Mind the Gap! Essays examining the impact of skill deficiencies on the UK economy at differing aggregation levels

by

David Peter Morris

A thesis submitted to the University of Sheffield in partial fulfilment of the requirements for the Degree of Doctor of Philosophy in the Department of Economics

September, 2014

Abstract

Workers that fulfil the current skill demands of firms are an essential component of a correctly functioning labour market. A recent survey conducted by the Prince's Trust and HSBC interviewed over 600 business leaders from firms with more than 500 employees in all major sectors. In excess of 40% of these business owners said they were already experiencing skills gaps within their firms, while more than half also faced difficulties filling vacancies. What is possibly more concerning for the UK economy is that over two-thirds of the business leaders (68%) said they held fears that skills shortages would slow down any form of economic recovery, with an alarming 35% harbouring fears that these skill shortages would cause their business to fail.

Skill deficiencies appear to be a major issue in the United Kingdom, both in terms of their number, and the lagging economic performance that is blamed on them. Despite anecdotal evidence, and logical arguments, that these deficiencies are having a negative impact on the economy, very little empirical evidence exists. The principle aim of this thesis is to examine the impact that skill deficiencies are having on the UK economy at differing levels of aggregation. This will include looking at the effect in terms of the economy as a whole, from the perspective of establishments and from the perspective of employees.

Chapter 2 investigates how establishments respond to Hard to fill Vacancies (HTFV) and Skill Gaps (SGs) within the premise of a low skill equilibrium framework. Generalized Poisson and Zero Inflated Negative Binomial models are fitted to a count of establishments' positive responses, using data from the 2005, 2007 and 2009 waves of the National Employers Skill Survey. The results indicate that most establishments do not respond in a negative way (cutting jobs) to HTFVs or SG. This suggests that a low skill equilibrium is not an economy wide issue. There are some occupations that seem less likely to respond to both internal and external skill gaps, and these areas are the intermediate skilled jobs where Skill Biased Task Change and job polarisation have been seen. Most noticeably a negative response has been seen in the following occupations; 'Administrative', 'Skilled Trades' and 'Machine Operatives'. In these intermediate skill occupations there does seem to be evidence of negative employer responses to skill deficiencies, which may suggest a downward skill trajectory exists for this subsection of the economy.

Chapter 3 constructs an establishment-level dataset by matching productivity data from the Annual Business Inquiry (ABI) to skill information from the National Employers Skill Survey (NESS). The effect of internal skill gaps on productivity is then examined for the successfully matched workplaces. An IV estimator is used to remove the endogeneity of inputs problem, with lagged labour costs used as an instrument. Result from this regression analysis are contrary to prior expectations and suggest that the selection issue, whereby establishments that are more likely to have a skills deficiency may be the more productive workplaces, may be more significant that first believed. Due to this selection issue both a Heckman two-step model and Propensity Score Matching are investigated. The results suggest that

while skill gaps are not having a positive effect on productivity (as was originally suggested by the IV results), they are not having any significant impact.

Chapter 4 attempts to establish whether there is a causal relationship between the level of skill deficiencies in a market and the remuneration workers receive in that market. In order to do this information on skill deficiencies from the Employers Skills Survey (ESS) is matched at an industry/occupation level to the Labour Force Survey (LFS) over the period 2006 to 2012. This information is used in a pseudo panel framework to test if wages in markets with greater skill deficiencies are higher than those in markets with lower skill deficiencies. To validate this result regressions were run on an individual's wage, with market level skill information matched in from the ESS. This allows us to further test if workers in markets with higher skill shortages receive a higher wage. An additional effort is made to decompose which workers gain (or lose) the most from being in a market with skill deficiencies. The results suggest that high levels of SSVs in a given market have a positive impact on a worker's wage. The analysis on SGs shows varying sized coefficients, but continuously suggests that SGs do have a significant negative effect on wages in the given market.

This thesis concludes that skill deficiencies do not have the clear negative effect on the UK economy that has been anecdotally claimed. There is no negative effect found on establishments' performance, and while intermediate occupations could be at risk of a downward spiral of skill attainment, the rest of the economy is not. It is believed that while a negative impact has not been seen in this work, this is more likely due to data limitations. Care should be taken when drawing implications from this work and further analysis should be conducted in regards to this topic.

Acknowledgements

I would like to take this opportunity to express my greatest thanks to my two supervisors, Dr. Steve McIntosh and Professor Andy Dickerson, for their guidance, time, and continual positive support throughout each stage of my PhD. Further to this, I would like to express my thanks to everyone else in the Department of Economics at the University of Sheffield for their friendship and support throughout my studies at Sheffield, especially those who always stopped by to share a conversation or a drink - Ian Gregory Smith.

In addition I would like to thank Genna Kik, Susannah Constable and all the other members of the UK Commission for Employment and Skills for their assistance and feedback throughout the process of writing this thesis, and the UKCES for the financial support provided during my PhD. I also feel it fitting to thank the ESRC for financial support both throughout my PhD and during my Masters degree which has allowed me to pursue my career of choice.

I thank all my friends that have kept me sane and reminded me there is a light at the end of the tunnel throughout this period, namely William Platt, Andreas Papamichail and Joshua Longshaw. I would also like to thank my friends and peers who have battled academia alongside me; Damon Morris, Luke Munford, Helen Bell in particular, who have laughed, cried, stressed, grown and more than anything succeeded with me during these past few years. Thanks also go to Guyanné Saul, for her continued support, motivation and care.

Finally, and by no means least, I thank my family: my parents, Harold and Adele, and my sister Kathryn, for their love and support - throughout my studies, but also throughout my life. Having a kind and caring family behind you makes the tough times in life seem much more surmountable and I will forever be grateful for that.

Notes and Disclaimers

All three of my chapters use data from the *Employers Skills Survey* (ESS). The data used in these chapters were made available by the UK Commission for Employment and Skills. The data were originally collected by the Commission and I am thankful for their use, and the guidance I have received throughout my time working with these data sources. The Commission for Employment and Skills do not bear any responsibility for the analysis or interpretations presented here.

Chapter 3 is based on data from the *Annual Respondents Database* (ARD) 1973-2009, produced by the Office for National Statistics (ONS) and supplied by the Secure Data Service at the UK Data Archive. The data are Crown Copyright and reproduced with the permission of the controller of HMSO and Queen's Printer for Scotland. The use of the data in this work does not imply the endorsement of ONS or the Secure Data Service at the UK Data Archive in relation to the interpretation or analysis of the data. This work uses research data sets which may not exactly reproduce National Statistics aggregates.

Contents

List of Tables

List of Figures

Glossary

This glossary gives a short guide to the key terms used throughout this thesis. Most of the skill definitions are taken from the Employers Skill Survey as the skills information in this work is predominately taken from this series of surveys.

Establishment -	A single location of an organisation with people working at it. Also referred to as workplace, business, employer, site
Hard-to-fill vacancies (HTFV) -	Vacancies which are proving difficult to fill, as defined by the establishment (from question: "Are any of these vacancies proving hard-to-fill?").
Skill-shortage vacancies (SSVs) -	Vacancies which are proving difficult to fill due to the establishment not being able to find applicants with the appropriate skills, qualifications or experience. This is sometimes referred to as an external skill shortage as the gap exists outside the establishment.

Skills gaps (SG) - A "skills gap" is where an employee is not fully proficient, i.e. is not able to do their job to the required level. This is sometimes referred to as an internal skill gap as the gap exists within the establishment. If respondents asked for clarification, then a proficient employee was described as 'someone who is able to do their job to the required level'. 'Proficient employee', however, is clearly a subjective and relative term and can vary given the perspective of the manager completing the survey and can vary as the job changes over time.

Skill deficiency - A generic term used to represent either an internal or external skill problem, i.e. can represent both SSVs and SGs.

Up-skilling - A need for employees to gain new skills, beyond the requirements of their current job role, in order to perform their role in future.

Product Market Strategy (PMS) - An establishment's PMS score is worked out from the combined answers of four questions: How customised their output is; How price dependent their offering is; How innovative the establishment is; Whether outputs are premium or basic quality. A high PMS score would indicate outputs are customised, not price-dependent, premium quality and the establishment often leads the way in product development.

Skill Gap Density - The number of staff reported as being not fully proficient as a proportion of all employment.

Sector - For definitions of the different sector groupings used in this work the 14 sectors of the 2007 one digit Standard Industrial Classification (SIC) SIC were used.

Occupation - For definitions of the occupational groups used in this work the nine major (one-digit) Standard Occupation Classification (SOC) 2010 categories were used.

Chapter 1

Introduction

1.1 Motivation and Aims

1.1.1 What are Skills?

Workplace skills are expertise and capabilities in a given activity or occupation. While there are a variety of differing skills it is possible to group them into three distinct categories; basic skills, generic skills and specific skills. Basic skills are those such as literacy and numeracy which provide a grounding for almost all jobs. Generic skills such as communication and team working, are applicable in most jobs. Specific skills are those that only apply to the given task being conducted and thus are less transferable between jobs and occupations. Labour market roles involve differing subsets of skills, with many jobs requiring some level of attainment in all of the previously mentioned categories.

While skills are easily defined as an individual's ability to undertake a task, they are more difficult to measure. Education and qualifications are very common measures of labour market skills, with those who have completed a higher level of qualification often being referred to as 'highly skilled'. It is important to note that information such as this is not a perfect measure for skills, and is rather a proxy. Education usually involves the acquisition or communication of information or knowledge, while skills refer to the ability to perform certain tasks. These concepts obviously have a large amount of overlap, with those individuals who have higher education or qualification levels often being able to perform or learn more complex tasks, but they do still have their differences.

Besides education and academic measures of a worker's ability numerous other skill proxies exist, such as work-related training, vocational qualifications, apprenticeships and professional qualifications. These all provide different, but useful, proxies for skills. These proxies play an important signalling role in a properly functioning labour market, not just matching suitable workers into the correct jobs, but also aiding career progression. Further to these formal types of skill measures there are also self-assessed task based skills such as computing ability, problem solving and communication and there is literature surrounding these elements which addresses the fact that not all skills can be measured by formal qualifications or exams.

While evaluating and assessing these different types of skills can be difficult, two facts seem clear from the large amount of literature in the area. Firstly, higher skills are related to more productive employees. Secondly, creating and balancing the right skills is one of the key drivers for economic growth in a nation.

1.1.2 Why Skill Deficiencies?

While the issue of skills is clearly important, the flip side of the coin, concern about skill deficiencies in an economy is both less investigated and more intriguing. By measuring and comparing skills in comparative economies, a useful picture of the education system and the incentives in the respective nation to up-skill begins to emerge. However, no knowledge is gained of what skills are actually needed.

On the contrary, measuring skill deficiencies on the contrary gives both a picture of the current skill situation in the economy as well as the desired skills in the labour market. For a skill deficiency to exist, demand for an employee with the stated skill level must exist, and this type of employee must not be available. This measure thus includes both the demand for skills in the economy from workplaces and also where improvements in skill levels need to be made. This alternative measure of skills is related more directly to current labour markets than standard measures such as skill levels.

Due to the difficulty in having to measure what skills are required as well as what skills are missing, very few pieces of work investigate the issue of skill deficiencies. To cope with these concerns, the preferred measure of information on skill deficiencies is workplace-reported. This allows the employers to give reliable information on the skills that are both currently unavailable and desired.

The dual nature of skill deficiencies does mean that their interpretation is not straightforward. While deficiencies are informally associated with diminished economic performance, there is an increasing likelihood of skill deficiencies appearing as a nation's output increases, as a growing economy creates an increasing demand for both workers and skills. Due to this, any analysis of skill deficiencies over time needs to be firmly grounded in the context of the current economic climate.

Two differing but complementary forms of skill deficiency exist and distinguishing between them is important. These measures reflect the interior and exterior nature that skill deficiencies can take. Skill deficiencies where a workplace or employer cannot find an adequately skilled worker to fill a vacancy are seen as external deficiencies and commonly referred to as skill shortage vacancies (SSVs). They are external as the deficiency exists outside of the workplace in question. The second type of deficiency is a form of internal deficiency and are referred to as skill gaps (SGs). They are internal as they are the deficiencies that exist within a workplace when an employee does not possess all the skills required to do their job.

The idea that skill deficiencies are an important measure of labour markets is further supported by the skill mismatch literature. In Europe, over-education is averaging around 30% and yet skill deficiencies are still seen (CEDEFOP, 2012). These contrary findings suggests that skill deficiencies can be seen due to classical reasons (such as insufficient training and education, skill-biased technological progress and business cycles), but also when people accept low-skilled jobs while continuing to search for a better match. The idea of skill mismatches being bad for productivity is also highlighted (CEDEFOP, 2012), where firms are forced to place lower-skilled workers into skilled positions, or with workers using their strong position in a market with skill deficiencies to alter their employment terms and conditions in a way that may harm productivity. It is thus suggested that "greater care in job matching may... not only improve the welfare of over-skillled employees, but also have positive effects on productivity and growth in the economy". This skill mismatch literature reinforces the idea that high levels of skills in an economy does not mean that there are a high number of skill matches. Thus, a measure of skill deficiencies can help to remove the issue of the supply of skills, and rather focus on how well the supply of skills meets the current demand for skills.

1.1.3 Skills and Skill Deficiencies in the UK

Workers that fulfil the current skill demands of firms are an essential component of a correctly functioning labour market. A recent survey conducted by the Prince's Trust and HSBC interviewed over 600 business leaders from firms with more than

500 employees in all major sectors. In excess of 40% of these business owners said they were already experiencing skills gaps within their firms, while more than half also faced difficulties filling vacancies. What is possibly more concerning for the UK economy is that over two-thirds of the business leaders (68%) said they held fears that skills shortages would slow down any form of economic recovery, with an alarming 35% harbouring fears that these skill shortages would cause their business to fail.

Prior work on the UK's skill situation was conducted by the Leitch Review of 2006, which considered the nation's longer term skill needs. The Review found that the nation's skills were not world class and this could play a role in undermining the UK's long term productivity- productivity which has lagged behind comparator countries. While the report was written as the UK was coming out of 14 years of continued growth and was sporting the lowest unemployment rate of the G7, the picture of skills in the nation was still not entirely positive.

While the review was not too shy to find fault with the UK's current situation, there are some positives that do come out of the work. Lord Leitch and the review believed that the possible up-skilling that is called for could result in a boost to the productivity growth rate of 15% and an increase in employment growth rate by around 10%. This is estimated to lead to decreased social deprivation, decreased poverty, reduced inequality as well as leading to a possible benefit of at least £80 billion to the UK economy over a period of 30 years.

One of the key principles that the review suggests should underpin this raise in ambition is a push towards demand-led skills whereby "the system must meet the needs of individuals and employers" with vocational skills being "demand-led not centrally planned". The fact that this is such a prominent point highlights that it is skill deficiencies, especially those identified by employers, rather than skill levels which are of interest to the UK's growth.

Other anecdotal reports have painted a bleaker visions of the UKs current skill situations. A report titled 'Skills for Growth' (2012) by the Engineers Employers Foundation (EEF) highlights that "manufacturers are reliant upon a shrinking population of highly skilled, older workers to fill the jobs that need to be filled now, are are staring at a yawning skills gap on the horizon". The organisation claims that despite weak UK growth and high unemployment, four in five firms are reporting recruitment difficulties and most of these are due to skills reasons (a lack of technical skills or experience). These skill deficiencies are seen to be doubly problematic. Firstly they prevent long term growth and recovery from the current recession. Secondly as firms looks to develop new markets, launch new products and introduce new services in aim to become more competitive in global markets the demand for skilled employers must be met.

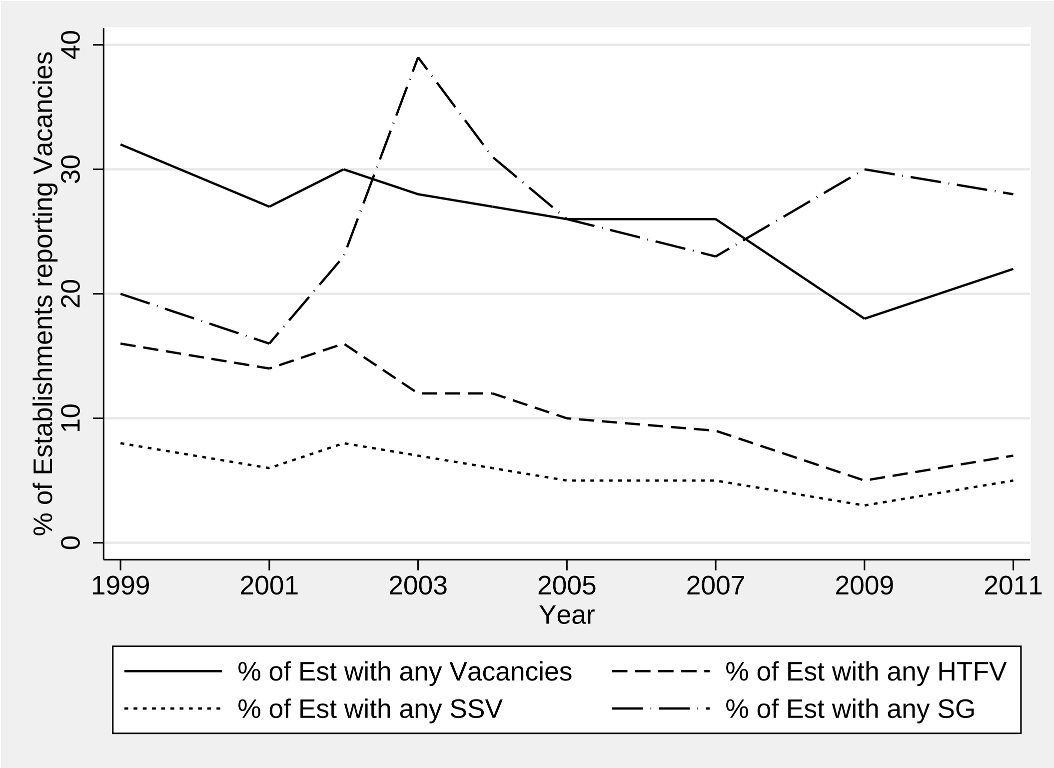
Similar reports have also recently been seen by the Institute of Directors (IoD) and by Confederation of British Industry (CBI). IoD's recent report, 'Shackled by the Skills Crunch' (2010), claims that UK business are being held back by skill weaknesses with three-fifths of employers experiencing skill deficiencies claiming this has impacted their growth, and 80% of employers that suffer from SGs believe this will impact their ability to capitalise on any economic recovery.

The recent CBI report, 'A better off Britain' (2014), makes the case that skill deficiencies due to rising skill requirements are having a negative impact on both individuals and economic growth in general. The report highlights the fact that the number of jobs requiring no formal qualification has halved over the past 10 years and by 22 50% of jobs will require employees to have completed at least some form of higher education. This rising skill requirement means that individuals are facing ever more pressing forces to up-skill. The fact that many firms report skill shortages suggests that people are not up-skilling and the CBI report claims this fault lies heavily on the vocational education system which makes it difficult for people to move into these jobs, either in the case of young people getting onto the employment ladder, and adults re-training. Adult retraining is seen by the CBI to be of particular importance as it will ensure that people are not left behind by the current up-skilling of jobs.

While the above reports provide an interesting insight into skills from an employers' perspective, the links they draw between skills and growth are untested. Evidence from the US (Cappelli, 2014) suggests that while concerns about the supply of skills are often raised from employer-associated organisations there is little evidence consistent with these complaints. Further to this, the relative weight of skills gaps vis-à-vis other factors such as poor investment in capital, research and development spending, monopsony power in markets and lagging UK infrastructure are missing.

The employers reports provide a good snapshot of the current thoughts in UK labour markets however when it comes to numerical evidence the current picture of skill deficiencies in the UK is far from simple. The Employers Skills Survey (ESS) provides invaluable information on the skills situation in the UK with information collected from a sample of all establishments which are larger than just a working propri-

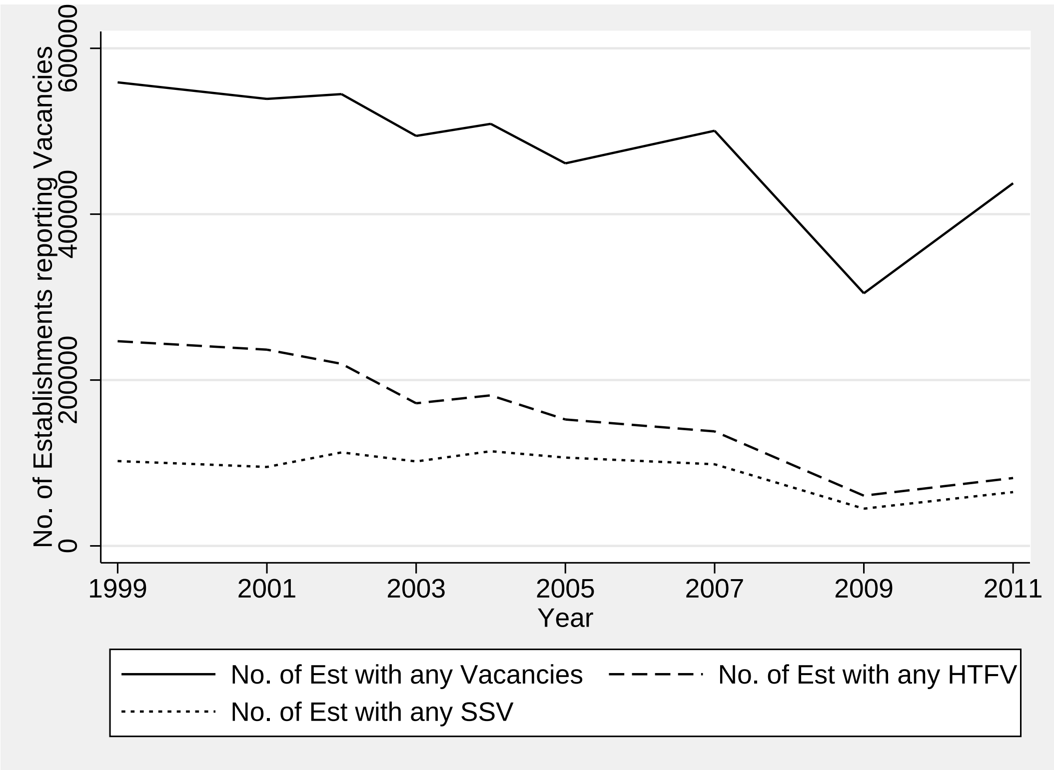
Figure 1.1: **Percentage of establishments with any skill deficiencies**



etor. The following three figures are created from the ESS data, which have been transformed to give a consistent survey strategy over the 12 year period of 1999 to 2011.

The proportion of establishments reporting vacancies over the period was fairly constant at around 30% from 1999 to 2007 with a significant drop seen from 2007 to 2009, and a slight recovery in 2011. A similar trend can be seen in the percentage of establishments that report HTFV with around 14% seen throughout the waves and a drop in 2007 to 9% and even further to 5% in 2009 before rising to 7% in 2011. The proportion of establishments reporting any SSV is again constant at around 7% with a smaller but still noticeable drop in 2009 to 3% rising again to 5% in 2011. The number of HTFV has to be lower than the number of vacancies as a vacancy must exist for it to be classified as hard to fill. In a similar way a HTFV must exist for a SSV to exist.

Figure 1.2: Number of skill deficiencies in the UK Economy

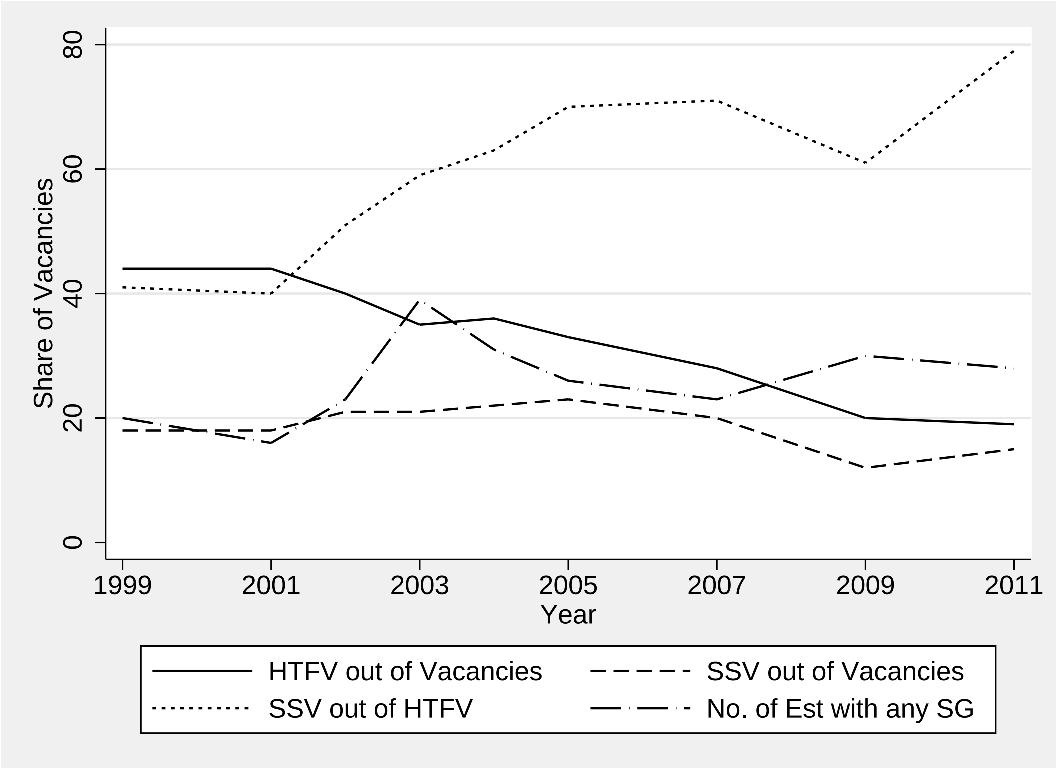


The decrease seen in all the categories addressed above could possibly be explained by the start of the recession in the UK. With establishments not looking to expand, vacancies are likely to decrease and as HTFV and SSV by nature depend on the number of vacancies we would expect these to decrease as well.

The total number of vacancies, HTFV and SSV reported, not just how many establishments reported them can be seen graphed in Figure 1.2. They show that there has been a fall in all three types of vacancy in a similar way to the number of establishments that reported vacancies until 2011 where all three start on an upwards trend.

Due to the interconnected nature of the vacancy types, with SSV being driven to an extent by HTFV (which are in turn driven by vacancies), an effort is made to disentangle the relative changes in these deficiencies. To do this the share of

Figure 1.3: Ratio of differing skill deficiencies



HTFV that are SSV and the share of vacancies that are HTFV are plotted in Figure 1.3. While most of these ratios are falling, the share of SSVs out of HTFVs has increased showing that while the actual number of SSV is falling. This means that SSVs are taking up a greater share of the HTFVs and are thus becoming a more significant problem when establishments are trying to fill vacancies. This is even more prominent in 2011 with a large increase in the share of SSV out of the total number of HTFV.

1.1.4 Possible effects of Skill Deficiencies

While it seems clear that skill deficiencies are a prominent factor in the UK's labour markets, very little work has been conducted on determining the impact that these gaps are having. This is particularly surprising given that skill deficiencies could be impacting the UK at every level of aggregation.

The Leitch Review and anecdotal evidence often blames skill deficiencies as one of the main reasons that the UK's productivity is consistently lagging behind comparator nations and thus a shortage of adequately skilled workers would have an impact on the UK economy as a whole.

Further to this, it is expected that a lack of skilled workers will have an impact on establishments, both in terms of their productivity and in an increasingly global economy, their survival. Skill deficiencies can stop establishments from pursuing new high-tech working practices and ultimately stop them from becoming producers of high-end, specialist goods.

The final level of aggregation at which skill deficiencies are likely to have an impact is on individual workers. Workers with gaps in their skills are likely to have greater difficulty both finding employment and staying employed. Those workers with skill gaps that do find employment are also likely to receive lower remuneration than their job normally entails due to not being able to perform the job to the full level required.

1.1.5 Aims of the Thesis

This thesis aims to provide evidence of the impact that skill deficiencies are having on the UK economy at the three levels of aggregation mentioned above. Hopefully this information will allow a more detailed picture of the UK's skill situation and skill needs to be developed, and will aid future decisions on both whether more should be done to tackle skill deficiencies and how this is best done in practice.

1.2 Structure and Content of Thesis

This thesis is divided into three main chapters, each looking at a different level of aggregation in the UK's labour markets, starting with the economy as a whole and moving onto look at establishments and finally workers. A brief outline of each of these three chapters can be found below.

1.2.1 Brief overview of Chapter 2

Chapter two aims to investigate the impact that skill deficiencies are having on the UK economy as a whole and thus investigates to see if there is evidence of a low skill equilibrium in the UK. It tests how establishments respond to HTFVs and SGs, as this is one of the key links in the low skill equilibrium framework. If establishments remove jobs that are characterised as being HTFV then a downward skill spiral could exist with workers having less incentive to train for these jobs and more HTFV being created. On the other hand if this link is not seen then it is unlikely that a low skill equilibrium can be formed.

1.2.2 Brief overview of Chapter 3

Chapter three attempts to provide evidence on the effect that skill deficiencies have on establishment productivity by creating an establishment level matched dataset of the Employers Skill Survey and the Annual Respondents Database. This chapter then tests if establishments with higher levels of SGs are more or less productive. A selection issue whereby those establishments which are more likely to have a SG may have different productivity levels could also exist and this effect is also investigated.

1.2.3 Brief overview of Chapter 4

Chapter four aims to identify the effect that skill deficiencies are having on workers' remuneration by using a pseudo panel set up after performing an industry/occupation match of the ESS and the Labour Force Survey. It tests to see if those individuals that are in markets with higher levels of skill deficiencies receive different wages. This is done through both individual level and market level analysis.

1.2.4 Why these particular questions

As has already been outlined, skill deficiencies are an under evaluated area of the UK's labour force. Due to this, there are more questions than could be considered in this piece of work. The questions outlined above are chosen for two main reasons.

Firstly, these topics where chosen due to the applied nature of this piece. As the work was supported financially in part by the UK Commission for Employment and Skills (UKCES) having questions that could have direct impacts on policy were key. These topics tied into the work done by the UKCES at the time of writing and helped to link academic research in labour economics to current policy debate.

Secondly, all three topics have a focus around the headline statistics of HTFV, SSV and SGs that were used in the ESS. This information is the most appropriate and readily available data on skill gaps and by using these reoccurring measures of skill deficiency a unique picture can be drawn of the UK's labour market.

While numerous other questions could have been investigated none of them seemed to have such broad reaching impact and obvious policy recommendations as these topics.

Chapter 2

How do Establishments Respond to Skill Deficiencies in the Labour Market? An Investigation into whether a Low Skill Equilibrium Exists in the UK

2.1 Introduction

The idea that a Low Skill (LS) equilibrium may exist in the United Kingdom has been the issue of a long standing debate throughout the 1980s (Gospel (1998) and Steedman et al. (1991), the 1990s (Finegold (1999)) and into the new millennium (Leitch (2006). While recent OECD findings (Indicators (2011) show that the UK is doing relatively well in terms of high skilled workers (level four and above) it is lagging behind comparator countries when intermediate skills are added (level two and above)[1]. While educational attainment is not a perfect proxy for skills the notion that an equilibrium could exist with fewer intermediate skills than is optimal in the UK is a real possibility. While previous work seems to be mainly theoretical this paper intends to add clarity to, firstly, whether a LS equilibrium does exist and secondly, what characteristics are associated with those firms that drive the downward spiral. It does so by investigating how establishments respond to both HTFVs and internal SGs.

A generalized Poisson model is estimated using establishments' positive responses to HTFVs, allowing the under-dispersion seen in the dependent variable to be accounted for. For the internal SG analysis a Zero Inflated Poisson model is used to take account of the large number of zero responses. The paper proceeds as follows. Section 2 presents the background associated with LS equilibrium. Section 3 gives an overview of the theory that may underline establishments' responses. Section 4

[1] "OECD education at a Glance" (2011) shows that the UK ranks 9th out of 34 OECD countries in the proportion of 25-64 year olds qualified to level 4+ with 37%, above the OECD average (30%). In terms of those qualified to level 2+ however the UK ranks 19th out of 33 with a percentage of 74%, only just above the OECD mean of 73%. While the ranking for those with 4+ level skills has been increasing over the past five years the UK's position in the rankings for 2+ level skills has remained between 17th and 19th.

details the National Employer Skills Survey data used in this work. The generalized Poisson model, zero inflated negative binomial model and binomial probit regressions used in the work are outlined in Section 5. Section 6 provides and discusses the resulting estimates. The paper concludes in Section 7.

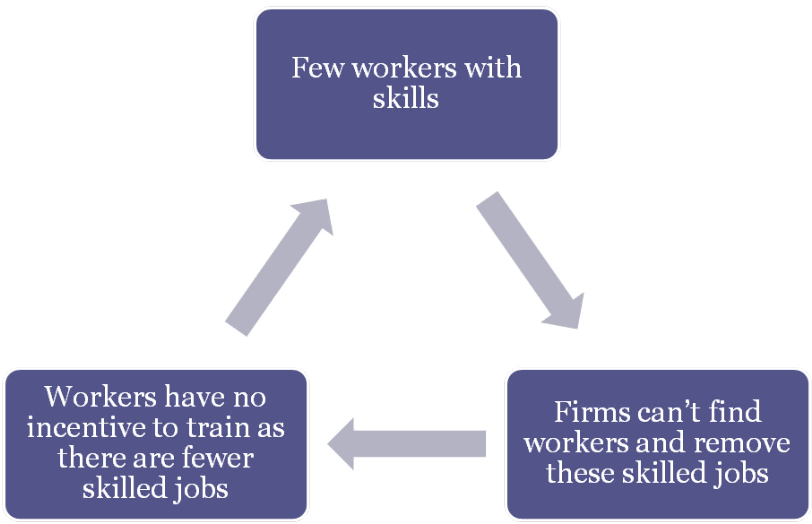
2.2 Background

2.2.1 Defining a Low Skill Equilibrium

The term "Low Skill Equilibrium", originally coined by Finegold and Soskice (1988), refers to the trap "in which the majority of enterprises are staffed by poorly trained managers and workers produce low-quality goods and services" (pg. 22). The low skills seen in the UK had originally been attributed to historical and cultural factors. The conceptualizations of a LS equilibrium provided a more logical and satisfying explanation for the circumstances the UK exhibited. While many (Donovan and Britain (1969) pg. 92), suggested that the reason for the UK's poor skill levels was due to supply problems (not enough skilled works) others argued that it was in fact demand problems (with firms not requiring skilled workers). The LS equilibrium framework tied together both the demand and supply side pressures to create a dynamic process by which a labour market moves along a spectrum of sorts towards fewer skilled workers. This is predicted to occur when the supply of trained workers in the market is insufficient to fill demand. Establishments cannot fill vacancies and reduce the number of skilled jobs available. As the demand for trained workers decreases so does the incentive for workers to gain these skills. This reduces supply further, creating a self-perpetuating cycle towards a LS equilibrium, Figure 2.1.

Individuals and employers react in a rational way to the environment they face, resulting in them acquiring low skills rather than high skills. This is a market failure as rational behaviour at the micro level leads to a large proportion of the UK economy being involved in low specification work at the macro level. Low specification work such as this is often associated with low wage jobs and little value added in the production process, thus having a negative effect on the economy. The Leitch Review estimates a possible net benefit of at least £80 billion over 30 years from up-skilling across the UK as well as a fairer society with a more competitive and mobile labour force.

Figure 2.1: Low Skill Equilibrium



2.2.2 Previous Findings

In their seminal work Finegold and Sockice (1988) outline the evidence that suggests Britain is in a low skill situation when evaluated next to its international comparators. Whilst most work investigating the failure to train details the free riding and poaching problems this does little to explain the variations between countries that has been noted by the authors. The reasons for the British system to have fallen behind its comparators are discussed by the authors as is the issue of why the government failed to take corrective action where the market was failing. These arguments are based around not only around the education system but also the political-economic institutions such as the financial markets and the organization of industry.

Finegold and Sockice note that the low skills seen in Britain are both "the cause and consequence" of the nation's poor economic performance. A consequence as the education and training systems in the country evolved to meet the demands of the world's first industrialized nation, and a cause as the absence of trained workers has made it difficult for the workforce to respond to changing economic conditions and challenges. This is the first suggestion that a cycle or trap may exist with the current situation dictating the future of skills in the nation.

Further to Finegold and Sockice's work suggesting that a trap could be apparent in the UK, Burdett and Smith (2002) outlined a simple matching model with rent sharing to better explain the skills trap. This model is based around the principle that while trained workers do not gain the full return to skill acquisition, due to rent sharing, they do still gain and thus still have some incentive to train. Firms also gain from skilled workers because of rent sharing and thus skill acquisition induces a higher rate of job creation and thus improves the probability of matching prospects.

The authors claim that this principle leads to two possible equilibria emerging. If workers up-skill then firms receive higher profits per worker and this higher return leads to an increase in the number of skilled vacancies. With skilled workers acquiring jobs quicker the return to having skills is higher, justifying the decision to pursue training. The alternative equilibrium occurs if workers fail to pursue training causing the relative profit per worker to be low. This limits a firm's incentive to post skilled vacancies positions, jobs became scarce and the returns to skills are lower. This is seen as the low skill equilibrium trap. A model such as this shows how Britain's past history of low skilled, mass production industry can mould future skill levels creating a low skill cycle where skills are not demanded and thus no incentive to train exists.

While the topic of debate and political concern, little academic work has looked to test empirically the idea of a low skill equilibrium. One of the few pieces that does aim to assess the available evidence on a low skill trap is Wilson and Hogarth (2003) in their review for the Department of Trade and Industry. In this work the authors conduct numerous company case studies to investigate product market strategies and skill deployment with a view to establish how the latter works to influence the educational system. This work has a both a regional and sectoral element, focusing on the food processing and business hotels and hospitality areas, to determine differences across the nation but its main UK wide aim was to determine how a shift can be made from lower to higher value-added production.

Their conclusions on this matter are not promising for the UK economy as a whole with no quick fixes being apparent. This is mainly due to the current product market strategies of firms being successful in terms of business and profits. The fact that firms are making rational decisions, but decisions that may not help the economy as a whole, makes it more difficult to change their trajectories and product market strategies. On top of this is a secondary problem where there is likely to be a trade off between a high skills trajectory and the low levels of unemployment seen currently in the UK, especially during the short run while the economy adapts. How strong this trade off is and how long the economy takes to react may determine how worthwhile pursuing a high skill equilibrium may be. Another point to note from this is that the authors suggest that supply is not the problem and the industries investigated in these case studies have processes and jobs that leave limited room for increased levels of skills. Therefore policies that just aim to increase skills, that seem to have been pushed for numerous years, are relatively pointless until the product market strategies of establishments are changed.

The only noticeable work that looks at actual firm responses to skill shortages is Fang (2009). He uses the Workplace and Employees Survey in Canada to investigate the adaptations of various workplace practices to both vacancies and skill shortages. Using both linear and probit models he finds that employers in this survey focus more on short-term and less costly solutions, with no evidence suggesting that either a raise in employee wages or fringe benefits would help reduce shortages. While knowing *how* workplaces respond is obviously important, *why* they respond in this way is one of the key, and possibly manageable, links in the LS equilibrium cycle. While it is not possible to explicitly test the LS equilibrium, if it is possible to identify the characteristics of firms that are responding in a certain way to vacancies policies can be put in place to avoid a LS equilibrium pathway.

2.3 Establishments' Response Decision

2.3.1 External Skill Gaps

Vacancies are not typically filled instantaneously. There is an amount of time spent by suitable employees 'searching' for the job and it is only after a position has been vacant for a reasonable[2] amount of time that it becomes a HTFV. Whilst there could be numerous reasons for a job to remain vacant, bad working conditions, poor salary and lack of future prospects to name but a few, some are undoubtedly left vacant due to the inability of the establishment to find appropriately skilled employees[3].

It is only at the point that vacancies become HTFV that establishments become aware that action may have to be taken to fill the vacancy. How establishments respond obviously dictates what type of trajectory the economy will embark on. The NESS defines these HTFV as any vacancies "that are proving hard to fill". This question is a follow up question asked to all establishments that report a vacancy.

Assuming that an establishment has a HTFV, as we will throughout this work, and that they operate in a rational manner then a response will be made if the expected

[2] HTFV in the National Employers Skills Survey are defined as vacancies that the establishment classifies as hard-to-fill. This therefore means that the length of time a vacancy is left before becoming a HTFV could differ by workplace based on their own expectations.

[3] The National Employers Skills Survey which is being used for this analysis suggests that around 2/3 of HTFV vacancies are actually due to skill shortages.

cost of the response is less than the expected cost of the vacancy remaining empty in that period. The establishment's decision is therefore an attempt to minimize costs around the level of output they wish to achieve.

As Meager (1986) notes, costs and responses should be considered together as the major cost of a shortage is responding to it (either by alleviating the shortage or accommodating it). The only other cost to establishments is a premium for over-responding, either by paying a higher wage than necessary or recruiting lower-skilled workers than desired.

It is assumed that there are four key elements that affect the outcome of the response decision. Firstly establishments are more likely to respond and respond in a stronger way, the further they are from their optimal employment level. Under the previous assumption that establishments only respond if the cost of responding is less than the cost not responding, then the filling of a vacancy must increase profits for the establishment. Obviously transpiring from this, the more workers an establishment is away from their optimal level, the further they are from profit maximizing. If an establishment is one employee short some of this loss in output may be covered by other employees etc. If, however, the establishment has multiple gaps, these are likely to accentuate each other meaning the firm has to bear the full cost of the vacancies. This is likely to increase the cost of not responding pushing more establishments to undertake some sort of action.

The second reason for establishments to respond in a stronger manner is the relative competitiveness of the environment. If an establishment is in a competitive environment it has a greater need to be close to the optimal level of employment. In a competitive environment any gap between optimal output and the establishments' real output will be filled by one of the competing firms. While this may only be a one period loss any decrease in market share is likely to be harder to regain in a more competitive market. Therefore the reduced output and loss associated with it in a competitive environment may be associated with greater costs in the long-run increasing the cost of not responding.

Another reason for establishments to respond is the simple matter of ease of response. Whilst this may seem almost too obvious there are numerous firm characteristics that could affect ease of response and thus the firm's likelihood of responding. For example a larger establishment may have more resources available to increase advertising for a given HTFV and may have a human resources team that are in a better position to direct a response. A contrary argument for establishment size is larger firms have more regulations in place to ensure that all decisions taken will benefit the firm. The paperwork and decision making process associated with this may reduce the ease of making a response and thus responses may be greater in smaller establishments.

Finally, it would be expected that those establishments that make the most profit per worker respond more to HTFVs. If an establishment makes more profit per worker then for each missing employee the firm is likely to react more as the cost of not responding is higher. While establishment size may affect the profit per worker due to monopsony power, so could the skill level of the employees at the establishment. It has been noted that establishments make greater profits from skilled workers in terms of their relative productivity to wage ratio[4]. This is possibly due to the difficulties quantifying their output giving them less wage bargaining power, wage compression, or other reasons (Acemoglu and Pischke (1999)). For these establishments, more profit is lost per job vacancy and thus the cost of responding is more likely to be outweighed by the cost of no action.

2.3.2 Internal Skill Gaps

While skill deficiencies can be external matters, as seen above with firms trying to recruit skilled workers, internal skill gaps also exist with members of staff not being fully proficient at their jobs. These SG pose similar problems to external skill shortages. There is once again a period of time before it is recognized that a SG exists and then the establishment has a chance to respond to the problem. The question in the NESS is "how many of your existing staff would you regard as fully proficient at their job?" and thus SG are classified as those where the individual is not fully proficient at their job.

Assuming once more that establishments have a SG then a rational response would be to undertake action if the cost of the response is less than the full cost of the SG. The same four factors that are seen to affect the HTFV response decision discussed earlier should hold true again for SGs as it is only the response decision that is being investigated. Establishments are more likely to respond or respond in a stronger manner if they are further away from their optimal level. Rather than as a

[4] Dearden et al. (2006) show that when training increases by 1% there is an associated rise in value added per worker of 0.6% and and increase in wages to the employee of only 0.3%, *half* of the increase in productivity.

percent of HTFV, this is dictated by the percentage of non-proficient staff employed at the establishment. Again it would be expected that a small SG can be covered by the rest of the workforce but multiple or major gaps will have a knock on effect to productivity swinging the response decision further towards responding.

