Essays on the demand and supply of labour

Wenchao Jin

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Department of Economics

University College London

March 3, 2022

I, Wenchao Jin, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work. Chapter 2 is based on joint work with Richard Blundell and David Green, which was published in Blundell et al. (2021). Everything else is my own work.

Statement of joint work

All chapters except Chapter 2 are sole-authored. Chapter 2 is based on joint work with Richard Blundell and David Green, which has been published recently as Blundell et al. (2021). All authors contributed equally to the development of ideas and research strategies. I carried out most of the data analysis.

Abstract

This thesis contains three chapters analysing labour supply and labour demand, and their interaction.

The first two chapters examine the labour market consequences of the rapid expansion of Higher Education in the UK since the 90s. The first chapter looks at the wages of graduates and non-graduates. It proposes a model of endogenous adoption of skill-biased technology to explain the remarkably flat trend of the college wage premium. The second chapter extends the theory by adding the occupational dimension and uses it to study quantitatively the historical phenomenon of job polarisation. It shows a large shift in employment from middle-skill occupations to the higher-skilled ones in the UK. It develops a multi-sector equilibrium model of occupational labour, featuring endogenous adoption of task-biased technology. This can explain not only the employment patterns, but also movements in occupational wage, and occupational trends within education groups.

The third chapter investigates the low employment rate of older women in urban China. Based on a wide range of descriptive correlations, I argue that there are two main reasons for the low employment rate. One is the early age at which urban women become eligible for pensions. One is financial transfers from their children. I build and calibrate a life-cycle model of female labour supply, incorporating income uncertainties and income-contingent transfers from children. The model can generate realistic patterns of employment, including a lot of bunching in the timing of labour market exit at the point of pension receipt. I use the model to simulate the effects of increasing pension age.

Acknowledgements

For most of my PhD, I have been funded by the Economic and Social Research Council scholarship through the Institute for Fiscal Studies. I am grateful for the funding extensions as a result of two spells of maternity leave and Covid. For some of the years, I have worked as a Research Economist at the IFS on a part-time basis. I have benefited a great deal from this both financially and intellectually.

I am indebted to my supervisors Professor Richard Blundell and Dr Suphanit Piyapromdee for their guidance and support. Thank you for giving me constructive comments on my work, encouraging me during periods of frustration, and supporting my professional development in general. I would also like to thank other academics at the UCL department, especially Uta Schönberg and Valérie Lechene, for giving me feedback on my research and career advice.

I have been privileged to work with many excellent co-authors and IFS colleagues over the years. Thank you to Professor David Green for working with me on Chapter 2, hosting me for a visit to the University of British Columbia, and giving me helpful comments on my job market paper. During the PhD, I have also learned a great deal from collaborations with Rachel Griffith, Valérie Lechene, Lars Nesheim, and Pedro Carneiro on other research projects. Thank you all for your kindness, support and patience while I worked on the PhD. I would also like to thank all my IFS colleagues for inspirational discussions and friendly chats.

Last but certainly not least, I would like to thank my husband Tian and my children Felix and Ivy for giving me love, perspectives and meaning.

Impact statement

The findings of this thesis are of interest to both academics and the general public. The findings in Chapter 2 are already published in the Review of Economic Studies (Blundell et al., 2021). An earlier version was made public as an IFS working paper in 2016, together with a short briefing note summarising the key findings. They were cited widely in the media, including on FT and BBC websites on 18 August 2016. The other two chapters have been presented or accepted at international conferences (SOLE 2022, EALE 2022 and various IZA workshops). I intend to revise and submit them to academic journals in the near future.

Chapters 2 and 3 contribute to the broader intellectual debate about technical change and wage inequality. Many developed countries have seen rising inequality in labour income since the 80s. It is important to understand to what extent it is caused by technological progress, and whether investment in the education system can alleviate it. While the empirical analysis in Chapters 2 and 3 is mostly based on UK data, this thesis has documented some descriptive facts for other European countries which are quite similar to the UK. Hence, the models proposed in this thesis may be applicable to the study of inequality and other labour market trends in European countries.

Chapter 4 contributes to the smaller academic literature on pension policies in China. Although this is an urgent public policy matter that would affect hundreds of millions of people, rigorous quantitative studies are rare. It is my hope that this chapter will be the basis for more in-depth economic studies of the Chinese pension system and realistic modelling of related household behaviours. Beyond the Chinese context, the chapter may be interesting to other academics who study retirement behaviour. In particular, Chapter 4 documents a lot of bunching in labour market exit at the exact point of pension receipt. This behaviour has been documented in other countries as well, and often attributed to irrationality.

Meanwhile, all three chapters in the thesis have direct relevance for public policies. Chapter 2 and Chapter 3 have direct relevance for policies on higher education expansion in the UK. The model developed in Chapter 3 can also be used to simulate the impact of any other policy that shifts the supply of skills. One notable example is the selection of immigrants based on their skills. Recent and future immigrants from Europe will likely be more skilled than those that came to the UK under the single market. This calls for a structural model like the one proposed in Chapter 3.

