representing the ATT for the IO border using different bandwidths. Using 1km as the bandwidth produces an estimate -0.158. Notably the estimate is of similarly large magnitude, negative and significantly different from zero at the 1% confidence level. Between 1.25km and 1.75km the model continues to produce stable negative estimates, all of similar magnitude, the highest in this range being

-0.249 and the lowest being -0.216. Each of these estimates are significant at the 1% level. Increasing the bandwidth beyond this 2km produces point estimates that, although negative, are not significantly different from zero.

Dependent variable: ln(vacancy duration)

	1km	1.25km	1.5km	1.75km	2km	2.25km	2.5km
Inner-Outer (IO)							
Treatment	-0.158***	-0.216***	-0.249***	-0.244***	-0.157 **	-0.088	-0.082
	(0.036)	(0.055)	(0.079)	(0.061)	(0.070)	(0.069)	(0.070)
N	14,882	18,778	20,167	21,970	34,546	37,800	38,206
R^2	0.1757	0.1747	0.1679	0.1717	0.1585	0.1555	0.1550
Outer-Fringe (OF)							
Treatment	-0.069**	0.133	0.134	0.264*	0.159*	0.159*	0.141***
	(0.030)	(0.102)	(0.071)	(0.264)	(0.082)	(0.082)	(0.031)
N	5,834	8,911	8,931	8,935	11,271	11,271	12,311
R^2	0.163	0.1498	0.1497	0.1493	0.1809	0.1809	0.1742

Table 12: Regression Outputs: Standard errors are clustered by postcode and are presented in parentheses. (***) (**) (*) denotes significance at the 1%,5% and 10% confidence levels.

The fourth row in table (12) reports the estimated coefficients of interest for hospitals close to the OF border. Using a bandwidth of 1km, the model produces an estimate of -0.069 which is significant at the 5% significance level. Between 1.25km and 2.5km the estimated coefficient is large and positive. Notably the estimates at 1.75km, 2km and 2.25km are significant at the 10% confidence level. Meanwhile the estimate using the 2.5km bandwidth is positive and significantly different from zero at the 1% significance level. However the reliability of these estimates remain questionable given the uncertainly surrounding its estimation around the OF, as discussed above.

The main implication of the bandwidth adjustment exerise is that the estimated coefficients on the treatment effect are mostly stable for bandwidths between 1km and 2.5km for both the IO and OF border. This is a common consideration in RD estimation designs. Choosing smaller bandwidth may act to reduce bias but will also increase the variance of the proposed estimators. Conversely a higher bandwidth may serve to reduce variance but may also introduce some bias into the proposed point estimates. In choosing a bandwidth We face another trade off between number of observations and minimising the probability that economic conditions vary between hospitals. Increasing the bandwidth, encompasses more observations due to the inclusion of more hospitals. Therefore, choosing a bandwidth too small reduces the number of hospitals and consequently the sample of job vacancies. However, choosing a bandwidth too large increases the chance of induced endogeneity as hospitals are likely to experience different conditions that are difficult to otherwise account for.

9 Limitations

In this section We critically evaluate the empirical methodology employed in the section above and provide some insight into possible directions of future research to account for these concerns. There are three limitations of the methodology We employ. Namely We discuss factors that may induce endogeneity in the treatment variable at the IO border, issues surrounding the construction of the dependent variable and possible improvements that could be made to the text classification model.

We start by discussing possible factors that may invalidate the sharp RD approach We employ. Notably We discuss other ways in which hospitals my differ from one another in ways that, because of a lack of available data, cannot be controlled. Hospitals on separate sides of the border may vary in other capacities, other than different HCAP requirements, that otherwise determine the desirability of job postings for job seekers. One such factor could be the existence of differences in economic and labour market conditions between hospitals. Such factors may include demand for NHS services which could result from local health conditions and crime rates. These factors are uncontrolled for and may differ between different PCT zones, given that PCT zones correspond with London borough boundaries. Therefore, the coefficient on the treatment variable could capture effects of crossing London borough boundaries rather than exogenous wage differences. Furthermore hospitals may differ in prestige which would naturally influence job applicants demand for a given hospital. However, We would argue that the use of a sharp RD mitigates these concerns relative to the use of OLS, as discussed in section 6.