Similar logic dictates that the remaining three factors; competitiveness of the environment, ease of response and the nature of the worker with the deficiency will all play roles in the SG response decision.

2.4 Description of Employer Data

The aforementioned hypotheses are tested using the ESS (also referred to as the NESS). The ESS is the largest employers' survey of its kind in the UK with the 2009 wave involving around 80,000 establishments. It is a representative cross sectional survey covering establishments of all sectors and sizes. The 2009 (most recent), 2007 and 2005 waves of the sample are used in this work. As the recession appears to decrease the number of vacancies reported by establishments in the 2009 wave of the survey this will hopefully allow trends to be seen before the recession as well as any effect on responses due to the recession. There are a total of 3,185 establishments reporting HTFVs in the 2009 wave of the data, with 6,323 and 6,838 in 2007 and 2005 respectfully. The total number of responses in each wave can be seen in Table 2.2. As this work investigates the impact of establishments responses to HTFV only those with HTFV are included. Obviously as a vacancy is required for this vacancy to be hard to fill there are no establishments included in the dataset that have no vacancies.

While an establishment may respond in numerous different ways to a HTFV (see Table 2.3) or SG (Table 2.4) all the responses in the survey can be grouped into two distinct sects. These are identified throughout as 'positive' and 'negative'. Positive responses are defined as those where the establishment attempts to fill the deficiency in some way. They are referred to as positive as in many cases they involve improving the characteristics associated with the job, and in all cases mean that the job still exists. This type of response repels the start of a LS equilibrium, as if a job still exists people still have the incentive to up-skill and seek employment there.

Negative responses are the opposite, where the reaction of the establishment results in the destruction of the job. When this occurs the labour force will have no motivation to gain skills, as there are fewer skilled vacancies to fill, kick starting the LS equilibrium cycle.

The list of responses establishments have given to HTFVs can be seen in Table 2.3, divided into both positive and negative responses with Table 2.4 showing the same for SGs. A few of the responses in the survey are vague and could therefore appear in either the positive or negative groups. These responses are grouped in a separate section at the end of each table. If both positive and negative responses were included in the response variable it would be measuring contrary effects and is likely to give confusing results. Due to this, and the fact that the number of negative responses is low, only the positive manner in which establishments respond to skill deficiencies will be used in this work.

To model an establishment's response, a count variable is created which represents the number of positive responses an establishment has to a HTFV. While responses are classified as either positive or negative a scale still exists with establishments able to respond in numerous strengths in both directions. Therefore, the strength of an establishment's response is measured by how many positive responses they undertake. The fact that the variable is a count, means that the values that the dependent variable can take are strictly non-negative integers; $y = 0, 1, 2...$, where y is the number of responses.

As previously mentioned, some of the responses could be classed as positive or negative. To ensure that the inclusion of these variables had no effect on the outcome of the results they were all included in a larger count variable, with a smaller count variable constructed only including the responses that were definitely positive. Results throughout will be based on the smaller of these counts with the larger count results and figures included in Appendix 2 for robustness. The smaller HTFV count variables can be seen from Figure 2.2-2.4 for the years 2009, 2007 and 2005. There are a low number of zeros in the count due to relatively few establishments not responding to HTFVs. This has caused both count variables to be under-dispersed, with a mean greater than their variance. This under-dispersion is seen as significant with the Cameron and Trivedi (2005) (pp.670-671) dispersion test.

Independent variables are included in line with the proposed theory with region and sector dummies also added as controls. The percentage of HTFVs in the establishment is used and expected to have a positive relationship to the count, as those with more HTFVs are likely to respond more. A private sector dummy variable (private=1, other=0) is also added as it is expected those in competitive environments have more responses. The type of market the establishment competes in is included as a series of dummy variables to help model this with International (International=1, non-international=0), National (National=1, non-national=0) and Regional (Regional=1, non-regional=0) used, leaving local as a base category.

A series of dummy variables were included in the same way as the market variables above to determine the impact of the quality of the product. The levels are defined as Very High Quality, High Quality, Medium Quality, Low Quality and the base category of Very Low Quality. It is expected that establishments that produce higher quality products have employees with higher skill levels, and thus may lose more per vacancy, giving higher quality a positive effect on responses.

Establishment size is included as a continuous variable to determine what impact this has on responses to HTFV. In theory, the effect of size is ambiguous with concerns that factors such as the amount of 'hoops' the establishment has to jump through could affect the ease of response. 'Human Resource Sophistication' and 'Job Description' are included as dummy variables. Human resource sophistication is a dummy variable that is equal to one if the establishment answers positively to any of the following; having plans for future training, having a budget for future training or a business plan (see Appendix 1 for full question details). Job description is again a dummy variable equal to one if more than 50 percent of the employees have a formal job description and zero otherwise. Using these should help to separate the effect of more organized establishments from establishment size.

The percentage of HTFVs in the establishment that are SSV (SSV rate) is included. The SSV rate is used to see if establishments with more skill shortages react differently to other establishments.

Finally, a set of occupational dichotomous variables are included to indicate whether the establishment has a SSV in the given occupation (Occupation has a SSV=1, Occupation has no SSV=0). These are included to try to establish if certain occupations are responding differently to skill shortages. Each establishment can report skill deficiencies in any one of the nine of the occupations, all the occupations, or none of them. Due to this, there is no base category excluded as the variables do not have any required interaction.

The details of all of these independent variables are summarized in Tables 2.4, 2.5 and 2.6. An identical procedure is undertaken with the internal SG problem, again with two counts being created to ensure that the allocation of ambiguous responses has no effect on the results. The distributions can be seen in Figures 2.5 to 2.7. The distribution of these counts is very different from the dispersion shown in the HTFV figures with a much greater number of zeros. This means that the count variable does not have the same under-dispersion that was previously seen.

Independent variables are again included in line with the proposed theory with very few differences to the HTFV equation. The only two variables that differ between the two analyses are the measure of how far an establishment is from its optimal level and the occupational dummy variables. In the HTFV work the percentage of HTFV is used to show how far the establishment is away from the optimal level. In the SG work a SG density is used instead, with the percentage of employees with a skills gap being used. Also, in the HTFV work a dummy is included to denote whether or not a SSV exists in the given occupation. In the internal work a dummy is instead included if the establishment has a SG identified within the given occupation. The details of all the SG variables can be seen in Tables 2.7 to 2.9.

2.5 Modeling

2.5.1 Generalised Poisson Model

As the dependent variable for the HTFV analysis (number of responses to a HTFV) is a count variable a natural starting point is the Poisson regression model. The main assumption in the Poisson model is that the mean of the count is equal to the variance, which has already been disproved for this count as with many "real life" applications. Winkelmann and Zimmermann (1994) show that the Poisson regression model is not appropriate if a data set shows over-dispersion and similar logic dictates the same for the rarer case of under-dispersed data.

Instead, the Generalized Poisson Model (GPM) developed in Famoye (1993) is used. While this model is similar to the normal Poisson model it allows the mean to vary by establishment and includes an alpha term, known as the dispersion parameter. With a normal Poisson model the count is expected to decrease monotonically from zero to the highest relevant number. This is not the case in this dataset with the low number of zeros causing the highest count to instead be one, see Figures 2 to 4. By using the GPM and allowing the mean to vary by establishment this can be accounted for creating a much better fitting model.

Following Famoye (1993), the probability density function of y_i, the number of positive responses an establishment has to an HTFV, is given by Equation 2.1;

$$f_i(y_i; \mu_i, \alpha) = \left(\frac{\mu_i}{1+\alpha\mu_i}\right)^{y_i} \frac{(1+\alpha y_i)^{y_i-1}}{y_i!} \exp\left(\frac{-\mu_i(1+\alpha y_i)}{1+\alpha\mu_i}\right) \tag{2.1}$$

With $\mu_i = \mu(y_i) = \exp(X_i\beta)$, and with X_i being a (k-1) dimensional vector of establishment variables and β is a k dimensional vector of regression parameters. In this model the mean is given by Equation 2.2;

$$E(Y_i|X_i) = \mu_i \tag{2.2}$$

And the variance by Equation 2.3;

$$V(Y_i|X_i) = \mu_i(1 + \alpha\mu_i)^2 \tag{2.3}$$

As can be seen from Equation 2.3 when $\alpha = 0$ then $V(Yi|Xi) = \mu_i$ and thus the variance reduces to equal the mean and the model reduces down to the Poisson Probability function. When $\alpha < 0$ the variance is less than the mean and the GPM represents data with under-dispersion. If $\alpha > 0$ then over-dispersion can be modeled. The alpha term is therefore known as the dispersion parameter and can be estimated simultaneously with the coefficients in the model.

To estimate (β, α) in the GPM we use the method of maximum likelihood (ML). This estimation technique identifies the parameter values which maximize the log-likelihood function. To do this the log-likelihood function thus needs to be derived and is done so by simply logging the probability density function in Equation 2.1 to give Equation 2.4;

$$LnL(\alpha, \beta; y_i) = \sum_{(i=1)}^{n} \left[y_i \log \left(\frac{\mu_i}{1 + \alpha\mu_i} \right) + (y_i - 1) \log(1 + \alpha y_i) - \frac{\mu_i(\alpha y_i)}{1 + \alpha\mu_i} - \log(y_i!) \right] \tag{2.4}$$

By applying the linear-form assumption, (Gould et al. (2006)) that the observations are not linked and so the log-likelihood contribution can be separately calculated, the log-likelihood function is all that is needed to derive parameter estimates.

Due to the GPM reducing to the Poisson model when $\alpha = 0$ a useful test of the GPM against the Poisson is a simple test of the significance of the dispersion parameter;

$$H_0 : \alpha = 0 \qquad against \qquad H_\alpha : \alpha \neq 0 \tag{2.5}$$

If H_0 is rejected then it is recommended to use the GPM rather than the Poisson model as while both may provide consistent results the Poisson Model will be inefficient when $\alpha \neq 0$. To conduct the test in Equation 2.5 the asymptotically normal Wald type t statistic can be used based on the estimate of α compared to its standard error.

The Akaike Information Criterion (AIC) is also used to test the different nested models against the Poisson. The AIC (Akaike (1998)) is a common test of the goodness of fit of a model which is defined as;

$$AIC = -2LnL + 2K \tag{2.6}$$

where K is the number of estimated parameters in the model and LnL is the log-likelihood value of the estimated model. Using this method of modeling goodness of fit the smaller the value of the AIC, the better the model as the better the log-likelihood has fitted given the number of parameters.

Numerous specifications are used throughout the work. The 2007 and 2005 waves of the data do not have all the variables seen in the 2009 wave and thus a smaller specification is made that is consistent across all the waves. A larger specification is also run on the 2009 wave of the survey to see if this affects any of the results. These varying specifications are also all run with and without both region and sector dummies to ensure that they are having the expected effect.

2.5.2 Zero-Inflated Poisson Model

It has been previously noted in this work that the SG data shows significant over-dispersion. Due to this the Poisson model, which is the base model for count data, will not be appropriate as this relies on the assumption that the mean is equal to the variance which is not true in this case. Instead a selection of other models were tested including a Negative Binomial (NB), a zero-inflated Poisson model (ZIP) and a zero-inflated Negative Binomial Model (ZINB). The NB model can be used for over-dispersed count data. It is a generalization of the Poisson regression, having the same mean structure as the Poisson regression but with an extra parameter to model for over-dispersion. The ZIP and ZINB models are further extensions to the Poisson and NB models where a term is introduced to model the probability of a zero occurring allowing the null responses to be modeled more effectively. While the

GPM used for under-dispersed data could have been used again with over-dispersed data numerous other models exist to model over-dispersion and these models appear more frequently in the count data literature so are tested here.

To test the fit of these models AIC and BIC were used, as well as likelihood ratio tests. The Bayesian Information Criterian (BIC), is another method of testing specifications or nested models and similar to the AIC. It again includes the log-likelihood value of the estimated model as well as a penalty term for the number of parameters in the model, meaning the likelihood cannot be increased too much by adding parameters and over-fitting. The penalty term is the BIC is much larger than the AIC and can lead to differing results. Again as with AIC a smaller BIC is better than a larger BIC. The formula for this can be seen below in equation 2.7.

$$BIC = -2LnL + KlnN \tag{2.7}$$

The results for all the models can be seen in Table 2.1, with the Generalised Poisson model included for comparisons sake. The results showed that the ZINB and ZIP models largely outperformed the Poisson and NB models in terms of AIC, BIC and plotted residuals. As all four of the models are nested, they come from the same family with just slight adjustments, they can also all be tested by likelihood ratio tests. While the GPM outperforms the Poisson and performs similar to the NB model, it lags behind the ZIP and ZINB models. This is expected as while the distribution is flexible in the GPM to allow for over and under-dispersion it is not able to handle the excessive number of zeros in the same way that the ZIP and ZINB models can.

Table 2.1: Testing Model Fit

Model	Poisson	NB	ZIP	ZINB	GPM
LL	-3668.675	-3619.702	-3469.106	-3463.896	-3619.73
AIC	7393.813	7303.934	7051.140	7048.786	7255.46
BIC	7351.349	7255.405	6966.213	6957.793	7303.99

When comparing the ZIP and the ZINB models the ZINB slightly outperforms the ZIP with an AIC of 2.354 and BIC of -18642.085 compared to the ZIP model's AIC of 2.187 and BIC of -18639.73. This shows that the difference between the two models is relatively small on both measures. While the ZINB model does give a slightly better fit for the data the ZINB method of maximum likelihood does not always converge to a maximum point with the data and thus the ZIP was chosen as the final model.

As previously noted, zero-inflated models can account for excess zeros by classifying some as "true zeros" and others as "excess zeros". Zero-inflated models then estimate both equations simultaneously, one for the count model and one for the excess zeros. This is how the ZIP works, with the count model based on the normal Poisson model and an extra weight added to measure the probability of predicting a zero value. This probability shows the difference between being a non-responder, with probability p, and a responder with probability $1-p$. This allows those establishments that would never respond and those that would respond but have not in this period to be accounted for.

The ZIP used here is based on Lambert (1992) and can be defined as follows;

$Y_i \sim 0$ with probability p_i

$Y_i \sim Poisson(\lambda_i)$ with probability $1 - p_i$

so that;

$Y_i = 0$ with probability $p_i + (1 - p_i) \exp^{\lambda_i}$

$Y_i = k$ with probability $(1 - p_i) \exp^{-\lambda_i} \lambda_i^k / k!$,

where k = 1, 2, ... and p is the proportion of establishments that will not respond whatever and due to the zero-inflated nature of the data will be assumed to be $0 \leq p < 1$.

2.5.3 Probit Model

To further understand how the responses may be correlated to each other a series of probit regressions are also estimated. A binomial probit model was used in this instance, as the dependent variable, whether the establishment had the given response, gives a dichotomous response of zero or one. The probit model which is based on the standard normal distribution constrains the predicted probability to be between zero and one and thus is a good fit for this type of dependent variable.

The top six responses that establishments have are each used in turn as the dependent variable with the other responses used as independent variables. This can be seen in the series of estimated equations below in equations 7 and 8.

$$\text{Prob} \quad y_i = \begin{cases} 1 : \text{Response A} \\ 0 : \text{Not Response A} \end{cases} = \beta_0 + \beta_1 \text{Response} B_i + \beta_2 \text{Response} C_i ...$$

$$... + \beta_3 \text{Response} D_i + \beta_4 \text{Response} E_i + \beta_5 \text{Response} F_i + \beta_6 X_i + \varepsilon_i \qquad (2.8)$$

$$\text{Prob} \quad y_i = \begin{cases} 1 : \text{Response B} \\ 0 : \text{Not Response B} \end{cases} = \beta_0 + \beta_1 \text{Response} C_i + \beta_2 \text{Response} D_i ...$$

$$... + \beta_3 \text{Response} E_i + \beta_4 \text{Response} F_i + \beta_5 \text{Response} A_i + \beta_6 X_i + \varepsilon_i \qquad (2.9)$$

Where X_i is a vector of establishment level characteristics and with the responses continuing to rotate through the equations until each have been the dependent variable.

These equations show the probability of the dependent variable occurring due to the other responses while controlling for a host of establishment characteristics. This was completed for both internal and external gaps.

While these regressions could obviously suffer from endogeniety due to the nature of the variables, simple correlations were run between the same variables to see if the results were of the same sign. Due to the results from these correlations showing almost the same signs to those predicted in the probit work any bias is likely to be small and a general impression of correlations can still be estimated from the probit results.

2.6 Results

2.6.1 Results over Time - External Skill Shortages

Looking first at the responses establishments have given, seen in Table 2.2 and Table 2.3, three important findings stand out. First, and possibly most importantly, the number of negative responses in the sample is much lower than the number of positive responses. Even if all the ambiguous responses are classed as negative, the number of positive responses still far outweighs them (3,034 to 513 in 2009 and a similar percentage in other waves). This suggests that the worry about a LS equilibrium forming should be relatively small as most establishments either do not respond to skill deficiencies or respond in a positive way, neither of which would result in an LS equilibrium.

The second point of note is that very few establishments in the entire sample respond in both positive and negative ways to HTFV or SGs (less than 1% in each wave).

While this makes sense in terms of not spending money advertising a vacancy if at the same time you are trying to automate the job, some overlap is still seen in the sample.

Finally, when comparing the responses seen in this survey to Fang (2007) there are considerable differences. In Fang's work he finds "there is no evidence that workplaces would raise employee wages or fringe benefits to alleviate shortages". The responses to HTFVs and SGs in Table 2.3 and Table 3 however show a very different pattern, with a large proportion of the responses involving fringe benefits such as training or raises in employee wages. This could be a reflection of the different circumstances the two nations face. The OECD (2011) findings show that the UK and Canada have similar levels of participation in more academic tertiary education with levels of 26% and 25% respectfully. However when looking at the more practical, technical side of tertiary education, where skills are gained for direct entry into the labour market Canada is well ahead of the UK, with a level of 24 to the UK's 10. The level of 24 seen by Canada is the highest amongst all recorded nations and suggests that maybe workplaces do not need to raise employee wages or benefits as the supply of skills is equal to or above the demand.

Table 10 present the results from the GPM for the smaller count variable. The table shows the parameter estimates, marginal effects and significance levels for all three waves of the data. It should be noted that the GPM could not find convergence on the 2005 wave of the data and thus the Poisson model was used instead with robust standard errors to give a prediction of the results. Estimates of the GPM and the Poisson model were assessed where both found convergence and the difference was negligible.

The alpha term in the regression, which represents the level of dispersion, has a negative and significant coefficient at the 1% level with both count variables in all the waves. The fact that this alpha is negative and significantly different to zero suggests that the GPM was the correct choice and is superior to a Poisson regression. This is further enforced by the fact that when Poissons were run for all key regressions the AIC was always smaller in the GPM than the Poisson model.

As can be seen from Table 10 many of the results are similar to the predictions made in the earlier theory about how establishments would respond, Section 3.1. 'HTFV' coefficient is negative and significant in the 2007 wave, but was insignificant in the 2009 wave. The marginal effects can be seen in the MFX column, with an average negative effect of around 0.2 for each HTFV an establishment has.

The SSV coefficient is not significant in either wave. Therefore, those establishments with a higher percentage of SSV do not respond more to HTFV. While both this and the HTFV were expected to provide positive results there may be some other factors that have caused this result. If, for instance, those firms that have the most gaps do so because of some inherent characteristics, such as lack of motivation, then this would be likely to cause the establishment to have fewer responses as well.

The one set of results that contrast the theory are those that refer to competitiveness. While the expectation was that those establishments in more competitive environments would respond more to HTFV the opposite appears to be true. The coefficient for private sector establishments is negative and is significant in 2009. This effect is large with those establishments in the private sector having 0.5 fewer responses to HTFV than those in other sectors. The use of a private sector dummy to capture competitiveness is obviously not perfect and this could be picking up other factors that are distinct between the two categories further to competition. No other variables were available, however and thus this was the best estimate of competition that could be provided.

Looking at the effect of establishment size over the waves the coefficient is always significant and positive. The marginal effects show an increase of between 0.6 and 1 response per 1,000 workers. Region and sector dummy variables were also tested but proved to have no impact. Due to this only a dummy for whether an establishment was in London was included in the final specification seen in Table 10. This London coefficient is seen to be insignificant in 2009 but positive and significant in the 2007 wave.

These regressions included dichotomous indicators to denote if a skill shortage vacancy existed in a given occupation. The results show that certain occupations are much more likely to respond to the SSV they face than others. Occupations such as Professionals, Associate Professionals, Personal Services and Elementary have large and significant positive effects on the response decision in both 2007 and 2009.

2.6.2 2009 Wave - External Skill Shortages

The previous specifications are limited to provide consistent results over the numerous waves of the survey. Regressions were also run separately on the 2009 wave of the data which had more variables available. These results can be seen in Table 11, with both count types shown.

The results appear similar to the earlier findings with the establishment size coefficient being positive and significant and the SSV percentage coefficient again not being significant. The coefficient on the private sector dummy variable, used to give a measure of competitiveness, is again negative, significant and large. All of these variables have marginal effects similar in size and of the same sign as those seen in the earlier regressions.

The interesting findings come with the inclusion of the quality dummies and the human resource sophistication dummy. Those establishments with high or very high quality products are seen as significantly different to those with very low quality products. These responses are again in the direction expected with establishments producing higher quality products having an estimated 0.5 to 0.6 more responses. The coefficient on the level of human resource sophistication is both positive and significant for both count variables. This fits the assumption that those establishments that find it easier to respond will respond more if it is assumed that having a HR department eases responding. This is again quite a large effect, with those establishments with human resource (HR) departments having an estimated 0.4 more responses.

The job description coefficient was also seen as positive but it was not significant in the large count. Also of interest is the fact that the coefficient on the HTFV variable is positive and significant in the smaller count model, the opposite of the previous results. None of the coefficients on the dummy variables used to capture what market the establishment competes in are seen as significant in either model.

2.6.3 Probit Analysis - External Skill Shortages

The results from the probit models can be seen in Tables 12, 13 and 14. The columns in these tables show the dependent variables with the independent variables displayed in the subsequent rows.

The 2005 results, Table 12, show most of the marginal effects to be negative for the responses. This means that by having one of the responses it makes it less likely that the dependent variable is reported. This is particularly true of the 'new recruitment channels' and 'new recruitment methods' responses which both have large, negative and significant coefficients and marginal effects for all the other responses. The only responses that gives a positive and significant response for the other responses is the 'increased advertising' dependent variable. This means that those establishments that respond in any of the other 5 main ways are also more likely to increase advertising as well.

The results from the 2007 and 2009 waves of the data show very similar findings with most of the responses having negative effects on the other responses. The 'London' and 'Private' dummy variables have an interesting relationship to some of the responses in 2007 and 2009 that they don't have in 2005. The private dummy variable has a negative and significant coefficient on half of the 2009 responses and two of the six responses in 2007. This is expected due to the private sector dummy being negative and significant in all of the GPM regressions. The London dummy variable's significance is slightly less expected though as it is not significant in the GPM work. The London dummy is also interesting as it has a negative and significant coefficient on 'New Recruitment Channels' but a positive and significant coefficient on 'New Recruitment Methods' in both 2007 and 2009. This means that establishments are less likely to respond by using new channels but more likely to use new channels.

The results suggest that very few of the responses are positively linked and those establishments that respond in more than one way are doing so more due to a desire to respond more than because one response leads to another. The opposite is in fact seen with most responses decreasing the likelihood of another response occurring.

Once again, dummies that identify if a SSV exists in a given occupation were added with the results seen in Table 16. Extremely similar results are seen again with the Managers, Administrative, Skilled Trades, Sales and Machine dummies being the only occupations that are not significant.

2.6.4 Results over Time - Internal Skill Gaps

The SG results, Table 15 show that the Establishment size coefficient has a positive and significant effect on the response decision, with larger workplaces responding more to a SG. The coefficient on the size squared variable used to pick up any non-linear effects is negative and significant. This shows that establishment size has a decreasing effect on responses, so while it is positive it is more important in smaller workplaces than larger ones. The private sector dummy is again negative and significant indicating those establishments in the private sectors respond less than those in the public sector.

The coefficient on the London dummy variable is of positive sign and can be seen as significant in 2007 but negative and significant in 2009. The SG density variable which shows how far a workplace is from its optimal level is negative and significant. This means that those establishments with the most SGs are responding less.

The results show the same occupations that had no significantly different response to having an external shortage again appear when investigating SGs. Most noticeably Managers, Administrative, Skilled Trades and Machine Operatives. These seem to be relatively consistent across both years but also count types.

2.6.5 2009 Wave - Internal Skill Gaps

The results from 2009 with a larger specification can be seen in Table 16. The establishment size coefficient is once again positive and significant, with size squared negative and significant. This suggests that workplaces are more likely to respond if they become bigger, though increasing at a decreasing rate. The coefficient on the private sector dummy variable is also again negative and significant. This means that establishments in the private sector are less likely to respond than those in the the public sector.

The coefficient on the four quality dummy variables that were included are all seen as significantly different to the base level of very low quality. Their sizes are also as expected the number of responses increasingly monotonically with quality. The job description coefficient was also seen to be positive and significant suggesting the those workplaces that have formal job descriptions are more likely to respond as expected.

Similar to the external skill gap analysis the market in which the establishment operates doesn't seem to affect the response decision. The SG density coefficient is significant and negative again, which is suggesting that those establishments with more internal gaps respond less.

The occupation results look extremely similar to those seen in the other SG work and close to those from the external skill shortage results. The coefficients on Administrative, Skilled Trades and Machine Operative occupations are all not significantly different from zero. This means that they do not respond more even when they have a SG in this occupation holding all other variables constant. While the managers occupation is seen as positive and significant, in these regressions its marginal effect is relatively small when compared to the other occupations. This suggests that the effect on the response decision is still small.

2.6.6 Probit Analysis - Internal Skill Gaps

The results from the probit models can be seen in Tables 17 and 18. The columns in these tables show the dependent variables with the independent variables displayed in the subsequent rows. There are no results in the increased training activity column as the probit regression would not converge with the same reason meaning that no table exists for the 2005 results. Most of the probit results for 2007 and 2009 show that there are negative relationships between the responses, meaning one response is likely to lead the establishment to not respond in another way. The few exceptions to this are 'more appraisals' and 'more supervision of staff' which have some significant and positive effects in both 2007 and 2009.

2.7 Conclusions

This work has investigated the ways establishments respond to HTFVs and the impact these findings have on the current diagnosis that the UK is caught in a LS equilibrium cycle.

The findings show that very few workplaces respond in a negative way and remove those vacancies that they cannot fill. While this response seems to being good news for the UK economy there is a possibility of a low skill equilibrium for intermediate skilled occupations, the same skills the UK is said to be missing.

When a more detailed view of how establishments respond is developed it seems that some key factors affect the response decision of workplaces for both internal and external skill gaps. Larger establishments and those establishments in the public sector are both likely to respond in a much more positive way to any deficiency. On top of this establishments that produce higher quality products are more likely to respond in a positive way to internal and external gaps.

Three occupations also always appear to respond less than any other in all the analysis. These are Administrative, Skilled Trades and Machine Operatives. The fact that the UK's low skill equilibrium was suggested to be in intermediate skill jobs, the same occupations that are seen to not respond, seems to suggest that something significant is happening in this area. What this is, however, remains unclear. Do these occupations not offer enough of an incentive to up skill? Is information not available for individuals identifying the gap in these areas and the benefits they could receive by up-skilling? Is there a shortage of intermediate skills that are leading establishments into a low-skill equilibrium? In a time when skills are increasing almost continuously, except amongst intermediate skill levels, the fact that this area of the UK economy is not positively responding to their skill deficiencies never mind moving forward and developing further jobs seems extremely worrying.

These occupations where responses seem limited are those areas that are currently being deemed "non business essential" and this is likely why no response is being made, the establishment can get by without them. This idea comes from Autor et al. (2001) and Goos and Manning (2007) that some middle level jobs that have typically required 'routine manual and cognitive skills', traits that can now be replaced with technology. While the jobs are not being noticeably dissolved it seems likely to suggest that they are being split between the higher skilled (for the more technical tasks) and the lower skilled (for the more manual tasks). This polarization of the workforce towards both ends of the spectrum with very few middle tier jobs is in line with the current literature around Task Biased Technological Change (TBTC (Autor et al. (2001))) and polarization (Goos and Manning (2007)).

The TBTC work adds further nuance to the idea of skill biased technical change (Autor et al. (2001) whereby the higher skilled gain from continuous technological advancements adapting the idea as growth in employment appears to have occurred in both the highest skilled occupations and the lowest skilled. The idea of task biased technological change is that technology can replace routine jobs, tasks that can be explained by a step-by-step procedure or a list of rules but it cannot (yet) replace non-routine human labour tasks. The new technology thus acts as a complement to skilled workers helping them with high end problem solving and complex communication activities but it is a substitute for those in lower skilled occupations.

The results in this work suggest that firms in the UK are taking this approach to the skill shortages they face, responding only to those jobs that are not replaceable at the top and bottom of the occupational spectrum. It is still unclear from this work whether TBTC and job polarisation have caused an intermediate skill level low-skill equilibrium to form, if the reverse is the case with the low-skill equilibrium driving polarisation or if the two are only apparent as thy are occurring alongside each other.

As the 'Great Recession' started in the period that this work investigates it was predicted to have an impact of some sort. With the economy shrinking it would be assumed that firms would not be expanding their operations and thus negative response were likely to be more common. This is definitely not the case and if anything the response coefficients are larger on most occupations in 2009 than they were in 2007 and 2005.

Table 2.2: Establishments Responding

Wave	2005	2007	2009
Total in wave	74,835	79,018	77,421
HTFV			
Total reporting HTFV	6,838	6,323	3,185
Total responding to HTFV	6,068 (89%)	5,70 (87%)	2,711 (85%)
Total not responding to HTFV	718 (11%)	735 (12%)	430 (14%)
Total with multiple responses to HTFV	1,817 (27%)	1,421(25%)	788 (25%)
Responding positively to HTFV	5,934 (87%)	5,329 (84%)	2,616 (82%)
Responding negatively to HTFV	139 (2%)	142 (2%)	96 (3%)
Responding both positively & negatively to HTFV	45 (1%)	43 (1%)	24 (1%)
SG			
Total reporting SG	16,176	15,754	19,857
Total responding to SG	14,060 (87%)	14,275 (91%)	18,342 (93%)
Total not responding to SG	1,642 (10%)	1,183 (8%)	1,351 (7%)
Total with multiple responses to SG	2,594 (16%)	3,276 (21%)	6,148 (34%)
Responding positively to SG	13,712 (85%)	14,216 (90%)	18,220 (92%)
Responding negatively to SG	416 (0%)	86 (0%)	220 (0%)
Responding both positively & negatively to SG	68 (0%)	27 (0%)	98 (0%)

Table 2.3: Responses HTFV

Number in Sample	2005	2007	2009
Positive Responses			
Increasing Salaries	322	294	106
Increasing the training given to your employees	693	737	318
Increasing advertising / recruitment spend	2,960	2,812	1,373
Increasing/expanding trainee programmes	611	506	268
Using NEW recruitment methods or channels	2,301	1,597	878
Recruiting workers who are non-UK nationals	-	-	96
Offering enhanced terms and conditions	70	47	45
Considering a wider range of applicants	86	69	48
Recruiting (additional) staff from overseas	102	94	4
Negative Responses			
Making existing staff work longer hours	176	181	120
Automating certain tasks	8	4	1
Ambiguous Responses			
Redefining existing jobs	515	374	272
Hiring (additional) part-time/temporary staff	144	184	68
Subcontracting (more) work to outside organizations	39	42	26
Non-positve, Non-negative			
Other	294	253	126
No response	713	734	440
Totals			
Total: Positive Responses	7,145	6,156	3,034
Total: Positive Plus Ambiguous Responses	7,843	6,756	3,400
Total: Negative Responses	223	227	147
Total: Negative Plus Ambiguous Responses	921	827	513
Total: Ambiguous Responses	698	600	366
Overall Total	9,034	7,928	4,189

Table 2.4: Responses SG

Number in Sample	2005	2007	2009
Positive Responses			
Increased Training Activity	9,605	11,821	15,537
Increased Recruitment Activity	987	729	807
More staff appraisals/performance reviews	700	1,739	3,341
Implement mentoring/buddying	173	1,125	2,408
More Supervision	272	1,867	3,366
Recruiting non UK workers			124
Changing work practices	132	100	147
Build up team spirit	170	93	153
Discipline action	312	150	229
Help/assistance/advice	242	159	284
Increased Salaries	164	16	5
Redefine work practices	500	54	192
Negative Responses			
Automate Certain Tasks	9	5	3
Make staff redundant	117	81	217
Ambiguous Responses			
Subcontract more	20	19	20
Other	392	599	246
Non-positive, Non-negative			
Nothing	1,642	1,183	1,351
Don't know	93	296	164
Totals			
Total: Positive Responses	13,257	17,853	26,593
Total: Positive Plus Ambiguous Responses	13,669	18,471	26,859
Total: Negative Responses	126	86	220
Total: Negative Plus Ambiguous Responses	538	704	486
Total: Ambiguous Responses	412	618	266
Overall Total			

Table 2.5: **Table of Variable Definitions**

Hard-to-fill vacancies (HTFV) -	Vacancies which are proving difficult to fill, as defined by the establishment (from question: "Are any of these vacancies proving hard-to-fill?"). This is measured as the percentage of HTFVs the establishment experiences.
Skill-shortage vacancies (SSVs) -	Vacancies which are proving difficult to fill due to the establishment not being able to find applicants with the appropriate skills, qualifications or experience. The SSV variable is measured as the percentage of SSVs the establishments experiences.
Skills gaps (SG) -	A "skills gap" is where an employee is not fully proficient, i.e. is not able to do their job to the required level (see Glossary for further details). This is measured as the percentage of SGs the establishment is currently facing.
Private -	A private sector dummy variable is used to indicate whether the establishment is in the private or non-private sectors (private=1, other=0)
Size -	Establishment size is included as a continuous variable
Job Description -	Job description is a dummy variable equal to one if more than 50 percent of the employees have a formal job description and zero otherwise.

HR Sophistication -	Human resource sophistication is a dummy variable that is equal to one if the establishment answers positively to any of the following; having plans for future training, having a budget for future training or a business plan (see Appendix 1 for full question details).
Competition Level (International, National, Regional) -	The type of market the establishment competes in is included as a set of dummy variables, with International (International=1, non-international=0), National (National=1, non-national=0) and Regional (Regional=1, non-regional=0) used, leaving local as a base category.
Quality (Very High, High, Medium, Low -	A series of dummy variables are used to determine the quality of the product. The levels are defined as Very High Quality, High Quality, Medium Quality, Low Quality and the base category of Very Low Quality. These quality measures come from the ESS question that asks establishments to indicate whether they compete in a market for a standard or basic quality product or service, or that they compete in a market for premium quality products or services.
London -	A dichotomous variable indicating whether or not the given establishment is located in London.

Table 2.6: Variables for 2005 Wave: HTFV

Variable	Obvs	Mean (SD)	Min	Max
Count (large)	6,786	1.109(0.78)	0	7
Count (small)	6,786	1.10(0.75)	0	6
Size	6,786	51.4(162.0)	2	7,726
Private	6,786	0.83 (0.38)	0	1
SSV	6,786	68.68(45.25)	0	100
HTFV	6,786	2.44(6.31)	1	400
London	6,786	0.12 (0.32)	0	1
Managers	6,838	0.011 (0.10)	0	1
Professionals	6,838	0.017 (0 .13)	0	1
Associate Professionals	6,838	0.024 (0.15)	0	1
Admin/Clerical	6,838	0.019 (0 .14)	0	1
Skilled Trades	6,838	0.031 (0.17)	0	1
Personal Services	6,838	0.035 (0.18)	0	1
Sales and Cust Services	6,838	0.029 (0.17)	0	1
Machine Operatives	6,838	0.020 (0.14)	0	1
Elementary	6,838	0.032 (0.18)	0	1

Table 2.7: **Variables for 2007 Wave: HTFV**

Variable	Obvs	Mean (SD)	Min	Max
Count (large)	6,205	1.13(0.72)	0	6
Count (small)	6,205	1.03(0.71)	0	6
Size	6,205	52.80(181.59)	2	8,500
Private	6,205	0.87	0	1
SSV	6,205	71.21(44.32)	0	100
HTFV	6,205	2.31(4.0)	1	120
London	6,205	0.19	0	1
Managers	4,588	0.089 (0.28)	0	1
Professionals	4,588	0.163 (0.37)	0	1
Associate Professionals	4,588	0.188 (0.39)	0	1
Admin/Clerical	4,588	0.102 (0.30)	0	1
Skilled Trades	4,588	0.206 (0.40)	0	1
Personal Services	4,588	0.086 (0.28)	0	1
Sales and Cust Services	4,588	0.104 (0.31)	0	1
Machine Operatives	4,588	0.081 (0.27)	0	1
Elementary	4,588	0.088 (0.28)	0	1

Table 2.8: **Variables for 2009 Wave: HTFV**

Variable	Obvs	Mean (SD)	Min	Max
Count (large)	3,141	1.14(0.79)	0	7
Count (small)	3,141	1.02(0.78)	0	6
Size	3,141	58.28(183.5)	2	5,000
Private	3,141	0.81	0	1
SSV	3,141	73.89(43.0)	0	100
HTFV	3,141	2.17(7.3)	1	350
Job Description	3,141	0.86	0	1
HR Sophistication	3,141	0.87	0	1
International Competition	3,141	0.13	0	1
National Competition	3,141	0.26	0	1
Regional Competition	3,141	0.17	0	1
Very High Quality	3,141	0.06	0	1
High Quality	3,141	0.25	0	1
Medium Quality	3,141	0.24	0	1
Low Quality	3,141	0.19	0	1
London	3,141	0.18	0	1
Managers	2,409	0.099 (0.30)	0	1
Professionals	2,409	0.139 (0.35)	0	1
Associate Professionals	2,409	0.222 (0.42)	0	1
Admin/Clerical	2,409	0.093 (0.29)	0	1
Skilled Trades	2,409	0.161 (0.37)	0	1
Personal Services	2,409	0.120 (0.32)	0	1
Sales and Cust Services	2,409	0.088 (0.28)	0	1
Machine Operatives	2,409	0.064 (0.25)	0	1
Elementary	2,409	0.095 (0.29)	0	1

Table 2.9: Variables for 2005 Wave: SG

Variable	Obvs	Mean (SD)	Min	Max
Count (large)	16,176	0 .774 (0.61)	0	5
Count (small)	16,176	0.748 (0.61)	0	5
Size	16,176	48.635 (130.1)	2	4,500
Private	16,176	0.846 (0.36)	0	1
SG Density	16,176	24.15 (20.3)	0.13	100
London	16,176	0.121 (0.33)	0	1
Managers	16,176	0.256 (0.44)	0	1
Professionals	16,176	0.070 (0.25)	0	1
Associate Professionals	16,176	0.086 (0.28)	0	1
Admin/Clerical	16,176	0.237 (0.43)	0	1
Skilled Trades	16,176	0.152 (0.36)	0	1
Personal Services	16,176	0.099 (0.30)	0	1
Sales and Cust Services	16,176	0.253 (0.43)	0	1
Machine Operatives	16,176	0.0850 (0.28)	0	1
Elementary	16,176	0.240 (0.43)	0	1

Table 2.10: Variables for 2007 Wave: SG

Variable	Obvs	Mean (SD)	Min	Max
Count (large)	15,754	1.172 (0.71)	0	6
Count (small)	15,754	1.133 (0.72)	0	5
Size	15,754	53.77 (180.18)	2	8,500
Private	15,754	0 .858 (0.349)	0	1
SG Density	15,754	24.94 (20.31)	0.13	100
London	15,754	0.175 (0.380)	0	1
Managers	15,754	0.269 (0.44)	0	1
Professionals	15,754	0.080 (0.27)	0	1
Associate Professionals	15,754	0.083 (0.28)	0	1
Admin/Clerical	15,754	0.242 (0.43)	0	1
Skilled Trades	15,754	0.154 (0.36)	0	1
Personal Services	15,754	0.076 (0.26)	0	1
Sales and Cust Services	15,754	0.254 (0.44)	0	1
Machine Operatives	15,754	0.077 (0.27)	0	1
Elementary	15,754	0.221 (0.42)	0	1

Table 2.11: **Variables for 2009 Wave SG**

Variable	Obvs	Mean (SD)	Min	Max
Count (large)	19,857	1.353 (0.85)	0	6
Count (small)	19,857	1.340 (0.85)	0	6
Size	19,857	59.192 (192.72)	2	8,000
Private	19,857	0.849 (0.36)	0	1
SG Density	19,857	23.25 (19.73)	0.15	100
Job Description	19,857	0.861 (0.35)	0	1
HR Sophistication	19,857	0.855 (0.35)	0	1
International Competition	19,857	0.138 (0.34)	0	1
National Competition	19,857	0.285 (0.45)	0	1
Regional Competition	19,857	0.183 (0.39)	0	1
Very High Quality	19,857	0.081 (0.27)	0	1
High Quality	19,857	0.292 (0.45)	0	1
Medium Quality	19,857	0.241 (0.43)	0	1
Low Quality	19,857	0.169 (0.38)	0	1
London	19,857	0.137 (0.34)	0	1
Managers	19,857	0.293 (0.46)	0	1
Professionals	19,857	0.086 (0.28)	0	1
Associate Professionals	19,857	0.097 (0.30)	0	1
Admin/Clerical	19,857	0.247 (0.43)	0	1
Skilled Trades	19,857	0.168 (0.37)	0	1
Personal Services	19,857	0.100 (0.30)	0	1
Sales and Cust Services	19,857	0.252 (0.43)	0	1
Machine Operatives	19,857	0.078 (0.27)	0	1
Elementary	19,857	0.221 (0.41)	0	1

Figure 2.2: 2009 Small Count

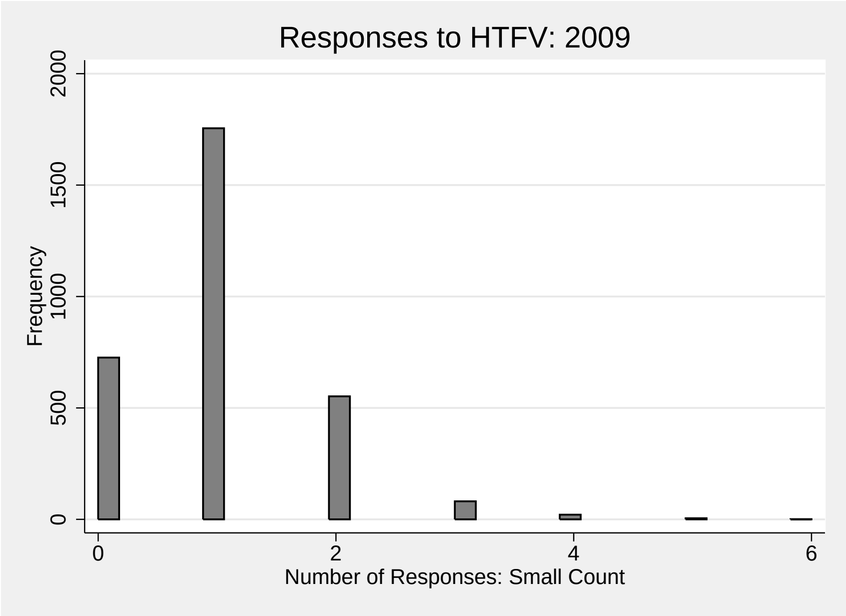


Figure 2.3: 2007 Small Count

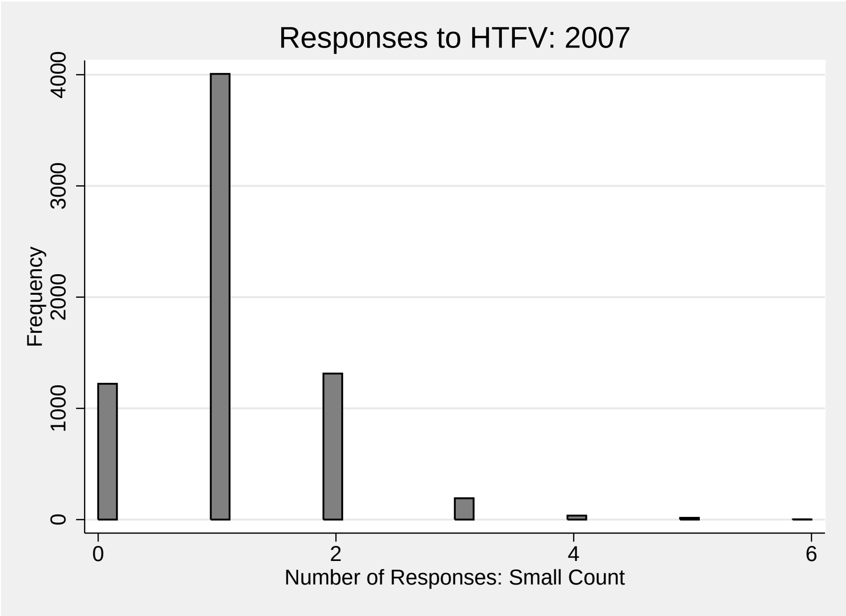


Figure 2.4: **2005 Small Count**

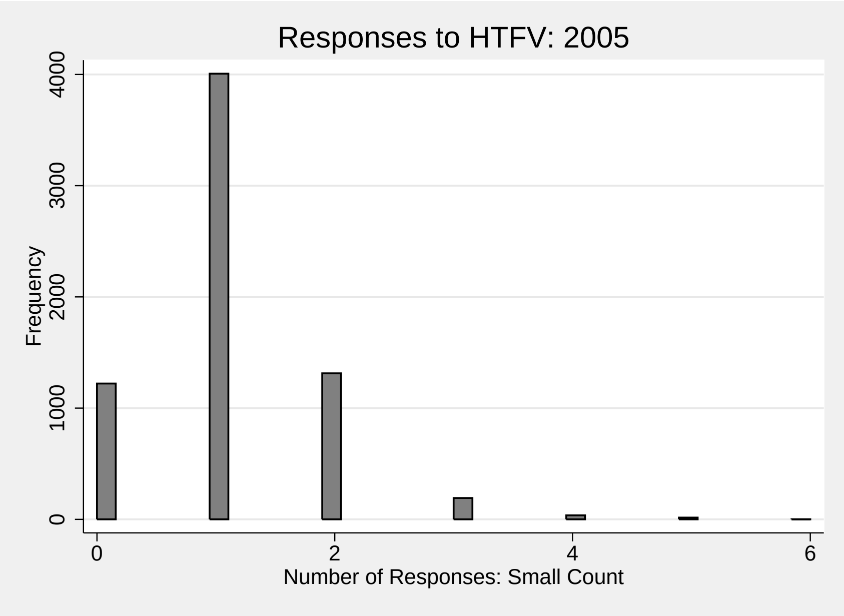


Figure 2.5: **2009 Small Count**

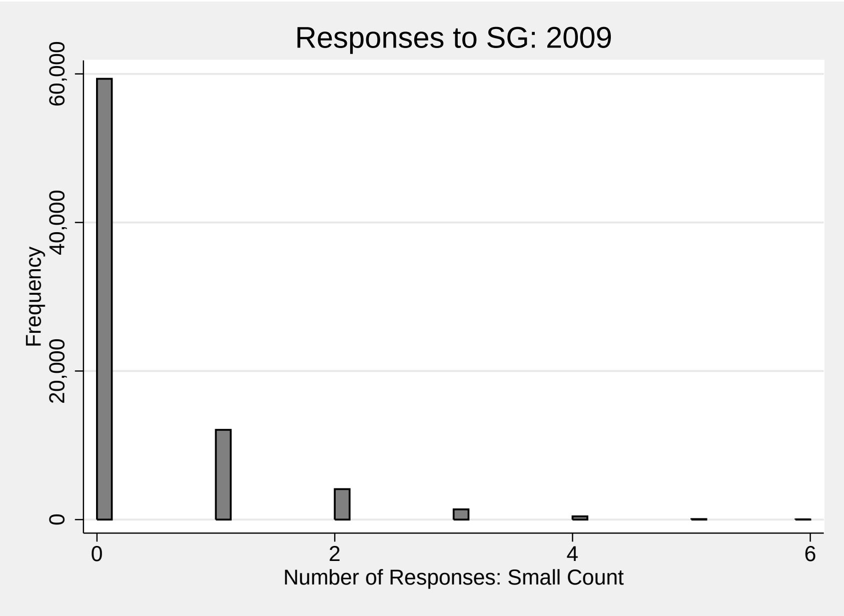


Figure 2.6: 2007 Small Count

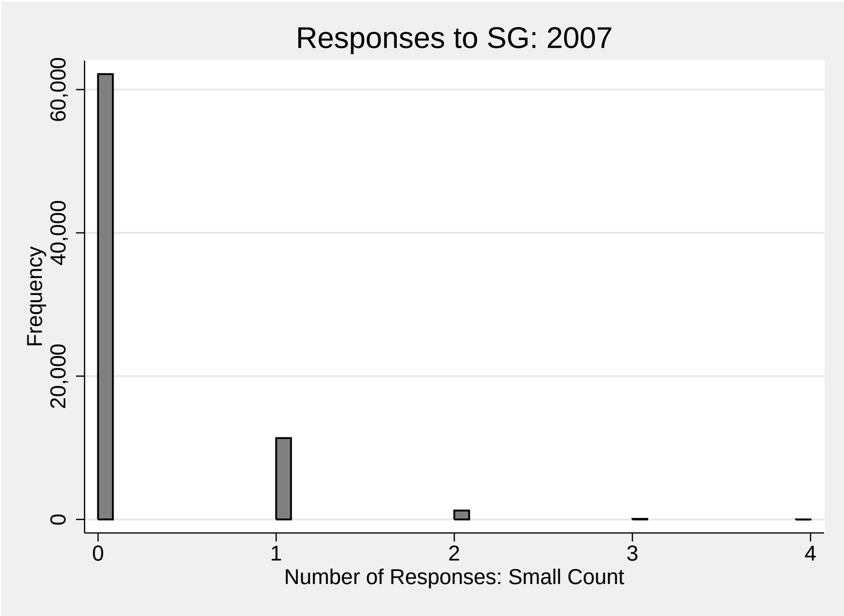


Figure 2.7: 2005 Small Count

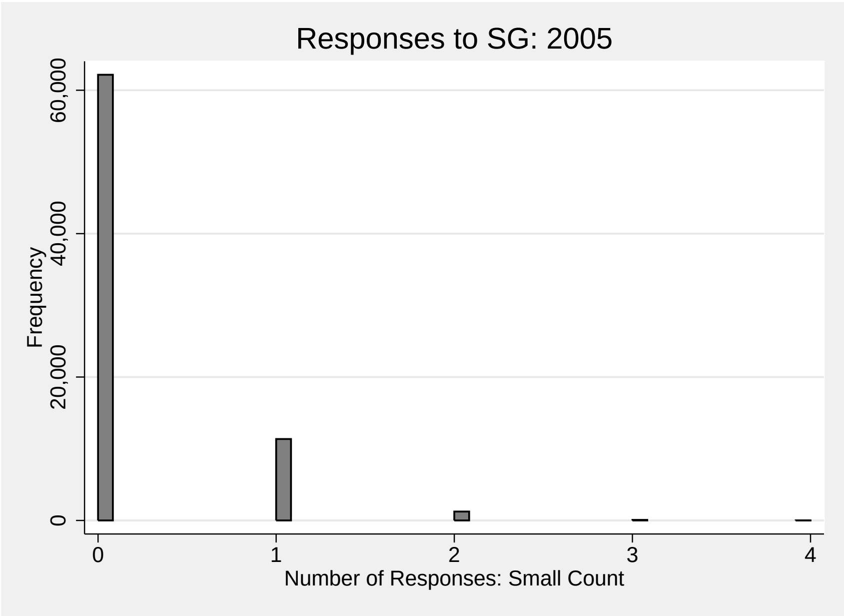


Table 2.12: GPM Results - Small Count Variable

	2005*		2007		2009	
	β / SE	MFX	β / SE	MFX	β / SE	MFX
Est Size (1,000)	0.199***	0.218***	0.208***	0.621***	0.360***	1.065***
	(0.045)		(0.063)		(0.098)	
Private	-0.087***	-0.095***	-0.045	-0.137	-0.154***	-0.479***
	(0.031)		(0.041)		(0.051)	
SSV Percentage	0.001***	0.002***	-0.001	-0.004	-0.001	-0.002
	(0.000)		(0.001)		(0.002)	
HTFV	-0.041	-0.045	-0.070**	-0.208**	0.017	0.050
	(0.035)		(0.035)		(0.055)	
London	0.013	0.014	0.056*	0.171*	-0.053	-0.154
	(0.036)		(0.031)		(0.045)	
Managers	0.032	0.035	0.058	0.177	0.132*	0.413*
	(0.108)		(0.052)		(0.076)	
Professionals	0.109	0.120	0.163***	0.513***	0.175**	0.551**
	(0.081)		(0.047)		(0.076)	
Associate Professionals	0.003	0.003	0.135***	0.422***	0.146**	0.450**
	(0.075)		(0.045)		(0.069)	
Admin/Clerical	-0.109	-0.120	0.047	0.143	0.027	0.027
	(0.087)		(0.051)		(0.076)	
Skilled Trades	-0.017	-0.018	0.110**	0.339**	0.073	0.222
	(0.068)		(0.046)		(0.073)	
Personal Services	0.036	0.040	0.148**	0.470**	0.224***	0.724***
	(0.062)		(0.058)		(0.081)	
Sales and Cust Services	-0.015	-0.017	0.132**	0.416**	0.057	0.173
	(0.070)		(0.054)		(0.081)	
Machine Operatives	-0.110	-0.121	0.122**	0.382**	0.024	0.072
	(0.088)		(0.058)		(0.087)	
Elementary	-0.056	-0.061	0.145***	0.460***	0.196**	0.629**
	(0.067)		(0.055)		(0.082)	
Constant	0.062*	*	1.111***	***	1.111***	***
	(0.033)		(0.127)		(0.185)	
Alpha	-	-	-0.162***	***	-0.152***	***
	-	-	(0.002)		(0.005)	
Observations	6,786		4,530		2,374	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

(Note: 2005 has figures from a standard Poisson regression)

Table 2.13: GPM Results - 2009 Variable

	Large Count		Small Count	
	β / SE	MFX	β / SE	MFX
Est Size (1,000)	0.513**	1.78**	0.574***	1.70***
	(0.201)		(0.188)	
Size Squared	-0.054	-0.19	-0.064	-1.19
	(0.052)		(0.049)	
Private	-0.299***	-1.14***	-0.233***	-0.75***
	(0.091)		(0.082)	
SSV Percentage	-0.002	-0.01	-0.000	-0.001
	(0.002)		(0.002)	
HTFV	0.023	0.08	0.060	0.18**
	(0.061)		(0.057)	
Job Des	0.047	0.16	0.108**	0.31**
	(0.058)		(0.053)	
HRsoph	0.139**	0.46**	0.139**	0.39**
	(0.062)		(0.057)	
Low Q	0.050	0.18	0.037	0.11
	(0.104)		(0.093)	
Medium Q	0.125	0.45	0.162**	0.50**
	(0.081)		(0.074)	
High Q	0.162**	0.58**	0.172**	0.54**
	(0.082)		(0.074)	
V High Q	0.178**	0.65**	0.196**	0.62**
	(0.085)		(0.077)	
London	-0.013	-0.05	-0.058	-0.17
	(0.049)		(0.045)	
Managers	0.176**	0.65**	0.110	0.34
	(0.083)		(0.076)	

Table 2.13: GPM Results - 2009 Variable ...Continued

	Large Count		Small Count	
	β / SE	MFX	β / SE	MFX
Professionals	0.200**	0.74**	0.140*	0.44*
	(0.085)		(0.077)	
Associate Professionals	0.163**	0.59**	0.124*	0.38*
	(0.077)		(0.070)	
Admin/Clerical	0.111	0.40	0.025	0.08
	(0.086)		(0.077)	
Skilled Trades	0.164**	0.60**	0.078	0.24
	(0.081)		(0.073)	
Personal Services	0.213**	0.80**	0.172**	0.55**
	(0.091)		(0.083)	
Sales and Cust Services	0.116	0.42	0.045	0.13
	(0.090)		(0.082)	
Machine Operatives	0.072	0.26	0.023	0.07
	(0.096)		(0.088)	
Elementary	0.264***	1.02***	0.180**	0.58**
	(0.092)		(0.082)	
Constant	1.165***	-	0.868***	-
	(0.222)		(0.198)	
Alpha	-0.134***	***	-0.153***	***
	(0.003)		(0.005)	
Observations	2,293		2,374	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

(International, National and Region dummies included but not shown)

Table 2.14: Probit 2005: HTFV

	(1) Increase Advertising	(2) Increasing training given	(3) Redefining existing jobs	(4) New Recrut Channels	(5) Increasing trainee prog	(6) New recruitment methods
Increase Advertising	N/A	0.10***	0.13***	-0.07	0.10***	0.03
Increasing training given	0.05***	N/A	0.08***	-0.48***	0.13***	-1.19***
Redefining existing jobs	0.09***	0.09***	N/A	-0.47***	0.01	-1.01***
New Recrut Channels	-0.02**	-0.14***	-0.11***	N/A	-0.08***	-0.74***
Increasing trainee prog	0.06***	0.14***	0.01	-0.29***	N/A	-0.42***
New recruitment methods	-0.01	-0.11***	-0.08***	-0.25***	-0.04***	N/A
Est Size	0.00	0.00*	-0.00	0.00	0.00***	0.00*
Private	-0.03**	-0.01	-0.08***	-0.11***	-0.00	-0.30**
SSV Percentage	-0.00*	0.00***	-0.00	0.00	0.00***	0.00***
HTFV	0.00	0.00	0.00	0.00	0.00	0.06***
London	-0.02	0.03	0.02	0.02	-0.00	0.17
Managers	0.06	-0.05	0.05	0.09	-0.03	0.78**
Professionals	0.05	-0.05	0.02	0.14	-0.00	0.35
Associate Professionals	-0.00	-0.06	0.01	0.03	-0.02	-0.13
Admin/Clerical	0.00	-0.13**	0.01	-0.05	-0.06	-0.04
Skilled Trades	0.05**	-0.03	0.01	0.04	-0.04	-0.39
Personal Services	0.00	-0.03	-0.05	0.12	-0.01	-0.01
Sales and Cust Services	0.03	-0.07*	0.06*	0.22***	-0.09**	-0.34
Machine Operatives	0.03	0.03	-0.03	-0.19*	-0.02	-0.97***
Elementary	-0.02	-0.08**	0.03	-0.00	-0.03	-0.04
Observations	6,838	6,838	6,838	6,838	6,838	6,838

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.15: Probit 2007: HTFV

	Increase Advertising	Increasing training given	Redefining existing jobs	New Recrut Channels	Increasing trainee prog	New recruitment methods
Increase Advertising	N/A	0.14***	0.08***	-0.15	0.04	-0.07
Increasing training given	0.06***	N/A	0.01	-0.64***	0.08***	-0.30***
Redefining existing jobs	0.07***	0.03	N/A	-0.43***	-0.00	-0.18***
New Recrut Channels	-0.03***	-0.21***	-0.09***	N/A	-0.09***	-0.16***
Increasing trainee prog	0.02	0.11***	0.00	-0.37***	N/A	-0.08*
New recruitment methods	-0.03*	-0.16***	-0.05***	-0.27***	-0.04**	N/A
Est Size	0.00	0.00	0.00	-0.00	0.00	0.00
Private	-0.03	-0.00	-0.07***	-0.06	-0.04*	0.02
SSV Percentage	-0.00	0.00	-0.00	-0.00	0.00	-0.00
HTFV	-0.00	0.01***	0.00	0.01***	0.00	0.01**
London	-0.01	0.00	-0.02	-0.07*	-0.03	0.17***
Managers	-0.00	-0.01	0.02	0.05	0.03	0.07
Professionals	0.02	0.02	0.01	0.10	0.01	0.14***
Associate Professionals	0.01	0.02	-0.02	0.17***	0.00	0.07*
Admin/Clerical	-0.00	0.02	0.03	0.03	0.01	0.03
Skilled Trades	-0.00	0.06**	-0.04*	0.03	0.03	0.01
Personal Services	-0.01	0.04	-0.02	0.21***	0.04	0.01
Sales and Cust Services	0.01	0.05	-0.01	0.19***	-0.02	0.06
Machine Operatives	0.04	-0.01	-0.03	0.10	0.03	-0.03
Elementary	0.03	0.03	-0.05	0.13*	0.02	0.04
Observations	4,588	4,588	4,588	4,588	4,588	4,588

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.16: Probit 2009: HTFV

	Increase Advertising	Increasing training given	Redefining existing jobs	New Recrut Channels	Increasing trainee prog	New recruitment methods
Increase Advertising	N/A	0.14***	0.09*	0.11	0.08	-0.01
Increasing training given	0.05**	N/A	-0.00	-0.49***	0.16***	-0.58***
Redefining existing jobs	0.04*	0.01	N/A	-0.53***	0.01	-0.54***
New Recrut Channels	0.00	-0.14***	-0.12***	N/A	-0.08***	-0.28***
Increasing trainee prog	0.03	0.19***	0.02	-0.29***	N/A	-0.20
New recruitment methods	-0.01	-0.13***	-0.09***	-0.21***	-0.04*	N/A
Est Size	-0.00	0.00	0.00	0.00	0.00*	0.00
Private	-0.06***	-0.05*	-0.10***	-0.30***	-0.00	0.08
SSV Percentage	-0.00	0.00	-0.00	-0.00	0.00	0.00
HTFV	-0.00	0.00***	-0.01*	0.01*	0.00	0.03***
London	0.00	-0.02	0.02	-0.13**	-0.02	0.15*
Managers	0.07***	-0.02	0.13***	0.08	-0.00	0.36**
Professionals	0.06*	-0.03	0.07**	0.19*	0.03	0.29**
Associate Professionals	0.04	0.05	0.06*	0.07	0.01	0.31**
Admin/Clerical	0.04	-0.01	0.13***	0.01	0.04	0.09
Skilled Trades	0.05*	0.04	0.08**	-0.09	0.04	0.02
Personal Services	0.04	0.09**	0.03	0.26**	0.05	0.19
Sales and Cust Services	0.00	-0.07	0.14***	0.16	0.04	0.04
Machine Operatives	0.02	0.03	0.02	-0.13	0.05	0.06
Elementary	0.09***	0.03	0.07*	0.08	0.10***	0.00
Observations	2,409	2,409	2,409	2,409	2,409	2,409

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.17: SG Results - Small Count Variable

	2005*		2007		2009	
	β / SE	MFX	β / SE	MFX	β / SE	MFX
Est Size (1,000)	-0.106	-0.092	0.199***	0.225***	0.429***	0.576***
	(0.078)		(0.048)		(0.050)	
Size Squared	0.008	0.007	-0.030**	-0.034**	-0.072***	-0.097***
	(0.033)		(0.014)		(0.013)	
Private	-0.049***	-0.044***	-0.099***	-0.116***	-0.085***	-0.117***
	(0.015)		(0.015)		(0.013)	
Density	0.001***	0.001***	-0.002***	-0.003***	-0.001***	-0.002***
	(0.000)		(0.000)		(0.000)	
London	-0.104***	-0.087***	0.074***	0.086***	-0.024*	-0.032*
	(0.017)		(0.013)		(0.013)	
Managers	0.076***	0.068***	0.007	0.008	0.035***	0.048***
	(0.012)		(0.012)		(0.011)	
Professionals	0.020	0.017	0.078***	0.091***	0.122***	0.172***
	(0.021)		(0.018)		(0.016)	
Associate Professionals	0.069***	0.061***	0.133***	0.160***	0.077***	0.106***
	(0.018)		(0.018)		(0.015)	
Admin/Clerical	0.047***	0.041***	0.070***	0.080***	0.005	0.007
	(0.012)		(0.012)		(0.011)	
Skilled Trades	0.015	0.013	0.062***	0.072***	0.013	0.017
	(0.015)		(0.015)		(0.012)	
Personal Services	0.038**	0.034**	0.210***	0.260***	0.187***	0.270***
	(0.018)		(0.020)		(0.015)	
Sales and Cust Services	0.055***	0.049***	0.170***	0.201***	0.115***	0.158***
	(0.013)		(0.012)		(0.011)	
Machine Operatives	0.053***	0.047***	-0.003	-0.003	0.011	0.014
	(0.018)		(0.020)		(0.017)	
Elementary	0.061***	0.054***	0.102***	0.119***	0.089***	0.122***
	(0.012)		(0.012)		(0.011)	
Constant	-0.175***	***	0.113***	***	0.267***	***
	(0.017)		(0.017)		(0.015)	
Observations	16,176		15,754		19,857	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: 2005 has figures from a standard Poisson regression)

Table 2.18: SG Results - 2009 Variable

	Large Count		Small Count	
	β / SE	MFX	β / SE	MFX
Est Size (1,000)	0.399***	0.542***	0.412***	0.553***
	(0.047)		(0.048)	
Size Squared	-0.064***	-0.087***	-0.068***	-0.091***
	(0.012)		(0.013)	
0 Private	-0.206***	-0.300***	-0.204***	-0.293***
	(0.022)		(0.022)	
SG Density	-0.000*	-0.001*	-0.001**	-0.001**
	(0.000)		(0.000)	
London	-0.028**	-0.037**	-0.025*	-0.033*
	(0.013)		(0.013)	
Job Des	0.160***	0.203***	0.165***	0.208***
	(0.014)		(0.015)	
Low Q	0.075***	0.105***	0.077***	0.106***
	(0.024)		(0.025)	
Medium Q	0.126***	0.176***	0.128***	0.177***
	(0.019)		(0.020)	
High Q	0.143***	0.201***	0.147***	0.205***
	(0.020)		(0.020)	
Very High Q	0.153***	0.218***	0.154***	0.217***
	(0.021)		(0.021)	
International	-0.006	-0.009	-0.009	-0.013
	(0.014)		(0.014)	
National	0.009	0.013	0.007	0.010
	(0.012)		(0.012)	
Regional	0.022	0.031	0.020	0.027
	(0.014)		(0.014)	
Managers	0.032***	0.043***	0.030***	0.040***
	(0.010)		(0.011)	
Professionals	0.112***	0.158***	0.113***	0.159***
	(0.016)		(0.016)	
Associate Professionals	0.068***	0.095***	0.066***	0.091***
	(0.015)		(0.015)	
Admin/Clerical	-0.001	-0.002	0.001	0.001
	(0.011)		(0.011)	
Skilled Trades	0.019	0.026	0.017	0.023
	(0.012)		(0.012)	
Personal Services	0.165***	0.238***	0.169***	0.242***
	(0.015)		(0.015)	
Sales and Cust Services	0.094***	0.131***	0.096***	0.132***
	(0.011)		(0.011)	
Machine Operatives	0.012	0.016	0.009	0.012
	(0.017)		(0.017)	
Elementary	0.082***	0.113***	0.085***	0.117***
	(0.011)		(0.011)	
Constant	0.128***	-	0.112***	-
	(0.020)		(0.020)	
Observations	19,857		19,857	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.19: Probit 2007: SG

	More training activity	More recruitment activity	More appraisals	Implement mentoring	More supervision of staff	Disciplinary action
More training activity	N/A	-0.09***	-0.13***	-0.04***	-0.09***	-0.07***
More recruitment activity	N/A	N/A	0.02	0.04**	0.05***	-0.03*
More staff appraisals	N/A	0.01	N/A	0.09***	0.15***	-0.00
Implement mentoring	N/A	0.04***	0.13***	N/A	0.20***	-0.03*
More supervision of staff	N/A	0.03***	0.15***	0.14***	N/A	-0.02**
Disciplinary action	N/A	-0.07*	0.00	-0.09**	-0.12**	N/A
Est Size	N/A	0.00**	0.00***	0.00	0.00	0.00
Size Squared	N/A	-0.03*	-0.01	-0.02	-0.01	-0.15
Private	N/A	0.02**	-0.06***	-0.03***	-0.04***	-0.00
SG Density	N/A	0.00***	-0.00***	-0.00***	-0.00	-0.00***
London	N/A	0.01	0.06***	0.02*	0.02*	0.04***
Managers	N/A	0.03***	0.07***	0.00	-0.05***	-0.01
Professionals	N/A	0.00	0.01	0.03***	0.02	-0.01
Associate Professionals	N/A	0.04***	0.07***	0.03**	0.02	0.00
Admin/Clerical	N/A	0.02**	0.03***	0.02**	-0.01	-0.00
Skilled Trades	N/A	0.02*	-0.02	0.02**	0.00	0.01
Personal Services	N/A	0.01	0.07***	0.01	0.09***	0.02**
Sales and Cust Services	N/A	0.02**	0.07***	0.03***	0.02**	0.02***
Machine Operatives	N/A	0.02	0.01	-0.02	-0.00	0.02**
Elementary	N/A	0.01	0.02**	0.01	0.05***	0.01**
Observations	N/A	15,754	15,754	15,754	15,754	15,754

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.20: Probit 2009: SG

	More training activity	More recruitment activity	More staff appraisals	Implement mentoring	More supervision of staff	Disciplinary action
More training activity	N/A	-0.07***	-0.09***	-0.05***	-0.09***	-0.07***
More recruitment activity	-1.30***	N/A	0.06***	0.04**	0.02	-0.07***
More staff appraisals	-0.59***	0.02***	N/A	0.13***	0.22***	-0.01
Implement mentoring	-0.32***	0.02***	0.17***	N/A	0.28***	-0.03***
More supervision of staff	-0.54***	0.01	0.22***	0.21***	N/A	-0.02***
Disciplinary action	-2.54***	-0.13***	-0.01	-0.11**	-0.11**	N/A
Est Size	0.00***	0.00***	0.00***	0.00*	0.00***	0.00
Size Squared	-0.41***	-0.03***	-0.04***	-0.01*	-0.02**	-0.03
Private	-0.05	0.01	-0.06***	-0.07***	-0.01	-0.01
SG Density	-0.01***	0.00***	-0.00***	-0.00***	0.00*	-0.00*
London	-0.42***	0.00	0.01	-0.00	-0.00	0.02***
Managers	-0.08	0.02***	0.07***	0.01*	-0.05***	0.01
Professionals	0.50***	0.01	0.07***	0.06***	0.00	-0.00
Associate Professionals	0.38***	-0.01	0.06***	0.06***	-0.02	0.01
Admin/Clerical	0.05	0.00	0.03***	0.01	-0.05***	-0.01
Skilled Trades	0.47***	0.00	0.01	-0.02*	-0.02	0.00
Personal Services	1.09***	0.02**	0.08***	0.01	0.10***	-0.01
Sales and Cust Services	0.54***	-0.02***	0.11***	0.01	0.02	0.01**
Machine Operatives	0.05	-0.02	0.01	0.02	-0.03	0.02*
Elementary	0.29***	0.01*	0.06***	0.02**	0.02	0.02***
Observations	19,857	19,857	19,857	19,857	19,857	19,857

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix 1

The human resources dummy variable is coded as one if the establishment responds in a positive manner to any of the following questions; Whether establishment has business plan that specifies the objectives for the coming year, Whether establishment has training plan that specifies in advance the level and type of training employees will need in the coming year, Whether establishment has a budget for training expenditures. If the answer to all of these questions is no then the establishment is coded as a zero.

Appendix 2

Figure 2.8: 2009 Large Count

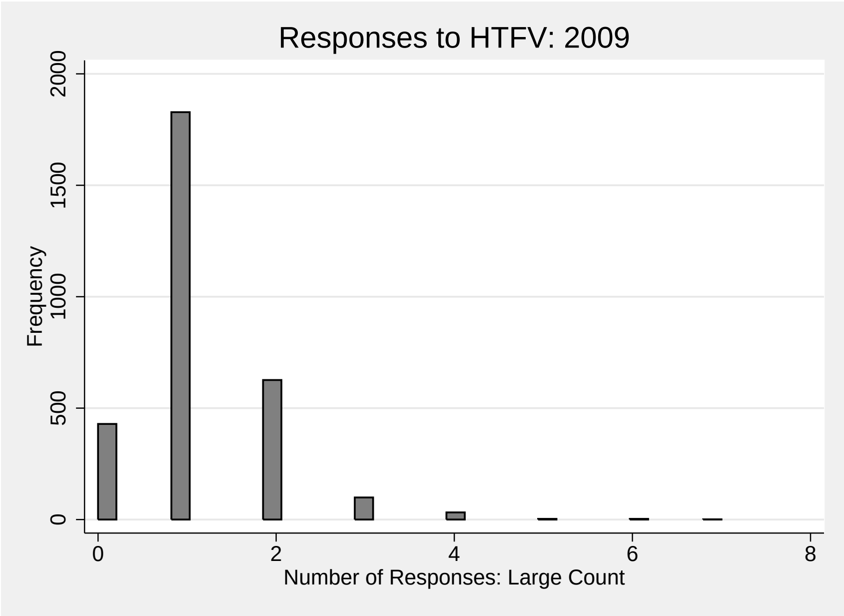


Figure 2.9: 2007 LargeCount

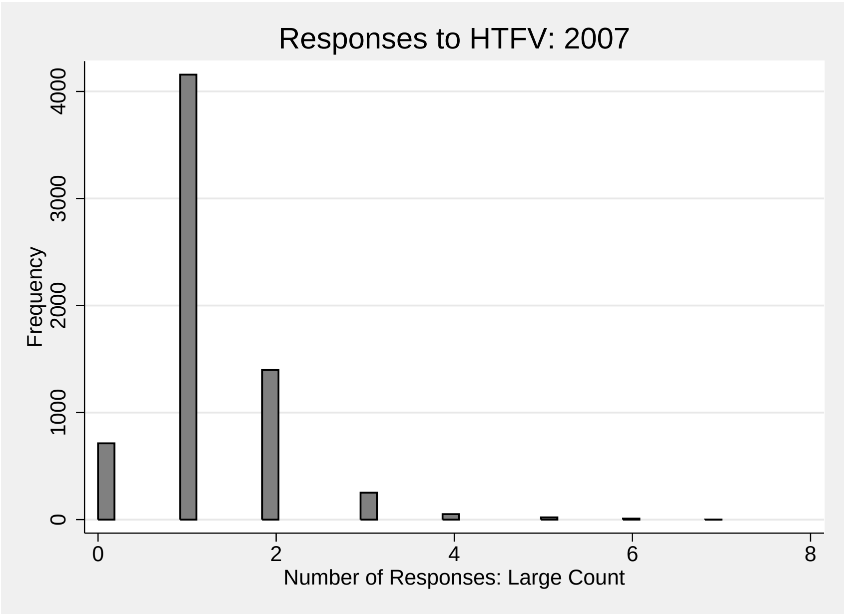


Figure 2.10: 2005 Large Count

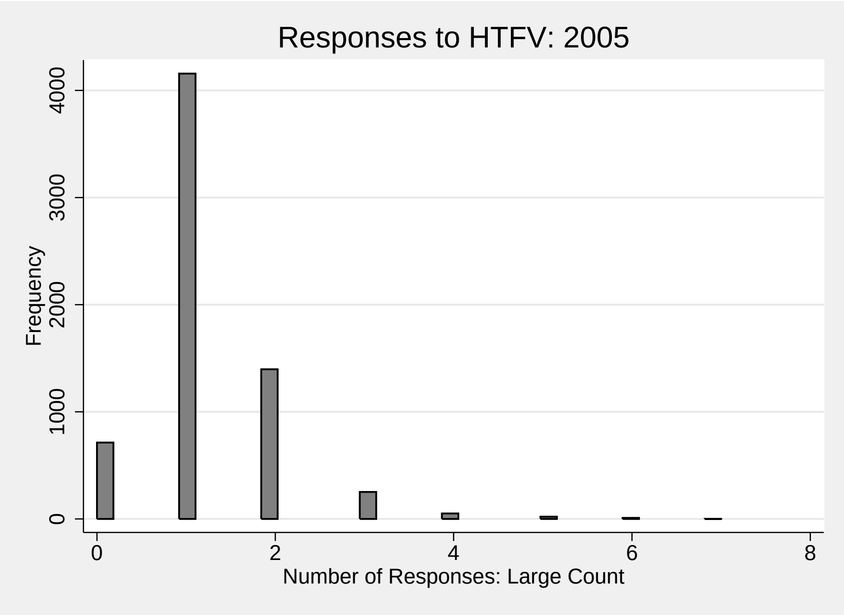


Figure 2.11: 2009 Large Count

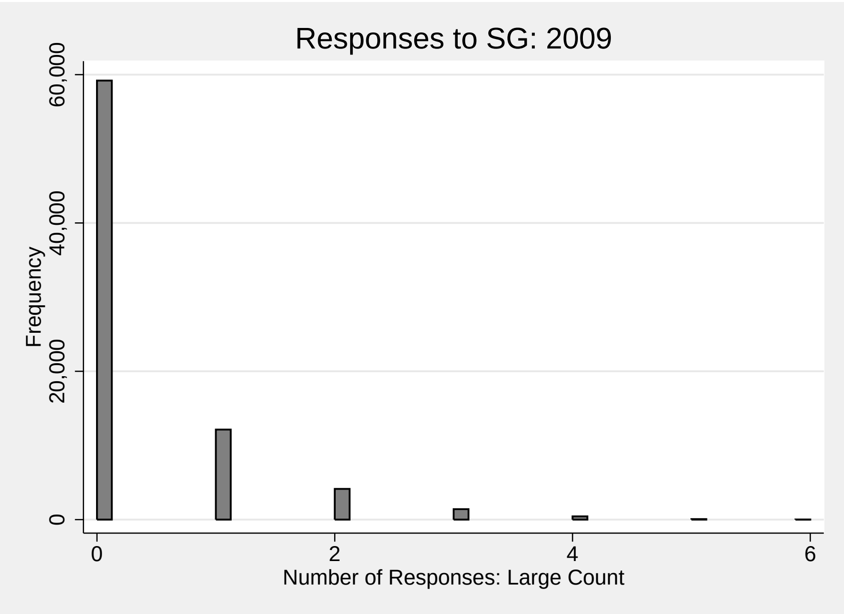


Figure 2.12: 2007 LargeCount

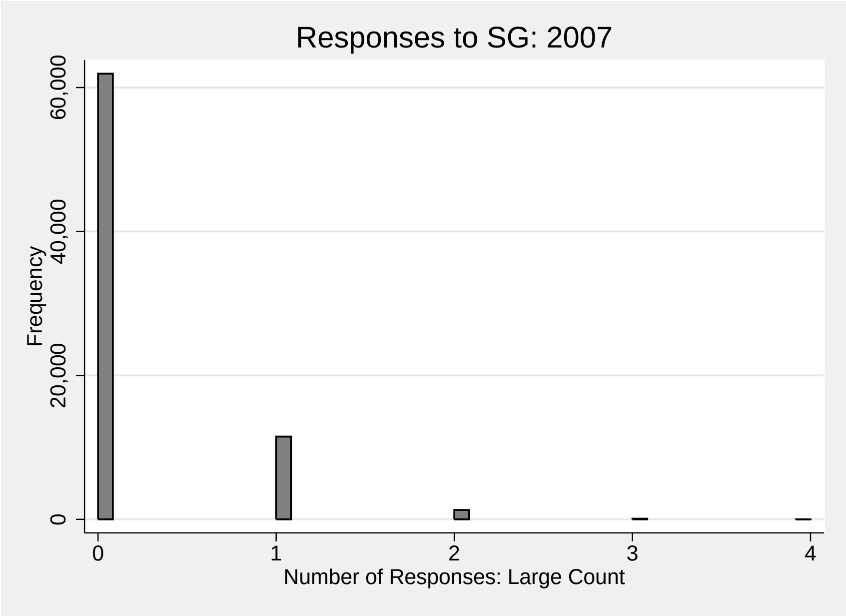


Figure 2.13: 2005 Large Count

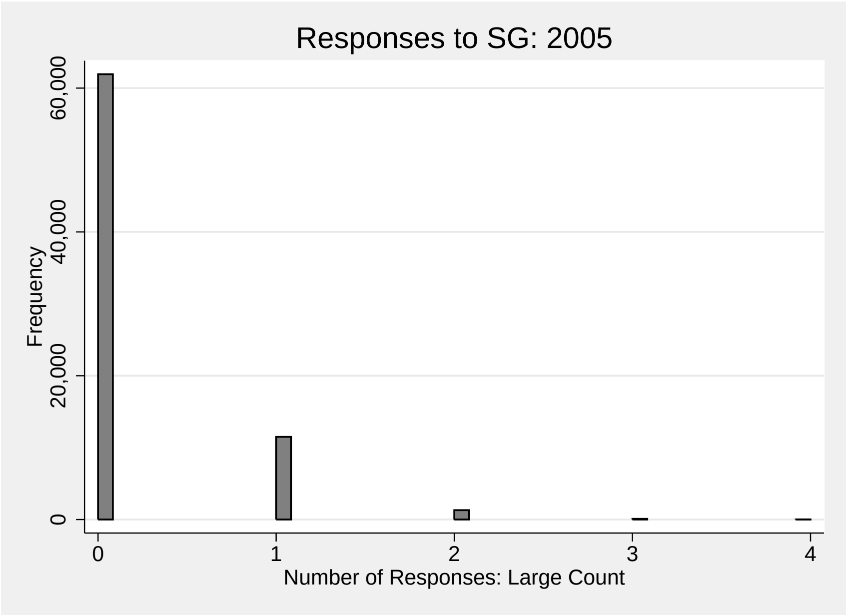


Table 2.21: **GPM Results - Large Count Variable**

	2007		2009	
	β / SE	MFX	β / SE	MFX
Est Size (1,000)	0.466***	1.582***	0.335***	1.156***
	(0.095)		(0.107)	
Private	-0.081*	-0.284 *	-0.212***	-0.779***
	(0.044)		(0.056)	
SSV Percentage	-0.002	-0.007	-0.002	-0.006
	(0.001)		(0.002)	
HTFV	-0.027	-0.091	-0.009	-0.033
	(0.042)		(0.059)	
London	0.016	0.055	-0.010	-0.033
	(0.033)		(0.049)	
Managers	0.081	0.283	0.198**	0.741**
	(0.055)		(0.083)	
Professionals	0.181***	0.652***	0.234***	0.879***
	(0.050)		(0.084)	
Associate Professionals	0.141***	0.500***	0.186**	0.679**
	(0.048)		(0.077)	
Admin/Clerical	0.107*	0.378*	0.120	0.436
	(0.055)		(0.085)	
Skilled Trades	0.110**	0.388**	0.163**	0.598**
	(0.049)		(0.081)	
Personal Services	0.133**	0.4777**	0.254***	0.966***
	(0.061)		(0.089)	
Sales and Cust Services	0.133**	0.476**	0.133	0.485
	(0.057)		(0.089)	
Machine Operatives	0.108*	0.383*	0.070	0.249
	(0.061)		(0.096)	
Elementary	0.115*	0.408*	0.275***	1.067***
	(0.058)		(0.092)	
Constant	1.309***	-	1.348***	-
	(0.136)		(0.208)	
Alpha	-0.152***	-	-0.133***	-
	(0.002)		(0.003)	
Observations	4,410		2,293	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.22: **GPM Results - 2009 Variable**

	Large Count		Small Count	
	β / SE	MFX	β / SE	MFX
Est Size (1,000)	0.554***	1.816***	0.615***	1.739***
	(0.178)		(0.167)	
Size Squared	-0.068	-0.222	-0.079*	-0.223*
	(0.049)		(0.046)	
Private	-0.233***	-0.819***	-0.202***	-0.606***
	(0.073)		(0.067)	
SSV Percentage	0.002***	0.007***	0.002***	0.006***
	(0.000)		(0.000)	
HTFV	0.039	0.129	0.060	0.171
	(0.053)		(0.049)	
Job Des	0.072	0.228	0.104**	0.282**
	(0.050)		(0.045)	
HRsoph	0.190***	0.578***	0.181***	0.478***
	(0.051)		(0.046)	
Low Q	-0.010	-0.033	-0.032	-0.089
	(0.087)		(0.077)	
Medium Q	0.118*	0.398*	0.132**	0.386**
	(0.068)		(0.061)	
High Q	0.156**	0.535**	0.139**	0.409**
	(0.068)		(0.062)	
V High Q	0.173**	0.601**	0.186***	0.560***
	(0.071)		(0.064)	
International	-0.027	-0.087	-0.036	-0.101
	(0.052)		(0.048)	
National	-0.078*	-0.252*	-0.126***	-0.346***
	(0.043)		(0.039)	

Table 2.22: GPM Results - 2009 Variable ...Continued

	Large Count		Small Count	
	β / SE	MFX	β / SE	MFX
Regional	-0.071	-0.229	-0.081*	-0.223*
	(0.050)		(0.046)	
London	-0.018	-0.057	-0.057	-0.157
	(0.044)		(0.040)	
Constant	0.885***	-	0.711***	-
	(0.075)		(0.068)	
Alpha	-0.136***	-	-0.155***	-
	(0.003)		(0.004)	
Observations	3,021		3,141	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.23: **SG Results - 2009 Variable**

	Large Count		Small Count	
	β / SE	MFX	β / SE	MFX
Est Size (1,000)	0.541***	0.734***	0.554***	0.742***
	(0.046)		(0.048)	
Size Squared	-0.083***	-0.113***	-0.087***	-0.117***
	(0.015)		(0.016)	
Private	-0.241***	-0.355***	-0.239***	-0.349***
	(0.022)		(0.022)	
SG Density	0.000	0.000	-0.000	-0.000
	(0.000)		(0.000)	
London	-0.021	-0.028	-0.018	-0.024
	(0.013)		(0.013)	
Job Des	0.183***	0.231***	0.189***	0.236***
	(0.014)		(0.015)	
Low Q	0.070***	0.098***	0.072***	0.099***
	(0.024)		(0.025)	
Medium Q	0.131***	0.182***	0.132***	0.183***
	(0.020)		(0.020)	
High Q	0.155***	0.218***	0.159***	0.222***
	(0.020)		(0.020)	
V High Q	0.163***	0.233***	0.164***	0.232***
	(0.021)		(0.021)	
International	-0.020	-0.027	-0.023	-0.031
	(0.014)		(0.014)	
National	0.004	0.005	0.001	0.002
	(0.011)		(0.012)	
Regional	0.018	0.025	0.016	0.021
	(0.014)		(0.014)	
Constant	0.203***	***	0.189***	***
	(0.019)		(0.020)	
Observations	19,857		19,857	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.24: SG Results - Large Count Variable

	2007		2009	
	β / SE	MFX	β / SE	MFX
Est Size (1,000)	0.193***	0.226***	0.417***	0.563***
	(0.048)		(0.048)	
Size Squared	-0.029**	-0.034**	-0.068***	-0.093***
	(0.013)		(0.013)	
Private	-0.104***	-0.126***	-0.088***	-0.123***
	(0.015)		(0.013)	
SG Density	-0.002***	-0.002***	-0.001***	-0.001***
	(0.000)		(0.000)	
London	0.060***	0.072***	-0.027**	-0.036**
	(0.013)		(0.013)	
Managers	0.010	0.012	0.038***	0.051***
	(0.012)		(0.010)	
Professionals	0.078***	0.094***	0.121***	0.171***
	(0.018)		(0.016)	
Associate Professionals	0.123***	0.152***	0.079***	0.110***
	(0.018)		(0.015)	
Admin/Clerical	0.062***	0.074***	0.003	0.004
	(0.012)		(0.011)	
Skilled Trades	0.058***	0.069***	0.015	0.021
	(0.014)		(0.012)	
Personal Services	0.188***	0.239***	0.182***	0.265***
	(0.019)		(0.015)	
Sales and Cust Services	0.152***	0.186***	0.113***	0.157***
	(0.011)		(0.011)	
Machine Operatives	0.006	0.007	0.013	0.018
	(0.019)		(0.017)	
Elementary	0.094***	0.113***	0.085***	0.118***
	(0.012)		(0.011)	
Constant	0.151***	-	0.278***	-
	(0.016)		(0.015)	
Observations	15,754		19,857	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Chapter 3

The effect of Internal and External Skill Gaps on Establishment Performance

3.1 Introduction

Improving the skills of the UK labour force has been, and will likely continue to be, a paramount objective of the government. This emphasis on skills is an attempt to raise the lagging productivity of the UK compared to some of its key international competitors. Although the link between increased skills and increased productivity is intuitively appealing there is little evidence of it at an establishment level in the UK, with even fewer works investigating the negative effect that skill gaps may have on productivity. Proving that this link exists and the strength of the link may help motivate workplaces to demand a higher quality of worker, creating incentives for the workforce to upskill.