We also note that the construction of the job vacancy duration τ_i as discussed in section 5.2 is open to critical discussion. The first issue with this variable is that the behaviour of vacancy durations does not solely depend on application numbers as some jobs have a fixed closing date which is independent of the number of applications. Therefore, this variable may not capture the number of applicants for a particular position, or the quality of those applicants. We do account for this possibility by including jobs that are closed early, which likely results from finding a suitable candidate, but this does not incorporate all jobs. Another argument in defence of this construction is that firms could anticipate the length of time a job vacancy should be opened to gain enough adequately qualified candidates. In this case some of the issues in using this variable are mitigated.

Finally We suggest that improvements could be made to the machine learning, text classification model used to derive job title controls. While Bayesian classification models often perform well for text classification tasks, Xu (2018), they rely on some strict assumptions that are unlikely to strongly hold. Notably Bayesian classification relies on the assumption that all features in $\mathbf{x}$ are mutually independent, conditional on the class C_k. This suggests that given a particular job category $k \in \{Admin, ..., Doctors, Nurses\}$, the probability of a given word occurrence must be independent of each other word in $\boldsymbol{x}$. Given that the construction of sentences in the English language follows predictable patterns, this assumption is unlikely to hold. For future research logistic regressions, as employed by Abi Adams-Prassl (2020), may be an alternative for this text classification task. This may help to improve the models recall, precision and F1 score, thereby improving the robustness of the analysis.

10 Conclusion

This paper looks to analyse the relationship between wages posted by firms and job vacancy durations. To investigate this question theoretically, We adapt a model of directed search, first derived by Shimer (2005) to introduce homogeneous jobs which reflects the fact that job seekers menu of jobs are strictly limited according to the NHS 'band' system. We also introduce exogenous wage, a random exclusion parameter and geographical immobility to better understand the dynamics between posted wages and job seeker behaviour in the structure of the NHS. By adapting Shimer's (2005) model We find that there exists a positive relationship between a jobs posted wage and the number of applicants that a job advert receives. This corresponds to a negative relationship between posted wages and vacancy durations. In the context of homogenous firms this relationship holds without the use of detailed job heterogeneity controls. The model also predicts that increasing wages may also act as a mechanism to improve the composition of high-productivity applicants that apply to an open vacancy.

This paper utilises NHS job vacancy data, collected from the job posting website *findajob.co.uk* between April 2021 and March 2023, in conjunction with a sharp border discontinuity in London High Cost Area Payments (HCAP's), to analyse the causal relationship between posted wages and the duration of vacancies. Evidence from the existing body of empirical literature reference difficulties in controlling for firm and job heterogeneities. By using a large firm with a strictly defined 'band' system, my identification strategy aims to bypass this concern.

We find that differences in centrally determined pay scales translate into actual differences in posted wages for NHS workers at both the inner-outer and outer-fringe border, although the effect is more pronounced at the former than the latter. This is a result of a poor balance of hospitals at the OF border. Using a parametric sharp border regression discontinuity design, We find vastly different estimates for the effect of posted wages on vacancy durations when using the inner-outer and outer-fringe border. At the inner-outer border We find large negative and significantly non-zero estimated wage-duration elasticites. Conversely, when using the outer-fringe border We find small, positive but non-significant estimates. Both reported estimates rely on a bandwidth of 1.5km and a linear relationship between the assignment and dependent variables. We suggest that the differences in estimates may be a result of NHS trust boundaries, which coincide with the OF HCAP zones, as well as factors negatively influencing geographical mobility between the Outer and Fringe PCT zones. This hypothesis is reflected by the directed search model We present in section 4 which suggests that when geographical immobility cannot be accounted for in econometric specifications, there could exist a positive relationship between wages and vacancy durations.

However, We note that the estimates presented are not robust to changes in functional form which may necessitate the use of non-parametric methods in future research. In the absence of such analysis We present the linear estimation as the most robust measure given the small bandwidth of 1.5km. Estimated elasticities are robust to changes in adjustments in the bandwidth.

We implement a placebo test to evaluate the validity of the negative estimates produced using the inner-outer border. In doing so We shifting both the inner-outer and outer-fringe border inwards uniformly by 1km and 2km. At both fictitious borders for each $h \in \{IO, OF\}$, the model does not replicate the results using the original boundary, where the wage discontinuity actually occurs. This lends weight to the argument that the negative significant result generated using the sharp RD approach at the inner-outer border is driven by the HCAP discontinuity.

References

Abi Adams-Prassl, Maria Balgova, M.Q., 2020. Flexible work arrangements in low wage jobs: Evidence from job vacancy data. *Iza discussion papers*, (13691).