This work looks to investigate the effect of skill gaps on the performance of establishments. Productivity is the chosen performance measure in this work due to the availability of data on numerous productivity measures in the Annual Business Inquiry (ABI) . While establishment survival was also contemplated as an alternative performance measure, it was not possible with the current data as the number of establishments 'dying' in the sample was too small.

To investigate this link, the ESS is matched at a one-to-one establishment level to the ABI and the Business Structure Database (BSD) for the years 2007, 2009 and 2011. The resulting dataset contains a wealth of information on both internal and external skills gaps and establishment productivity. Establishment level information on internal skill gaps (where employees in a particular job are not fully proficient) and external skill gaps (where an establishment cannot fill a vacancy due to workers of an adequate skill level not being available) is then used to estimate the impact of skill deficiencies on productivity, whilst controlling for endogeneity of inputs and possible selection bias.

A simple model of internal skill gaps is created in this work whereby effort and proficiency can be traded off against each other. Thus a higher share of unproficient workers is estimated to only reduce the productivity of a firm when effort can be constrained. This hypothesis is then tested by investigating the difference in productivity by industry.

3.2 Background on Performance

Two distinct methods of investigating skill gaps exist. Firstly the skill gaps in the market can be investigated at a local level as done by Haskel and Martin (1993). They lay a simple theoretical groundwork to show the effect of higher levels of external skill shortages on productivity, suggesting it works through two channels. These two channels are as follows: firstly the small number of skilled workers causes larger hiring costs for firms and secondly these gaps also cause establishments to substitute away from skilled labour towards unskilled labour. After explaining these channels the authors estimate the effect on productivity using a panel of 81 industries matched in to the Quarterly Trends Survey from 1980-1986. Using Generalised Method of Moments (GMM) they find that the increased skill shortages in the 1980s had a negative effect on productivity growth of 0.7% per annum. While this paper is key in identifying that a link does exist it only looks at skill shortages in the market as a whole and not at an establishment level, and also only investigates external skill shortages.

The second method of investigating this relationship involves measuring the skill gaps and productivity at the actual firm. This gives much richer data but up until now only appears to have been conducted for a subset of industries. Bennett and McGuinness (2009) look at the IT, Electrical Engineering and Mechanical Engineering industries in Northern Ireland. Their results suggest that even though 47% of firms have unfilled vacancies that are reported to decrease productivity no statistically significant relationship was seen in between those with unfilled vacancies and those without when OLS regressions were estimated. The firms with Hard to Fill Vacancies (HTFV) were in fact seen to have higher productivity. This could be due to the human resource management policies and organizational restructuring undertaken in the firms with HTFV having an additional impact above and beyond alleviating the productivity fall expected from a HTFV. The authors also note that the probability of firms experiencing skill shortages is not randomly distributed and thus they also used a Heckman (1979) two step model to account for any differences in the characteristics of firms experiencing skill shortages. The results from these selection models suggest that firms with skill shortages have ex ante productivity levels approximately 50% higher than average. The effect of HTFVs are then seen to be around a 65% decrease in productivity and

75% for unfilled vacancies. This is extremely important as it suggests that a selection issue may occur when looking at firms with skill deficiencies.

Forth and Mason (2004) investigate enterprises' difficulties in recruiting scientists and engineers with ICT related skills and the effect that these have on output. Using the 1998 Technical Graduates Employers Survey they find that 'quality related difficulties had no impact on output' but quantity based difficulties in recruitment did have a significant impact on the average level of sales per employee. The authors also tested the hypothesis that ICT skill shortages may be delayed and become more severe over time, however, there was no difference seen between the immediate and delayed impacts of skill shortages on output per employee.

Further work by Forth and Mason (2006) again looks at ICT skills shortages, this time combining the 1999 UK International Benchmarking Survey (IBS) with Dun & Bradstreet post survey financial data. This gives a total of 459 companies with full information across the combined dataset. Using instrumented variables to remove the endogeneity between performance and ICT investment, the authors find that ICT skill shortages do have negative effects on the performance experienced by these companies.

These works indicate that while a negative relationship has been seen between firm level skills and productivity it is only established in a few specialized industries. This work looks to contribute to the literature by investigating at a wider sample of workplaces to establish if this relationship holds. While doing this, an effort is made to understand what type of gap (internal or external) is more important, which until now has been neglected in the literature.

While the above literature focuses on productivity, this is not the only way that performance can be measured. One of the other frequently used measures of firm performance is the length of survival. This is easily computed from just the entry (birth) and exit (death) of the firms. Research shows the firm survival is closely related to other measures of performance, including size, growth and productivity, with some suggesting that survival is the most comprehensive of these measures (Stigler (1958)). While this area of research has received considerable attention (Audretsch and Mahmood (1995); Audretsch (1991); Klepper (2002) to name a few) the idea that skill gaps could have an effect on the likelihood of a firm surviving has not been investigated in the literature.

Collier et al. (2008) investigate a similar question but regarding training and survival rather than skill gaps and survival. They use the Workplace Employer Relations Survey (WERS) in 1998 and 2004 to see if firms that trained previously were more likely to survive into 2004. The results suggested that training was lowering the chance of "death" by around nine percentage points.

While an alternative measure of performance such as this would help to fully capture the effects of skill deficiencies on performance, it is beyond the capabilities of this work. The Business Structures Database (BSD) was matched to the ESS/ABI sample to see if skill deficiencies have an impact on the survival of workplaces in the near future. While this would give an alternative outcome measure to productivity and allow consideration across a larger time horizon, the match between the ABI and ESS does not provide many establishments and those that were matched were larger workplaces that were less likely to fail. Due to this, the BSD information on deaths provided little extra information, with only a handful of establishments failing in each year, not providing enough observations for any type of thorough survival analysis to be undertaken. Productivity is therefore the only performance measure used in this work.

3.3 Theory of Productivity

This work uses a standard Cobb-Douglas production function with an effective labour function built in. As unproficient workers are used as a measure of internal skill gaps the effective labour function is additive rather than multiplicative. This is logical as workplaces do not need these unproficient workers in order to produce. Assuming two types of worker, those who are fully proficient at their jobs (ψ_H), where there is no internal skill gap, and those workers who are not fully proficient (ψ_L), and an internal skill gap exists. This gives the following production function and effective labour capital formulae for firm i :

$$Y_i = A_i N_i^{\alpha} K_i^{1-\alpha} \tag{3.1}$$

where output depends on technological development (A), effective labour inputs (N) and capital (K). Where effective labour input is defined by the effort of high proficiency workers (e_H), their level of proficiency (ψ_H), the number of high proficiency workers (L_H) and the same for low proficiency workers;

$$N_i = e_H \psi_H L_H + e_L \psi_L L_L \tag{3.2}$$

Combining these two equations gives:

$$Y_i = A_i [e_H \psi_H L_H + e_L \psi_L L_L]^\alpha K_i^{(1-\alpha)} \tag{3.3}$$

Dividing by the labour force (L) to give output per capita and substituting in θ which is equal to $\frac{L_L}{L}$, the share of skill gaps out of total labour, gives the following equation:

$$\frac{Y_i}{L} = A_i [(1-\theta)(e_H \psi_H) + \theta(e_L \psi_L)]^\alpha \frac{K_i}{L}^{(1-\alpha)} \tag{3.4}$$

Under the assumption that $\psi_H > \psi_L$ then $\frac{Y_i}{L}$ will increase as θ decreases, as long as effort (e) does not decrease by a larger amount.

This framework allows us to show the three major ways in which productivity can be affected by skill gaps.

Firstly, if effort is not perfectly substitutable with proficiency then any increase in the share of skill gaps a workplace has, θ, would cause a fall in productivity ceteris paribus. This is the intuitive idea that increased skill gaps result in lower productivity.

The second possibility is that effort and productivity are perfectly substitutable. Assuming this we model employees' effort using a simple shirking constraint framework. Individuals produce output based upon effort and their proficiency level:

$$N_j = e_j \psi_j \tag{3.5}$$

Thus making the sum of all workers equal to the workplace's labour function:

$$\sum N_j = N_i \tag{3.6}$$

Effort (e) is variable in the period (between 0 and 1) and the employee's proficiency (ψ) is fixed (again between 0 and 1). Effort is considered a dis-utility to workers and thus they will try and minimize effort while still producing the level of output required. Due to the complex output associated with most modern jobs the level of output a worker produces can only be measured compared to other workers, and thus a worker is expected here to produce the average output of the other workers at the establishment or be fired. This expected output will be equal to the share of high proficiency workers $(1 - \theta)$ multiplied by their effort and proficiency and the share of unproficient workers (θ), again multiplied by their effort and proficiency:

$$\bar{N_j} = (1 - \theta)(e_H \psi_H) + \theta(e_L \psi_L) \tag{3.7}$$

If an employee does not meet the average level of output they will be fired which results in a greater disutility than supplying any level of effort. This constraint means that a worker's output should be equal to their expected output, which is thus equal to the average expected output of the other employees at the establishment. In other terms equation 3.5 should equal 3.7, giving the following:

$$e_j \psi_j = (1 - \theta)(e_H \psi_H) + \theta(e_L \psi_L) \tag{3.8}$$

From 3.8 it is possible to see that an unproficient employee must put in higher levels of effort to achieve their expected product, with the opposite applying to proficient workers. Assuming the worker is unproficient (ψ_L), by differentiating with respect to the share of proficient workers $(1-\theta)$ it is possible to see that the level of effort given will *decrease* with θ if;

$$e_H \left(\frac{\psi_H}{\psi_L} \right) > e_L \tag{3.9}$$

This thus means that the level of effort, e, will increase with an increased share of proficient workers (ψ) if the difference in effort between proficient and unproficient workers is smaller than the difference in proficiency. This is likely to occur in circumstances where workers' effort, which is more observable than proficiency, is monitored and thus unable to fall in proficient workers. Further to this, for complex jobs the expected level of ability would be higher, thus meaning that the gap between proficient and unproficient workers could be greater possibly widening the gap. In these circumstances, a higher share of proficient workers causes an increase in the effort supplied by lower proficiency workers whilst their own level of effort remains constant.

Referring back to equation 3.4, this suggests that firms' production will increase with a lower share of skill gaps,(θ), only if there is a method of observing and controlling effort or if the gap between the proficient and unproficient workers is suitably large. Thus decreased θ may not actually increase output but rather lower workers' effort across the establishment.

The final possibility is that an increased share of skill gaps increases productivity due to an unseen increase in technology, the A term in the Cobb-Douglas production function. When establishments face skill gaps it is possible that they restructure the workforce in order to counteract these gaps. These new working practices may have a large enough effect to cancel out the negative effect of the skill gaps, and instead cause productivity to increase. While this is inefficient and the workplace could produce more if it kept its new working practices and removed its skill gaps this could still be a possible outcome for workplaces.

While these theories seem mutually exclusive this is not necessarily true and numerous equilibria could exist, depending on the level of skill gaps that the establishment currently has. For example, if a workplace suffers from a small number of SGs it may restructure causing productivity to increase above its original level. However, if the skill gaps are too great then the productivity loss would be greater than the increase from restructuring and the establishment would suffer a fall in productivity. The extent of the SGs can thus lead to different equilibrium emerging, with contrary effects on productivity possible. This should be taken into account when estimating this relationship empirically, to ensure the positive effects and negative effects of skill gaps do not appear to cancel each other out and hide their true impact.

While the above framework assumes workers are homogeneous in all regards except for proficiency this is obviously not true. There are workers with different skill levels in the labour force, with each of these able to be proficient or unproficient. Whilst we do not have a measure of the worker's background we can control for this heterogeneity to some degree by basing our estimates on the share of proficient workers in each occupation rather than just the share of proficient workers in the establishment as a whole. This makes no difference to the effort-proficiency trade-off observed earlier, which is proved below in an example with two occupations, managers and skilled trades.

With both managers and skilled trades equation 3.8 can be rewritten as follows:

$$e_i^j \psi_i^j = (1-\theta^M)(e_H^M \psi_H^M) + \theta^M (e_L^M \psi_L^M) + (1-\theta^{ST})(e_H^{ST} \psi_H^{ST}) + \theta^{ST}(e_L^{ST} \psi_L^{ST}) \tag{3.10}$$

Where superscript M denotes Managers and superscript ST denotes Skilled Trades.

Again, assuming the worker is unproficient and differentiating with respect to the share of proficient managers will give the following:

$$e_H^M \left(\frac{\psi_H^M}{\psi_L^M} \right) > e_L^M \tag{3.11}$$

Where the productivity of managers will only increase with a higher share of proficient managers if there is again some constraint put on effort.

The next section of this work will look to empirically test this model using the matched NESS and ABI dataset. If effort and proficiency are perfectly substitutable, then it would be expected that a higher share of proficient workers would only increase productivity in industry/occupations where effort can be constrained. If this is not seen and productivity increases are observed in all areas from an increase in the share of proficient workers, it is clear that effort is not substitutable with proficiency, or not substitutable to such a degree. If it arises that the depth of the skill gaps is important then the issue of separate equilibria is important and steps should be taken to accurately investigate it.

3.4 Estimation

We estimate the relationship between skill gaps and productivity based on equation 3.4 using the matched dataset of the NESS and the ABI discussed in Section 5.

Estimating establishment productivity is not a new idea, with Von Thuenen collecting data at his farm to measure the marginal product of inputs and substitutability in the early 1800's. The pitfalls of this type of estimation have also been well discussed with perhaps the most important of these issues, endogeneity of inputs, being raised by Marschak and Andrews (1944). For a standard OLS regression to give reliable estimates the inputs in the production function, mainly capital and labour, must be exogenous (determined independently from the firm's efficiency level). Maschake & Andrews noted that inputs in the production function do not have this trait and are rather decided by the firm's characteristics and efficiency. This is expressed intuitively in the work of Griliches and Mairesse (1995), where they state that the choice of inputs is not under the control of the economatrician, but rather determined by the individual firms.

Because inputs decisions may be partly determined by the firm's productivity beliefs (Olley and Pakes (1996); Ackerberg et al. (2007)) then a negative productivity shock will likely lead to decreased variable input usage and a downward bias on the coefficients of the input variable. With numerous input variables calculating the effect of this simultaneity problem on productivity is difficult.

This endogeneity problem, often referred to as the simultaneity problem, has been the focus of a large amount of methodological work in the area. Techniques ranging from fixed effects to instrumental variables (Griliches and Mairesse (1995)) to more recent models created by Olley and Pakes (1996), Arellano and Bond (1991), Blundell and Bond (1998) and Levinsohn and Petrin (2003) have been used in an effort to combat this.

Fixed effects estimators have long been used in the production function literature (Mundlak (1961)) and by only including within-firm variation in the sample they remove the endogeniety problem discussed previously.

Secondly to this, fixed effects estimation can overcome some of the selection bias that can occur from endogenous exits in the sample. This bias exists when firms that enter or exit within a period are omitted from the analysis. There are several theoretical models (Jovanovic (1982)) that suggest the exit of firms is driven to a large extent by the productivity differences at the firm level with empirical work backing up this finding (Farias and Ruano (2005)). This can lead to biased capital coefficients and can cause in-firm productivity estimates that are upwards biased. If, however, exit decisions are determined by time-invariant, firm-specific effects then fixed effects can eliminate the selection bias.

Despite the theoretical appeal of the fixed effects estimator it does not seem to perform well in the empirical literature. Olley & Pakes (1996) use the fixed effects estimator on both an unbalanced and a balanced sample and find large difference in the two sets of coefficients estimated, suggesting that some of the underlying assumptions do not hold in practice. The poor performance is further supported by Ackerberg et al. (2007).

Due to concerns like this, several other models have been created in the productivity literature including Arellano and Bond (1991) and Blundell and Bond (1998). Both of these methods use lags as instruments which are then inserted into a first difference model, ensuring that the instruments are not correlated with the error term in the model and there is no endogeneity. While this type of estimation has been popular in the recent literature it relies on having suitable panel data in a similar manner to the previously discussed fixed effects estimator. A more detailed description of the classic estimation approach will be given below.

3.4.1 Standard Productivity Estimation and Limitations in this work

As previously mentioned, a large array of work has looked at firm performance and a standard practice has become apparent when using the ABI, both in terms of the production function estimated and the estimator used (Griffith (1999); Harris and Robinson (2003); Harris (2002)). This practice will be outlined below before reasons are given for why it was not used in this work.

The commonly used production function is derived from a simple Cobb-Douglas function of the form:

$$Y_{it} = A_{it} K_{it}^{\alpha} L_{it}^{\beta} X_{it}^{\gamma} \tag{3.12}$$

This function is then transformed to be log linear,

$$y_{it} = \alpha k_{it} + \beta l_{it} + \gamma x_{it} + a_{it} \tag{3.13}$$

Whereby the residual, a_{it}, is interpreted as total factor productivity and can decomposed to give the following,

$$a_{it} = \eta_i + t_t + e_{it} \tag{3.14}$$

where η_i represents establishment specific differences in productivity that are fixed over time, t_t captures common macro productivity shocks and e_{it}, captures the establishment specific shocks to productivity which are assumed to be idiosyncratic and serially uncorrelated. This augmented production function can then be lagged and estimated.

While this approach is common, with the resulting Cobb-Douglas function being adapted to contain the variable of interest in a given paper, it is not appropriate for this work. When investigating skill gaps the assumption that a unit of labour is a constant factor that can input into a production function obviously does not hold. This was discussed previously when it was outlined that workers were an effort of both their effort and their ability. Due to this, measuring labour in the traditional sense of 'number of employees' is not correct. Instead, a production function is used which contains labour input costs rather than units of labour. Also, as the units of labour do not function heavily in the production function used in this work over aspects of the function were transformed to be in per capita terms to aid with interpretation.

The second major way that this work differs from the standard methodology is in the estimator used. The traditional view is that establishment level unobservable factors may be correlated with the right hand side regressors and thus estimates from an OLS regression will be biased. To counter this first-difference GMM is traditionally used (Arellano & Bond (1991, 1998)). This method will allow consistent estimates to be obtained but will suffer from finite sample bias due to the fact that the levels of variables are used as weak instruments for differences. The alternative is to use a system estimator (Blundell & Bond (1998a, b). This method should provide consistent estimates even if the firm-specific component of the error, η_{it}, is correlated with other independent variables and if independent variables are persistent and thus provide weak instruments in the earlier described first-differenced model. While these models are sophisticated and remove the standard endogeneity of inputs problem seen in this area, they rely on having suitable panel data. While the hope was that by matching the ABI and ESS together a suitable panel could be created, this was not possible in this work (see the data section for further details). Due to this an alternative approach was implemented exploring the use of instrumental variables estimation to remove endogenous issues and Heckman/PSM models to remove selection problems.

3.4.2 Instrumental Variables

Instrumental variables (IV) estimation involves, as the names suggests, instrumenting the independent variables that are the cause of the endogeneity issue, in this case the inputs in the production function. Unlike the fixed effects estimator previously discussed, IV estimation does not rely on strict endogeneity of the inputs for consistent estimation Wooldridge (2010).

For the sake of illustration, consider the simplified example whereby productivity, y, is being estimated by inputs, x and an error term, u. In this simple model it is assumed that x is uncorrelated with u, the error term. The only effect that x has on y can be seen as the direct effect via the measurable βx term. This is displayed diagrammatically below.

$$\begin{array}{ccc} x & \longrightarrow & y \\ & \nearrow & \\ u & & \end{array}$$

In the above diagram there is no directional arrow between x and u as there is no relationship between the two. This means that a simple OLS estimator with give a consistent estimate of β. However, due to the endogenity of inputs seen in production functions a more appropriate diagram for the relationship is given by;

$$
\begin{array}{l}
x \longrightarrow y \\
\uparrow \nearrow \\
u
\end{array}
$$

whereby there is an association between the inputs (x) and the error term (u). In this case the OLS estimate of β is a combination of the true β value and the indirect effect that is included due to the association between u and x. This means there is endogeneity bias in the estimate.

IV gives a possible solution to this approach. A new variable, z, is introduced which is an instrument. This means that z is associated with changes in x but does not lead to changes in y, except through the indirect effect via x. This leads to the following diagram;

$$
\begin{array}{l}
z \longrightarrow x \longrightarrow y \\
\quad\;\; \uparrow \nearrow \\
\quad\;\; u
\end{array}
$$

which provides a consistent estimate for β provided that the instrument z is uncorrelated with the error term u, and correlated with the regressor x.

This can be seen more formally below. Given the equation;

$$Y_i = \beta_0 + \beta_1 x_i + u_i \tag{3.15}$$

where β_0 is the constant term, x_i is an endogenous regressor and u_i is an error term. Two stage least squares can be seen to break the x_i term into two parts, one part that might be correlated with u_i, and a part that is not. By isolating the part that is not correlated with u_i it is possible to estimate β_1. To do this, an instrument, z_i, which is uncorrelated with u_i, is used. This instrument picks up changes in x_i that are uncorrelated with u_i and uses these to estimate β_1.

The first stage of a two stage IV isolates the part of x_i that is uncorrelated with u_i, by estimating the following regression;

$$x_i = \pi_0 + \pi_1 Z_i + v_i \tag{3.16}$$

Due to Z_i being uncorrelated with u_i then $\pi_0 + \pi_1 Z_i$ will also be uncorrelated with u_i. Once the $\pi_0 + \pi_1$ terms have been estimated it is possible to predict the values of X_i, which are denoted as $\hat{X}_i$.

The second step is thus estimated where the regression of interest is run with the X_i terms replaced with their predicted values, $\hat{X}_i$, giving;

$$Y_i = \beta_0 + \beta_1 \hat{X}_I + u_i \tag{3.17}$$

This allows a consistent estimate of β_1 to be given by OLS.

To ensure robustness in the IV estimator the instruments used must meet the following three requirements (Greene, 2008):

1) The instruments need to be correlated with the endogenous inputs in the production function.

2) The instruments cannot be part of the production function directly.

3) The instruments cannot be correlated with the error term, and thus in this case with productivity.

From these above requirements it is clear to see that the strength of any IV work is reliant upon the choice of instruments.

One of the possible and often used instruments are the lagged levels of inputs. This can be done through using a standard production function or, more frequently used when the production function has been first-differenced (Wooldridge (2010)). The use of lagged instruments is common practice as the increased productivity in the current period cannot cause increases in the previous years inputs, thus the independent variables (Z_{it-1}) are uncorrelated with the disturbance term (ε_{it}). Also Z_{it-1} is likely to be highly correlated with Z_{it}, meaning that it is likely to give efficient estimates. It is possible, however, that ε_{it} is serially correlated, and if this is the case then Z_{it-1} will still be correlated with ε_{it}, making the instrument invalid. For this reason, a two-year lag of the inputs is also often used as an instrument (Z_{it-2}).

The model used contains the following: a measure of the establishments' productivity in per capita terms (Gross Value Added), the total share of proficient workers at the establishment, the share of skill shortage vacancies the establishment has out of total employment, interaction terms to pick up if SGs and SSVs have more of an effect on those establishments operating in global markets (INTSG and INTSSV), a dummy for if managers in the establishment have SGs or SSVs, the share of managers with SGs or SSVs, region dummy variables, industry dummy variables, a measure of capital stock per capita, labour costs per capita, as well as establishment level characteristics.

3.4.3 Selection Bias

Due to the interesting results found by Bennett and McGuinness (2009) the issue of selection bias cannot be ignored when investigating skill deficiencies. If there are unobserved characteristics that determine if a workplace has skill deficiencies then it may be that the observed effect on performance is biased as it is also capturing this selection effect. For example, if those establishments that are more likely to have skill deficiencies are those that are more innovative then they are also likely to have higher productivity than the counterfactual group that has no skill gaps. This will of course provide a bias in the results as the skill measure will pick up both the true effect of skills and also this selection issue. This appeared to be the case in Bernett & McGuiness where skill gaps were having a positive effect on productivity in their investigative regressions but a negative effect when the selection issue is controlled for.

3.4.4 Heckman Selection Model

To control for this selection issue, two possible methods exist. Firstly, a Heckman (1979) selection model could be estimated, the approach taken by Bernett & McGuiness. This technique uses a two-step approach to remove the selection issue, which Heckman describes as simply an alternative form of omitted variable bias. The first step in this model involves a probit model based up the key characteristics that could effect selection into the sample. The second step of the model involves then running a standard regression with a control term (the inverse Mills ratio) included as a regressor to remove the selection bias. In this case the first step can be seen as factors that effect that likelihood of having a skills gap and the second step can be seen as the productivity regression with the control variable included and so conditional on selection. The inverse mills ratio is given by;

$$\frac{\phi(Z_i\gamma/\sigma_0)}{\Phi(Z_i\gamma/\sigma_0)} \tag{3.18}$$

where ϕ represents the normal density, Φ represents cumulative normal density and γ_i is obtained from the first step probit model, thus allowing the second step to be estimated as follows;

$$Y_i = X_i\beta + \frac{\phi(Z_i\gamma/\sigma_0)}{\Phi(Z_i\gamma/\sigma_0)}\tilde{\sigma} \quad (3.19)$$

As the inverse mills ratio captures the effect of the omitted variable by including it the omitted variable bias is lost (see Heckman (1979) for a full explanation of this process).

In order to ensure that the Heckman model is correctly specified at least one regressor from the first step must be omitted from the specification in the second step (Hilmer (2001)). In order to theoretically choose the variable to leave out (often refered to as the instrument) it is necessary to find a variable that both determines the likelihood of having a skill deficiency and does not have an effect on the productivity of the establishment. Finding a variable that fits these criteria is difficult and Bernett & McGuiness avoid this issue by empirically choosing the variable to remove from the second step (i.e. use a variable that is significant in the first step but is not in the second).

Whilst this approach is possible, it is better if there is a theoretical reason for the variable to be selected as an instrument. No perfect variable exists but a measure of whether the establishment formally assesses if workers have a SG or not seems to be the best possible variable in this dataset. While this variable does not pick up the true effect of variation around whether establishments have SGs, it provides a measure of the variation that may be seen in the data. For example, any SGs that are observed in the data are comprised of both a true SG and an establishment reporting effect.

$$\text{Measure of SGs} = \text{True SG} \times \text{Establishment Reporting Effect} \quad (3.20)$$

The instrument that is tested picks up the latter part of this, the variation within the establishments reporting habits rather than differences within the true SG number, but this is still enough to hopefully translate into differences in the measure of SGs seen in the dataset. If there is a correlation between the "true SG" level and the "reporting effect" then this instrument will work well. Obviously, if these two elements are not correlated then this method will have severe limitations (it will only pick up the variation in SGs across establishments due to reporting issues. Actual productivity will be affected by real SGs, whether reported or not.

An alternative to this is to not select an instrument at all but rather let the model identify itself based on functional form. This is not ideal but could be used as a test to see how much the instrument alters the results.

3.4.5 Propensity Score Matching

The second method that could be used to remove this selection issue is Propensity Score Matching (PSM). PSM is a very popular estimation method in recent years, normally used to evaluate the effect of interventions which can be traced back to Rosenbaum and Rubin (1983). With non-random assignment into treatment and control groups the challenge is to create a credible counterfactual to give an estimate of the outcome for the participating group if they had not participated. Obviously, it is impossible to observe the outcomes of the treated if they were not treated and non-random assignment means that the treated may differ in their characteristics to a chosen control group. The idea behind matching is simple, to find a non-treated unit that is 'similar' to a participation unit and to use this counterfactual to estimate the impact of the treatment. By finding a match to compare with for all units and taking an average of the treatment effect the mean impact on the dependant variable can be estimated.

While the process of matching sounds relatively straightforward, it is often difficult to both define and justify what a 'similar' observation is. If the matching is going to prove successful and remove any potential bias then a full range of covariates which may differ between the treatment and counterfactual group must be considered and controlled for. Rosenbaum and Rubin (1983) show that if potential outcomes are independent of treatment conditional on covariates X, they are also independent of treatment conditional on a balancing score $B(X)$. This allows the list of covariates to be reduced into a manageable single dimension, a propensity score, defined as the probability of a unit in the combined sample of treated and non-treated groups receiving the treatment, given all the relevant controls.

Denoting the impact of a treatment as D, it is possible to see the effect of a treatment as the difference between the potential outcome for those who received the treatment (Y_1) and for those that did not receive the treatment (Y_0);

$$D = Y_1 - Y_0 \tag{3.21}$$

When trying to evaluate the impact of policies a common standard of measurement is the Average Treatment Effect on the Treated (ATT);

$$ATT = E(Y_1 - Y_0|D = 1) \tag{3.22}$$

This can be re-written as;

$$ATT = E(Y_1|D = 1) - E(Y_0|D = 1) \tag{3.23}$$

The problem with this equation is that there is an unobserved element that is by nature an unseen counterfactual (the $E(Y_0|D = 1)$ term is not observed as the outcomes that the treated individuals would have obtained in absence of treatment cannot occur). It is possible to observe the term $E(Y_0|D = 0)$, which under random assignment should equal $E(Y_0|D = 1)$ and thus the ATT can be written as;

$$ATT = E(Y_1|D = 1) - E(Y_0|D = 0) \tag{3.24}$$

if there are no differences between the treated and the non-treated groups (i.e. $E(Y_0|D = 0) = E(Y_0|D = 1)$ or in other words if there is no selection bias).

Under non-random selection these two terms are not likely to be equal and thus there is a bias.

$$E[Y(1)|D = 1] - E[Y(0)|D = 0] = \tau_{ATT} + E[Y(0)|D = 1] - E[Y(0)|D = 0] \tag{3.25}$$

The role of PSM is to try and keep $E(Y_0|D = 1) = E(Y_0|D = 0)$ by finding the best match for each observation and thus allowing an accurate estimate of the outcome to be made.

In order to ensure matching can provide unbiased results two criteria must be met.

1) *Conditional Independence Assumption (Selection on Observables)*

If treated units are to be matched to untreated units that are similar in all relevant characteristics then it is necessary that all of these characteristics can be observed. The inability to measure one or more characteristic of importance in the selection process can result in biased estimates. The CIA dictates that after controlling for the the relevant characteristics, X, the potential outcomes are independent of the treatment status, or formally;

$$(Y_1, Y_0) \perp D|X \tag{3.26}$$

2) *Common Support Condition (Overlap Condition)*

The Common Support Condition ensures that there is sufficient overlap in the characteristics of the treated and un-treated units so that adequate matches can be found for each observation. This relies on the fact that for each value of X there is a positive probability of being both treated and un-treated,

$$0 < P(D = 1|X) < 1 \tag{3.27}$$

and thus due to the laws of probability $P(D = 0|X)$ is also between the values of 0 and 1 Heckman et al. (1999).

In order to estimate a Propensity Score a logit or probit function is normally used as they only give values between zero and one, an important characteristic when predicting probabilities. The specification used when calculating a propensity score is also important as all variables that determine participation in treatment should be included.

Once a propensity score has been calculated the next important step is to choose a matching algorithm. There are numerous different ways matches can be formed from a propensity score, all with their own benefits and pitfalls.

Nearest neighbour matching (NNM) is perhaps the simplest of these matching techniques where a comparison for a treated unit is found by choosing the observation with the closest propensity score. Nearest neighbour can be conducted with replacement, without replacement and can also be carried out with more than one neighbour being matched to each treated individual (k-nearest neighbours). Whilst NNM benefits from being efficient and creating a match for each observation there is a trade off as poor matches can be found where a unit may be the 'closest' match but may not actually be that 'close' to the treated observation.

To combat the issue of the closest match not being 'close' enough radius matching can be conducted. With radius matching, a caliper approach is used whereby there is a maximum difference of propensity scores set and any matches inside this radius will be used and all matches outside the radius are rejected. This ensures that there are no bad matches and that as many comparisons as possible are used within the preset caliper.

Further techniques exist such as Kernel matching, a non-parametric way of matching together treated and non-treated observations. This works by comparing the outcome of each treated person to a weighted average of the outcomes of all untreated units, with the highest weight put on those with propensity scores closest to the treated unit. This method obviously uses more information but again faces the issue of some of the matches being poor. This bias/efficiency tradeoff is common in PSM as the bias is minimised by using only the nearest observations but this is less efficient as lots of information has to be disregarded.

In this work both NNM and Kernel matching are used. Whilst it would be ideal to also compare these results to caliper matches this sort of matching is computationally difficult and thus was impractical to conduct on the Secure Data Service.

3.5 Data

This study draws upon the Employers Skills Survey (ESS), the Annual Business Inquiry (ABI) and the Business Structure Database (BSD). These surveys will be discussed in turn below before the process of matching the datasets together is outlined.

ESS

The Employers Skills Survey (ESS), formally known as the National Employers Skills Survey (NESS), is a nationally representative cross sectional survey of establishments that commenced in England in 1999 and underwent major reform in 2003 becoming the NESS. The survey was recently extended in 2011 to include Wales, Scotland and Northern Ireland in order to give the first truly UK wide skills survey. While skills have obviously always been the main focus of the survey, the consistency that can be seen from the 2003 wave onwards allows robust analysis to be conducted and consequently these waves are the focus of this study. The survey is conducted biennially and covers all sectors, regions and establishment sizes above working proprietors. The survey contains around 80,000 workplaces per wave, thus covering around four percent of the establishments in the

UK, with a slight over sampling of larger workplaces.

Of particular interest in this work is the detailed information the survey includes on both internal and external skill gaps. Employers are asked whether at the time of the interview they have any vacancies, whether any of these vacancies are proving hard to fill and if any of the vacancies are proving hard to fill due to lack of skills (where lack of skills is defined as a lack of appropriate skills, qualifications or experience in the labour market). Further to this information on external skill shortages, workplaces are also asked at an occupational level, "how many (staff) do you think are fully proficient at their job? A proficient employee is someone who is able to do the job to the level required". These questions provide a wealth of information on the current skill situation of the firm both in terms of the workers that the establishment already employs and the recruitment situation the establishment faces.

In addition to the information on skills, the ESS also contains a wide range of information about the establishment, which includes workplace size (number of employees), the sector and region the establishment is in, whether the establishment is part of a larger organization and whether the establishment is in the private sector, the public sector or is a charitable organization. A breakdown of the above variables from the final dataset can be seen in Table 3.6.

ABI

The Annual Business Inquiry (ABI) is a large survey of firms in the UK based on the Inter-Department Business Register (IDBR) similar to the ESS. The ABI contains a population of firms with more than 250 employees and a sample of firms which are smaller than this, stratified by size, region and sector. The ABI is an annual survey of businesses covering the production, construction, distribution and service industries in the UK. These sectors account for about two thirds of the UK's whole economy in terms of Gross Value Added. The main industries excluded are: Agriculture, Financial Intermediation, Public Administration and Defense, Education, and Health, with the finance sector only being partly covered (Insurance and Reinsurance only) in the ABI from 2008 onwards. A detailed description of the ABI can be found in Criscuolo et al. (2003).

It is important to note that the ABI contains three distinct levels of firm aggregation, making it different to the ESS that is just carried out at the establishment level. The three levels covered in the ABI are as follows: the firm or enterprise which is the whole organisation, individual workplaces (which collectively would form an organisation) and various groups of workplaces that the firm may choose to report in, is referred to as reporting units throughout.

The ABI contains information on output (in numerous forms) employment, input materials, investments, wage costs and many other firm characteristics. While the ABI has many variables of interest, it does not contain a direct measure of capital stock, instead including details on each firm's capital expenditure during each year. From this, it is possible to estimate the capital stock of the firm using the Perpetual Inventory Method (PIM). This PIM involves deflating capital stock, adjusting for depreciation and summing over time to give an estimate of the total capital stock.

While the method is relatively simple in theory several problems occur when trying to implement it using the ABI. As the ABI contains a population of large firms but only a sample of smaller firms these smaller firms may not be included in each year of the survey. While a firm may not be sampled in the ABI in that particular year it seems reasonable to assume that these firms may still be investing in capital during the year. This can thus lead to missing values in the capital expenditure and thus cause the final capital stock value to be underestimated. Due to this, it is required to impute these missing observations in order to give a reliable estimate of the final capital stock for firms.

The first step in the process of creating an accurate capital stock variable is to adjust the reported capital expenditure values for inflation. This is done by using the Volume Index of Capital Services (VICS) from the Office of National Statistics (ONS) which is split by year, industry and sector. Once the values for each firm's expenditure have been deflated across the years of the survey the task of filling in the missing observations where firms have been included in one wave but were not included in all waves of the survey is undertaken. The missing values are imputed in a two-step approach based upon the number of employees in the firm. Firstly, missing data in the number of employees variable is imputed using linear interpolations of the firm and a localised average for missing values in the first year of the firm's existence. The second step then involves calculating an investment per employee variable for the years in which capital expenditure is recorded, to give an average investment for each worker. The missing capital expenditure figures are then replaced with the interpolated number of employee data multiplied by the estimated average investment per worker. Finally, once the estimates of capital expenditure have been created for all firms with missing figures then the PIM method of capital estimation can be conducted for each firm. Where negative results are estimated it is assumed these are incorrect and underestimates due to the missing years coinciding with a large 'lump' of investment from the firm. Due to this, action is taken to reduce the number of negative series in the data. Further details on the PIM method of capital stock estimation can be found in Martin (2002).

BSD

The Business Structure Database (BSD) contains almost all business organisations in the UK with only a limited number of variables per observation. The BSD is mainly derived from the IDBR with further information added from the ONS business surveys. This means that the BSD includes any business that is liable for VAT (turnover exceeds the VAT threshold) and/or has at least one member of staff registered for the PAYE tax collection system. It was estimated that in 2004 the IDBR accounted for almost 99% of economic activity in the UK with only extremely small businesses such as the self-employed being excluded.

While the IDBR and the BSD are frequently updated the snap shot available in this data is taken annually in April with reporting periods given as the financial year. The dataset contains details on both enterprises and local units, with information on employment, turnover and foreign ownership included.

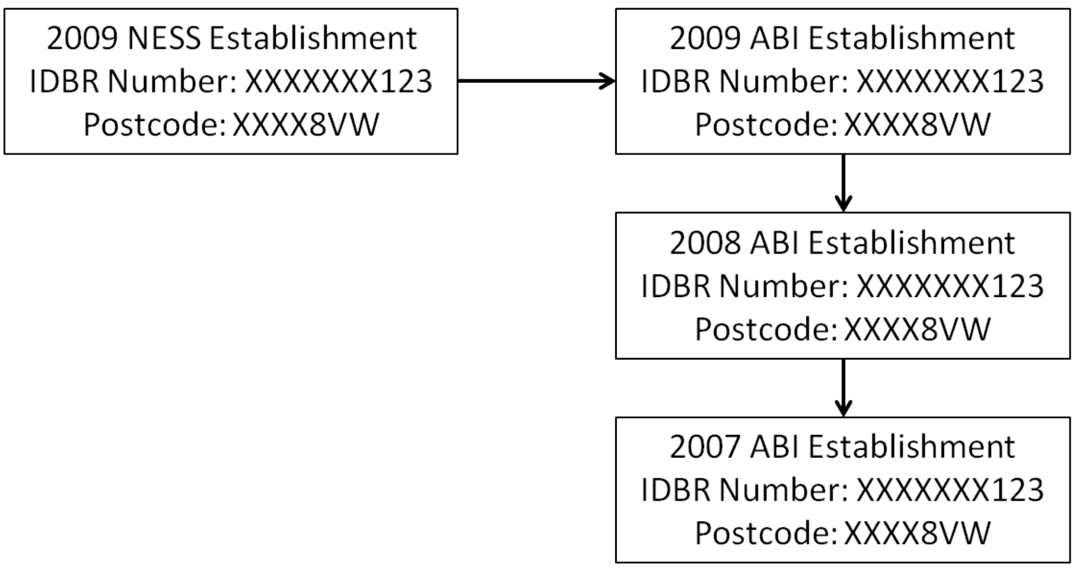
Matching

The fact that the ABI and NESS have different levels of aggregation causes some difficulties when matching them together. In this work, reporting units are matched to establishments in the NESS using a two step approach. This will provide a final dataset at the establishment level, but with some of the business information being at the level of the reporting unit that the establishment belongs to.

The first step of the match involves linking the establishments in the NESS to single reporting units in the ABI using their unique IDBR number. Where reporting units' IDBR numbers are not unique, and thus did not represent the whole firm in the ABI, they were matched to the appropriate NESS establishment using both the IDBR number and the site's postcode. Once the three waves of ABI data were matched with the NESS then postcode information only helped match less than ten establishments. Due to a fear that these matches may differ from those matched using the simpler method, these establishments were excluded from the final analysis.

In order to avoid endogeniety of inputs, past values of inputs are used as instruments in the work as previously discussed. Due to this, the 2009 dataset was also matched to the 2008 and 2007 waves of the ABI. As the ABI is not a panel dataset, this means that the sample size decreases with each match, as those firms that have not been re-sampled in the ABI are dropped. Once the 2009 NESS has been matched to the 2009, 2008 and 2007 waves of the ABI only 2,170 establishments remain, each with three years of ABI information. This set up can be seen in Figure 3.1.

Figure 3.1: Matching numerous waves of the ABI to the NESS



While the numerous waves of matching reduced the final dataset down to 2,170 establishments the analysis conducted in this work is on a reduced sample of 1,497. This final reduction was due to some establishments in the sample being located in industries that are classified as 'public'. While these establishments cover a range of areas (all two digit SIC codes from 75-99) none of these industries are primarily concerned with making profits. Due to this, the establishments in these areas were dropped as GVA may not be a matter of interest for the establishments and thus they may bias the results.

The dataset that is used for the analysis is made from matching the ABI, the BSD and the NESS and thus the size of the sample is dependent upon the number of matches obtained. The number matched with the BSD does not have any impact on the sample size as the the BSD contains all workplaces included in both the ABI and the NESS. The match between the NESS and the ABI is likely to generate a final dataset that has a higher proportion of large workplaces than the general UK population due to the larger firms being included as a population in the ABI rather than a sample meaning they are more likely to be in consecutive waves of the survey. This is likely to be further exaggerated by the fact that numerous waves of the ABI are being matched together so that lagged values of inputs can be used as instruments in the analysis. This was investigated in Tables 3.2-3.4, where the original NESS sample is compared to the final matched dataset.

Table 3.2 looks to investigate the bias from matching by comparing the matched dataset (three years of ABI matched together and to the NESS) to the original NESS sample. As can be seen the final matched dataset is biased towards larger workplaces as expected with the 2-4 establishment size band being significantly smaller in the matched sample than in the NESS. There is very little that can be done about this but it is worth noting that analysis on this dataset may be driven by this bias.

The middle column of Table 3.2 shows the share of establishments of varying sizes if only two waves of the ABI are matched together, 2009 and 2008. This was done in an effort to see if the benefits of matching three waves together and thus having two lags of inputs available for the productivity estimation was worth the trade off in bias of having a dataset that is skewed towards larger establishments. As Table 3.2 shows the bias towards larger workplaces is relatively strong by the time two years of the ABI are matched together and to the NESS, and thus very little extra bias in introduced to the dataset. The final column of Table 3.2 shows the share in each size band when the NESS is weighted to the population. As can be seen, this further emphasises the fact that our final dataset is biased towards larger establishments.

Table 3.3 does exactly the same as 3.2 but looking at region rather than at size. As can be seen from the table there is no significant bias in terms of region caused from our matching as we expected as both surveys are conducted within the same region framework.

Table 3.4 is the same as 3.2 and 3.3 but looking at industry rather than size or region. In similar fashion to Table 3.3, there is very little difference in terms of the percentage of establishments in the different industry bands framework. Some cells are left blank as limited numbers of observations in the cell meant the data could be disclosive.

While little can be done about the distortion of the establishment size in the dataset an effort is made to reweight the dataset back to population percentages for some of the following analysis.

3.6 Descriptive Statistics

Instruments

In terms of measuring capital inputs two possible measures exist; capital expenditure and capital stock. When testing the variables in the merged dataset for endogeneity, it appears that while capital expenditure is endogeonous, capital stock is not (both in terms of the Hausman test of OLS against IV and the Durbin-Wu-Hausman test). This makes sense in terms of capital stock having a pre-determined element (past investment in capital stock which will be exogenous) and a smaller element of capital investment in this period which may be endogenous. In comparison, capital expenditure is entirely determined in this time period and thus suffers from the endogeneity of inputs as discussed previously. Due to this, capital stock is used in this work as it not only avoids an endogeneity problem but also provides a better estimate of the true capital input as capital is not seen to have a 100% depreciation from one period to the next, and previous capital investments can still be used in production in this period.

Endogeneity in labour inputs is less easily avoided however. While there is a choice of ways to measure labour inputs in the ABI, the total labour cost per capita was chosen as the key variable for this work. While classic production functions tend to use the number of workers as the labour aspect this seems like a very impractical way of measuring labour when we are fundamentally concerned with the quality of labour in this work. While adding an extra worker is likely to affect the performance of the establishment this could be by very different amounts depending on the worker that is added. On the contrary, by using labour costs the ability of the worker is taken into account in the wage bill and an increase in the wages paid will be expected to provide a fairly constant increase in productivity. While this variable suits the purpose of the work it suffers from endogeneity of inputs. While capital does not depreciate fully from one time period to the next, labour costs do and thus a useful instrument needs to be used to avoid this endogeneity bias. Both the one year lag (LC_{t-1}) and the two year lag of labour costs (LC_{t-2}) were tested, but did not pass Hansen's J test for overidentification. Due to this, the second lag was dropped and thus labour costs are instrumented with the previous period's lagged value.

The base line specification used in this work is thus as follows;

$$\text{GVA per Capita}_i = \beta_0 + \beta_1 \bar{x}_i' + \beta_2 \text{SG}_i + \beta_3 \text{SSV}_i + \beta_4 \text{Capital Stock}_i ...$$

$$... + \beta_5 \text{Total Labour Costs 2008}_i + \varepsilon_i \tag{3.28}$$

where Total Labour Costs in 2008 is used to instrument for total labour expenditure, β_0 is the intercept, $\bar{x}$ is a K-dimensional row vector of independent variables, β is a K-dimensional column vector of parameters and ϵ is the error term.

Measures of Productivity and Skill Deficiencies

The main dependent variable in the majority of this work is productivity which is measured as Gross Value Added (GVA). The ABI contains information on GVA allowing an accurate figure to be used with no ambiguity in terms of different researchers creating the variable in different ways.

In terms of measuring skill deficiencies, the percentage of skill shortage vacancies (SSV) is the main variable for measuring external skill gaps in this work, with the share of non-proficient workers (SG) defining the internal skill gaps for a workplace. The percentage of SSV has a minimum of zero, where the establishment faces no external skill gaps, and has no upper bound as the establishment can have more external skill gaps than current employees. Internal skill gaps, SG's, on the other hand are capped at both zero and 100 as a workplace can only have as many unproficient workers as the number of workers it employs.

Table 3.5 shows a breakdown of the two main skill deficiencies measured in the NESS for the relevant matched dataset. The first two columns in Table 3.5 represent the number of establishments with any SSV, with 0 relating to no SSV and 1 representing having some SSV. The final column shows the total number of establishments in the row. The rows of Table 3.5 represent if the establishment has any SG, again with 0 being no SG and 1 meaning that the establishment has a positive number of SGs. Table 3.5 shows that the number of establishments with SG is much larger than the number of establishments with SSV and due to this the analysis conducted in this work focuses on SGs. The split between those establishments with SGs and those without is fairly even in the dataset with 58 percent of establishments having no form of skill gaps and 42% having SGs. The figures for SSV are also fairly evenly split in terms of percentages but the absolute size of both is obviously significantly smaller.

Further measures of skill shortages are also included in the analysis and can be seen summarised in Table 3.6. The 'AnySg' variable is a dummy variable, representing one if the workplace has a SG and zero otherwise. 'AnySSV' is the same but for if the establishment has any SSV at all. These variables are included to determine if having a gap is more important than the depth of the gap. Similarly, to determine if the occupation the deficiencies occur in is important to the productivity of the establishment, the measure of SGs for managers is also tested in the analysis. This variable is included in the same way as the standard SG variables and shows if there is a SG in the managers occupation (Variable titled: SGmanagers).

Furthermore, banded measures of SG and SSV are created to interpret if there are separating equilibria with different levels of shortage having differing effects on productivity. 'BandedSG0' represents those with no SG's, 'BandedSG1' represents the first quartile of SG's, and with 'BandedSG2', 'BandedSG3' and 'BandedSG4' showing the second, third and fourth quartiles of SG respectively.

The final skill measure used is titled INTSG. INTSG is an interaction term where a dummy variable for whether the establishment competes in a global market is interacted with the share of SGs the establishment has. This is included to see if the effect of skill deficiencies are greater on those workplaces that face increased competition due to being in worldwide markets.

Independent Variables

To control for other variations between the establishments that may explain some of the productivity differences seen numerous independent variables are included in the regression analysis. These variables will be discussed below along with their reasons for inclusion.

The size of establishments is included because different sized workplaces may face different skill deficiencies and establishment size may also influence productivity through factors such as economies of scale. Further to this, both industry and regions are also included due to similar reasoning. The mean size of establishments is 83 employees but there is a large standard deviation around this with some establishments being both very small and very large. Larger Org is a dummy variable that represents whether the establishment is part of a larger organisation. As can be seen here almost 95% of the sample are part of larger groups but that may be expected due to this analysis being conducted at the smallest possible level of aggregation, establishment level. FO is again a dummy, equal to one if the establishment is foreign owned and zero otherwise. Foreign ownership is included as foreign owned firms are often found to be more productive than local firms (Harris and Robinson (2003)).

The next string of variables represent the self reported quality level of products made at the establishment. These are again all dummy variables and as Table 3.6 shows there are very few establishments that report producing low quality goods/services or very low quality goods/services. Int Markets is a dichotomous variable that is equal to one if the establishment conducts business in international markets and zero otherwise. IntSG is an interaction variable whereby the Int Markets variable is combined with the SG variable to see if skill gaps are having a distinctly different effect on those establishments that are competing in international markets.

The lead variables are dichotomous measures of how often the establishment feels they are leading the market, varying from very often to very rarely. An interaction has also been included for those establishments that have SG and are leaders as it was expected that SGs in these workplaces may matter more.

The first two skill measures in Table 3.6 show the percentage of the Skill Gaps (SG) an establishment has and the percentage of Skill Shortage Vacancies (SSV) that an establishment has. 8% of workers in establishments suffer from SGs on average with SSV being a much smaller deficiency in this dataset. The following two variables are dichotomous variables that indicate if the establishment has any SGs or any SSV rather than the level of skill deficiency. Around 40% of establishments have an internal SG of some sort in the final dataset with less than one percent having a SSV (this is substantially less than in the ESS as a whole and is likely a function of the matched dataset). Due to this, the main focus of the analysis is on SGs are there is enough observations for robust statistical analysis to be conducted. The final skill measure included in this table is SG Man which is again a dichotomous variable equal to one if the establishment has someone in a managerial occupation that has a SG.

Finally, included in Table 3.6 are the measures of establishment inputs that are controlled for in the following analysis. Capital Stock has been derived from the information available in the ABI and is thus presented for the establishment in 2009, 2008 and 2007 as a measure of capital. Labour is measured for the establishment in terms of Total Labour Costs and this is simply taken from the ABI for the years 2009, 2008 and 2007. GVA is also included even though this is the dependent variable in the regressions.

The impact that SGs were having upon establishment productivity was investigated by banded measures as well as a continuous variable. Table 3.7 shows the distribution of establishments with SGs across these bands. As can be seen, the number of establishments with no SGs is much larger than the other bands with 1000 establishments. The rest of the bands only have around 200 establishments each.

Figure 3.2 shows the above bands plotted against mean productivity. The red line shows the mean productivity of all the establishments as a whole, at around 33,000. The 0 represents the establishments with no SGs, and is seen to be to the north of the red line, suggesting that establishments without SGs are more productive. This is obviously just descriptive evidence. The other values show the varying bands of skill gaps with 1 representing a small percentage of employees having skill gaps and increasing up through to 4. It is clear to see that the establishments that report the higher percentage of SGs are not those with the average lowest productivity. Again, this is only descriptive analysis but it may suggest that there is more to the picture of SGs than just a linear relationship. Due to the trends seen in the Figure 3.2 it makes sense to investigate how SGs are affecting productivity through banded measures rather than just a simple dichotomous variable, and as such this will be investigated further in the regression analysis.

3.7 Results

3.7.1 IV Results

Table 3.8 shows the results from the preliminary Instrumental Variables analysis for various specifications. The dependant variable in all of the regressions was GVA with both labour costs and capital stock instrumented, with two lags of labour costs and two lags of capital stock respectfully.

The first cells for each variable represent the coefficients on the variable with standard errors below.

The varying specifications tested are displayed in columns in Table 3.8 with the first column, X, containing a simplified specification, the second column, X+, containing a more detailed specification with dummies for quality, a dummy for if the establishment is in an international market and two more skill measures included. The final three columns build upon this extended specification, by including region dummies, industry dummies and finally both region and industry dummies.

The table shows that both of the main inputs into the production function, labour and capital, have positive coefficients in all the tested specifications. While the labour inputs' coefficient is significant at 1% in all of the specifications the capital coefficient is not significantly different to zero, and is very small in size in all the specifications. It is hard to know how accurate this figure is as no other work seems to use a production function with similar form to this (an specification without logs and with labour defined as labour costs). Work on production functions tends to find capital coefficients as smaller than labour coefficients (Eisner (1967)) however it has been noted that measurement error can play a large role in determining the size of capital coefficients, often making it appear smaller than expected (Lizal and Galuščák (2012)). Further to this, most of the work using the ABI has found that the coefficient on capital stock is insignificant or extremely small in terms of its size (see Griffith (1999), Disney et al. (2003), Harris (2002)).

The establishment size coefficient, measured in terms of employees, is negative and significant in most specifications suggesting that larger establishments are less productive. While this seems surprising, labour costs are being held constant in this regression. This therefore means that for a given level of labour costs more employees produce less, i.e. lots of cheaper employees are less productive than a few expensive workers that cost the same in total.

The foreign owned coefficient is the only other establishment characteristic that is significant, with a positive sign in all regressions. This is supported by the literature with foreign owned firms on the whole being more productive than native firms (Harris and Robinson (2003)). None of the coefficients on the extended set of independent variables are significant in multiple specifications.

The coefficient on SSV is insignificant in most specifications, however there are very few establishments with SSV and thus this was expected. The coefficient on SG is positive and significant, suggesting that those establishments with higher levels of internal SGs have higher productivity. This is contrary to what was expected, but there are numerous reasons why that this may be seen. Firstly, those establishments with higher skill gaps may be those that are more productive and growing faster and thus are struggling to fill vacancies with adequately skilled workers, resulting in SGs. This suggests that those firms with SGs are distinctly different to those without them and this issue is investigated in the latter Heckman and PSM work. The second alternative to explain this positive coefficient is that action is taken by establishments with higher SGs to work around these problems and this action is having an effect above and beyond the negative of the original productivity loss. For example, if an establishment puts into place new working practices to counter an internal skill gap it may be much more efficient than with the old practices and may actually increases productivity even when taking the SG into account.

The only other interesting result from Table 3.8 is that the interaction term between those establishments with SGs and those establishments that are international markets. The coefficient on this interaction term is negative and significant in almost all of the specifications, suggesting that skill gaps may be having a negative impact on those establishments that are global or at least operate in international markets.

Table 3.9 shows similar regression analysis to Table 3.8 however the skill measures are slightly different in this table. While the specifications tested in each column are the same as the previous table the external skill measure in Table 3.8 that represented the percent of establishments workforce with external skill gaps is replaced with dummy variables indicating if the establishment has a skill deficiency of this type in a given occupation. This was done in an effort to see if SGs in certain occupations are more important than the level of the SGs in the establishment as a whole.

The results suggest that this may be the case for the sales occupation (Sales) as the coefficient is positive and significant in all of the specifications tested. This shows that having an external skill gap in this occupation is having a positive impact on performance. This again seems to be unlikely and further supports the idea suggested above that it is not SGs that are positively impacting performance but rather that the establishments with SGs are different to those without SGs in some unseen way, and this factor is also related to productivity.

The coefficients on labour and capital remain similar to those in the previous table, with the coefficient on size being of a similar magnitude to Table 3.8 but no longer significant. The coefficient on Foreign owned remains positive and significant in all specifications. A result that was not seen in the previous table is shown here whereby the coefficient on the variable that indicates if an establishment produces a "Very High Quality Good or Service" is both positive and significant. This result suggests that establishments that produce high quality goods/services tend to be more productive.

Table 3.10 uses the same specification as Table 3.8 however it re-weights the results back to the population size bands that can be seen in Table 3.2.

The results from this table are very similar to Table 3.8 showing that the influence of the establishments size bias created when matching datasets together is relatively small in the analysis.

Table 3.11 shows similar results to the rest of the IV regression results but with the skill gap measure broken down into bands rather than as a continuous variable. This is done as SGs may have a non-linear effect which can be captured by including dummies for different levels of SG bands. For example, small levels of SGs could actually increase productivity as new working practices are promoted, whereas larger SGs result in losses to productivity as the workers are not skilled enough to undertake their jobs properly. If this is the case the results from the continuous measure of SGs that had been used in the previous IV regression results could be misleading. If the effect at one end of the scale is negative and the other is positive this could lead the linear approximation to see no significant result.

The results in 3.11 suggest that this is not a problem. The banded SG measures show very little evidence of this non-linear impact, as most of them have insignificant coefficients. Two of the 8-15% SG measures have a coefficient that is significant at the 10% level but this drops out as region and industry dummies are added to the specification. The rest of the coefficients seem to be a similar size, sign and significance to the previous IV results.

3.7.2 Heckman Regression Results

Table 3.12 shows the results from the Heckman selection model. In this two-step model the first step uses the question whether establishment formally assesses whether individual employees have gaps in their skills as an instrument that will affect the likelihood of an establishment reporting a SG but will not affect productivity, allowing the two impacts to be estimated separately. This variable is titled 'instrument' in the first step probit model.

The results in Table 3.12 are from the second step of the model where an IV regression was run using GVA in market prices as the dependent variable. The Inverse Mills ratio taken from the first step is included to control for the selection issue and capital and labour inputs are instrumented with their own lags. As the errors are not correct due to the two-step nature of this estimator they have been bootstrapped with 1000 repetitions each in both Tables 3.12 and Table 3.13.

The columns in Table 3.12 are the same as previous tables with the first column, X, containing a simplified specification, the second column, X+, containing a more detailed specification and the final three columns building upon this extended specification by including region dummies, industry dummies and finally both region and industry dummies.

The results from this Heckman model are very similar to the previous IV regressions with the coefficient on labour being significant, positive and almost equal to one. The coefficient on capital stock is very similar in size to previous results but is not significant here, presumably due to some efficiency being lost in the two step approach. The foreign ownership coefficient is one of the only other parameters that is significant in the regressions, again showing a positive and significant effect on productivity. The coefficient on the international SG measure is negative and significant in the Heckman model as was seen in the previous IV results.

The Inverse Mills ratio is not significant in this model, indicating that a two-step approach may not have been needed, with the coefficient on the SG variable that was previously positive and significant remaining positive and significant. This suggests that while there may be some selection issue at play whereby those establishments that have SGs are fundamentally different to those without SGs this is not being picked up here, possibly due to the instrument being weak. If this is the case then the effect should be noticeable in the propensity score matching (PSM) analysis.

Table 3.13 shows the results from the first step of the Heckman model. The coefficient on the instrument used in the first step of the Heckman model is not significant in any of the specifications tested. This is concerning as the instrument is obviously not picking up the desired correlation and is just identifying the measurement error. To account for this further PSM work will be done to investigate the difference between those establishments that have SGs and those that do not.

3.7.3 Propensity Score Matching Results

The following results are from the Propensity Score Matching analysis that was conducted to try to control for any selection issue in this work.

Table 3.14 shows the average treatment effect on the treated (ATT) and standard errors for numerous different matching methods on the chosen matching specification. The first column of interest (NN) shows the results for nearest neighbour matching with the next columns showing nearest neighbour with the closest 5 neighbours used, a Calliper match with a radius of 0.01 and a Kernel match. The ATT results from these four columns are all positive however none of them are significantly different from zero.

The quality of the match can be seen in Table 3.15, where the variables matched on, and their relevant test statistics, can be seen. As can be seen from the mean treated and the mean control columns of the table there was quite a large difference between the subgroup of establishments with SGs and those without before the matching took place, with some of the unmatched p values showing that the means are statically significantly different, (i.e. size, FO, etc). However, the match results show no significant difference on any of the hypothesis tests suggesting that the two groups now share the same characteristics and none of the biases in the matched results has an absolute value of greater than 5, further supporting the strength of the match.

Table 3.16 shows comparable results to Table 3.14 however, this time controlling for regions in the match as well as the characteristics controlled for in Table 3.14. As can be seen the results do not differ much with the ATT being positive but insignificant with all the different matching types. These results suggest that there was a selection bias in the previous IV results and the instrument in the Heckman model was too weak to control for this. Whilst there is not a positive effect of SGs on productivity as the earlier IV results suggested no negative effect of SGs has been found in this work.

3.8 Conclusions and Discussion

The 2013 wave of the Employers Skills survey, created by the UK Commission for Employment and Skills, shows that skill gaps are at a ten year low. While this at first seems positive (with fewer skill gaps meaning establishments are being more productive) skill gaps could be a characteristic of more innovative and more productive workplaces, where gaps exist as jobs are constantly evolving and new working practices are being created.

This work aims to provide evidence of the causal relationship between skill gaps and establishments' productivity. To do this it investigates the following hypotheses;

Hypothesis 1) Skill gaps will have some form of negative causal relationship with SGs.

Hypothesis 2) This relationship is likely to be negative and of greater significance when workers' effort can be constrained (i.e. in industries/occupations where output is more observable).

Hypothesis 3) The impact of skill gaps on productivity will not be linear, with different volumes of deficiency leading to differing impacts.

Hypothesis 1

While there appears to be a relationship between SGs and productivity, this appears to be positive in the standard IV regression results. This is contrary to the theory discussed in this work and the evidence from other authors. Due to the confusing nature of this result selection models were investigated to try to remove any existing selection bias. While the Heckman selection model use was uninformative, the propensity score matching appeared to remove this bias and show that rather than a negative relationship between the two variables, no relationship exists.

Hypothesis 2

The IV analysis investigating the impact of SGs in different occupations provided little evidence that SGs impacted establishments differently. This may be as the ability to measure a worker's output does not vary as much as expected, or given the result for hypothesis 1 that no relationship exists, this relationship cannot differ across occupations.

Hypothesis 3

The non-linear analysis investigating if differing levels of SGs have different effects on productivity was again inconclusive. While the preliminary descriptive statistics showed that a non-linear trend may exist this was unseen in the full analysis.

Discussion

This work finds no real impact of SGs on establishments' productivity. There could be numerous reasons why this result was not found however.

Firstly the final matched dataset that is used in this work is both small in size and is biased towards larger establishments. This lack of sample size and possible lack of variation in the type of establishments sampled may have washed out any negative effect.

The other possible explanation is that there is no effect of SGs on productivity. Establishments could adapt their working practices and make do with the employees that they have available, adapt the workloads of the workers that they have effectively, or it is possible that establishments are not impacted by skill gaps. If the latter of these is the case it may bring into question the self reported nature of the SG measure in the ESS. Any worker that is classed as not being fully proficient at their job should be less productive than a fully proficient worker and if the SG measure is not able to pick this up there could be problems with using a self reported measure.

Given the limitations discussed above, it seems premature to state that skill deficiencies are not impacting the performance of UK workplaces. While no impact is found in this work, a larger matched dataset may reveal more results of interest, and may also allow the assessment of external skill shortages and other forms of productivity.

One point of note from this work is that some of the analysis showed that the impact of SGs may be more apparent in more innovative establishments. This suggests that low levels of skill gaps are not necessarily a good indicator of a flourishing UK economy as has recently been the message. It is important where the skill gaps are and what impact this may have on establishments.

Table 3.1: **Table of Variable Definitions**

Skill-shortage vacancies (SSVs) -	Vacancies which are proving difficult to fill due to the establishment not being able to find applicants with the appropriate skills, qualifications or experience. The SSV variable is measured as the percentage of SSVs the establishments experiences.
Skills gaps (SG) -	A "skills gap" is where an employee is not fully proficient, i.e. is not able to do their job to the required level (see Glossary for further details). This is measured as the percentage of SGs the establishment is currently facing.
Any SG -	Dummy variable that indicates whether the establishment is suffering from any internal skill gaps (The establishment has at least one SG=1, other=0)
Any SSV -	Dummy variable that indicates whether the establishment is suffering from any external skill shortages (The establishment has at least one SSV=1, other=0)
SG Manager -	Dummy variable that indicates whether the establishment is suffering from internal skill gaps in the managers occupation (The establishment has at least one manager with a SG =1, other=0)

Larger Org -	Dichotomous variable to indicate if the establishment is owned by a larger organisation (The establishment is owned by a larger organisation=1, other=0)
Size -	Establishment size is included as a continuous variable
International Market -	The type of market the establishment competes in (International=1, non-international=0).
International SG -	Interaction term that indicates the SGs an establishment faces if the establishment competes in an international market.
Quality (Very High, High, Medium, Low) -	A series of dummy variables are used to determine the quality of the product. The levels are defined as Very High Quality, High Quality, Medium Quality, Low Quality and the base category of Very Low Quality. These quality measures come from the ESS question that asks establishments to indicate whether they compete in a market for a standard or basic quality product or service, or that they compete in a market for premium quality products or services.

Lead - (Very Rarely, Rarely, Lead, Very Often)	A series of dummy variables are used to determine to dynamic nature of the firm. Self reported based on the question "Compared to others in your industry this establishment...Very often leads the way, often leads the way, leads the way, rarely leads the way, very rarely leads the way"
FO -	A dichotomous variable indicating whether or not the given establishment is owned by a foreign company (Owned by foreign company=1, otherwise=0).
GVA Market Prices 2009 -	Gross Value Added in market prices for the firm in 2009.
Capital Stock (Year) -	Capital stock for the workplace in the given year as created using the PIM.
Total Labour Costs (Year) -	Total labour costs reported by the workplace in the given year.
SG leader -	Interaction term that indicates the SGs an establishment faces if the establishment identifies itself as a leader in the market.

Table 3.2: Matching Bias by Size Band

Size Bands	090807 Match	0908 Match	NESS 2009	NESS Reweighted to Population
2-4	7.65	8.25	28.47	51.29
5-19	14.84	15.04	23.55	18.87
10-24	21.01	21.05	21.19	17.72
25-49	22.4	22.17	13.71	6.35
50-99	13.41	14.2	6.84	3.15
100-199	9.63	9.75	3.38	1.47
200-250	3	2.93	0.96	0.4
251-499	5.3	4.49	1.19	0.48
500+	2.76	2.13	0.7	0.27
Total	2170	3478	79152	100

Table 3.3: Matching Bias by Region

Region	09/08/07 Match	09/08 Match	NESS 2009	NESS Reweighted to Population
East of Eng	10.05	10.64	10.8	11.43
Easst Midlands	8.25	8.37	9.27	8.42
London	13.55	13.89	15.16	15.78
North East	6.96	6.7	7.17	3.91
North West	13.82	13.46	12.53	12.37
South East	15.9	16.16	13.95	17.41
South West	10.46	11.13	10.99	11.43
West Midlands	11.43	10.87	10.34	10.04
Yorkshire/Humber	9.59	8.8	9.78	9.21
Total	2170	3478	79152	100%

Table 3.4: Matching Bias by Industry

Industry	09/08/07 Match	09/08 Match	NESS 2009	NESS Reweighted to Population
Agriculture	0.69%	0.60%	2.97%	4.34%
Mining	.	.	0.15%	70.07%
Manufacturing	9.26%	10.72%	11.84%	6.95%
Electrical	0.51%	0.55%	0.29%	0.13%
Construction	2.44%	2.90%	6.67%	7.46%
Personal	27.65%	28.18%	19.59%	21.25%
Hotels & Catering	15.94%	13.05%	7.09%	9.02%
Transport	11.11%	8.60%	5.69%	4.57%
Financial	.	0.66%	3.10%	2.44%
Real Estate	11.94%	14.43%	16.90%	23.16%
Public	.	0.60%	2.97%	0.81%
Education	4.38%	.	0.15%	4.12%
Health	7.42%	10.72%	11.84%	7.34%
Other	7.97%	0.55%	0.29%	8.34%
Total	2170	3478	79152	1,215,395

Table 3.5: Breakdown of Skill Deficiencies

	Any SSV		
Any SG	0	1	Total
0	1,222	39	1,261(58%)
1	855	54	909(42%)
Total	2,077(96%)	93(4%)	2,170(100%)

Table 3.6: Summary of Independent Variables

Variable	Mean	S.D	N
Size	83.81	189.31	2,170
Skill Gap %	8.36	15.16	2,170
SSV %	0.19	1.33	2,170
Any SG	0.42	0.49	2,170
Any SSV	0.01	0.11	2,170
SG Managers	0.14	0.34	2,170
Larger Org	0.94	0.23	2,050
FO	0.32	0.46	2,170
V Low Qual	0.02	0.14	2,170
Low Qual	0.05	0.22	2,170
Medium Qual	0.27	0.45	2,170
High Qual	0.27	0.45	2,170
V High Qual	0.21	0.41	2,170
Int Market	0.21	0.41	2,140
Int SG	0.04	0.19	2,170
GVA Market Prices 2009	27,745	159,378	2,168
Capital Stock 2009	88,554	535,634	1,900
Capital Stock 2008	89,541	537,477	1,900
Capital Stock 2007	91,010	515,904	2,064
Total Labour Costs 2009	24,397	151,034	2,169
Total Labour Costs 2008	25,407	154,396	2,170
Total Labour Costs 2007	25,692	156,250	2,170
V Rarely Lead	0.09	0.29	2,076
Rarely Lead	0.1	0.3	2,076
Lead	0.46	0.5	2,083
V Often Lead	0.53	0.5	2,087
SG Leader	0.87	5.71	2,076

Table 3.7: Breakdown of SG Banded Variables

BandedSG	Freq.	Percent	Cum.
0 (SG=0)	1,033	58.26	58.26
1 (SG: 0-8%)	187	10.55	68.81
2 (SG: 8-15%)	177	9.98	78.79
3 (SG: 15-25%)	189	10.66	89.45
4 (SG:25%+)	187	10.55	100
Total	1,773	100	-

Figure 3.2: Banded Skill Deficiencies Plotted against productivity

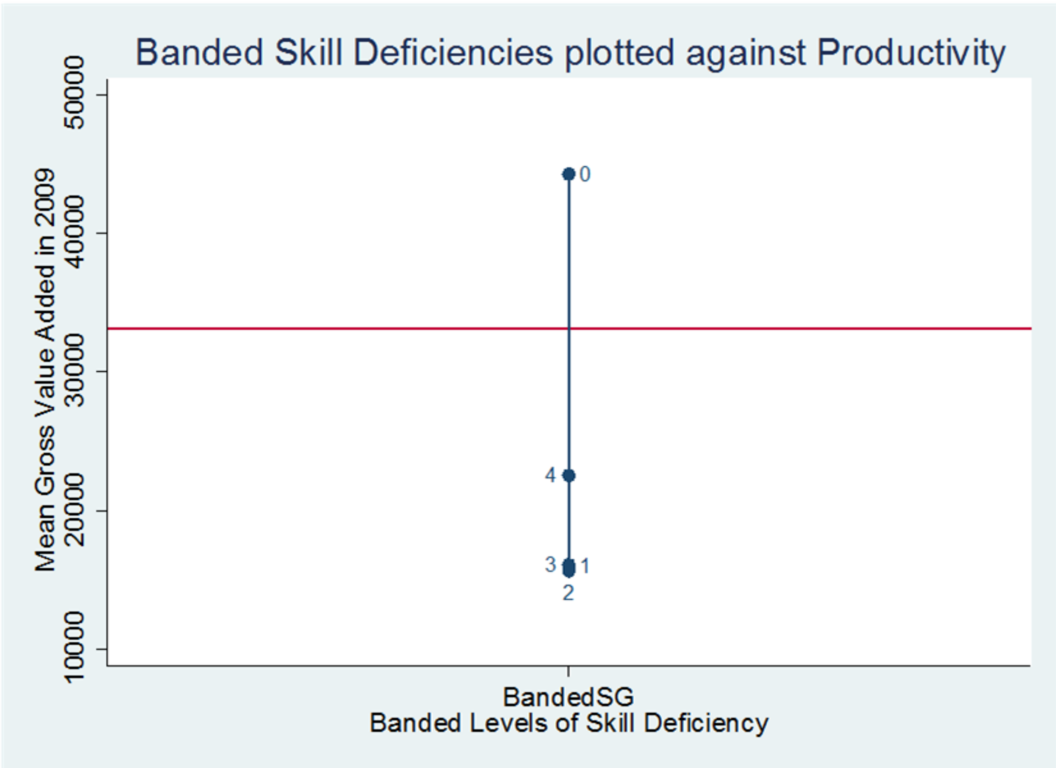


Table 3.8: **Standard IV Results with GVA per capita as the dependent variable**

	X	X+	X+, Region	X+, Indust	X+, Region , Indust
	b/se	b/se	b/se	b/se	b/se
Labour Costs	0.96***	0.95***	0.99***	0.94***	0.98***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Capital Stock	0.02	0.02	0.01	0.02	0.01
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Size	-8.08**	-8.21**	-8.56**	-10.76*	-13.51**
	(3.86)	(4.01)	(3.95)	(5.54)	(6.52)
SG %	223.69*	257.51*	235.42	272.70*	263.42*
	(129.45)	(151.73)	(155.81)	(146.29)	(149.90)
SSV %	-195.99	-47.63	-436.50**	117.22	-236.70
	(223.78)	(206.37)	(216.40)	(234.29)	(238.02)
FO	10773.18***	11138.86***	9337.93***	13386.21***	12949.31***
	(3150.86)	(3268.91)	(3090.91)	(3842.77)	(4022.12)
Larger Org	738.46	1596.91	-1524.62	-218.84	-2791.76
	(2170.22)	(2203.50)	(2131.34)	(2293.20)	(2254.75)
Int Market	3684.07	3833.44	1269.39	337.99	-1817.06
	(2587.30)	(2706.66)	(2815.94)	(3016.15)	(3127.75)
SG for Man	611.44	1113.35	1408.46	696.43	376.06
	(2680.22)	(2869.16)	(3155.75)	(2786.18)	(3120.91)
InterSG	-10071.20**	-11121.96**	-8950.80*	-8131.37*	-5809.33
	(4466.13)	(4731.16)	(4984.27)	(4445.20)	(4726.70)
V Low Qual		-1291.10	-231.22	684.37	1233.61
		(4394.66)	(4800.10)	(4365.94)	(4755.37)
Low Qual		-1846.71	-1473.57	3117.48	3402.34
		(3164.76)	(3301.14)	(3728.29)	(3834.31)
High Qual		-1937.18	-2777.19	2792.23	2029.93
		(3486.04)	(3617.53)	(4116.79)	(4214.33)
V High Qual		3466.20	4488.11	2970.66	4739.15*
		(2872.23)	(2893.74)	(2822.61)	(2868.88)
SGleader		-243.92	-180.62	-167.47	-101.90
		(186.16)	(190.45)	(180.58)	(182.31)
Very rarely lead		3474.90	799.82	3275.78	670.40
		(3244.74)	(3278.93)	(3280.31)	(3310.09)
Rarely lead		4497.58	4233.44	4181.86	4341.26
		(3184.92)	(3471.77)	(3103.87)	(3408.78)
Lead		-1132.82	-818.46	-1078.54	-391.98
		(2088.89)	(2095.07)	(2054.59)	(2065.26)
Often lead		-2380.80	-1941.67	-2291.35	-1849.03
		(2070.61)	(2098.05)	(2048.00)	(2091.09)
Observations	1497	1434	1373	1434	1373

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.9: IV Results looking at SGs in different Occupations with GVA per capita as the dependent variable

	X	X+	X+, Region	X+, Indust	X+, Region, Indust
	b/se	b/se	b/se	b/se	b/se
Labour Costs	1.11***	1.14***	1.14***	1.01***	1.02***
	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)
Capital Stock	-0.01	-0.02*	-0.02**	-0.02*	-0.02**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Size	-5.49	-5.13	-4.46	-7.56	-7.24
	(5.48)	(5.67)	(5.86)	(5.54)	(5.72)
Any SSV	261.11	1223.33	527.66	2028.37	1305.36
	(4711.48)	(4991.03)	(5119.04)	(4787.49)	(4900.76)
FO	9943.15***	10731.52***	10166.54***	11033.93***	11934.09***
	(2305.80)	(2472.16)	(2645.77)	(2485.72)	(2644.53)
Larger Org	1907.85	1878.47	1214.97	-390.20	-1253.20
	(4454.53)	(4833.13)	(4981.16)	(4652.18)	(4786.32)
		(3347.32)	(3482.18)	(3236.49)	(3355.77)
Int Market		-1966.09	-3250.15	-4827.12	-5659.13
		(3469.00)	(3600.68)	(3378.49)	(3498.60)
InterSG		-2760.60	-3017.36	-130.62	-63.79
		(5518.45)	(5728.26)	(5317.93)	(5507.86)
V High Qual		6259.42*	6671.95*	6543.54**	7260.39**
Other Quality Dummies	No	Yes	Yes	Yes	Yes
Lead dummies	No	Yes	Yes	Yes	Yes
Have any skill gaps for:					
Managers	1459.55	2643.64	2988.70	2125.51	1874.23
	(2379.71)	(2878.52)	(3079.15)	(2780.31)	(2977.29)
Professionals	-1127.84	-2537.26	-2272.25	-4248.77	-4239.18
	(4703.57)	(4988.07)	(5125.00)	(4929.88)	(5061.08)
Assoc prof	-907.68	74.11	-485.96	644.46	530.54
	(4345.68)	(4588.67)	(4729.32)	(4585.55)	(4715.75)
Administrative	-2249.80	-2808.19	-2308.19	-2771.45	-2651.31
	(2798.65)	(2931.92)	(3103.52)	(2855.21)	(3012.61)
Skilled trades	-2343.06	-1933.76	-848.73	635.01	1531.29
	(2985.93)	(3115.96)	(3259.39)	(3041.24)	(3169.16)
Personal serv	-3823.36	-4771.20	-3629.72	-7542.69	-6280.19
	(4788.20)	(4956.60)	(5281.81)	(5093.88)	(5418.39)
Sales	5572.11**	5441.87**	5756.73**	8903.79***	8851.94***
	(2453.53)	(2574.91)	(2740.62)	(2673.62)	(2809.84)
Machine oper	-2591.38	-2355.78	-632.02	-5613.36	-4903.93
	(3561.39)	(3742.42)	(3884.68)	(3867.42)	(4011.31)
Elementary	-565.59	-1069.86	-2154.04	-1827.48	-2955.13
	(2439.72)	(2562.95)	(2775.24)	(2751.13)	(2963.52)
Observations	771	732	694	732	694

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.10: IV results with GVA per capita as the dependent variable, when the dataset is re-weighted back to the original sizebands

	X	X+	X+, Region	X+, Indust	X+, Region, Indust
	b/se	b/se	b/se	b/se	b/se
Labour Costs	0.99***	0.99***	1.13***	0.93***	1.08***
	(0.09)	(0.08)	(0.08)	(0.06)	(0.06)
Capital Stock	0.01	0.01	-0.03	0.02	-0.03*
	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)
Size	-523.94**	-507.53*	-634.41**	-586.06**	-769.71***
	(264.19)	(264.24)	(281.85)	(256.41)	(262.62)
SG %	1455.29*	1690.50**	1368.16*	1150.69*	824.41
	(798.69)	(845.51)	(804.77)	(668.10)	(590.31)
SSV %	-1635.55	-1151.86	-2771.57**	-1069.16	-1750.72*
	(1168.35)	(1149.70)	(1296.58)	(1016.58)	(1046.29)
FO	72625.63**	77172.08***	79378.16***	60964.27***	66728.19***
	(28700.14)	(28239.66)	(28937.41)	(18844.09)	(19982.26)
Larger Org	-20557.64	-20106.32*	-10237.79	-25595.44	-8120.94
	(14396.26)	(11487.61)	(11181.18)	(17527.12)	(18264.73)
Int Market	13484.27	7943.61	-18651.60	11363.19	-16941.42
	(24804.79)	(22865.39)	(25113.15)	(17235.15)	(18165.28)
SG for Man	10848.32	-3991.31	-5961.17	2024.23	1919.95
	(39113.72)	(41050.66)	(40094.91)	(30144.56)	(27815.67)
InterSG	-37172.16	3177.81	5850.31	5243.63	10198.73
	(42861.26)	(39588.83)	(39022.41)	(30381.71)	(28307.41)
V Low Qual		10269.37	22447.89	17484.30	34267.92
		(17972.87)	(18862.44)	(21356.69)	(22849.69)
Low Qual		-38314.82*	-29187.35	-6630.37	7508.95
		(22760.37)	(24522.40)	(36737.88)	(38545.50)
High Qual		-4808.76	-5372.98	9211.97	15918.63
		(31122.71)	(32111.33)	(44649.56)	(44920.85)
V High Qual		30719.49	32769.71*	27849.85	34510.79**
		(21077.93)	(19725.33)	(18143.15)	(14734.17)
SGleader		-2006.55**	-1772.62*	-1154.68	-1034.79
		(912.75)	(997.08)	(766.03)	(786.16)
Very rarely lead		108.44	-17468.83	5264.90	-13481.31
		(13429.35)	(13933.24)	(16489.14)	(17036.14)
Rarely lead		4357.35	-7739.96	2770.73	-8879.72
		(10939.91)	(12813.01)	(12252.51)	(12486.33)
Lead		3509.26	7345.22	14902.52	15139.06
		(13032.37)	(11316.16)	(11849.77)	(9681.34)
Often lead		-17742.98	-14417.77	-10861.29	-6675.40
		(14821.44)	(13315.01)	(12854.14)	(10513.81)
Observations	1497	1434	1373	1434	1373

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.11: IV results with GVA per capita as the dependent variable, with banded measures of SGs

	X	X+	X+, Region	X+, Indust	X+, Region, Indust
	b/se	b/se	b/se	b/se	b/se
Labour Costs	0.96***	0.95***	0.99***	0.94***	0.98***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Capital Stock	0.02	0.02	0.01	0.02	0.01
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Size	-8.13**	-8.41**	-8.43**	-10.74**	-13.21**
	(3.76)	(3.93)	(3.82)	(5.32)	(6.24)
SG 0-8%	-1097.48	-738.55	-2692.95	-1494.90	-3500.27
	(1699.46)	(1844.02)	(2023.24)	(1924.33)	(2135.99)
SG 8-15%	-3164.03*	-3219.00*	-2951.66	-3026.23	-2983.79
	(1792.69)	(1903.91)	(1968.05)	(1952.16)	(2102.98)
SG 15-25%	1449.07	2133.31	-92.44	3358.71	1650.33
	(3577.66)	(3781.06)	(3925.16)	(3639.09)	(3779.01)
25+	5734.92	6493.22	4907.72	7165.05	6178.06
	(4349.73)	(5012.60)	(5134.05)	(5008.72)	(5057.60)
SSV %	-164.45	-31.41	-382.69	94.52	-231.54
	(231.64)	(223.31)	(239.63)	(250.00)	(253.02)
FO	11157.07***	11556.04***	9818.17***	13732.32***	13330.96***
	(3270.70)	(3387.59)	(3211.72)	(3945.97)	(4129.19)
Larger Org	464.31	1230.98	-1979.09	-629.76	-3265.47
	(2046.34)	(2068.13)	(2017.41)	(2178.94)	(2167.18)
Int Market	3529.22	3716.86	1151.39	217.86	-1963.41
	(2609.28)	(2737.71)	(2854.94)	(3058.28)	(3172.54)
SG for Man	3229.42	3885.45	4622.77	3347.33	3463.04
	(3746.97)	(3912.17)	(4259.24)	(3733.97)	(4081.72)
InterSG	-9392.84**	-10410.53**	-8282.13*	-7299.53*	-5014.74
	(4258.99)	(4501.84)	(4773.66)	(4247.07)	(4569.31)
V Low Qual		-1206.86	-149.65	767.94	1321.50
		(4379.81)	(4795.43)	(4348.46)	(4744.60)
Low Qual		-1507.79	-1055.25	3410.78	3790.52
		(3118.03)	(3258.66)	(3709.46)	(3830.34)
High Qual		-1653.49	-2457.35	3057.85	2343.70
		(3465.10)	(3602.20)	(4108.99)	(4215.95)
V High Qual		3237.42	4226.92	2754.54	4499.08
		(2835.89)	(2853.53)	(2791.78)	(2837.66)
SGleader		-123.35	-49.91	-51.24	24.35
		(146.98)	(149.02)	(149.04)	(148.68)
Very rarely lead		2284.63	-395.34	2055.45	-573.82
		(3176.63)	(3205.01)	(3232.60)	(3258.41)
Rarely lead		4427.95	4030.81	4117.66	4150.36
		(3191.83)	(3477.57)	(3116.40)	(3413.55)
Lead		-1038.94	-609.15	-988.41	-172.68
		(2114.83)	(2119.42)	(2082.09)	(2092.63)
Often lead		-2457.14	-1984.87	-2348.83	-1855.19
		(2054.88)	(2088.79)	(2033.95)	(2080.70)
Observations	1497	1434	1373	1434	1373

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.12: Heckman Two Step Model: Second Step, IV regression results with GVA per capita as the dependent variable

	X	X+	X+, Region	X+, Indust	X+, Region, Indust
	b/se	b/se	b/se	b/se	b/se
Labour Costs	0.93***	0.92***	0.92***	0.90***	0.91***
	(0.07)	(0.09)	(0.08)	(0.08)	(0.08)
Capital Stock	0.02	0.03	0.03	0.03	0.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Size	9.42	9.43	10.26	6.45	7.60
	(17.64)	(20.53)	(18.14)	(24.98)	(23.43)
SG %	220.10*	254.26*	254.19*	270.39*	271.53*
	(129.32)	(148.74)	(145.37)	(143.32)	(140.85)
SSV %	1191.25	1070.60	1148.55	709.20	758.01
	(1152.43)	(1365.39)	(1149.85)	(1052.42)	(929.41)
FO	16781.18***	17531.87**	17847.52***	16671.51***	16768.68***
	(6396.48)	(7213.23)	(6902.84)	(5444.32)	(5464.87)
Larger Org	-4950.45	-5727.40	-3782.80	-6188.63	-4687.21
	(5836.54)	(7688.16)	(5929.73)	(6559.77)	(4999.82)
Int Market	2459.23	3122.75	3476.98	-326.50	136.39
	(3965.96)	(3931.95)	(3682.17)	(3779.49)	(3543.63)
SG for Man	1792.63	1899.44	2439.34	1332.42	1701.11
	(2592.11)	(2864.69)	(2777.91)	(2874.82)	(2810.97)
InterSG	-10835.83**	-11570.32**	-12060.67**	-8520.56*	-8757.32*
	(4352.36)	(4720.36)	(4857.12)	(4557.56)	(4676.27)
Mills	132647.10	137016.03	134700.89	115216.83	113958.84
	(94401.58)	(123638.10)	(105757.14)	(116008.74)	(106898.18)
V Low Qual		665.88	710.50	1792.67	1831.62
		(6919.50)	(6567.79)	(5863.71)	(5552.84)
Low Qual		8184.46	7507.83	10578.83	10099.23
		(11305.35)	(9404.34)	(9915.02)	(8916.17)
High Qual		8117.10	7541.34	10045.39	9673.29
		(11883.15)	(10432.85)	(10640.23)	(10101.26)
V High Qual		-1293.88	-1211.30	-1425.10	-1394.38
		(6008.69)	(5722.64)	(5896.24)	(5793.93)
SGleader		-237.60	-243.54	-159.47	-168.60
		(190.28)	(191.38)	(191.42)	(193.03)
V Rarely Lead		1424.30	1815.22	1709.26	2077.18
		(5671.92)	(5068.90)	(4836.96)	(4481.36)
Rarely Lead		4472.57	4086.21	3862.17	3601.72
		(5638.97)	(5397.46)	(4989.47)	(5086.91)
Often Lead		3725.08	4203.38	2359.76	2693.50
		(6207.55)	(5352.06)	(4917.88)	(4627.52)
V Often Lead		-2896.27	-2975.44	-2667.29	-2672.50
		(3860.86)	(3559.53)	(3504.71)	(3149.43)
Observations	1485	1424	1424	1424	1424

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.13: Heckman Two Step Model: First Step; Probit regression results with 'likelihood of an establishment reporting a SG' as the dependent variable

	X	X+	X+, Region	X+, Indust	X+, Region, Indust
	b/se	b/se	b/se	b/se	b/se
ANY SG					
Instrument	0.10	0.02	0.02	0.02	0.02
	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)
Capital Stock	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Size	0.00	0.00***	0.00***	0.00***	0.00***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
SSV %	0.05	0.04	0.04	0.03	0.03
	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)
FO	0.20**	0.21***	0.24***	0.13	0.16*
	(0.08)	(0.08)	(0.08)	(0.09)	(0.09)
Larger Org	-0.12	-0.25	-0.19	-0.24	-0.17
	(0.18)	(0.17)	(0.18)	(0.18)	(0.18)
Int Market	-0.09	-0.04	-0.03	-0.04	-0.03
	(0.09)	(0.08)	(0.08)	(0.09)	(0.09)
Labour Costs	-0.00	-0.00	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
V Low Qual		0.08	0.08	0.05	0.05
		(0.15)	(0.15)	(0.15)	(0.15)
Low Qual		0.37**	0.36**	0.34*	0.33*
		(0.17)	(0.17)	(0.18)	(0.18)
High Qual		0.37**	0.36*	0.32*	0.31
		(0.18)	(0.18)	(0.19)	(0.19)
V High Qual		-0.17*	-0.17*	-0.20*	-0.20*
		(0.10)	(0.10)	(0.10)	(0.10)
V Rarely Lead		-0.06	-0.04	-0.05	-0.03
		(0.13)	(0.13)	(0.13)	(0.13)
Rarely Lead		-0.00	-0.01	-0.01	-0.03
		(0.12)	(0.13)	(0.12)	(0.13)
Often Lead		0.18**	0.19**	0.15*	0.16*
		(0.08)	(0.08)	(0.08)	(0.08)
V Often Lead		-0.02	-0.03	-0.02	-0.02
		(0.08)	(0.09)	(0.09)	(0.09)
Observations	1282	1425	1425	1425	1425

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.14: ATT for Propensity Score Match where the dependent variable is GVA per capita

	NN b/se	NN (5) b/se	Caliper (0.01) b/se	Caliper (0.0001) b/se	Kernel b/se
ATT	1344.38	2224.54	1354.78	-993.64	3304.74
	(4353.57)	(3694.57)	(4402.85)	(12130.54)	(3079.70)
Observations	1284	1284	1284	1284	1284

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.15: PSM Match Summary

Variable	Sample	Mean Treated	Mean Control	%Bias	%reduct absolute bias	p>t
Size	Unmatched	103.59	59.377	23.2		0.000
	Matched	94.813	92.453	1.2	94.7	0.785
Size2	Unmatched	66381	20647	7.4		0.163
	Matched	29837	28966	0.1	98.1	0.911
Size3	Unmatched	1.8e+08	1.8e+07	6		0.256
	Matched	1.8e+07	1.8e+07	-0.0	99.7	0.968
FO	Unmatched	.37455	.28237	19.7		0.000
	Matched	.37117	.36741	0.8	95.9	0.897
Larger Org	Unmatched	.94444	.96419	-9.5		0.089
	Matched	.94775	.9537	-2.9	69.9	0.647
SSV	Unmatched	.21854	.10649	9.1		0.096
	Matched	.21972	.2062	1.1	87.9	0.876
CS2009	Unmatched	42334	1.1e+05	-13.4		0.024
	Matched	42562	39045	0.7	94.5	0.739
Labour Costs 2009	Unmatched	12442	32964	-14.4		0.015
	Matched	12509	10896	1.1	92.1	0.521
Low Qual	Unmatched	.05556	.05234	1.4		0.800
	Matched	.05586	.05312	1.2	15.0	0.841
Medium Qual	Unmatched	.40502	.37052	7.1		0.208
	Matched	.4036	.39955	0.8	88.2	0.891
High Qual	Unmatched	.57348	.58953	-3.3		0.563
	Matched	.57477	.58118	-1.3	60.1	0.829
V High Qual	Unmatched	.21326	.26309	-11.7		0.039
	Matched	.21441	.21817	-0.9	92.5	0.879
INT Market	Unmatched	.27957	.26584	3.1		0.584
	Matched	.27928	.29227	-2.9	5.3	0.632
V Rarely Lead	Unmatched	0	0	.		.
	Matched	0	0	.	.	.
Rarely Lead	Unmatched	.10753	.10744	0		0.996
	Matched	.10811	.10867	-0.2	-527	0.976
Often Lead	Unmatched	.5681	.47796	18.1		0.001
	Matched	.56577	.56482	0.2	98.9	0.975
V Often Lead	Unmatched	.53763	.60055	-12.7		0.024
	Matched	.53694	.54183	-1.0	92.2	0.870

Table 3.16: ATT for Propensity Score Match where the dependent variable is GVA per capita - including Region Controls

	No Rep	NN (5)	Caliper (0.1)	Caliper (0.01)	Caliper (0.0001)
	b/se	b/se	b/se	b/se	b/se
_bs_1	4399.66	4717.84*	4399.66	3863.81	2990.98
	(3135.49)	(2718.00)	(3137.83)	(5018.04)	(2645.88)
Observations	1497	1497	1497	1497	1497

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Chapter 4

The Impact of Skill Shortages on Worker Remuneration

4.1 Introduction

The impact that education and skills have on both an economy and on individual workers are relatively well researched topics in economics, and rightfully so. With an increasingly global economy, staying productive is more important than ever, keeping the issue of education and skills at the forefront of many countries' policies. The flip side of this coin, namely the effect of skill deficiencies rather than skills, is seldom investigated. This may be due to the difficulties encountered trying to quantify what is missing rather than what is present, but whatever the cause the literature is sparse.