Aggarwal, C.C. and Zhai, C., 2012. A survey of text classification algorithms. *Mining text data*, pp.163–222.

Aiken, L.H., Clarke, S.P., Sloane, D.M., Lake, E.T. and Cheney, T., 2008. Effects of hospital care environment on patient mortality and nurse outcomes. *The journal of nursing administration*, 38(5), p.223.

Angrist, J.D. and Lavy, V., 1999. Using maimonides' rule to estimate the effect of class size on scholastic achievement. *The quarterly journal of economics*, 114(2), pp.533–575.

Armstrong, T.B. and Kolesár, M., 2018. Optimal inference in a class of regression models. *Econometrica*, 86(2), pp.655–683.

Banfi, S. and Villena-Roldan, B., 2019. Do high-wage jobs attract more applicants? directed search evidence from the online labor market. *Journal of labor economics*, 37(3), pp.715–746.

Belot, M., Kircher, P. and Muller, P., 2018. How wage announcements affect job search-a field experiment. *Cepr discussion paper no. dp13286*.

Black, D.A., 1995. Discrimination in an equilibrium search model. *Journal of labor economics* [Online], 13(2), pp.309–334. Available from: `http://www.jstor.org/stable/2535106`.

Buchan J, Gershlick B, C.A.S.I., 2019. *Falling short: the nhs workforce challenge. health foundation* [Online]. Available from: `https://www.health.org.uk/publications/reports/falling-short-the-nhs-workforce-challenge`.

Burdett, K. and Mortensen, D.T., 1998. Wage differentials, employer size, and unemployment. *International economic review* [Online], 39(2), pp.257–273. Available from: `http://www.jstor.org/stable/2527292` [Accessed2023-08-17].

Calonico, S., Cattaneo, M.D. and Titiunik, R., 2014. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6), pp.2295–2326.

Calonico, S., Cattaneo, M.D. and Titiunik, R., 2015. Optimal data-driven regression discontinuity plots. *Journal of the american statistical association*, 110(512), pp.1753–1769.

Carter, C., 2023. Wages and vacancy durations. *Univeristy of bath*, may.

Chen, S.F. and Goodman, J., 1999. An empirical study of smoothing techniques for language modeling. *Computer speech & language*, 13(4), pp.359–394.

Cubel, M., Nuevo-Chiquero, A., Sanchez-Pages, S. and Vidal-Fernandez, M., 2016. Do personality traits affect productivity? evidence from the laboratory. *The economic journal*, 126(592), pp.654–681.

Eeckhout, J. and Kircher, P., 2010. Sorting and decentralized price competition. *Econometrica* [Online], 78(2), pp.539–574. Available from: `https://doi.org/https://doi.org/10.3982/ECTA7953`.

Faberman, R.J. and Menzio, G., 2018. Evidence on the relationship between recruiting and the starting wage. *Labour economics*, 50, pp.67–79.

Francis, R., 2013. *Report of the mid staffordshire nhs foundation trust public inquiry: executive summary*, vol. 947. The Stationery Office.

Frölich, M. and Huber, M., 2023. Including covariates in the regression discontinuity design. *Iza discussion papers*, (11138).

Greaves, E. and Sibieta, L., 2019. Constrained optimisation? teacher salaries, school resources and student achievement. *Economics of education review*, 73, p.101924.

Greene, W.H., 2000. Econometric analysis 4th edition. *International edition, new jersey: Prentice hall*, pp.201–215.

Hahn, J., Todd, P. and Klaauw, W. Van der, 2001. Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1), pp.201–209.

Health and Social Care Committee, 2022. *Workforce: recruitment, training and retention in health and social care.*

Holzer, H.J., 1990. Wages, employer costs, and employee performance in the firm. *Ilr review*, 43(3), pp.147–S–164–S.

Ihsaan Bassier, A.M. and Petrongolo, B., 2023. Vacancy duration and wages. *Iza discussion papers*, (16371).

Imbens, G.W. and Lemieux, T., 2008. Regression discontinuity designs: A guide to practice. *Journal of econometrics*, 142(2), pp.615–635.

Kelly, E., Stoye, G. and Warner, M., 2022. *Factors associated with staff retention in the nhs acute sector* [Online]. Institute for Fiscal Studies. Available from: `https://doi.org/10.1920/re.ifs.2022.0216`.