This work tries to establish whether there is a causal relationship between the level of skill deficiencies in a market and the remuneration a worker receives in the market. In order to do this, information on skill deficiencies from the Employers Skills Survey (ESS) is matched at an industry/occupation level to the Labour Force Survey (LFS) over the period 2006 to 2012. This provides a wealth of information on both skills and workers' characteristics across the relevant time period.

Information on the percentage of internal skill gaps (where employees in a particular job are not fully proficient) establishments have and the percentage of skill shortage vacancies (where an establishment cannot fill a vacancy due to workers of an adequate skill level not being available) establishments face are then used in a pseudo panel framework to establish if wages in markets with greater skill deficiencies are higher than those in markets with lower skill deficiencies. To validate this effect regressions were run on individuals' wages with market level skill information matched in from the ESS to establish if workers in markets with higher skill shortages receive a higher wage.

Further to this, an effort is made to distinguish which workers gain/lose the most from these skill deficiencies. Is it that workers who are moving jobs have more knowledge of the relevant skill shortages in the market and thus can demand a higher wage? Or do those staying in jobs have an equal amount of information and power to demand adequate pay rises relative to the skills (or lack of skills) in the market? To do this, workers are divided into two cohort: those that have moved jobs in the last twelve months and those that have not, suitably titled 'movers' and 'non-movers'. The difference in the worker's expected wage and actual wage is then calculated and used as a dependent variable in a second regression for both of these groups to try to capture if either group are being paid relatively more than expected in markets where skill deficiencies are higher. The same analysis is also conducted for those individuals who are educated, and those who are uneducated, to investigate if it is only a certain subset of the labour force that gain a reward from being in a market with skill deficiencies, as well as for the two subgroups females and males.

4.2 Background & Literature

While the economics literature is relatively sparse in terms of work on skill deficiencies there are a few papers that aim to investigate the impact that they have on workers' outcomes.

Frogner (2002) uses information from the Employers Skills Survey in an ONS report. The data show that out of establishments that claim to experience skill shortages, only half indicated any willingness to raise their wages. While this indicates there may be a link between the two variables it is obviously not clear cut. Further to this, when plotting seasonally adjusted average earnings growth against lagged skills shortages (taken from the CBI Industrial Trends Survey) it is apparent that the two series move together. While the correlation seems strong, there appears to be a one year lag between the two whereby skill shortages moved before earnings do. This is confirmed in the regression analysis with lagged skill shortages having a significant impact on earnings over the period 1972 to 2001.

Wallis (2003) also uses information from the CBI Industrial Trends Survey (ITS) to conduct analysis on both unemployment and real wage growth. To do this, a structural model approach is used whereby two equations are estimated at the same time (based on the approach by Manning (1993) and Layard and Nickell (1985)). The first of these equations is a wage setting equation, with the second equation being a labour demand curve or a pricing equation. The model suggests that skill shortages are linked to both increased real wage growth and to reduced unemployment. The two stage least squares (2SLS) estimation predicts that a one percentage point increase in the level of skill shortages will cause an immediate 0.09 percentage point increase in real wage growth, with the long run effect being a 0.018 percentage point increase. Zellner (1962)'s 'Seemingly unrelated regressions technique' (SURE) was also used with results very close to the 2SLS results. While the effect of skill shortages on real wage growth is small when compared to variables such as GDP this is expected and the result is not negligible.

While these works are useful in part showing that a link does seem to exist between skills shortages and wages there are some draw backs. As Wallis discusses, the ITS only covers manufacturing firms in the UK and its skill measure is less informative than the ESS. The ITS asks, *"What factors are likely to limit your output over the next four months"* with one of the options being *'skilled labour'*. This does not divide skill deficiencies into internal or external gaps and also introduces a degree of speculation on what the market will look like in the forthcoming months in regards to what factors will be needed in this period. While Wallis admits that the ESS provides richer data, it was not available for more than two waves at the time of his work and thus cannot be used for the time series analysis he conducts.

Further to these works, there are also two papers by Jim Allen that are of interest here. Both of these papers look at skill mismatches, which seems to be a good proxy for internal skill gaps whereby a worker is in a job they are not efficient at. The first of these two works, Allen and Van der Velden (2001) looks at the effect that educational mismatches and skill mismatches have on wages, job satisfaction and also on-the-job search. Data were used from the 'Higher Education and Graduate Employment in Europe' project where over-skilled matches occurred when people refute the claim *'My current job offers me sufficient scope to use my knowledge and skills'* and skills deficits are where people agree with the claim *'I would perform better in my current job if I possessed additional knowledge and skills'*. Skill deficits here appear to have no effect on wages and under-utilisation of skills shows the expected negative effect on wages. The skill mismatches seem to have less explanatory power towards wages than the education mismatch variables also investigated in the paper.

The second paper, Allen and de Weert (2007)), looks at education and skill mismatches again, this time using five countries that participated in the 1998 Careers after Higher Education: A European Research Study (CHEERS); Spain, Germany, the Netherlands, the UK and Japan. This survey asks questions on the skill mismatches a worker feels they have with a 5 point scale being used. These are as follows; *'Job at a higher level than own education'*, *'Job at own level and within own field'*, *'Job at own level but in a different field'*, *'Job at a lower tertiary level'* and *'Job below tertiary level'*. This scale allows both under and over skilled matches to be investigated. While the work finds a strong effect of over-education on wages (also seen in the over education literature discussed in the next section) the wage effects of skill mismatches were much weaker. Furthermore, in both the UK and in Germany there was a positive effect of skill shortages on wages. This may suggest that skill mismatches do not represent below-par workers and instead may be an indication of high-powered jobs but so far this is unclear.

4.3 Theory

While there are numerous paths through which the impact of skill shortages on workers' remuneration could be investigated, this work will focus on the different matching equilibrium that could occur in the market and how skill shortages may affect the share of each match that occurs.

To do this, evidence is drawn from the over-education literature. The field surrounding over-education grew mainly from Duncan and Hoffman's (Duncan and Hoffman (1981)) work investigating the returns to required education, over-education and under-education. The authors find that those with 'more education than they need for a job' do not have as large a coefficient on their education level return as those individuals with similar characteristics working where they have 'the required education for the job'. These results suggest that while there is still a premium for being over-educated, workers in over-educated matches are earning less than those with the same level of human capital in a perfectly matched job. Similar, but reverse, findings can be seen for the under-educated side of the market, with under-educated workers earning more than their peers with the equivalent amount of human capital in perfectly matched jobs but less than those with the same characteristics but the required education for the job they are currently working.

While Duncan & Hoffman's findings started the literature similar results have been found repeatedly since. Results from Rumberger (1987), Sicherman (1991) and a host of other authors suggest similar conclusions for both over- and under-education. Groot and Maassen van den Brink (2000)), re-analysed by Rubb (2003), give a full review of the comparable results in the over-education literature, showing that the premium to under-education is approximately equal to the penalty to over-education. Further to this, they also outline some of the different definitions of 'required education' and show how these impact the results in the literature. This is by no means a full review of the over-education literature as this is not the main concern of this work. For a more updated review of the literature see Leuven and Oosterbeek (2011). The reason for raising the over-education literature is due to education widely being used as a proxy for skill (whether this is correct or not is an issue beyond the scope of this work) and it is proposed here that similar equilibria will occur with skills as have been seen in the education literature. This would suggest that the following three matches are likely to exist;

1) Perfect Match: Workers have the skills necessary to perform the job that they are employed to do and receive a wage relating to their human capital.

2) Over-skilled Match: Workers have more skills than necessary to perform the job that they are employed to do and receive a wage lower than their human capital counterparts in perfectly matched jobs. However, they earn a higher amount than others in the same job, holding all other personal characteristics constant.

3) Under-skilled Match: Workers have fewer skills than necessary to perform the job that they are employed to do and receive a wage higher than their human capital counterparts in perfectly matched jobs. However, they earn a lower amount than others in the same job with the required skills, holding all other personal characteristics constant.

To see how the level of skill deficiencies in the labour market impact the distribution of these matches the case of a typical workplace is discussed which produces based upon labour (L), captial (K) and technology (A);

$$Y = f(A, L, K) \tag{4.1}$$

where the labour term is a function of the desired number of workers (L^*), the share of vacancies the workplace currently has (V) and the proportion of Skill Gaps (SGs, whereby a worker is not fully proficient at their job) the workplace has (θ). Having vacancies (V) or skill gaps (θ) both cause labour inputs to be below the desired level of input (L^*). V and θ are both range between 0 and 1;

$$L = L^*(1 - V)(1 - \theta) \tag{4.2}$$

Thus when the firm is profit maximising it produces an output of;

$$Y^* = f(A, L^*, K^*) \tag{4.3}$$

where $L^* = L$ as both the share of vacancies the workplace has (V) and the proportion of skill gaps the workplace has (θ) are equal to zero. This gives profits as just a function of the optimal level of output;

$$\pi^* = \gamma(Y^*) \tag{4.4}$$

If the firm is not in equilibrium and V, θ or V and θ are positive then the profits can be characterised by the following equation;

$$\pi = \gamma(Y^*) - (C_1 + C_2)V - C_3\theta \quad (4.5)$$

where C_1 is the cost associated with the share of vacancies the workplace currently has which means that output is below its potential level. C_2 is the cost of searching for a new worker and thus is related to the size of V (i.e. the proportion of vacancies). C_3 is a similar cost but related to the share of SGs the firm currently has.

From these various costs we can start to analyze decision making when different levels of skill shortage vacancies (SSVs, whereby an establishment cannot fill a vacancy as a worker of the required skill is not available) exist.

If the workplace faces a vacancy and the C_1 term is extremely low (lower than the C_2 term in any time period) then the search costs associated with hiring a new worker are greater than the costs of the vacancy and the firm will not look to hire another worker.

If on the other hand $(C_1 + C_2)V > C_3\theta \times Pr(SSV)$, where Pr(SSV) is the probability of having an unskilled match (as a SG can occur if a skilled match cannot be found), then the costs associated with the vacancy are greater than the probability of a bad match multiplied by the cost of a unskilled match, and thus the workplace should hire a new worker. As can be seen here vacancies are known quantities as they exist within the workplace. SSVs, on the other hand, are express as probabilities as they exist outside the firm and thus full information regarding the current SSV level is unlikely.

The third and final case of interest occurs if $(C_1 + C_2)V < C_3\theta \times Pr(SSV)$. In this case, the firm will reject the match that is found in the period if the worker is not skilled enough for the job, and will continue to search for a new worker in the next period. It is worth noting that while C_2 and C_3 are likely to be constant over time it is likely that

C_1 is increasing with time. C_2, the cost associated with searching for a new worker, is likely to be fixed as advertising a job and interviewing etc. take up a certain amount of effort and expenditure which is unlikely to vary over time. Large macroeconomic shocks that impact unemployment levels in the economy could, of course, impact on the cost of hiring a worker, however these shocks are likely to also impact the demand for the good and thus labour input demands and the employment decision will have adapted. The C_3 term which represents the cost of having an unproficient worker in a job is also likely to be constant over time. While the worker may not possess the skills necessary to complete tasks in the manner envisioned by that particular job, this will not vary over time. In terms of C_1, while a vacancy may be relatively easy and cheap to cover in the short-run, either by outsourcing the work or by increasing the working hours of other employees these are only temporary measures and are likely to become less cost efficient over time. With this in mind if $(C_1 + C_2)V < C_3\theta \times Pr(SSV)$ and the firm does not find an adaquetly skilled worker they are likely to continue searching for a better match until either a skilled worker is found, or C_1 has increased so that $(C_1 + C_2)V > C_3\theta \times Pr(SSV)$ and they hire the next worker through the door. (Note the subtle difference between a SSV and a vacancy. A vacancy has a negative impact on an establishment's performance whereas a SSV decreases the likelihood of a good match and can push up the cost of hiring a new worker).

With this framework in place it is easy to show how the SSV rate impacts the likelihood of different matches occurring. With a higher number of SSVs in the labour market, the chance of finding a perfect match with an adaquetly skilled worker in any time period decreases (there is a greater pool of unskilled workers to randomly draw from). Further to this, the firm is more likely to choose the first worker they find rather than searching for a better match in the next period as the chance that $(C_1 + C_2)V < C_3\theta \times Pr(SSV)$ is satisfied will increase as the level of SSVs increase. Thus increasing SSVs can only help to cause bad matches from the firm's side of the matching relationship. These bad matches can be seen as internal skill gaps and the effect that they have on wages can be tested for within this work.

Secondly, the level of skill gaps in the market could have a possible effect on wages from the supply side of the market. Without formalising the workers' side of the matching equation, there are two possible results, depending on whether information on the level of SSVs in the market is known.

If a worker knows that there is an increased number of SSVs in a given market and adapts their relative expectations about finding a suitable job they have greater bargaining power, and thus can choose to only accept offers whereby they have a perfect match or an under-skilled match (i.e. where they are paid relative to their human capital or above it).

On the other hand, if the individual has no knowledge of the current SSV level in the labour market and thus does not adapt their expectations, then the level of SSV will have no effect on the supply side of the match and no more/no less perfect matches will occur.

Combining the two elements of the matching equilibria it is possible to see that by increasing the share of SSV in the relevant labour market the theory suggests that there can only be an increase in the likelihood of an under-skilled match occurring. This increase in under-skilled matches could have differing effects on wages depending on the level of aggregation that is investigated. The over-education literature suggests that workers will receive a higher wage than their counter parts with the same skills working in a perfectly matched job, but less than a fully proficient worker doing the same job. Due to this, depending on what the reference group is, the effect of increased SGs could have either a positive or negative effect. If the workers are compared to those around them in the same job they will be making less money and the average market wage will be pulled down. If the workers are compared to those with the same human capital as themselves (or their own expected wage) they are receiving a higher wage.

This gives the following testable hypotheses;

Hypothesis 1: Higher levels of SSVs will increase the wage of workers both at an individual and market level, as the scarcity of these skilled workers will push wages up.

Hypothesis 2: Higher levels of SGs will decrease the market wage as workers will be making less money than those around them doing the job proficiently.

Hypothesis 3: Higher levels of SGs will increase the individual wage of workers as they will be making more than their human capital would suggest.

4.4 Datasets

To test the above hypotheses, two different forms of analysis are conducted. Firstly a pseudo panel dataset is created by matching together industry/occupation means from the two datasets discussed below. The second stage of analysis involves matching information on the relevant skill information for the market cell to a relevant individual.

Employers Skill Survey

The Employers Skill Survey (ESS), formally known as the National Employers Skills Survey (NESS) is the largest survey of its kind to date in the UK. It is a nationally representative cross sectional survey of establishments in the UK. It originated in 1999 in England and underwent major reforms in 2003, becoming the NESS. In 2011, the survey was updated and extended to cover all of the UK and now includes Scotland, Wales and Northern Ireland. Whilst the survey has always been centered around skills the consistency of the survey after the 2003 wave allows for robust econometric analysis to take place and consequently these waves are the focus of this paper's analysis.

The ESS is a biennial telephone survey which covers all regions and sectors of the economy and all establishment sizes above the working proprietor level. The result is about 80,000 establishments being surveyed in each wave, roughly four percent of the establishments in the UK. Due to its sampling framework, it tends to over sample larger workplaces but includes weights to bring results back to population estimates.

Of particular interest in this work is the detailed information the survey includes on both internal and external skill gaps. Employers are asked whether, at the time of the interview, they have any vacancies, whether any of these vacancies are proving hard to fill and if any of the vacancies are proving hard to fill due to lack of skills (where lack of skills is defined as a lack of appropriate skills, qualifications or experience in the labour market). Further to this information on external skill shortages, workplaces are also asked at an occupational level "how many (staff) do you think are fully proficient at their job. A proficient employee is someone who is able to do the job to the level required". These questions provide a wealth of information on the current skill situation of the firm both in terms of the workers that the establishment already employs and the recruitment situation the establishment faces.

In addition to the information on skills, the ESS also contains a wide range of information about the establishment, which includes workplace size (number of employees), the sector and geographic region the establishment is in and whether the establishment is in the private sector, the public sector or is a charitable organisation.

Labour Force Survey

The Labour Force Survey (LFS) is a quarterly survey of households in the UK which has been running in some form or another since 1973. The survey, conducted by the Office of National Statistics, is again the largest of its kind in the UK with between 40,000 and 50,000 households responding (or imputed) each quarter, giving information on over 100,000 individuals.

Sampling in the LFS is conducted in a rotational design, whereby once a household is initially selected for interview, they are retained for a total of five quarters. These interviews take place exactly 13 weeks apart, resulting in the 5th interview taking place one year after the first. This sampling format allows the LFS to not only keep a constant moving picture of the labour force in the UK but also to track individuals in a short panel situation allowing for multiple types of analysis to be conducted. Due to the sampling method described above, the same number of wave one addresses are selected each quarter meaning that in any given quarter roughly one-fifth of the addresses are in their wave one sampling, one-fifth are in their wave two sampling and so on. While there are five waves of information, not all the relevant information for this work is included in each wave. For instance, an individual's earnings are only recorded in the first and fifth waves of the survey.

Sampling for the survey occurs regionally ensuring that a geographic spread of addresses are selected, with no address that has been sampled being selected for interview again for at least two years after the final interview. Both face to face interviews as well as telephone interviews are used in the surveying, with almost all of first time wave interviews being conducted face to face.

Whilst the LFS provides invaluable information on the labour market it is often criticised over its proxy response levels. The survey allows for interviewers to take answers from proxies if the respondent is unavailable, usually meaning that a related adult from the same household answers for them (although there are a few exceptions). While this is not necessarily much of a problem the number of proxy responses in the LFS is relatively high. Around one third of responses are proxied, both from the method described above and from two other groups; 1) those who were unavailable and did not have a proxy made for them in this wave but had a proxy or response in the previous wave which is carried forward to this wave, 2) economically inactive individuals aged 70 or more. While this high number of proxy responses is not ideal there is little that can be done about it when using the dataset.

Matching

For the first stage of the analysis the LFS and ESS are matched together at an industry/occupation (referred to as markets from here onwards) level. This means that averages are taken for each industry (i) and occupation (o) where both industry and occupation are measured at the one digit level from their appropriate classifications (SIC and SOC). This gives 14 industries and 9 occupation classifications and thus a matrix of 126 market cells in both the LFS and the ESS to match between. While the LFS does not start out as yearly data the quarters are merged together first and then averages are taken from these yearly samples. Not all the relevant information is included in each wave as already noted, with earnings only occurring in the first and fifth waves and if the person is in a trade union only occurring in the third wave. Summary statistics from this dataset can be found in Table 4.3.

The wage analysis conducted in the second stage of the analysis is done using a slightly different approach. Information on individuals in the LFS is matched to the relevant industry/occupation skill information from the ESS. This allows each individual's wage to be estimated based on their personal characteristics and the information on the skill shortages that impact them. Due to the skill information only being matched in at a market level it does not vary for people within a market. Again, descriptive statistics can be found tabulated, this time in Table 4.4.

The summary of observations in each industry/occupation cell can be seen in Table 4.1. Aggregating at this level the cell size is probably sufficiently large to reduce the importance of measurement error as seen in Deaton (1985).

Table 4.1: Average Cell Sizes in Pseudo Panel Dataset

Cohort	LFS Average Cell Size	ESS Average Cell Size
2011	1,046	466
2009	1,093	440
2007	1,225	436
2005	1,255	417

4.5 Key Variables

The key variables in the following analysis are discussed here, with descriptive figures included where necessary.

Remuneration

One of the key variables in this work is a measure of workers' remuneration. While various measures of workers' pay could be used varying from earnings, income and wages the obvious choice for this paper was to use wages as this method of measuring pay removes the effect of hours. When looking at earnings an increase in hours due to any reason would lead to an increase in earnings and hours are not the variable of interest here.

The LFS has no wage information in it, however it does have details on both gross earnings and the number of hours worked allowing an estimated wage to be calculated as seen below;

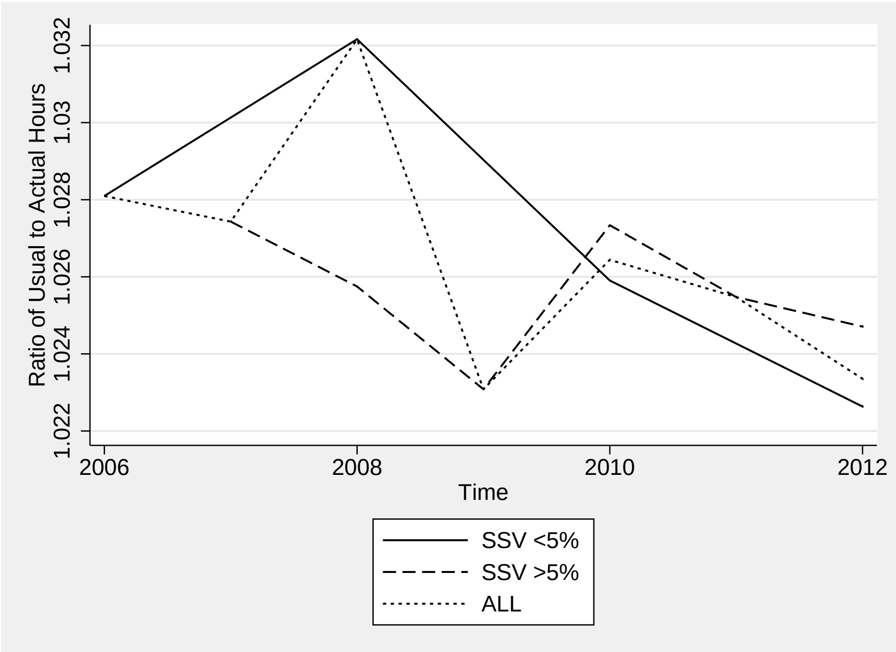
$$Wage = \frac{Earnings}{Hours} \tag{4.6}$$

While this gives a simple wage measure the LFS contains information on two different hours measures, both usual hours worked and actual hours worked.

Due to the nature of this work actual hours have been chosen instead of usual. The reason for this was simple; if skill shortages are being investigated and one of the main ways firms respond to these deficiencies is by increasing the hours people work then to get a fair measure of the wage effect we need to include the change in hours. By using the usual hours information this effect will be neglected and any change in the measured wage could be put down to an unseen hours change for the individual rather than a true wage effect.

To investigate this further actual and usual hours were compared. Usual hours always exceed actual hours. This can be seen graphed in Figure 4.1 when the ratio of usual hours to actual hours is shown and is always above 1. As the issue of concern is that actual hours may capture some of the change in SSV that usual hours does not, the hours ratio has also been plotted for just those individuals who are in areas with SSVs less than or equal to 5% and those in areas with SSVs more than 5%. The graph shows that while the hours ratio does vary between the two different levels of SSV this variance is within a small scale. The actual hours is still always less than the usual hours and thus while using actual hours still seems valid the difference between the two measures is insubstantial.

Figure 4.1: Actual vs Usual Hours

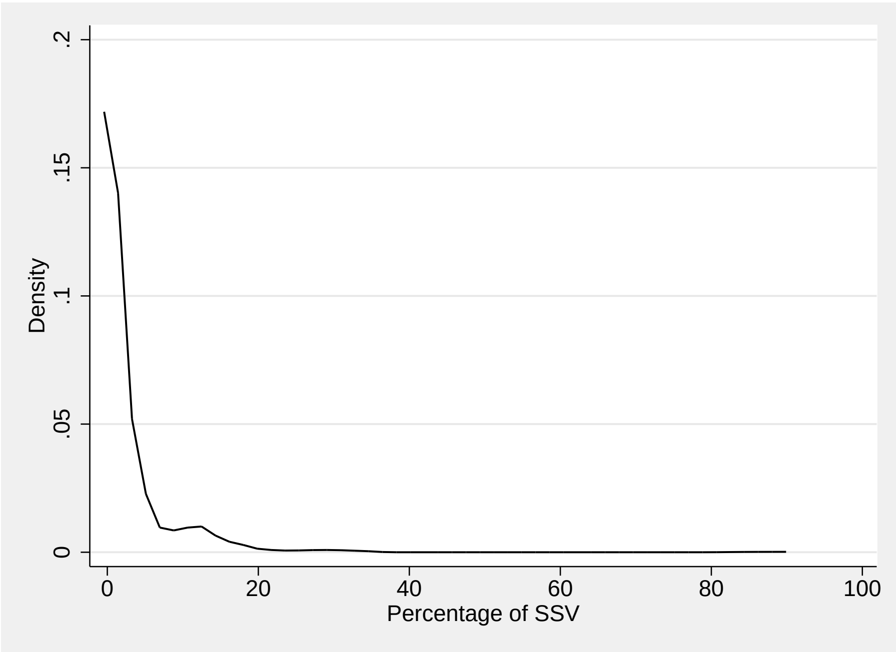


Skill Shortage Vacancies

Skill shortage vacancies (SSV), where an establishment cannot fill a vacancy due to workers of an adequate skill level not being available, are the measure of external skill deficiencies used throughout this work. This measure captures any gaps in the supply of skilled workers to firms in the given market and is measured in percentage terms throughout. While it is theoretically possible for SSVs to exceed 100 percent, where there are more vacancies that cannot be filled than workers, this is not a feature of the dataset.

The distribution of SSVs has implications for the analysis however, with a large number of observations at the bottom end of the distribution and very few markets having higher levels of SSVs. This can be seen in the kernel density plot of SSV seen below in Figure 4.2.

Figure 4.2: SSV Distribution



Due to the fact that SSVs are not normally distributed the percentage of SSVs in a given market are divided into two groups so the different impacts that the two subgroups may have can be interpreted separately. To do this a semi-parametric method known as splining was used, where a function can be split into numerous segments with a linear relationship then estimated for each of them. The points where a function is split (sometimes referred to as junction points) are more commonly called knots. The success of any spline is thus a function of the the number of knots and their position. A simple spline where a variable is split in two, as used in this work, can be seen as;

$$y = \beta_0 + \beta_1 X + \beta_2 (X - K)_+ + \epsilon \tag{4.7}$$

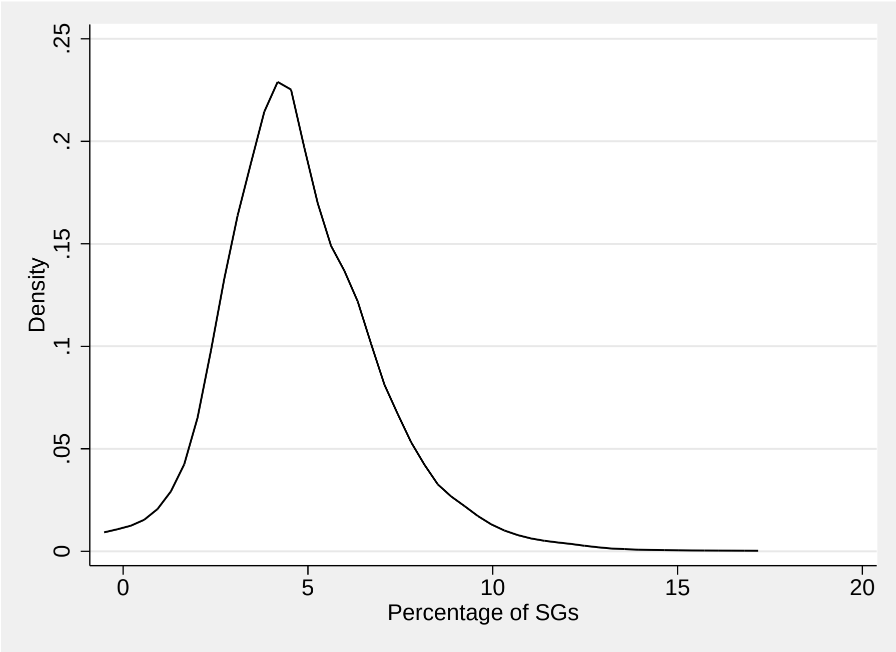
where $(X-K)_+$ is a function equal to $X-K$ if the given expression is positive and 0 otherwise. Thus, SSVs ranging from 0-15% are classified as one group, and those from 15% to 100% form a second group. This allows a greater understanding of the impact SSVs have to be estimated, whereby it is possible to distinguish if it is only large levels of SSVs that impact wages or if both groups are equally important.

Skill Gaps

Skill Gaps (SG), where employees in a particular job are not fully proficient, are the measure of internal skill deficiencies used throughout. These SGs capture the percent of employees that are not fully capable of performing their job in a given market. Unlike SSVs, SGs have an upper limit 100 percent as it is impossible to have more unproficient workers than the number of workers employed. While the SSV measure previously discussed was split into two due to its non-normal distribution the same is not done for SGs. This is due to SGs being much more normally distributed, as can be seen from Figure 4.3.

Between the two measures, a good profile of the skill deficiencies in the market can be drawn, with SGs capturing any 'poor' hires that may be made as a response to long standing SSVs. Due to this, it would be expected that an increase in the number of SSVs in a market would lead to an increase in SGs. Looking at Figure 4.4, this relationship does not appear to be clear. While SGs do seem to be higher in years when SSVs are higher, if SSVs are driving the number of SGs then SGs are not very responsive.

Figure 4.3: Skill Gap Distribution



Other Labour Market Variables

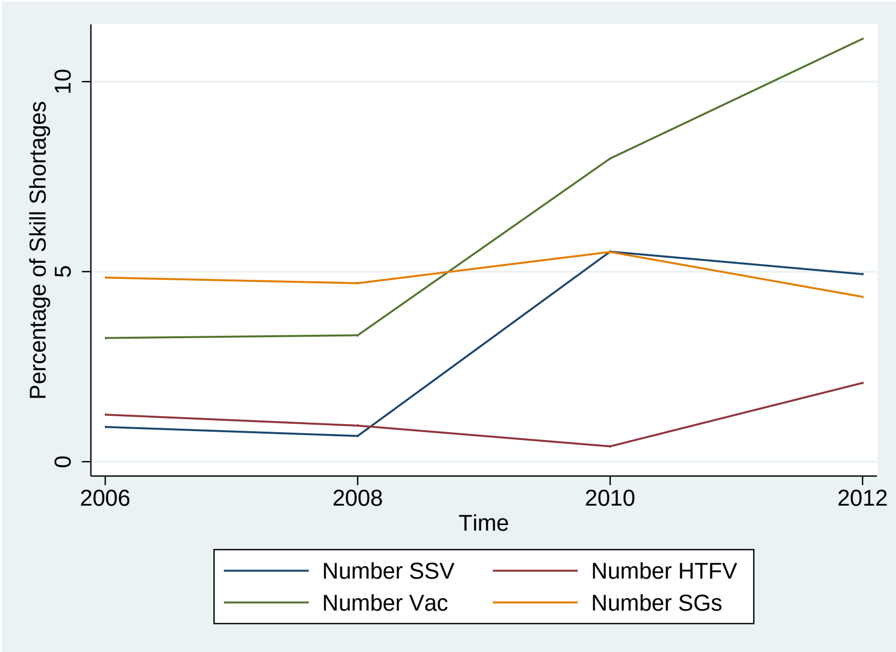
While the main two skill measures are SSV and SGs (discussed above), three other measures of the labour market are also used in sections of this work.

The first of these is a measure of vacancies that an establishment currently has, again like all the labour market variables measured as a percentage.

The second is a measure of Hard to Fill Vacancies (HTFV), these are any vacancies that the establishment has which are proving hard to fill for any reason. Both of these measures are not always included alongside SSVs as they are highly dependent upon each other. An establishment must have a vacancy for it to be hard to fill, and must have a hard to fill vacancy which is caused by skill deficiencies to have an SSV.

The final labour market measure that is included in some of the work is a ratio of the number of HTFV that are SSV. This is included to see if its the level of skill shortages that are important or their relative share.

Figure 4.4: Skill Gaps by Year and Type



4.6 Empirical Testing: Do workers gain from higher skill deficiencies in the market?

To test whether those workers in industries/occupations (referred to as markets from here on) with greater skill deficiencies receive a higher wage two distinct methods are used.

4.6.1 Panel Data Techniques

Firstly, the matched dataset is turned into a pseudo panel running over four consecutive periods, 2006, 2008, 2010 and 2012, which allows for numerous panel data estimators to be used on the sample including 'fixed effects' and 'random effects'. The dataset is a pseudo panel which means that it is not a true panel dataset, rather repeated cross-sections combined together. The first work of this sort is often credited as being Deaton (1985) with recent developments in the area coming from Antman and McKenzie (2007) and Dang et al. (2014). This approach of creating a mean-based pseudo panel involves tracking a cohort of individuals over a repeated cross-section, where the cohort is defined in this work as a market (industry/occupation cell). This use of pseudo panel data allows some of the limitations of longitudinal data to be overcome. The most important of these are;

1) Non-random attrition: This is no longer an issue as each household or individual only needs to be observed once in pseudo panel work, and thus non-random attrition cannot occur in the dataset.

2) Available information: Cross-sectional data is widely available and thus allows the construction of pseudo-panels where information was not previously available or for longer periods of time than existing real panels can provide.

3) Panel Conditioning: Genuine panel dataset can suffer from conditioning effects. Recent work around panel data has found increasing evidence that people adapt their behavior around the subject of the questionnaire. Zwane et al. (2011) split their survey sample in two, with half randomly allocated to the health section of the questionnaire and half allocated to the household finances section. The authors find that the conditioning effects change people's behaviour to a large enough extent to change the mean of the outcome variable as well as the estimated coefficients from the regression analysis. For example, being in the health half of the survey increased the take-up of medical insurance and the use of water treatment products. Similar findings have been found by Crossley et al. (2014) with saving behaviour while Das et al. (2011) show a panel conditioning effect by comparing refreshment samples to those that have been in the panel for longer and find a conditioning effect for those questions that are based around knowledge.

The pseudo panel dataset used in this work means that markets are observed at several points in time, namely every two years from 2006 to 2012. Panel data of this sort is useful when it is suspected that the dependent variable (i.e. the the wage) is dependent on explanatory variables that are not observed but that are correlated with observed variables. If these unobserved variables do not vary over time then panel data will allow a consistent estimate of the effect of the observed explanatory variables. For example, if the wage difference depends on the unobservable characteristics of a given market as well as skill gaps in the market a panel data estimator will still allow a consistent estimate of the impact of skills to be calculated.

More formally a standard multiple linear regression model for markets (i = 1, ... , N) observed over time (t= 1, ..., T) can be written as;

$$y_{it} = \beta_0 + x'_{it}\beta + v_i + u_{it} \tag{4.8}$$

where y_{it} is the dependent variable, β_0 is the intercept, x is a K-dimensional row vector of independent variables, β is a K-dimensional column vector of parameters, v_i is a market specific effect and u_{it} is an idiosyncratic error term.

This can be turned into a pseudo panel cohorts to obtain consistent estimators for β as seen in 4.9, even if v_i is correlated with one or more of the explanatory variables. Let us define C cohorts, which are groups of individuals sharing some common characteristics. These groups are defined such that each individual is a member of exactly one cohort, which is the same for all periods. Stemming from this, it is important that the variables that define each cohort to be observable for across the whole sample.

$$\overline{y}_{it} = \beta_0 + \overline{x}_{ct}\beta + \overline{v}_{ct} + \overline{u}_{ct} \tag{4.9}$$

$c = 1, ..., C; t = 1, ..., T$,

Where $\overline{y}_{ct}$ is the mean of all observed y_{it} values in the given cohort (C) in time period t. In a similar way, the other variables in the model represent the mean of the given variable for the given cohort in that time period. In this work, the cohort is defined as the market, i.e. the given industry occupation cell.

The main problem with estimating β from the above equation is that $\overline{v}_{ct}$ depends on t, is not observed, and is likely to be correlated with $\overline{x}_{ct}$, i.e. $\overline{v}_{ct}$ is correlated with $\overline{u}_{ct}$. Therefore the $\overline{v}_{ct}$ term cannot be combined into the error term as it will lead to inconsistent estimates and panel data methods need to be used to allow for v_i to be controlled for in the regression analysis.

Fixed Effects

Fixed effects estimation is a simple transformation of the above regression whereby time averages are subtracted from the initial model. This can be seen by taking $\bar{y}_{it} = 1/T \sum_t y_{it}$ away from equation 4.8 to give the model;

$$\ddot{y}_{it} = \ddot{x}'_{it}\beta + \ddot{u}_{it} \tag{4.10}$$

where $\ddot{y}_{it} = y_{it} - \bar{y}_i$, $\ddot{x}_{itk} = x_{itk} - \bar{x}_{ik}$ and $\ddot{u}_{it} = u_{it} - \bar{u}_i$.

From this it is possible to see that several important changes occur when the model is transformed. Firstly, the individual-specific effect (v_i) and the intercept term a are both cancelled out as they do not vary over time. This is important as the individual-specific effects, in this case the unobservable market characteristics, are removed and allow an accurate estimate of the impact of skill gaps.

Secondly, and also worthy of note, is that any time-invariant regressors also cancel out. Since regressors are written as $x_{itk} - \bar{x}_{ik}$ then when they do not vary x_{itk} and $\bar{x}_{ik}$ will be the same resulting in $\ddot{x}_{itk}$ being equal to $x_{itk} - \bar{x}_{ik} = 0$. Due to this, the effect of time-invariant regressors cannot be estimated by a fixed effects model. While the effect of these will be controlled for in terms of the v_i term in the regression this only works if it is believed that these effects are constant over time. If the wage premium for a particular market is expected to be higher than another this can be estimated but it will not be possible to separate out their effects into industry effects and occupation effects.

With panel data estimation, effects are usually described in terms of 'within group' and 'between group' effects. 'Within group' picks up the change within the group over time, and thus in this case the FE model will pick up the effect of changes that a market is having over time. The 'between groups' picks up on the change between the groups i, i.e. if the markets have wages that are different to each other based on unobserved characteristics.

Random effects

The underpinning of the random effects model is that the variation across groups (markets) is assumed to be random and uncorrelated with the independent regressors used in the model. Thus, random effects should only be used if there is a belief that the groups have some influence on the dependent variable in the regression but which is uncorrelated with the observed characteristics. It is commonly assumed in regression analysis that all factors that affect the dependent variable, but have not been included as regressors, can be summarised by a random error term (α_i).

This model can be seen as;

$$y_{it} = \beta_0 + x'_{it}\beta + \alpha_i + u_{it} \tag{4.11}$$

where $\alpha_i + u_{it}$ is an error term consisting of two components, an individual specific component that is constant over time (α_i) and a remainder part, which is assumed to be uncorrelated over time (u_{it}). Any correlation of the error terms over time is attributed to the individual effects term (α_i). It also assumes that α_i is purely random, implying thus that it is uncorrelated to any of the regressors. Estimation is then possible by feasible generalized least squares (FGLS).

While random effects does have the advantage of allowing for time-invariant regressors to still play a role as explanatory variables, unlike in a fixed effects model, it will provide inconsistent estimates when the fixed effects model was appropriate rather than random effects.

Fixed Effects or Random Effects?

The random effects estimator is an appropriate tool to use if it is believed the unobserved regressors are uncorrelated with the independent variables in the model. If this is the case then random effects will give an unbiased estimate of the coefficients and will use all available information making it efficient.

Fixed effects controls for this omitted variable bias as markets act as their own controls. This only works if the effect of the omitted variables is constant over time and thus their effect is constant. The other worry with fixed effects occurs if there is little variation within a market over time, as markets are used as their own controls there needs to be within-subject variation. Fixed effects will almost always provide unbiased results however they may not always be efficient as information may be lost in the differencing process if fixed effects was not appropriate.

To decide which model is appropriate a Hausman test can be run (Greene (2003)) which tests whether the fixed effects and random effects coefficients are significantly different, with the null hypothesis being that they are not. While this does not explicitly test whether the unique errors (V_i) are correlated with the regressors or not it does give an implication. In order to ensure the correct model is used both are estimated, as well as an OLS regression to allow for further comparison of the estimates.

4.6.2 Individual Wage Regressions

The second technique used to asses the impact that skill shortages may be having on workers' pay is by matching the market skill information to each individual and estimating a worker's wage based upon their characteristics and the skills in their relative market, see equation 4.12.

$$Wage_i = \beta_0 + \bar{x}_i'\beta_1 + \beta_2 SG_m + \beta_3 SSV_m + \epsilon_i \tag{4.12}$$

This analysis is conducted using OLS regressions with cluster adjusted standard errors to allow for the fact that some variables are taken at higher levels of aggregation. While an individual is observed twice in the LFS, both of these observations are within the time span of one skills survey and thus no panel can be comprised out of the available information. Further to this, the skills data is only matched in at a market level and thus both industries and occupations cannot both be controlled for throughout this estimation as they would not vary for a worker in a given market. Thus, the analysis is conducted mainly on just the worker's personal characteristics and the relevant skill information with some analysis containing either industry dummy variables or occupation variables but not both. While this is not ideal it is the best that is possible with the available data.

4.7 Empirical Testing: Which workers gain from higher skill deficiencies in the market?

In order to test who wins and who loses in a market with higher skill shortages, a useful counterfactual is needed. For example, to know if someone who moves job has greater returns than someone that does not we need to have the wage for the mover given that they have moved and the wage of a mover if they had not moved. This is of course not possible as the individual either moves or does not move, they cannot be both a non-mover and a mover. In order to create a useful couterfactual an estimate of the individual's wage is created from regression analysis and then their observed wage is compared to this. Thus, the difference between a worker's wage and their expected wage is calculated and this is then differenced between the two groups (here explained in terms of 'movers and non-movers' but also looked at for the 'uneducated' and 'educated' and also for the two groups 'females' and 'males'). The results from this analysis were averaged at the market level and then used in the second step of the analysis whereby it is tested if these averages vary more in markets with higher skill gaps. This approach has two main advantages over any form of standard analysis.

Firstly, by taking the difference in this way any bias from poor wage predictions should be removed as long as the same factors make the predictions poor across both groups. For example, any omitted variable bias would affect both the 'movers' and 'stayers' subgroups and thus be cancelled out. While there is obviously a selection bias where both movers and non-movers may have self selected into their groups on the basis of unobserved characteristics this should be controlled for by the movers dummy variable in the wage equation. This variable should pick up both the difference in wage due to moving rather than staying and the other unobserved characteristics that cause the two to be paid different wages. If biases did affect the residuals from the two groups wage equations differently this would still obviously be a problem but there seems to be no theoretical reason for this to be the case.

Further to this, this method also means that the differences in characteristics in the two subgroups become inconsequential. By differencing between a worker's wage and their expected wage the two subgroups are not directly being compared and thus the fact that their characteristics are not strictly comparable (See Table 4.11) does not cause any bias. A more comprehensive step by step review of the method conducted can be found below.

Step One

Step one involves calculating the difference between a worker's true wage and their expected wage, Equation 4.14. To do this an estimate of each worker's wage is formulated given their personal characteristics; namely experience, experience squared, their highest level of education achievement, gender, full time/ part time status and region dummies. A dummy is also included to identify if a worker has moved in the year prior to being surveyed as those who moved are likely to receive a lower wage due to losing firm specific human capital and tenure that they had at their previous job. This can be seen in equation 4.13 with the results from this preliminary regression seen in Table 4.12. While this negative effect may seem like it would stop workers from moving there are two reasons why moves may still occur. Firstly, the move may not be the choice of the individual, their contract may expire or be terminated. Secondly, there is a possibility that moving opens up higher future earnings, i.e. moving to a new job where with experience the wage would be higher than their current wage.

This 'moved' dummy also serves to divide the workers into two distinct groups aptly called 'Movers' and 'Non-movers'. These groups are used as it may be that it is only those that have recently moved who are able to take full advantage of the current SSV levels (due to being able to negotiate their contracts on current information) in the market and thus are more able to demand a wage higher than their human capital if SSVs are high.

$$Y_i = \beta_0 + \beta_i \bar{X}_i + \beta_2 MOVED + \epsilon_i \tag{4.13}$$

$$Wage_i - W\hat{a}ge_i \tag{4.14}$$

By taking the difference between a worker's actual wage ($Wage_i$) and their predicted wage ($W\hat{a}ge_i$) it is possible to estimate if workers are getting paid more than their human capital suggests they should, i.e. the difference is positive. This gap between a worker's true wage and their estimated wage is then differenced between the two subgroups that are analyzed, movers (M) and non-movers (NM) as seen in equation 4.15.

$$(Wage_{NM} - \hat{Wage}_{NM}) - (Wage_M - \hat{Wage}_M) \tag{4.15}$$

The average of this difference in difference is taken for each market to estimate if the difference between human capital and wage is greater for movers than for non-movers or vice versa.

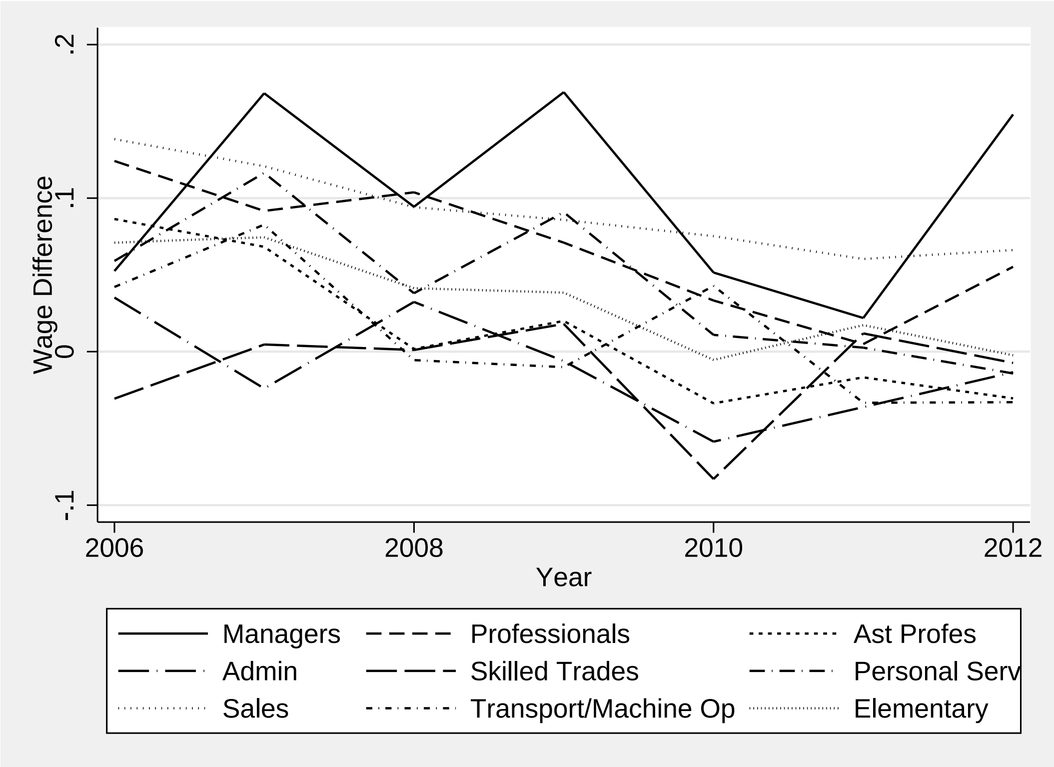
The difference calculated in equation 4.15 varies by industry and occupation and also year, as can be seen in figures 4.5 and 4.6. While both of these Figures initially show positive wage differences they have both shown steady decline over the sample years with most industries and occupations ending up being negative. This suggests that circumstances have changed to favour the movers relative to the non-movers.

This difference calculated in equation 4.15 then forms the dependent variable in the second step of the analysis. As data exist at the first digit occupational classification (SOC) level and the first level of the standard industrial classification (SIC) level a 9 by 12 matrix exists for the dependent variable in each time period, providing 126 observations for each year.

Step Two

The second step of the analysis involves using the industry/occupation wage difference average computed in step one as the dependent variable in a second regression. This regression aims to show the difference in the wage differential that is caused by market

Figure 4.5: Wage Differences $[(Wage_{NM} - \hat{Wage}_{NM}) - (Wage_M - \hat{Wage}_M)]$ by Occupation and Year

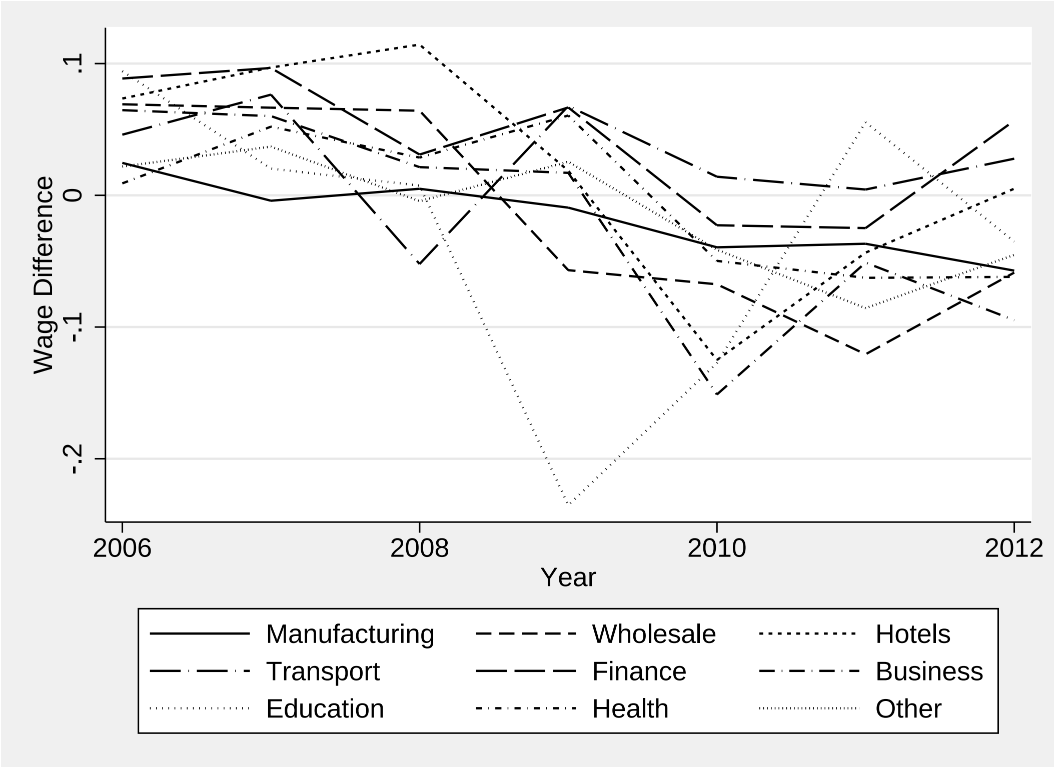


specific characteristics, including various skill shortage measures.The full list of variables in this regression is as follows; the average percentage of SSVs, the average percentage of HTFVs, the share of HTFVs that are SSVs, the average percentage of workers with SGs, the average percentage of vacancies (used as a measure of unemployment in the market), the average number of employees per establishment, the percentage of establishments that offered training, the percent of establishments that are in the private sector, the percent of workers in the market that are in a trade union and the gender balance in the industry/occupation. A simplified version of this regression can be seen in equation 4.16.

$$WageDifference_{I,O} = \beta_0 + \beta_1 SGs_{I,O} + \beta_2 SSVs_{I,O} + \bar{X}_{I,O} + \epsilon_i \tag{4.16}$$

This specification is then estimated by weighted least squares (WLS). WLS is used as the

Figure 4.6: Wage Differences $[(Wage_{NM} - \hat{Wage}_{NM}) - (Wage_M - \hat{Wage}_M)]$ **by Industry and Year**



variables in this regression are formulated from market averages in both the LFS and the ESS. Due to this, where markets do not have a large number of observations the values are likely to vary more and thus these observations do not warrant the same weight as the larger markets which are likely to give more precisely-estimated averages. In a traditional OLS regression it is assumed that each data point provides equally precise information about the total process variation. This can be seen as assuming that the standard deviation of the error term (ϵ_i) is constant for all values. This assumption does not hold here and more weight should be given to more precise measurements and less weight should be given to less precise measurements when estimating the unknown parameters in the model. By using weights that are inversely proportional to the variance yields more precise parameter estimates. The WLS is thus weighted by both the number of establishments supplying information in each market in the ESS and the number of individuals supplying information in each market in the LFS. Thus, rather than maximising the sum of squares as in a traditional OLS (seen below);

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{4.17}$$

$$= \sum_{i=1}^{n} \left[y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_{i,1} + \hat{\beta}_2 X_{i,2} + ... + \hat{\beta}_k X_{i,k})\right]^2 \tag{4.18}$$

the weighted sum of squares is minimized;

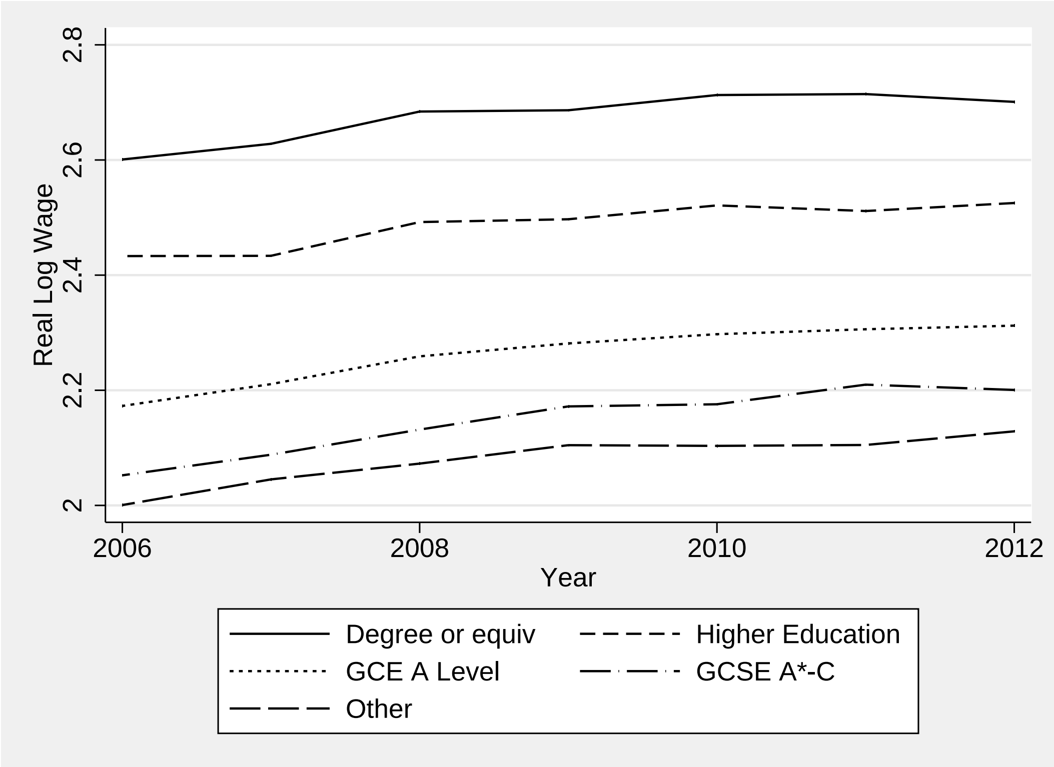
$$WSSE = \sum_{i=1}^{n} W_i (y_i - \hat{y}_i)^2 \tag{4.19}$$

$$= \sum_{i=1}^{n} W_i \left[y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_{i,1} + \hat{\beta}_2 X_{i,2} + ... + \hat{\beta}_k X_{i,k})\right]^2 \tag{4.20}$$

Full details on the final cell counts for both the LFS and the ESS can be found in Tables 4.5 and Table 4.6.

A similar process was used to analyse whether 'uneducated' or 'educated' individuals gained more from being in areas where skill gaps were higher as well as for 'females' and 'males'. Whilst the same method was followed educational levels obviously are not as easy to decompose into two groups as those moving jobs. Looking at the mean log wage of individuals with each of the five qualification measures available, Figure 4.7, it is clear the best way to divide them into two groups is with 'Degree or equivalent' and 'Higher Education' comprising one group and the other three qualifications in the second. Thus, the 'educated' group consists of the top two lines in this figure and the 'uneducated' group consists of the bottom three. For robustness, analysis results are also calculated based on splitting the 'uneducated' and 'educated' so that GCE A Level is also included in the 'educated' group rather than the 'uneducated' group.

Figure 4.7: Mean Log Wage by Qualification Type



4.8 Results

Pseudo Panel Dataset

The results from the pseudo panel dataset can be seen in Tables 4.7 and 4.9. Table 4.7 shows numerous specifications, each estimated with an OLS, a fixed effects (FE) and a random effects (RE) estimator. In all of these regressions the dependent variable is the log of the real wage.

The first three columns of Table 4.7 show the results when the only skill measures included were SSVs and SGs. While the size and significance of some of the coefficients vary between the different estimators here the SSV coefficients stay very similar between the models. The coefficients on both the SSV measure between 0% and 15% and the SSV measure for over 15% are positive and significant. As the model is log-linear this suggests that a one percentage point increase in SSVs increases the mean wage in the market by around 0.7 percentage points. While the coefficient on the larger SSV measure is significant at the 1% confidence level with all the estimators the same is not true of the smaller SSV measure. While the coefficient on this is still always positive and significant it varies from being significant at the 5% confidence level and at the 10% confidence level when different estimators are used.

The coefficients on the SGs variable do not show the same consistency across estimators. The OLS estimator gives a coefficient of -0.04 with RE giving -0.015 and FE giving a positive coefficient of 0.002. The OLS estimator is likely to be biased as it does not take into account the unobserved market specific characteristics. As these differences between markets are likely to be linked to an individual's wage, by not including these terms it is causing the effect to be misinterpreted as the effect of SGs, thus causing the effect of SGs to be overestimated.

When choosing between the fixed effects and random effects models it is important to understand that the RE model will give biased estimates if FE was applicable, however FE should never be biased due to unobserved heterogeneity, but may not be efficient as it could be disregarding important time invariant information. A formal test to choose the most appropriate model was created by Hausman (1978) which was explained in an earlier section of this work (Section 6.1). The null hypothesis is that the two estimation methods are both suitable and therefore they should yield 'similar' coefficients. The alternative hypothesis is that the fixed effects estimation is suitable and the random effects estimation is not; which would mean there is a difference in the two sets of coefficients. This works due to the random effects estimator making the assumption that the random effects are orthogonal to the regressors, which the fixed effects estimator does not. If this assumption is wrong, the random effects estimator will be inconsistent, but the fixed effects estimator is unaffected. Hence, if the assumption is wrong, this will be reflected in a difference between the two set of coefficients. The bigger the difference, the bigger the Hausman statistic. The Hausman statistic in this case is Prob $> Chi2 = 0.000$. This means there is a large and significant difference between the coefficients and thus the null hypothesis can be rejected, FE will be more appropriate in this case.

Focusing more on the FE results it is possible to see that aside from the two SSV measures, and the number of employees variable, the coefficients on the other regressors are insignificant. Columns 4 to 6 show the effect of adding the percentage of vacancies as a regressor. This has almost no effect on the sign, size and significance of any of the other regressors in the three columns and itself seems to be of little importance.

The final columns, 7 to 9, show the same regressions but including the percentage of HTFV and also the share of HTFV that are SSVs. While these variables do not have too much of an effect on the OLS and RE results they do wash out the significance of the SSV coefficient in the FE model. This was likely to happen as SSVs are a small subset of those that have HTFV and this high correlation was likely to soften the observed impact of SSV.

Table 4.9 shows similar regressions, this time using dichotomous variables for the skills measures rather than continuous measures. The results show very different findings to Table 4.7 with the coefficient on SSV not being significant. Again, looking closely at the FE model, as this is the most appropriate, it is clear to see that only two variables have significant coefficients, 'any vacancies' and the 'percentage growth in the last two years'. While the coefficient on the percentage growth variable is small the coefficient on the any vacancies dichotomous variable is both significant at the 10% confidence level and very large. Comparing these results to Table 4.7 the results suggest that having a gap of some sort is not as important to the wage as the level of the deficiency.

The results from the pseudo panel dataset seem to suggest there is a pay premium due to being in a market with SSVs, with being in a market with higher SGs proving to have no effect on wages. This supports the first hypothesis from the earlier theory, that higher levels of SSVs will lead to the wage of workers, both at an individual and a market level, increasing. The fact that SGs do not have an impact on wages does not contradict the two other hypotheses as these predict opposing outcomes. By no effect being found this could mean there is no impact from SGs, or the predicted decrease in the market wage due to higher levels of SGs is equally balanced by the predicted increase in the individual wage in markets with higher SGs. Disentangling these two ideas is impossible given the data available here.

Individual Wage Dataset

The results from the individual wage dataset can be seen in Table 4.10. The dependent variable in these regressions is an individual's real log wage, calculated as their pay divided by actual hours. Columns one to three in Table 4.10 show the same specification in terms of skills measures included with industry dummies added to the main specification in column two and occupations added in column three. Columns four to six repeat this process but with more skill measures included in the specification.

Looking at the explanatory regressors, the results are as expected. Experience has a positive and significant coefficient with the experienced squared coefficient being small, negative and significant. This is as expected as experienced workers earn a higher wage on average but the relationship is often described as an inverse 'u' shape which explains the negative square term. The positive and significant coefficient on full time workers suggests they earn more than part time workers ceteris paribus, and again as expected. The male dichotomous variable also has a positive and significant coefficient, suggesting that men earn more than women holding all other characteristics constant. The size of this coefficient, between 5 and 7 percent, is relatively small for a gender pay gap. While it would be reassuring from an equality perspective to believe that the presence of the skill variables included here show that the true gender pay gap was much smaller than previous findings this is not the case and with a standard wage regression the effect of an individual's gender is still always smaller than expected.

The education dummy variables that were included also give standard results with each higher tier of education having a greater, positive coefficient than the last and all forms of education giving a wage rise over those with no education, ceteris paribus. The fact that these results are in line with economic theory and previous empirical work suggests that the regressions are appropriate and correctly specified.

The main variables of interest in Table 4.10 are of course the skills measures, however the results here are less clear than in the previous pseudo panel regressions. Looking first at columns one to three the coefficient on SGs is both negative and large for the first two columns, without any extra controls and where industry dummy variables are added. However, in column three, where occupation dummies are added, the significance on the coefficient disappears and the magnitude of the coefficient drops considerably. This may suggest that an individual's occupation has some impact on their wage which is thus affecting the SG coefficient. This is hard to tell however as the the skill measures in Table 4.10 only vary at the industry and occupation level. Thus, if the main variance in the skill measure is at the occupation level then including the occupation dummies may be washing this effect out and thus causing the coefficients to become both small and insignificant.

Similar results can be seen with both the SSV measures. The coefficient on the SSV measure between 0% and 15% is both positive and significant when no extra controls are included and when industry controls are included but loses size and significance when occupation dummy variables are added. The larger SSV measure, that goes from 15% upwards, shows a similar trend but the coefficient loses significance when industry controls are added rather than occupation controls. Due to the problems with including the extra controls the first column may be the most appropriate to interpret. When comparing figures from this column to Table 4.7 it is clear to see that both the sign and size of the coefficients on both SSV measures seem to be similar at around 0.007 or 0.008 and with both tables showing a negative and large coefficient on the SG variable. As the model is log-linear again this means that a one percentage point increase in SSVs in the market would lead to a wage increase for an individual of 0.7 percent, ceteris paribus.

The final three columns of Table 4.10 show the same as the first three columns but with further skill deficiency regressors included. The results for these are similar to columns one to three but with the coefficients on the larger SSV measure being greater and the coefficients on the smaller SSV measure being insignificant in all three of these columns. This is likely as the HTFV measure that is included in the specification will be highly correlated with SSV. While the percentage of HTFV in a market does not seem to be having an effect on an individual's wage the share of HTFVs that are SSVs does. The coefficient on the share of HTFVs that are SSVs is both large and significant at the 1% level in all three specifications in which it is included, suggesting that when a greater proportion of the HTFV are SSVs workers in the market receive a higher wage. This coefficient is relatively large as well, with a size of between 0.68 and 0.39.

The individual wage analysis seems to support the pseudo panel results seen previously with similar sized significant coefficients being found on the SSV measures in both forms of analysis. This work also confirms the negative effect that SGs appear to have on an individual's wage and the fact that this effect is bigger than the positive effect of being around SSV. These results again seem to support the first hypothesis from the earlier theory, that higher levels of SSVs will lead to the wage of workers. The fact that SGs have a negative impact on individual's wages at first glance seems to contradict hypothesis three, but this is not the case. As it is not possible to identify which individual workers have SGs, no conclusions can be drawn about hypothesis three. It may be that this theory is incorrect, or it might be that the effect on the market wage is larger than the effect on an individuals wage and thus we are only observing hypothesis two here.

Movers and Non Movers

Table 4.13 shows the WLS results for the 'movers' vs 'non-movers' analysis. The first two columns show different skill specifications for if the worker moved jobs in the last 12 months, with columns three and four showing the same but for 18 months, and five and six 24 months. The dependent variable in these regressions is;

$$(Y_{NM} - \hat{Y}_{NM}) - (Y_M - \hat{Y}_M) \tag{4.21}$$

the relative ratio of the difference between the expected and actual wage for workers who have moved and for those that have not moved. As the wage is logged this can be seen as;

$$[\log(Y_{NM}) - \log(\hat{Y}_{NM})] - [\log(Y_M) - \log(\hat{Y}_M)] \tag{4.22}$$

and thus can be transformed to give;

$$\log\left[\frac{(Y_{NM}-\hat{Y}_{NM})}{(Y_M-\hat{Y}_M)}\right] \tag{4.23}$$

or;

$$\log\left[\frac{\left(\frac{Y_{NM}}{\hat{Y}_{NM}}\right)}{\left(\frac{Y_M}{\hat{Y}_M}\right)}\right] \tag{4.24}$$

Thus the dependent variable is given as;

$$\log\left[\frac{\text{Ratio that expected wage exceeds actual wage for non-movers}}{\text{Ratio that expected wage exceeds actual wage for movers}}\right] \tag{4.25}$$

To simplify the interpretation of the results, the marginal effects have been calculated whereby Marginal Effects = exp^{β}, thus giving a straight interpretation of the results in Table 4.13 as the ratio that expected wage exceeds actual wage for non-movers to the ratio that expected wage exceeds actual wage for movers. If the wage for non-movers equals their expected wage and the same occurs for movers then the dependent variable will be equal to zero, as can be seen in Equation 4.26.

$$\log\left[\frac{\left(\frac{X}{X}\right)}{\left(\frac{Y}{Y}\right)}\right] = \log\left[\frac{(1)}{(1)}\right] = \log[1] = 0 \tag{4.26}$$

Likewise, if those that do not move have actual wages twice that of their expected wage (i.e. their wage doubles due to being in an area with skill gaps) but the movers receive a wage equal to their expected wage the result will be a positive value (0.7 in this case). See Equation 4.27.

$$\log\left[\frac{\left(\frac{2X}{X}\right)}{\left(\frac{Y}{Y}\right)}\right] = \log\left[\frac{(2)}{(1)}\right] = \log[2] = 0.7 \tag{4.27}$$

With the opposite being true, i.e. the variable being negative, if the movers' actual wage exceeds their expected wage but the non-movers earn their expected wage, Equation 4.28.

$$\log\left[\frac{\left(\frac{X}{X}\right)}{\left(\frac{2Y}{Y}\right)}\right] = \log\left[\frac{(1)}{(2)}\right] = \log[0.5] = -0.7 \tag{4.28}$$

Thus if the SSV coefficient is positive then it is suggesting the non-movers are benefiting more due to the skill shortages in the market than the movers and vice versa.

Looking at columns one and two, where the moving period is classified as 12 months, the coefficient on SSV (0-15%) is small and insignificant, suggesting that there is no relative difference in the wage workers receive in a market if they have moved in the last 12 months or not. The larger SSV coefficient (15%+) is negative and insignificant in both columns, suggesting the movers and non-movers do not have different wages in markets where there are external gaps. While on average movers still lose wages (likely due to losses in firm specific human capital and being in high turnover areas) they do not lose out more where there are high levels of skill gaps, in comparison to their counterparts.

Skill gaps seem to have no differential effect on workers in either of the specifications. The percentage of vacancies does have a positive and significant coefficient. This suggests that non-movers earn more than their expected wage relative to movers in markets with higher levels of vacancies. Though this coefficient is significant it is smaller in size than the effect of SSVs.

The final coefficient that is significant in both of the first two specifications is gender balance, whereby non-movers in markets with a higher ratio of men receive a wage relatively higher than their human capital when compared to movers.

Looking at columns three and four, where movers are classified as anyone who changed jobs in the last 18 months, many of the findings from the first two columns still hold. Both the SSV coefficents are small and insignificant. in both specifications and the coefficient on the percentage of vacancies regressor is still positive, significant and smaller than the SSV coefficient.

The final two columns of Table 4.13 shows the results if movers are classified as anyone who changed jobs in the last 24 months. Here the SSV coefficients again seem to be insignificant, but this would would be expected due to movers two years prior not necessarily facing the same skill deficiencies that are being seen now.

The Uneducated and the Educated

Table 4.14 shows the regression results for the uneducated and educated decomposition, where the dependent variable is the log of the actual wage for uneducated workers minus their expected wage expressed as a ratio to the log of the actual wage for the educated minus their expected wage, equation 4.28. This can be interpreted in a similar way to the 'movers' and 'non-movers' analysis previously conducted, but with negative coefficients representing a ratio in favour of educated employees here and a positive coefficient representing a ratio in favour of the uneducated.

$$\log[(Y_E - \hat{Y}_E) - (Y_{HE} - \hat{Y}_{HE})] \tag{4.29}$$

The first two columns of Table 4.14 show results from the analysis whereby the educated group are defined as those with a 'Degree or equivalent' and 'Higher Education'. Columns three and four are included as robustness whereby the 'educated' group also includes 'GCE A Levels' to determine if the choice of where to split educational levels has a large effect on the outcome of the results.

Column one shows the reduced skills deficiency specification. The results show a negative and significant coefficient on the smaller skills measure. While the smaller SSV coefficient is significant the larger coefficient is not. The SG coefficient is also significant and negative, suggesting that being in a market where there are skill gaps causes the 'educated' to gain either a larger wage than their expected wage relative to the 'uneducated' or a smaller gap between their expected wage and actual wage (if expected wage is greater than the actual wage) relative to the uneducated.

The second column shows the effect of the SG variable has been washed out to some extent, with it again being negative but insignificant. There are numerous other variables that have consistent, significant coefficients across the two columns, including the number of employees which is negative, whether employer offered training, whether the firm is part of a union and the gender balance variable. The larger of the two SSV measures is negative and significant in the second column of this table, suggesting that once the other vacancies are taken into account only large skill shortages play a role. This coefficient is negative meaning the wage is in favour of educated employees when compared to their uneducated counterparts in markets with higher levels of SSVs.

The robustness analysis in columns three and four show that while the significance of some coefficients does depend on the cut off point between education groups, many of the coefficients are significant across both splits tested. The percentage of SGs is seen to have a positive coefficient, suggesting that the results do differ based on what cut off point is chosen.

The coefficients on both of the SSV measures are insignificant in the robustness analysis. A variable worth some mention due to its size and consistency is the dummy to represent if the employer offered training. The coefficient on this variable is extremely large and is significant at the 1% level in all four specifications. The coefficient is positive suggesting that uneducated workers in markets which offer more training do relatively better than the 'educated', with the size of the coefficient, around 0.35.

The 'uneducated' vs 'educated' analysis seems to suggest that there is no difference in who wins and loses in terms of SSV in the market, however, SG do have a differential effect. It appears the educated do better in areas where skill gaps are more prevalent, and while they may still lose out from the fact that SGs seem to push wages down the educated may not lose out by as much as the uneducated.

Females and Males Comparison

Table 4.15 shows the WLS results from the females and males comparison analysis, where the dependent variable is the log of the actual wage for females minus their expected wage as a ratio to the log of the actual wage for males minus their expected wage. This can be seen in equation 4.30. Column one shows the reduced form of the specification where not all of the skill deficiency measures are included with column two showing the full specification. The results suggest that SSVs do favour males to females (i.e. there is a negative coefficient) for the smaller SSV variable in column one and the larger SSV variable in column two. SGs have a positive and significant coefficient in column one, but this seems to be washed out in column two in a similar way to the results in the education analysis.

The results from this analysis show that there is very little consistent gender difference in the effect that both internal and external skill gaps are having. Establishments that offer training are again interesting with large positive coefficients in both columns, suggesting women do better than their male counterparts when in establishments that offer training.

$$\log[(Y_F - \hat{Y}_F) - (Y_M - \hat{Y}_M)] \tag{4.30}$$

4.9 Findings and Discussion

The evidence from both the pseudo panel analysis and the individual wage regressions suggests that a higher level of SSVs in a market has a positive impact on a worker's wage. While the impact is not large it is not negligible, with a one percentage point increase in SSVs in the market causing a 0.7% increase in a worker's wage. This is in line with Hypothesis one that the scarcity of a given skill in the market will cause higher wages to be offered for this skill, increasing a worker's wage.

The analysis on internal skill deficiencies, skill gaps, shows varied and interesting results.

While both the pseudo panel and the individual wage regressions show a negative coefficient on SGs the size of the coefficient varies from 0.065 to 0.015 depending on the estimator, specification and dataset used. Although the size of the coefficient does vary the evidence suggests that SGs in a market do have a significant negative effect on wages.

The results from this are in line with Hypothesis two, but not Hypothesis three, proposed in the theory section of this work. The impact of SGs at the market level was expected to be negative, with workers being paid less than those around them performing the same job proficiently, thus bringing down the average wage in the market as observed here. However, it was expected that SGs would push wages up for individuals in a similar way as the over/under education literature shows with workers with less education receive a higher wage than expected, as they are typically working in a harder job. This was not seen with the coefficient on SGs being negative suggesting that higher SGs actually decrease an individual's wage (though it is not possible to tell if an individual is underskilled themselves or just working in a market where others are more likely to be underskilled). The results found here instead suggest that workers are paid less in markets with higher SGs as they are paid relative to their human capital and SGs suggest a lower human capital. Workers must either be singled out individually and paid a lower wage (bringing down the market average wage), or all the workers in the market are paid less as their expected human capital is lower due to the increased SGs in the market. There is also a possibility that by working with low skill co-workers has a spillover effect reducing high productivity workers down to low productivity workers. While the idea that the same outcomes are apparent when looking at skill gaps to looking at over-education in terms of the general result the size of these results cannot easily be compared due to the difference in the measures.

Further to this, the effect of SSVs and SGs was broken down across different pairs of subgroups; 'movers and non-movers', the 'uneducated and educated' and 'male and female'.

The first of these two splits, ‘movers and non-movers’, reveals little.

The ‘uneducated and educated’ analysis shows no consistent effects for SSVs or SGs. While the ‘educated’ may still lose out as SSV are seen to be having a negative effect on a worker’s wage they are doing relatively better than the ‘uneducated’. The analysis comparing females and males showed both SGs and SSVs had very little difference between the two subgroups.

The results from this work show that the UK labour market appears to be functioning properly. Those workers in an area with higher SSVs receive a higher wage, suggesting that incentives exist to move into these industries/occupations. With the right information made available and the relevant (re)training courses made accessible then individuals have incentives to move to areas where SSVs exist allowing the market to stabilise itself. The results from the internal skill gaps analysis are harder to interpret as it is impossible to tell in this work if an individual has a SG themselves or if they are just in a market with higher SGs. This may be as markets with lower average human capital (due to increased SGs) pay a lower wage or individuals are singled out and paid a lower wage due to having a skill gap themselves. If it is the former there is no incentive to move into the market if you have the relevant skills but if the latter is true then there may be a premium for having the right skills and being in a higher SG area that cannot be identified in this work.

Table 4.2: Table of Variable Definitions

Pseudo Panel Dataset	
Log Real Wage -	Real wage in logarithms.
Percentage of Hard-to-fill vacancies (HTFV) -	Vacancies which are proving difficult to fill, as defined by the establishment (from question: "Are any of these vacancies proving hard-to-fill?"). This is measured as the percentage of HTFVs the establishment experiences.
Percentage of Vacancies -	Vacancies the establishment self reports. This is measured as the percentage of vacancies the establishment experiences.
Skill-shortage vacancies (SSVs) -	Vacancies which are proving difficult to fill due to the establishment not being able to find applicants with the appropriate skills, qualifications or experience. The SSV variable is measured as the percentage of SSVs the establishments experiences.
Share of HTFV that are SSV -	The percentage of HTFV that are due to skill reasons in the establishment. Measured as a percentage.
Percentage of Skills gaps (SG) -	A "skills gap" is where an employee is not fully proficient, i.e. is not able to do their job to the required level (see Glossary for further details). This is measured as the percentage of SGs the establishment is currently facing.

Number of employees -	Establishment size is included as a continuous variable
Perc Growth in last 2 years -	The number of percentage employment has grown by over the last two years.
Gender Balance -	The percentage of males working in the establishment. 0 represents an entirely female workforce and 1 represents an entirely male workforce.
Whether work offered training -	Whether the workplace offered its employees training opportunities in the last year.
Individual Wage Dataset	
Part of Union Indicator -	The percentage of workers who are part of a union in the establishment. 0 represents no union members in the workforce and 1 represents an unionised workforce.
Experience -	Continuous variable that indicates the amount of experience a work has.
Full time -	Dichotomous variable that signifies if a worker is employed full time (Full time contract=1, otherwise=0).
Male -	Dichotomous variable that signifies if a worker is male (Male=1, Female=0).

Private -	Dichotomous variable that identifies if a worker is employed in the private sector (Private=1, otherwise=0).
Permanent -	Dichotomous variable that identifies if a worker is employed on a permanent contract (Permanent=1, otherwise=0).
Education (Degree or Equi, Higher Education, GCE A Level, GCSE, Other) -	A series of dummy variables are used to determine the level of education the worker has. The levels are defined as Degree or Equivalent, Higher Education, GCE A Level, GCSE A8-C, Other. These education measures come from the LFS question that asks workers what their highest level of education is.
Moved (12m) -	Dichotomous variable that identifies if a worker has moved in the last 12 months (Moved=1, otherwise=0).
Moved (24m) -	Dichotomous variable that identifies if a worker has moved in the last 24 months (Moved=1, otherwise=0).