Kibriya, A.M., Frank, E., Pfahringer, B. and Holmes, G., 2005. Multinomial naive bayes for text categorization revisited. *Ai 2004: Advances in artificial intelligence: 17th australian joint conference on artificial intelligence, cairns, australia, december 4-6, 2004. proceedings 17*. Springer, pp.488–499.

Kim, S.B., Han, K.S., Rim, H.C. and Myaeng, S.H., 2006. Some effective techniques for naive bayes text classification. *Ieee transactions on knowledge and data engineering* [Online], 18(11), pp.1457–1466. Available from: `https://doi.org/10.1109/TKDE.2006.180`.

Klaauw, W. Van der et al., 1997. *A regression-discontinuity evaluation of the effect of financial aid offers on college enrollment*, 97. New York University, Faculty of Arts and Science, Department of Economics.

Koopmans, T.C. and Beckmann, M., 1957. Assignment problems and the location of economic activities. *Econometrica: journal of the econometric society*, pp.53–76.

Korting, C., Lieberman, C., Matsudaira, J., Pei, Z. and Shen, Y., 2023. Visual inference and graphical representation in regression discontinuity designs. *The quarterly journal of economics*, p.qjad011.

Kuhn, P. and Mansour, H., 2014. Is internet job search still ineffective? *The economic journal* [Online], 124(581), pp.1213–1233. Available from: `https://onlinelibrary.wiley.com/doi/abs/10.1111/ecoj.12119`.

Larrabee, J.H., Janney, M.A., Ostrow, C.L., Withrow, M.L., Hobbs, G.R., Burant, C. et al., 2003. Predicting registered nurse job satisfaction and intent to leave. *Jona: The journal of nursing administration*, 33(5), pp.271–283.

Lee, D.S. and Lemieux, T., 2010. Regression discontinuity designs in economics. *Journal of economic literature*, 48(2), pp.281–355.

Lewis, D.D., 1992. *Representation and learning in information retrieval*. University of Massachusetts Amherst.

Marinescu, I. and Wolthoff, R., 2020. Opening the black box of the matching function: The power of words. *Journal of labor economics*, 38(2), pp.535–568.

McHugh, M.D., Kutney-Lee, A., Cimiotti, J.P., Sloane, D.M. and Aiken, L.H., 2011. Nurses' widespread job dissatisfaction, burnout, and frustration with health benefits signal problems for patient care. *Health affairs*, 30(2), pp.202–210.

Menzio, G., Telyukova, I.A. and Visschers, L., 2016. Directed search over the life cycle. *Review of economic dynamics*, 19, pp.38–62.

Moen, E., 1997. Competitive search equilibrium. *Journal of political economy* [Online], 105(2), pp.385–411. Available from: `https://doi.org/10.1086/262077`.

Montgomery, J.D., 1991. Equilibrium wage dispersion and interindustry wage differentials. *The quarterly journal of economics*, 106(1), pp.163–179.

Morgan, B., 2022. NHS staffing shortages. *The kings fund* [Online]. Available from: `https://www.kingsfund.org.uk/publications/nhs-staffing-shortages`.

Mueller, A.I., Osterwalder, D., Zweimüller, J. and Kettemann, A., 2018. Vacancy durations and entry wages: Evidence from linked vacancy-employer-employee data. *National bureau of economic research*.

NHS, 2023. Nhs terms and conditions of service handbook [Online]. Available from: `https://www.nhsemployers.org/publications/tchandbook`.

NHS Pay Review Body, 2022. Nhs pay review body thirty-fifth report 2022 [Online]. Available from: `www.gov.uk/government/publications/nhs-pay-review-body-thirty-sixth-report-2023`.

Rennie, J.D., Shih, L., Teevan, J. and Karger, D.R., 2003. Tackling the poor assumptions of naive bayes text classifiers. *Proceedings of the 20th international conference on machine learning (icml-03)*. pp.616–623.

Rivett, G., n.d. *The history of the nhs*. Available from: `https://bolt.nuffieldtrust.org.uk/health-and-social-care-explained/the-history-of-the-nhs`.

Shanafelt, T., Balch, C., Bechamps, G., Shanafelt, T., Bechamps, G., Boone, S., Tan, L., Shirom, A., Nirel, N., Vinokur, A. et al., 2019. 79 shiftwork and health. *Cambridge handbook of psychology, health and medicine*, 11, p.357.

Shapley, L.S. and Shubik, M., 1971. The assignment game i: The core. *International journal of game theory*, 1, pp.111–130.

Shi, S., 2002. A directed search model of inequality with heterogeneous skills and skill-biased technology. *The review of economic studies*, 69(2), pp.467–491.

Shields, M.A., 2004. Addressing nurse shortages: what can policy makers learn from the econometric evidence on nurse labour supply? *Economic journal* [Online], 114(499), November, pp.464–498. Available from: `https://ideas.repec.org/a/ecj/econjl/v114y2004i499pf464-f498.html`.

Shimer, R., 2005. The assignment of workers to jobs in an economy with coordination frictions. *Journal of political economy*, 113(5), pp.996–1025.

Statista, 2022. The world's biggest employers. Available from: `{https://www.statista.com/chart/3585/the-worlds-biggest-employers/}`.

The Health Foundation, 2020. 2019 annual report and financial statements. Available from: `https://www.health.org.uk/publications/2019-annual-report-and-financial-statements`.

Thistlethwaite, D.L. and Campbell, D.T., 1960. Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of educational psychology*, 51(6), p.309.

Wolthoff, R., 2018. Applications and interviews: Firms' recruiting decisions in a frictional labour market. *The review of economic studies*, 85(2), pp.1314–1351.

Xu, S., 2018. Bayesian naïve bayes classifiers to text classification. *Journal of information science*, 44(1), pp.48–59.

Ying, Y., Mursitama, T., Shidarta, S. and Lohansen, 2021. Effectiveness of the news text classification test using the naïve bayes' classification text mining method. *Journal of physics conference series* [Online], 1764, 02, p.012105. Available from: `https://doi.org/10.1088/1742-6596/1764/1/012105`.

Zhang, H. and Li, D., 2007. Naïve bayes text classifier. *2007 ieee international conference on granular computing (grc 2007)* [Online]. pp.708–708. Available from: `https://doi.org/10.1109/GrC.2007.40`.

A Included Hospitals

Inner London	Camden Mews Day Hospital, Charing Cross Hospital, Chelsea and Westminster Hospital, Great Ormond Street Hospital, Guy's Hospital, Hammersmith & Fulham Mental Health Unit, Hammersmith Hospital, Homerton Univerity Hospital, Hospital Of St John & St Elizabeth, Hospital Of St John & St Elizabeth, King's College Hospital, Lambeth Hospital, Lewisham Heather Close, Maudsley Hospital, Mile End Hospital, Queen Mary Hospital, Queen Mary's House, Royal Brompton Hospital, Royal Free Hospital, Royal Marsden Hospital, Springfield University Hospital, St Bartholomew's Hospital, St Charles Hospital, Mental Health Unit St George's Hospital, St Leonard's Hospital, St Mary's Hospital, St Pancras Hospital, St Thomas' Hospital, The Gordon Hospital, The Royal London Hospital, The Royal London Hospital For Integrated Medicine, The Royal Marsden Hospital (London), The Whittington Hospital, University College Hospital, University College Hospital at Westmoreland Street, University Hospital Lewisham
Outer London	Wanstead Hospital, Barking Community Hospital,Barnes Hospital, Barnet Hospital, Bethlem Royal Hospital, Bridgeways Day Hospital, Cassel Hospital, Central Middlesex Hospital, Chase Farm Hospital, Clayponds Rehabilitation Hospital, Croydon University Hospital, Ealing Hospital, East Ham Care Centre, Edgware Community Hospital, Elderly Day Hospital, Eltham Community Hospital, Erith and District Hospital, Finchley Memorial Hospital, Gateway Surgical Centre, Goodmayes Hospital, Green Parks House, Hadley Wood Hospital, Harefield Hospital, Hillingdon Hospital, Hornsey Central Neighbourhood Health Centre, King George Hospital, Kingston Hospital, Memorial Hospital (Greenwich), Mount Vernon Hospital, New Victoria Hospital, Newham General Hospital, Nightingale Hospital, North Middlesex University, Hospital Orpington Hospital, Princess Royal University Hospital, Purley War Memorial Hospital, Queen Elizabeth Hospital, Queen's Hospital, Royal National Orthopaedic Hospital, Spire London East Hospital, St Mark's Hospital, St Michaels Hospital, St. Ann's Hospital, St. Helier Hospital, Sutton Hospital, Teddington Memorial Hospital, The Royal Marsden Hospital (Sutton), Tolworth Hospital, Victoria Hospital (Romford), West Middlesex University Hospital, Whipps Cross Hospital Wilson Hospital
Fringe London	Ashford Hospital, Basildon Hospital, Brentwood Community Hospital, Broadmoor Hospital, Cheshunt Community Hospital, Cheshunt Community Hospital, Darent Valley Hospital, Dorking Hospital, East Surrey Hospital, Epsom & Ewell Cottage Hospital, Epsom General Hospital, Farnham Hospital & Centre For Health, Frimley Park Hospital, Gravesend And North Kent Hospital, Harpenden Memorial Hospital, Haslemere Hospital, Heatherwood Hospital, Hemel Hempstead Hospital, Kingsley Green Hospital, Lakeside Medical Diagnostics - Billericay Community Hospital, Leatherhead Community Hospital, Milford Hospital, Molesey Hospital, New Queen Elizabeth II Hospital, Princess Alexandra Hospital, Queen Victoria Memorial Hospital, Royal Surrey County Hospital, Royston Hospital, St Albans City Hospital, St Mark's Hospital, St Peters Hospital, The Old Cottage Hospital, Thurrock Community Hospital, Upton Day Hospital, Watford General Hospital, West Park Hospital, Wexham Park Hospital, Woking Hospital, Wokingham Hospital.