Table 4.3: Descriptive Statistics for Pseudo Panel Dataset

	2012 mean	2010 mean	2008 mean	2006 mean
Log Real Wage	2.295	2.275	2.256	2.151
SSV (0-15%)	4.046	3.610	0.584	0.670
SSV (15%+)	0.137	0.525	0.591	0.000
Percentage of HTFV	1.702	0.318	1.400	0.958
Percentage of Vacancies	8.736	6.676	4.042	2.463
Share of HTFV that are SSV	0.746	0.763	0.691	0.691
Percentage of SG	4.537	5.533	4.498	4.567
Number of Employees	4.730	6.741	5.417	5.266
Perc Growth in last 2 years	1.274	6.869	0.055	0.083
Gender balance	0.559	0.557	0.557	0.556
Whether work offered training	0.181	0.578	0.586	0.590
Part of a Union Indicator	0.196	0.195	0.243	0.256
Observations	126	126	126	126

Table 4.4: Descriptive Statistics for Individual Wage Dataset

	2012 mean	2010 mean	2008 mean	2006 mean
Experience	24.078	24.139	23.265	22.827
Full time	0.746	0.745	0.744	0.740
Male	0.502	0.504	0.503	0.500
Private	0.646	0.641	0.670	0.668
Degree or equiv	0.304	0.272	0.237	0.226
Higher Education	0.103	0.107	0.097	0.097
GCE A Level	0.226	0.223	0.223	0.228
GCSE A*-C	0.223	0.221	0.234	0.241
Other	0.084	0.107	0.125	0.120
Moved (12m)	0.146	0.145	0.188	0.143
Moved (24m)	0.271	0.286	0.340	0.260
SSV (0-15%)	4.824	5.130	0.632	0.915
SSV (15%+)	0.110	0.397	0.045	0.000
Percentage of SG	4.339	5.519	4.696	4.844
Percentage of HTFV	2.073	0.404	0.949	1.238
Share of HTFV that are SSV	0.719	0.761	0.703	0.697
Percentage of Vacancies	11.126	7.977	3.327	3.254
Observations	31734	34264	41547	41487

Table 4.5: Cell Counts (LFS) - 2012/2010/2008

	2012	2010	2008	Total Sample	> 5	> 10
	mean/min/max	mean/min/max	mean/min/max	mean/min/max	mean/min/max	mean/min/max
Stayed in Job 12m	249	273	310	1898	2146	2375
	1	0	2	13	34	92
	2731	2731	2282	12369	12369	12369
Stayed in Job 18m	233	255	283	1760	1989	2202
	1	0	2	10	30	84
	2604	2592	2133	11680	11680	11680
Stayed in Job 24m	212	228	253	1590	1797	1989
	1	0	1	10	26	77
	2434	2388	1954	10752	10752	10752
Moved Job 12m	42	45	69	343	388	430
	0	0	0	4	6	14
	521	590	833	2235	2235	2235
Moved Job 18m	57	63	96	482	544	603
	0	0	0	7	10	20
	709	781	1099	3041	3041	3041
Moved Job 24m	78	89	126	652	737	816
	0	0	0	7	14	24
	921	1068	1392	4069	4069	4069
Observations	123	123	125	126	110	98

Table 4.6: Cell Counts (ESS) - 2012/2010/2008

	2012	2010	2008	Total Sample	> 5	> 10
	mean/min/max	mean/min/max	mean/min/max	mean/min/max	mean/min/max	mean/min/max
Number of SSVs	117	69	1420	3000	3421	3808
	0	4	1	6	48	48
	741	508	11309	22718	22718	22718
Number of HTFVs	404	1191	1420	4408	5001	5559
	0	31	1	33	77	77
	2321	10592	11309	34557	34557	34557
Number of Vacancies	404	298	1420	3515	4002	4452
	0	9	1	11	55	55
	2321	1796	11309	25521	25521	25521
Share of SSV/HTFV	35	23	48	159	182	203
	0	0	0	0	5	11
	264	210	418	1164	1164	1164
Number of SGs	1569	1495	1420	5878	6703	7454
	0	0	1	2	75	85
	10518	10592	11309	43043	43043	43043
Observations	126	126	126	126	110	98

Table 4.7: Levels Regression: The effect of Continuous Skill Gap Measures on Mean Log Actual Wage

	OLS	RE	FE	OLS	RE	FE	OLS	RE	FE
	β / SE	β / SE	β / SE	β / SE	β / SE	β / SE	β / SE	β / SE	β / SE
SSV (0-15%)	0.009**	0.007**	0.006*	0.009**	0.007**	0.006*	0.007*	0.005*	0.004
	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)
SSV (15%+)	0.008***	0.007***	0.007***	0.009***	0.008***	0.007***	0.009	0.007*	0.006
	(0.002)	(0.001)	(0.001)	(0.003)	(0.002)	(0.002)	(0.006)	(0.004)	(0.004)
Percentage of SG	-0.040***	-0.015*	0.002	-0.040***	-0.015*	0.002	-0.050***	-0.018**	0.006
	(0.007)	(0.008)	(0.009)	(0.007)	(0.008)	(0.009)	(0.006)	(0.008)	(0.009)
Percentage of Vacancies				-0.001	-0.000	0.000	-0.001	-0.001	0.000
				(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
Percentage of HTFV							-0.001	0.001	0.001
							(0.006)	(0.003)	(0.003)
Share of HTFV that are SSV							0.296***	0.097*	0.046
							(0.079)	(0.051)	(0.045)
Number of Employees	0.005***	0.001	-0.004*	0.005***	0.001	-0.004*	0.004**	0.000	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Perc Growth in last 2 years	-0.007***	-0.003**	0.000	-0.007***	-0.003**	0.000	-0.007***	-0.002*	0.001
	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.003)	(0.001)	(0.002)
Gender balance	0.341***	0.269***	0.181	0.339***	0.269***	0.181	0.273***	0.244***	0.192**
	(0.048)	(0.075)	(0.120)	(0.048)	(0.076)	(0.120)	(0.047)	(0.064)	(0.085)
Whether work offered training	1.215***	0.592***	0.282	1.214***	0.591***	0.282	1.116***	0.619***	0.399***
	(0.163)	(0.181)	(0.182)	(0.164)	(0.181)	(0.181)	(0.158)	(0.149)	(0.149)
Part of a Union Indicator	-0.306***	0.007	0.114	-0.304***	0.008	0.114	-0.179*	0.081	0.155
	(0.102)	(0.109)	(0.112)	(0.102)	(0.109)	(0.111)	(0.105)	(0.093)	(0.103)
Observations	489	489	489	489	489	489	439	439	439

* (p<0.10), ** (p<0.05), *** (p<0.01)
Year dummies also included but not displayed

Table 4.8: Levels Regression: The effect of Continuous Skill Gap Measures on Mean Log Actual Wage with Mincer Variables included

	OLS	RE	FE	OLS	RE	FE	OLS	RE	FE
	β / SE	β / SE	β / SE	β / SE	β / SE	β / SE	β / SE	β / SE	β / SE
SSV (0-15%)	0.006*	0.006**	0.005*	0.006*	0.006*	0.005	0.004	0.003	0.003
	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
SSV (15%+)	0.008***	0.007***	0.007***	0.009***	0.007***	0.006**	0.012*	0.009**	0.007
	(0.002)	(0.001)	(0.001)	(0.003)	(0.002)	(0.002)	(0.006)	(0.004)	(0.004)
Percentage of SG	-0.030***	-0.011	0.004	-0.030***	-0.011	0.004	-0.037***	-0.012	0.009
	(0.009)	(0.008)	(0.008)	(0.009)	(0.008)	(0.008)	(0.007)	(0.008)	(0.009)
Percentage of Vacancies				-0.001	0.000	0.001	-0.001	-0.001	0.001
				(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Percentage of HTFV							-0.002	0.000	-0.000
							(0.006)	(0.004)	(0.004)
Share of HTFV that are SSV							0.153*	0.055	0.027
							(0.085)	(0.065)	(0.055)
Number of Employees	0.003	0.001	-0.003	0.003	0.001	-0.003	0.002	-0.000	-0.001
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
Perc Growth in last 2 years	-0.005***	-0.003***	0.000	-0.005***	-0.003***	0.000	-0.005**	-0.003***	-0.001
	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)
Gender balance	0.010	0.067	0.114	0.010	0.068	0.115	0.021	0.034	0.096
	(0.081)	(0.102)	(0.146)	(0.082)	(0.102)	(0.147)	(0.066)	(0.075)	(0.103)
Whether work offered training	0.911***	0.496***	0.260	0.911***	0.495***	0.259	0.698***	0.447***	0.290**
	(0.187)	(0.190)	(0.181)	(0.188)	(0.190)	(0.180)	(0.160)	(0.137)	(0.123)
Part of a Union Indicator	-0.263**	-0.083	0.051	-0.261**	-0.083	0.050	-0.251**	-0.080	0.014
	(0.118)	(0.122)	(0.129)	(0.118)	(0.122)	(0.129)	(0.109)	(0.106)	(0.118)
Experience Squared	-0.000	-0.000	0.000	-0.000	-0.000	0.000	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Experience	0.007	0.010	0.006	0.007	0.010	0.006	0.047***	0.045***	0.039***
	(0.020)	(0.015)	(0.012)	(0.020)	(0.015)	(0.012)	(0.014)	(0.011)	(0.009)
Private	-0.018	-0.019	-0.014	-0.019	-0.019	-0.015	-0.121**	-0.078	-0.064
	(0.065)	(0.069)	(0.083)	(0.065)	(0.070)	(0.083)	(0.061)	(0.063)	(0.073)
Full Time	0.737***	0.461***	0.188	0.736***	0.460***	0.187	0.618***	0.435***	0.212**
	(0.133)	(0.115)	(0.121)	(0.133)	(0.115)	(0.121)	(0.120)	(0.089)	(0.092)
Observations	489	489	489	489	489	489	439	439	439

* (p<0.10), ** (p<0.05), *** (p<0.01)
Year dummies also included but not displayed

Table 4.9: Levels Regression: The effect of Dichotomous Skill Gap Measures on Mean Log Actual Wage

	OLS	RE	FE	OLS	RE	FE	OLS	RE	FE
	β / SE	β / SE	β / SE	β / SE	β / SE	β / SE	β / SE	β / SE	β / SE
Any SSV	-0.100	-0.111	-0.429	-0.086	-0.093	-0.391	0.097	0.203	0.837
	(0.211)	(0.222)	(0.431)	(0.208)	(0.215)	(0.423)	(0.703)	(0.729)	(0.828)
Any HTFV							-0.149	-0.242	-1.082
							(0.572)	(0.601)	(0.759)
Any Vac				-0.470	-0.513	-0.750*	-0.469	-0.513	-0.740*
				(0.394)	(0.407)	(0.419)	(0.395)	(0.407)	(0.423)
Any SG	0.071	0.060	-0.003	0.066	0.056	-0.017	0.069	0.060	-0.022
	(0.107)	(0.111)	(0.123)	(0.107)	(0.109)	(0.122)	(0.108)	(0.111)	(0.123)
Number of Employees	-0.002	-0.002	-0.003	-0.002	-0.002	-0.003*	-0.002	-0.002	-0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Perc Growth in last 2 years	-0.002***	-0.002***	-0.001***	-0.002***	-0.002***	-0.001***	-0.002***	-0.002***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Gender balance	0.220**	0.219**	0.183	0.213**	0.211**	0.166	0.212**	0.210**	0.165
	(0.092)	(0.098)	(0.124)	(0.091)	(0.095)	(0.120)	(0.091)	(0.095)	(0.120)
Whether work offered training	0.318**	0.312*	0.286	0.294**	0.287*	0.243	0.294**	0.286*	0.239
	(0.145)	(0.163)	(0.183)	(0.145)	(0.163)	(0.182)	(0.145)	(0.164)	(0.183)
Part of a Union Indicator	0.073	0.079	0.114	0.082	0.088	0.131	0.082	0.089	0.129
	(0.084)	(0.087)	(0.112)	(0.084)	(0.086)	(0.109)	(0.084)	(0.086)	(0.109)
Observations	491	491	491	491	491	491	491	491	491

* ($p<0.10$), ** ($p<0.05$), *** ($p<0.01$)
Year dummies also included but not displayed

Table 4.10: Impact of Skill Shortages on Individuals' Wages (Pay/Actual Hours)

	X β / SE	X+Ind β / SE	X+Occ β / SE	X β / SE	X+Ind β / SE	X+Occ β / SE
SSV (0-15%)	0.007*	0.005*	0.002	0.003	0.001	0.003
	(0.004)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
SSV (15%+)	0.008**	0.002	0.006***	0.013***	0.005	0.009***
	(0.004)	(0.004)	(0.002)	(0.005)	(0.006)	(0.003)
Percentage of SG	-0.065***	-0.064***	-0.004	-0.057***	-0.059***	-0.008
	(0.008)	(0.008)	(0.008)	(0.007)	(0.007)	(0.008)
Percentage of HTFV				-0.011**	-0.007	-0.004
				(0.005)	(0.005)	(0.004)
Share of HTFV that are SSV				0.675***	0.543***	0.388***
				(0.077)	(0.074)	(0.067)
Percentage of Vacancies				0.002*	0.002*	-0.000
				(0.001)	(0.001)	(0.001)
Experience	0.023***	0.022***	0.022***	0.022***	0.022***	0.022***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Experience SQ	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Full time	0.123***	0.099***	0.083***	0.093***	0.080***	0.076***
	(0.014)	(0.013)	(0.011)	(0.012)	(0.011)	(0.011)
Male	0.076***	0.054***	0.069***	0.066***	0.056***	0.061***
	(0.017)	(0.012)	(0.011)	(0.014)	(0.011)	(0.011)
Private	-0.000	-0.108***	-0.012	-0.053***	-0.109***	-0.042***
	(0.020)	(0.015)	(0.016)	(0.019)	(0.014)	(0.015)
Degree or equiv	0.605***	0.589***	0.431***	0.546***	0.538***	0.428***
	(0.028)	(0.023)	(0.024)	(0.024)	(0.019)	(0.024)
Higher Education	0.433***	0.428***	0.298***	0.381***	0.381***	0.297***
	(0.027)	(0.026)	(0.020)	(0.024)	(0.023)	(0.020)
GCE A Level	0.282***	0.273***	0.211***	0.238***	0.236***	0.207***
	(0.020)	(0.019)	(0.015)	(0.018)	(0.017)	(0.015)
GCSE A*-C	0.183***	0.174***	0.134***	0.150***	0.146***	0.132***
	(0.015)	(0.013)	(0.010)	(0.012)	(0.011)	(0.010)
Other	0.089***	0.090***	0.077***	0.077***	0.078***	0.076***
	(0.015)	(0.015)	(0.012)	(0.014)	(0.014)	(0.011)
Moved (12m)	-0.049***	-0.048***	-0.044***	-0.046***	-0.045***	-0.044***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.007)	(0.008)
Moved (24m)	-0.111***	-0.103***	-0.091***	-0.101***	-0.097***	-0.089***
	(0.008)	(0.007)	(0.006)	(0.007)	(0.007)	(0.007)
Observations	148639	148639	148639	147079	147079	147079

* (p<0.10), ** (p<0.05), *** (p<0.01)

Table 4.11: Movers and Non-movers Characteristics

	Moved (12m)	Non-mover	Moved (18m)	Non-mover	Moved (24m)	Non-mover
Log Wage	1.94	2.23	1.96	2.25	1.99	2.27
Experiance	15.39	25.04	15.95	25.67	16.56	26.49
Full time	0.64	0.74	0.65	0.74	0.67	0.75
Male	0.48	0.49	0.47	0.49	0.47	0.49
Private	0.76	0.63	0.75	0.62	0.74	0.61
Degree or equiv	0.25	0.26	0.26	0.26	0.26	0.25
Higher Education	0.07	0.11	0.08	0.11	0.08	0.11
GCE A Level	0.23	0.22	0.23	0.22	0.23	0.22
GCSE A*-C	0.25	0.23	0.25	0.22	0.24	0.22
Other	0.12	0.11	0.12	0.11	0.12	0.11
In Union	0.11	0.31	0.12	0.33	0.14	0.35
Permanent	0.81	0.97	0.84	0.97	0.87	0.98
Observations	49583	259992	69491	240084	93476	216099

Table 4.12: Individual Wage Estimation (Wage= Pay/ Actual Hours)

	Moved in past 12m β / SE	Moved in past 18m β / SE	Moved in past 24m β / SE
Experience	-0.014***	-0.014***	-0.014***
	(0.000)	(0.000)	(0.000)
Experience SQ			
Full time	0.166***	0.164***	0.163***
	(0.003)	(0.003)	(0.003)
Male	0.106***	0.105***	0.105***
	(0.003)	(0.003)	(0.003)
Private	-0.033***	-0.031***	-0.028***
	(0.003)	(0.003)	(0.003)
Permanent	0.101***	0.096***	0.094***
	(0.006)	(0.006)	(0.006)
Degree or equiv	0.610***	0.610***	0.610***
	(0.006)	(0.006)	(0.006)
Higher Education	0.456***	0.456***	0.455***
	(0.007)	(0.007)	(0.006)
GCE A Level	0.309***	0.308***	0.306***
	(0.006)	(0.006)	(0.006)
GCSE A*-C	0.223***	0.223***	0.221***
	(0.006)	(0.006)	(0.006)
Other	0.097***	0.098***	0.098***
	(0.006)	(0.006)	(0.006)
Moved (12m)	-0.119***		
	(0.004)		
Moved (18m)		-0.128***	
		(0.003)	
Moved (24m)			-0.138***
			(0.003)
2011	0.000	0.000	0.000
	(0.005)	(0.005)	(0.005)
2010	-0.009*	-0.008	-0.006
	(0.005)	(0.005)	(0.005)
2009	-0.023***	-0.021***	-0.017***
	(0.005)	(0.005)	(0.005)
2008	-0.026***	-0.024***	-0.023***
	(0.005)	(0.005)	(0.005)
2007	-0.071***	-0.070***	-0.068***
	(0.005)	(0.005)	(0.005)
2006	-0.101***	-0.098***	-0.095***
	(0.005)	(0.005)	(0.005)
Observations	215718	215718	215718

* (p<0.10), ** (p<0.05), *** (p<0.01)

Table 4.13: Weighted Least Squares Regression. Dependant variable is: The ratio that expected wage exceeds actual wage for non-movers to the ratio that expected wage exceeds actual wage for movers

	12m MFX / SE	12m MFX / SE	18m MFX / SE	18m MFX / SE	24m MFX / SE	24m MFX / SE
SSV (0-15%)	-0.004	0.001	-0.001	0.002	-0.001	0.002
	(5.5e+05)	(9.3e+05)	(0.00)	(2.4e+10)	(0.00)	(0.00)
SSV (15%+)	-0.007	-0.016	-0.006	-0.013*	-0.003	-0.007
	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)
Percentage of SG	0.015	-0.007	0.005	-0.009	0.003	0.004
	(6962.37)	(8.67)	(0.00)	(1.8e+06)	(0.00)	(0.00)
Percentage of HTFV		-0.014		-0.005		-0.005
		(4.74)		(5.1e+07)		(0.01)
Percentage of Vacancies		0.008		0.005		0.004***
		(6.69)		(1.5e+07)		(0.00)
Share of HTFV that are SSV		0.089		0.086		0.071
		(8.66)		(8.0e+05)		(0.05)
Number of Employees	0.002	0.004	0.001	0.003	0.003*	0.005***
	(4.7e+07)	(39.06)	(0.00)	(4.2e+09)	(0.00)	(0.00)
Perc Growth in last 2 years	-0.001	0.002	0.000	0.002	0.000	0.001
	(5.2e+14)	(319.87)	(0.00)	(5.4e+09)	(0.00)	(0.00)
Whether work offered training	0.415	0.172	0.184**	-0.034	0.144**	0.071
	(31845.91)	(15.28)	(0.07)	(5.2e+05)	(0.06)	(0.07)
Private Sector Indicator	-0.153	-0.130	-0.186**	-0.173	-0.093	-0.147*
	(12855.00)	(12.15)	(0.08)	(1.3e+06)	(0.07)	(0.08)
Part of a Union Indicator	0.168	0.126	0.147***	0.103	0.082***	0.077**
	(13867.70)	(11.50)	(0.04)	(9.2e+05)	(0.03)	(0.03)
Gender balance	0.205	0.050	0.080***	-0.028	0.038	0.004
	(17221.68)	(5.70)	(0.03)	(5.8e+05)	(0.03)	(0.03)
Observations	461	421	469	427	471	428

Marginal effects
* (p<0.10), ** (p<0.05), *** (p<0.01)
Year dummies also included but not displayed

Table 4.14: **Weighted Least Squares Regression. Dependant variable is: The ratio that expected wage exceeds actual wage for un-educated workers to the ratio that expected wage exceeds actual wage for educated workers**

	HE = $1/2^a$ MFX / SE	HE = $1/2^a$ MFX / SE	HE = $1/2/3^a$ MFX / SE	HE = $1/2/3^a$ MFX / SE
SSV (0-15%)	-0.008**	-0.002	-0.002	-0.002
	(0.00)	(0.00)	(0.00)	(3.2e+14)
SSV (15%+)	-0.007	-0.015**	-0.001	0.008
	(0.01)	(0.01)	(0.00)	(0.01)
Percentage of SG	0.022***	0.001	-0.008*	-0.002
	(0.01)	(348.80)	(0.00)	(6.0e+13)
Percentage of HTFV		-0.005		-0.010
		(0.01)		(2.2e+08)
Percentage of Vacancies		0.006***		0.000
		(0.00)		(6.5e+49)
Share of HTFV that are SSV		0.062		-0.025
		(0.06)		(6.5e+07)
Number of Employees	0.001	0.002	-0.003**	-0.004
	(0.00)	(0.00)	(0.00)	(3.0e+10)
Perc Growth in last 2 years	-0.002	0.002	-0.002	0.001
	(0.00)	(0.00)	(0.00)	(1.0e+29)
Whether work offered training	0.520***	0.193**	0.289***	0.330
	(0.09)	(0.08)	(0.06)	(2.2e+08)
Private Sector Indicator	-0.190*	-0.084	0.022	0.079
	(0.11)	(0.09)	(0.07)	(7.3e+07)
Part of a Union Indicator	0.181***	0.091**	0.173***	0.103
	(0.05)	(0.04)	(0.03)	(8.9e+07)
Gender balance	0.225***	0.047	0.082***	0.106
	(0.04)	(0.03)	(0.03)	(8.7e+07)
Observations	461	421	469	425

Marginal effects
* ($p<0.10$), ** ($p<0.05$), *** ($p<0.01$)
a - Where: 1= Degree or equiv; 2= Higher Education and 3= GCE A Levels
Year dummies also included but not displayed

Table 4.15: Weighted Least Squares Regression. Dependant variable is: The ratio that expected wage exceeds actual wage for females to the ratio that expected wage exceeds actual wage for males

	A MFX / SE	B MFX / SE
SSV (0-15%)	-0.008**	-0.002
	(0.00)	(0.00)
SSV (15%+)	-0.007	-0.015**
	(0.01)	(0.01)
Percentage of SG	0.022***	0.001
	(0.01)	(348.80)
Percentage of HTFV		-0.005
		(0.01)
Percentage of Vacancies		0.006***
		(0.00)
Share of HTFV that are SSV		0.062
		(0.06)
Number of Employees	0.001	0.002
	(0.00)	(0.00)
Perc Growth in last 2 years	-0.002	0.002
	(0.00)	(0.00)
Whether work offered training	0.520***	0.193**
	(0.09)	(0.08)
Private Sector Indicator	-0.190*	-0.084
	(0.11)	(0.09)
Part of a Union Indicator	0.181***	0.091**
	(0.05)	(0.04)
Gender balance	0.225***	0.047
	(0.04)	(0.03)
Observations	461	421

Marginal effects
* (p<0.10), ** (p<0.05), *** (p<0.01)
Year dummies also included but not displayed

Chapter 5

Conclusions

5.1 Conclusions

This thesis set out to provide evidence of the impact skill deficiencies are having on the UK economy at three distinct aggregation levels. Each of these aggregation levels will be discussed in turn here with an overview of their findings, the implications of the work, the limitations of the approach and the future work that is still required.

5.1.1 UK Economy

Chapter two of this thesis investigated the impact that skill shortages were having on the UK economy as a whole by testing the idea that a low skill equilibrium may exist. The idea of a low skill equilibrium was first proposed by Finegold and Soskice (1988) and suggests that an economy can end up on a low skill trajectory. This occurs when skill deficiencies exist and thus firms cannot hire the desired workers and thus they cut the jobs they cannot fill. The fact that these jobs no longer exist leads to no incentive for workers to gain the required skills and further skill deficiencies arise. This cycle reinforces itself leading to a situation whereby the level of skills, and subsequently the level of productivity, in the economy drops. While it is hard to concretely prove that this phenomenon exists the cycle cannot occur if one of the links does not exist. This is the approach taken in this work whereby it is investigated whether establishments have a negative response to skill deficiencies and cutting jobs. Very little evidence is found of this in the UK, however some occupations do respond in a negative way to skill shortages. These occupations are the intermediate skill occupations and those which are suggested to be struggling in the task biased technological change and the job polarisation literature.

While this work finds no evidence of a low skill equilibrium for most occupations, the results are worrying for the intermediate skill occupations. It has already been suggested that these occupations are struggling by the tasked technological change literature and as the jobs are often no longer deemed “business essential” and these routine tasks are therefore replaced by either machines or broken into their skilled and unskilled components and distributed amongst the other workers. While it is unfeasible to show that all the links of a low skill equilibrium exist, thus guaranteeing that the UK is on a low skill trajectory, the most important piece of the chain has been shown to exist for intermediate skill occupations. With establishments cutting jobs where skill deficiencies are found there exists the possibility that workers will have less incentive to gain these skills and thus the cycle is formed. One of the more interesting questions that arises from this is whether the polarisation and task biased technological change has started this cycle or if the gaps in intermediate skills in the UK have led to these occupations being seen as "non business essential".

This work provides valuable information on the way establishments respond to skill shortages though there is still one key link missing from the chain of the low skill equilibrium framework: how workers respond to these job cuts. Without evidence to support the prospect that workers have less incentive to up-skill and that this causes the cycle this cannot be known for sure. Further to this it would be interesting to see how the issues of job polarisation and the low skill equilibrium seen in the UK link together as mentioned above. Did one come first and cause the other? Did the two issues only become prominent due to existing at the same time? And do the answers to the problem of job polarisation lie in increasing the number of intermediate skills in the economy showing that these jobs are valuable and "business essential" when performed by adequately skilled workers?

5.1.2 Establishments in the UK

Chapter three investigated the impact that skill deficiencies were having on the productivity of UK establishments. To do this, the ESS was matched to the ARD at an establishment level. This dataset is weighted towards larger establishments due to the sampling framework of the ARD. Instrumental variables regressions were run on the dataset with lagged inputs in an effort to control for exogeneity in the production function. These results give a positive and significant coefficient on the SG variable, suggesting that establishments with higher levels of SGs are more productive. As this result was contrary to the theory and anecdotal evidence it was believed that a selection bias may exist whereby those establishments that were more likely to have SGs were the more productive establishments. In an effort to remove this bias both a Heckman two-step model and Propensity Score Matching (PSM) were used. While the instruments in the Heckman model were weak the results from the PSM were interesting. Once those establishments with SGs and those without were matched on all relevant criteria then there is no significant difference seen between the two in terms of Gross Value Added (GVA).

This work finds no evidence of internal skill gaps having a negative effect on establishments' productivity however there could be numerous reasons for this. The final matched dataset that is used in the analysis is both relatively small in size and also biased towards larger establishments due to the matching process. This lack of sample size and possible lack of variation in the type of establishments contained in the sample may have caused any negative effect to wash out. The other possible explanation is that there is no effect of SGs on productivity. Establishments could either make do with the workers they have, adapting their working practices and the workloads of the workers that they have effectively, or establishments do not feel the effect of skill gaps. If the latter of these two is the case then maybe the self reported nature of the SG measure in the ESS is not useful for this analysis as any worker that is classed as not being fully proficient at their job should be less productive than a fully proficient worker. Further to this the small sample size meant that nothing could be done to investigate the impact of external skill shortages as the number of workplaces reporting them was just too small.

One of the main results from this work is that the firm level match between the ABI and ESS does not provide a good enough data source for the typical complex productivity analysis. Standard methods for productivity estimation (see Griffiths, 1999) use GMM analysis to control for both endogeneity and selection issues. The dataset created with this match did not give enough time periods to implement this method. This limited the analysis and further work may be experience more success working at a more aggregated level, either industry or region, to avoid this problem.

While the measure of performance chosen for this work was productivity this is not the only way that performance can be measured. Another important domain which was considered for investigation was establishment survival, i.e. were those establishments with greater levels of skill gaps more likely to fail in the near future. Although this issue was of interest there seemed to be no way to measure this effect with the current data as when the BSD was matched to the dataset the small sample size meant that less than ten establishments failed in the relevant period, too small of a number for any sort of robust statistical analysis.

Given the above limitations it seems premature to state that skill deficiencies are not impacting the performance of UK workplaces. While no impact has been found in this work a larger matched dataset may reveal more results of interest, and may also allow the assessment of external skill shortages and establishment survival instead of the limited scope of this work. Given the above statement if it is true that skill gaps are having no impact on the performance of UK workplaces then perhaps it should the current product market strategies of establishments should be questioned. If workers with skill gaps are as productive as normal workers then it may be that none of these workers are really performing to the peak of their ability and thus nationwide productivity improvements could be made.

5.1.3 Workers in the UK

Chapter four investigates the impact that skill deficiencies have on the final level of aggregation, workers. This is done by investigating the wages that workers receive in markets with differing levels of skill deficiencies. To do this, the ESS is matched to the LFS at an industry/occupation (referred to as market) level. This market level match allows a pseudo panel dataset to be created. The impact that skill deficiencies are having is then investigated in two separate ways. First panel data techniques were carried out on the pseudo panel dataset to see if markets with higher skill deficiencies pay different wages to those without. The second method of analysis runs individual worker wages regressions with the data from the LFS and the relevant market level skill information matched in from the ESS. Both of these forms of analysis give similar results for the impact of external skill shortages, with workers in markets with one percentage point higher levels of skill shortages receiving a 0.7% increase in wage. While the results for the internal skill gap analysis both suggest that markets with higher levels of SGs have lower wages the size of this effect seems to vary with the method of analysis used. Further to this work an effort was made to see which subgroups won and lost in terms of remuneration with a unique two step model. The results from this work suggest that workers who have moved jobs in the last 12 months do better in areas with higher SSV and the 'educated' do better than the 'uneducated' in markets with higher levels of SGs. The results show no difference across genders.

While the measure of impact in this work was workers' remuneration there are other forms of analysis that could also have been undertaken. While wages are an important dimension it may also be beneficial to understand how skill deficiencies impact on the ease with which workers move into employment, how long they are employed for, and how they progress through the job market. This cannot be done with the data used in this work as information is only available for the gaps that establishments have, and not what gaps individual workers have. With individual level data a better understanding could be gained of the way that skill deficiencies may impact workers outside of just their wage.

The implications from this work are clear for external skill shortages but rather confusing for the internal skill gap analysis. The results from the external skill shortage analysis are positive for the UK. The fact that workers are paid higher wages in markets with skill deficiencies suggests that firms are willing to pay a higher wage to attract skilled workers into the market and thus markets should be able to stabilise themselves, moving towards an equilibrium over time. For skill shortages to exist in a given market then either the wage premium to work in the market found here to be about 0.7% for a one percentage point increase in skill shortages is not high enough, or workers are not aware of the rewards the given market has on offer and/or the skills needed to move into the market. If this is the case these issues are easily solved by making the relevant information available.

The results from the internal skill gap analysis are less clear cut. While SGs can be seen to have a negative effect on workers' remuneration in a given market it is impossible to tell if this negative is just for those with SGs, or for the market as a whole as workplaces cannot easily distinguish who has SGs and who does not. If it is the case that it just those with SGs that are penalised then it may be that workers with the relevant skills in a market gain a positive reward for this, but the effect is disguised by a larger negative to those who are not fully proficient. This would again mean that markets were functioning properly however not enough information is available to distinguish if this is the case. The alternative, that markets with higher levels of SGs pay a lower wage to all the workers in the market is less positive for the UK. If this is occurring then workers with the required skills will have no incentive to move into these markets and fill the skill gaps. A negative response such as this from establishments will only exaggerate the problem and could lead to a low skill trajectory as discussed in Chapter Two. While further work needs to be conducted to distinguish which of these is actually occurring intervention of some sort may be required if the latter is found to be true. This can be done relatively easily in theory by introducing some form of signal in these markets to allows workplaces to differentiate between workers with SGs and those without so each can be paid their respective wage.

5.1.4 Forward looking Policy Implications for the UK

Starting with a caveat it should be mentioned that this work does not look for solutions to skill issues in the UK, the analysis is entirely focused on determining in the impact of the skill deficiencies on the economy. Due to this any discussion of future policy is only supported by the evidence read and the amassed knowledge from conducting this work, rather than any of the empirical analysis that was investigated.

Given the above, the empirical findings suggest that the UK economy is not suffering as heavily as expected from skill deficiencies. While there are still skill shortage vacancies and skill gaps apparent in the UK these are not having a negative impact on the UK economy at every level. Firstly, a low skill equilibrium has not been seen and thus the worry that the UK skill demand might collapse inwards on its self seems out of place. Likewise, the wages individuals receive in markets with skill deficiencies are higher, meaning standard market forces are playing their role to encourage supply to move in line with demand. Whether these wage increases are enough to make people train, retrain, or more geographically is an interesting question and one that requires more thought. This will be imperative to policy decisions, as if markets are providing enough of an incentive then no further action should be taken and markets only require time to readjust. If this is not the case then an effort should be made to improve these incentives or reduce the costs of training/retraining so labour is more mobile.

The only level of aggregation investigated that did not bring positive news for the UK economy is the impact that skill deficiencies have on firms. While no link was found at this level this was likely due to the data issues faced in this work rather than due to no link existing. The idea that skill deficiencies have a negative impact on firm performance is not difficult to theorise and again while evidence was not presented here that supports this link, it is believed that this link will emerge given the correct analysis. This is of particular importance as the question of how skill deficiencies impact firms is perhaps the important links to have in place. With this link established it is possible to show business owners the true impact of skills and this should help in trying to shift both the demand for skills and product market strategies forward across the UK, leading to higher growth. While little can be said of the true strength of this link a few overarching comments can be made. Skill deficiencies are very much a real problem in terms of their existence, with around 30% of establishments reporting a SG in the 2011 wave of the ESS. The recent report by the Prince's Trust and HSBC also finds that over two-thirds of business leaders said they had fears about skill shortages slowing down any economic recovery and 35% of business leaders harboured fears that they may cause the business to fail. The link between skill deficiencies and workplace performance has been studied in several other smaller subgroups and it is seen to be a strong and significant relationship. This topic needs more analysis which is outlined below.

In terms of policy direction the route suggested here is twofold. Firstly, an effort should be made to statistically identify this link between skill deficiencies and performance. By quantifying this link and disseminating this information to business owners it should allow them to more accurately assess what skill demands they have. Without taking this action information on the skills desired by businesses they may mask their concerns about skill deficiencies leading to unrealistic demands. This information can then be fed into the second step of the route, addressing the important skill deficiencies and bridging these gaps so firms can improve performance and, ultimately, increase their product market strategy.

Many solutions to the UK skill problems have been highlighted in the past with both apprenticeships and immigration being involved in most discussions. During the recent election all the major political parties made claims that apprenticeships were the answer to the UK's problems. However, many previous attempts have been made at reforming the UK's apprenticeship system to try and hit the lofty achievements of the German system and other well functioning vocational systems. None of these attempts have been extremely successful and the intricacies of creating a successful apprenticeship, that is valued by employers is beyond the remit of this work. It is worthy of note that the skill deficiencies are seen to be having slightly different effects to 'non business essential' jobs and these are an area that could, in theory, be addressed well by substantive changes to the apprenticeship system.

This work therefore only serves to direct policy discussions towards the firm perspective and alleviate concerns about low skill-equilibriums and individual workers. Hopefully future work can build upon this to prove the link between firm performance and skill deficiencies, and to better inform on the legitimacy of solutions rather than just the impact of skill deficiencies, as was done in this thesis.

5.1.5 Final Remarks

This thesis has investigated the impact that skill deficiencies have had on the UK economy. The results are assorted. While there are positives to take away such as the fact that labour markets seem to be properly functioning and workers in markets with skill shortages are paid higher wages there are plenty of negatives too. The most important of these is perhaps that there seems to be some evidence that a low skill equilibrium could exist for intermediate skilled jobs. Further to this there are still areas where the impacts are still unclear, mainly in terms of what the impact is on establishments' performance.

Bibliography

Acemoglu, D. and J.-S. Pischke (1999): "Beyond Becker: Training in imperfect labour markets," *The Economic Journal*, 109, 112–142.

Ackerberg, D., C. Lanier Benkard, S. Berry, and A. Pakes (2007): "Econometric tools for analyzing market outcomes," in *Handbook of Econometrics*, ed. by J. Heckman and E. Leamer, Elsevier, vol. 6 of *Handbook of Econometrics*, chap. 63.

Akaike, H. (1998): "Information theory and an extension of the maximum likelihood principle," in *Selected Papers of Hirotugu Akaike*, Springer, 199–213.

Allen, J. and E. de Weert (2007): "What do educational mismatches tell us about skill mismatches? A cross-country analysis," *European Journal of Education*, 42, 59–73.

Allen, J. P. and R. Van der Velden (2001): "Educational mismatches versus skill mismatches: Effects on wages, job satisfaction, and on-the-Job search," *Oxford Economic Papers*, 53, 434–52.

Antman, F. and D. J. McKenzie (2007): "Earnings mobility and measurement error: A pseudo-panel approach," *Economic Development and Cultural Change*, 56, 125–161.

Arellano, M. and S. Bond (1991): "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," *Review of Economic Studies*, 58, 277–97.

Audretsch, D. B. (1991): "New-firm survival and the technological regime," *The Review of Economics and Statistics*, 73, 441–50.

Audretsch, D. B. and T. Mahmood (1995): "New firm survival: New results using a hazard function," *The Review of Economics and Statistics*, 77, 97–103.

Autor, D., F. Levy, and R. J. Murnane (2001): "The skill content of recent technological change: an empirical exploration," Tech. rep., National Bureau of Economic Research.

Bennett, J. and S. McGuinness (2009): "Assessing the impact of skill shortages on the productivity performance of high-tech firms in Northern Ireland," *Applied Economics*, 41, 727–737.

Blundell, R. and S. Bond (1998): "Initial conditions and moment restrictions in dynamic panel data models," *Journal of Econometrics*, 87, 115–143.

Bosworth, D., A. Dickerson, P. Elias, A. Green, S. Hallam, C. Hasluck, T. Hogarth, H. Livesey, J. Shury, S. Thomas, et al. (2004): "Skills in England 2003: Volume 2, research report," .

Burdett, K. and E. Smith (2002): "The low skill trap," *European Economic Review*, 46, 1439–1451.

Cameron, A. C. and P. K. Trivedi (2005): *Microeconometrics: Methods and applications*, Cambridge University Press.

Collier, W., F. Green, Y.-B. Kim, and J. Peirson (2008): "Education, training and economic performance: Evidence from establishment survival data," Studies in Economics 0822, Department of Economics, University of Kent.

Criscuolo, C., J. Haskel, and R. Martin (2003): "Building the evidence base for productivity policy using business data linking," *Economic Trends*, 600, 39–51.

Crossley, T., J. Bresser, L. Delaney, and J. Winter (2014): "Can survey participation alter household saving behavior?" IFS Working Papers W14/06, Institute for Fiscal Studies.

Dang, H.-A., P. Lanjouw, J. Luoto, and D. McKenzie (2014): "Using repeated cross-sections to explore movements into and out of poverty," *Journal of Development Economics*, 107, 112–128.

Das, M., V. Toepoel, and A. van Soest (2011): "Nonparametric tests of panel conditioning and attrition bias in panel surveys," *Sociological Methods & Research*, 40, 32–56.

Dearden, L., H. Reed, and J. Van Reenen (2006): "The impact of training on productivity and wages: Evidence from British panel data*," *Oxford Bulletin of Economics and Statistics*, 68, 397–421.

Deaton, A. (1985): "Panel data from time series of cross-sections," *Journal of econometrics*, 30, 109–126.

Disney, R., J. Haskel, and Y. Heden (2003): "Restructuring and productivity growth in uk manufacturing*," *The Economic Journal*, 113, 666–694.

Donovan, T. N. and G. Britain (1969): *Royal Commission on Trade Unions and Employers' Associations 1965-1968: Report presented to parliament by command of Her Majesty, June 1968*, HM Stationery Office.

Duncan, G. J. and S. D. Hoffman (1981): "The incidence and wage effects of overeducation," *Economics of Education Review*, 1, 75–86.

Eisner, R. (1967): "Capital and labor in production: some direct estimates," in *The Theory and Empirical Analysis of Production*, Columbia University Press, 431–476.

Famoye, F. (1993): "Restricted generalized Poisson regression model," *Communications in Statistics-Theory and Methods*, 22, 1335–1354.

Fang, T. (2009): "Workplace responses to vacancies and skill shortages in Canada," *International Journal of Manpower*, 30, 326–348.

Farias, J. C. and S. Ruano (2005): "Firm productivity, heterogeneity, sunk costs and market selection," *International Journal of Industrial Organization*, 23, 505–534.

Finegold, D. (1999): "Creating self-sustaining, high-skill ecosystems," *Oxford Review of Economic Policy*, 15, 60–81.

Finegold, D. and D. Soskice (1988): "The failure of training in Britain: analysis and prescription," *Oxford Review of Economic Policy*, 21–53.

Forth, J. and G. Mason (2004): *The impact of high-level skill shortages on firm-level performance: evidence from the UK Technical Graduates Employers Survey*, National Institute of Economic and Social Research.

——— (2006): "Do ICT skill shortages hamper firms performance? Evidence from UK benchmarking surveys," *National Institute of Economic and Social Research, Discussion Paper.*

Frogner, M. L. (2002): "Skills shortages," *Office for National Statistics, Labour Division: Special Feature. www.ons.gov.uk/ons/rel/lms/labour-market.../skills-shortages.pdf. Accessed 17/06/2012.*

Galindo-Rueda, F. and J. Haskel (2005): "Skills, workforce characteristics and firm-level productivity: Evidence from the matched ABI/Employer Skills Survey," IZA Discussion Papers 1542, Institute for the Study of Labor (IZA).

Glynn, S. and H. Gospel (1993): "Britain's low skill equilibrium: a problem of demand?" *Industrial Relations Journal*, 24, 112–125.

Goos, M. and A. Manning (2007): "Lousy and lovely jobs: The rising polarization of work in Britain," *The Review of Economics and Statistics*, 89, 118–133.

Gospel, H. (1998): "The revival of apprenticeship training in Britain?" *British Journal of Industrial Relations*, 36, 435–457.

Gould, W., J. Pitblado, and W. Sribney (2006): *Maximum likelihood estimation with Stata*, Stata Press.

Greene, W. H. (2003): *Econometric analysis*, Pearson Education India.

Griffith, R. (1999): "Using the ARD establishment level data to look at foreign ownership and productivity in the United Kingdom," *The Economic Journal*, 109, 416–442.

Griliches, Z. and J. Mairesse (1995): "Production functions: The search for identification," Tech. rep., National Bureau of Economic Research.

Groot, W. and H. Maassen van den Brink (2000): "Overeducation in the labor market: a meta-analysis," *Economics of Education Review*, 19, 149–158.

Harris, R. (2002): "Foreign ownership and productivity in the United Kingdomsome issues when using the ARD establishment level data," *Scottish Journal of Political Economy*, 49, 318–335.

Harris, R. and C. Robinson (2003): "Foreign ownership and productivity in the United Kingdom estimates for UK manufacturing using the ARD," *Review of Industrial Organization*, 22, 207–223.

Haskel, J. and C. Martin (1993): "Do skill shortages reduce productivity? Theory and evidence from the United Kingdom," *Economic Journal*, 103, 386–94.

Hausman, J. A. (1978): "Specification tests in econometrics," *Econometrica*, 46, 1251–71.

Heckman, J. J. (1979): "Sample selection bias as a specification error," *Econometrica*, 47, 153–61.

Heckman, J. J., R. J. LaLonde, and J. A. Smith (1999): "The economics and econometrics of active labor market programs," *Handbook of Labor Economics*, 3, 1865–2097.

Hilmer, M. J. (2001): "A comparison of alternative specifications of the college attendance equation with an extension to two-stage selectivity-correction models," *Economics of Education Review*, 20, 263–278.

Indicators, O. (2011): "Education at a glance 2011," *OECD Publishing.*

Jovanovic, B. (1982): "Selection and the evolution of Industry," *Econometrica: Journal of the Econometric Society*, 649–670.

Katz, L. F. and H. David (1999): "Changes in the wage structure and earnings inequality," *Handbook of labor economics*, 3, 1463–1555.

Klepper, S. (2002): "Firm survival and the evolution of oligopoly," *RAND Journal of Economics*, 33, 37–61.

Lambert, D. (1992): "Zero-inflated Poisson regression, with an application to defects in manufacturing," *Technometrics*, 34, 1–14.

Layard, R. and S. Nickell (1985): "The causes of British unemployment," *National Institute Economic Review*, 111, 62–85.

Leitch, S. (2006): *Prosperity for all in the global economy - World class skills: Final report*, The Stationery Office.

Leuven, E. and H. Oosterbeek (2011): "Overeducation and mismatch in the labor market," *Handbook of the Economics of Education*, 4, 283–326.

Levinsohn, J. and A. Petrin (2003): "Estimating production functions using inputs to control for unobservables," *The Review of Economic Studies*, 70, pp. 317–341.

Lizal, L. and K. Galuščák (2012): "The Impact of Capital Measurement Error Correction on Firm-Level Production Function Estimation," Tech. rep., William Davidson Institute at the University of Michigan.

Manning, A. (1993): "Wage bargaining and the phillips curve: The identification and specification of aggregate wage equations," *Economic Journal*, 103, 98–118.

Marschak, J. and W. Andrews (1944): "Random simultaneous equations and the theory of production," *Econometrica*, 12, pp. 143–205.

Martin, R. (2002): "Building the capital stock," *CeRiBA mimeograph; Notes*, 514, 1000.

Meager, N. (1986): "Skill shortages again and the UK economy," *Industrial Relations Journal*, 17, 236–248.

Mundlak, Y. (1961): "Empirical production function free of management bias," *Journal of Farm Economics*, 43, 44–56.

Olley, G. S. and A. Pakes (1996): "The dynamics of productivity in the telecommunications equipment industry," *Econometrica*, 64, 1263–97.

Redding, S. (1996): "The low-skill, low-quality trap: Strategic complementarities between human capital and R & D," *The Economic Journal*, 458–470.

Rosenbaum, P. R. and D. B. Rubin (1983): "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 70, 41–55.

Rubb, S. (2003): "Overeducation in the labor market: a comment and re-analysis of a meta-analysis," *Economics of Education Review*, 22, 621–629.

Rumberger, R. W. (1987): "The impact of surplus schooling on productivity and earnings," *Journal of Human Resources*, 22, 24–50.

Sicherman, N. (1991): ""Overeducation" in the labor market," *Journal of Labor Economics*, 9, 101–22.

Steedman, H., G. Mason, and K. Wagner (1991): "Intermediate skills in the workplace: deployment, standards and supply in Britain, France and Germany," *National Institute Economic Review*, 136, 60–76.

Stigler, G. J. (1958): "The economies of scale," *Journal of Law and Economics*, 1, pp. 54–71.

Wallis, G. (2003): "The effect of skill shortages on unemployment and real wage growth: A simultaneous equation approach," Royal Economic Society Annual Conference 2003 217, Royal Economic Society.

Wilson, R. A. and T. Hogarth (2003): "Tackling the low skills equilibrium: A review of issues and some new evidence," *Department of Trade and Industry*.

Winkelmann, R. and K. F. Zimmermann (1994): "Count data models for demographic data," *Mathematical Population Studies*, 4, 205–221.

Wooldridge, J. M. (2010): *Econometric analysis of cross section and panel data*, vol. 1 of *MIT Press Books*, The MIT Press.

Zellner, A. (1962): "An efficient method of estimating seemingly unrelated regression equations and tests for aggregation bias," *Journal of the American Statistical Association*, 57, 348–368.

Zwane, A. P., J. Zinman, E. Van Dusen, W. Pariente, C. Null, E. Miguel, M. Kremer, D. S. Karlan, R. Hornbeck, and X. Giné (2011): "Being surveyed can change later behavior and related parameter estimates," *Proceedings of the National Academy of Sciences*, 108, 1821–1826.