Table 13: Hospitals in inner,outer and fringe London

B Sharp Border Discontinuity Theory

Assume we want to determine the effect of some treatment D_i on a continuous variable Y_i. Consider a continuous variable X_i which defines the individuals who are treated, denoted by D_i. Therefore, D_i is a deterministic function of X_i,$D_i = \mathbf{1}(X_i >= c)$, where c is the cutoff value.

Consider the linear relationship between Y and X:

$$Y_i = \beta_0 + \beta_1 D_i + \beta_3 X_i + \epsilon_i \tag{62}$$

We can center the assignment variable at c using the specification:

$$Y_i = \beta_0 + \beta_1 D_i + \beta_3 (X_i - c) + \epsilon_i \tag{63}$$

We can therefore, define:

$$E[Y_i|D_i = 1, X_i = c] = \beta_0 + \beta_1 \tag{64}$$

$$E[Y_i|D_i = 0, X_i = c] = \beta_0 \tag{65}$$

Therefore, the average treatment effect at the border $X_i = c$is defined by:

$$E[Y_i|D_i = 1, X_i = c] - E[Y_i|D_i = 0, X_i = c] = \beta_0 \tag{66}$$

Note that in this case there is no interaction term. If we account for the varying effect of the X on vacancy durations. The specification can be written as:

$$Y_i = \alpha_0 + \alpha_1 D_i + \alpha_2 (X_i - c) + \alpha_3 (X_i - c) \times D_i + \eta_i \tag{67}$$

In this case α_1 is the treatment effect at $X_i = c$ since at $X_i = c$, $X_i - c = 0$.

C An Example Job Advert on findajob.gov.uk

Clinical Research Nurse - Critical Care | Guy's and St Thomas' NHS Foundation Trust

Posting date:	08 August 2023
Salary:	Not specified
Additional salary information:	£42,471 - £50,364 p.a. inclusive of HCA
Hours:	Full time
Closing date:	07 September 2023
Location:	London, SE1 7EH
Company:	Guys and St Thomas NHS Foundation Trust
Job type:	Permanent
Job reference:	5424427/196-BRC1992

Summary

Are you looking for a new challenge to progress you career within Research nursing? Do you think you could contribute to a busy and dynamic clinical research team? We are looking for a dynamic and highly motivated nurse to join our friendly Critical Care research team. You should have excellent communication and interpersonal skills and have the ability and initiative to work independently. You should be able to manage your workload flexibly across a range of projects and deadlines in a fast-pace working environment.

You will be an essential member of the Critical Care research team, leading a growing portfolio of observational research and clinical trials of investigational medicines & devices within this specialty. The role will include supporting the delivery of commercial and non-commercial studies. You will work closely with a dynamic multi-disciplinary research team to deliver our portfolio of studies. You will be required to liaise with trial site coordinators for site initiation visits, site visits and site closures as appropriate. You will be required to attend research meetings on a regular basis.

There are excellent opportunities for further education and research development as part of these roles, with access to CPD funding and revalidation support. Opportunities for developing new research and clinical skills are available and actively encouraged.

Figure 6: An Example Job Advert on *findajob.gov.uk*