Chapter 4 is the first paper to use a life-cycle model to simulate the impact of increasing pension age on female labour supply in Chinese cities, to the best of my knowledge. Due to the rapid ageing of the population, increasing pension age has been under consideration of the Chinese government for some time now. For policy purpose, it would be important to know to what extent the affected individuals will delay their labour market exit and the spillover effects on their family members.

Contents

List of Figures

List of Tables

Chapter 1

Introductory Material

How does government policy affect the quantity and quality of labour supplied by individuals? How does a skill supply shift affect employment and wages across different groups, and how does it interact with technology-induced demand shifts?

This thesis examines these questions in three chapters, using micro-data on households in the UK and China.

Chapter 2 and Chapter 3 examine the labour market consequences from the large expansion of higher education in the UK. Between the early 90s and the mid-2010s, the proportion of the UK population with a Bachelor's degree or above has tripled. This was mostly driven by government policy, as the number of undergraduates places was capped by the government for most of the period. Both policymakers and the public are concerned that graduates would increasingly end up in lower-paid lower-skilled jobs. Chapter 2 and chapter 3 explain why this has not been the case.

Chapter 2 looks closely at the wages of graduates and non-graduates. It investigates the variation in the proportion of graduates both over time and the variation across age groups and region. We found that the supply shift has essentially zero impact on the skilled wage premium. This is surprising because in standard economic models, an increase in the supply of one input factor would typically reduce its price.

Chapter 2 proposes a model where firms choose between two technologies, and the choice depends on the skilled and unskilled wages. This means that a shift in skill supply would cause an endogenous shift of technology, and during the transition phase, the skilled wage and the unskilled wage would remain constant. These implications for wages are different from other mainstream economic models about skill-biased technical change.

In models with exogenous technical change, the skilled wage premium reflects the race between technology and education (Goldin and Katz, 2008), where technology does not respond to educational shifts. Thus, an increase in the supply of graduates will reduce their relative wage, and for the observed trend to be flat, it would require the exogenous skill-biased technical change to speed up and slow down precisely to offset the supply shift. While this is not impossible, it is clearly less plausible than the model of endogenous adoption of technology, which implies no impact of the skill shift on the relative wages.

The third type of models for technical change allows the inventions of new technologies to respond to changes in the supply of different input factors (Acemoglu (1998), Kiley (1999)). In such models, a policy-driven supply increase of skilled labour would lead to more investment and innovations in skilled-biased technology, which in the longer run, will alleviate or even offset the negative impact on the skilled wage premium. However, in such models, the skilled wage premium still falls in the short run, and the sign of impact in the longer run is ambiguous.

Chapter 2 uses a general empirical framework that nests all three types of models of technical change to formally distinguish between them. The analysis uses variation in skill supply across region-age groups, instrumented with the local population's birth cohort composition. The results reject both the model with exogenous technical change and the one with endogenous innovations, and are consistent with the model of endogenous adoption of technology.

The concept of technology is abstract and somewhat intangible in Chapter 2. It hypothesises that the new skill-complementary technology features a more decentralised way of decision-making in the organisation. This is supported by the analysis of worker autonomy in the Workplace Employment Relations Survey (WERS). In subsection 2.5.2, employees' reports of how much influence they have in the workplace are summarised into an influence index, and regressions show that the local supply of graduates significantly and positively predicts the influence index. This correlation is in fact stronger for non graduates's influence index than for graduates. This is consistent with the idea that decision-making power was concentrated among a minority of workers (mostly skilled), and after switching to the decentralised form of organisation, less skilled workers now gain more influence.

Chapter 3 continues the theme of endogenous adoption of technology, but applies it to the context of occupational polarisation. It refers to the phenomenon where middle-paying occupations are declining as a share of aggregate employment, while both higher-paying and lower-paying occupations grow. This phenomenon has been documented for many developed countries since at least the 90s and has been the subject of a large literature. It is often interpreted as a result of Routine Biased Technical Change (RBTC). In Chapter 3, I propose a different yet complementary explanation that emphasises the increasing supply of skills and the endogenous adoption of task-biased technology. This explanation is not only consistent with the observed trends in occupational employment, but also two additional facts about occupations in the UK laid out below.

First, the UK pattern of polarisation is essentially a shift from the middle to the top. As a share of aggregate employment, middle-paying occupations (admin and secretary, skilled trades, and machine operatives) have shrunk by more than 10 percentage points between the mid 90s and mid 2010s. At the same time, the share of higher-paid abstract occupations (managerial, professional and technical) increased by about the same amount, so the share of lower-paid occupations has seen little change. In my framework, the employment trends reflect a shift in the technology adopted by firms, towards one that's less intensive in routine tasks and more intensive in abstract tasks. This shift in technology is a result of the increasing supply of skilled labour in the UK economy.

Second, there has been no wage polarisation in the UK. Subsection 3.2 shows the movements in occupational wages are small and uncorrelated with those in occupational employment in the UK. The lack of wage polarisation does not directly support the hypothesis of exogenous Routine Biased Technical Change (RBTC). In my model, new technology is adopted endogenously in response to the supply-side shift, so there's less impact on occupational wages.

Third, the huge increase in educational attainment since the early 90s has not resulted in much occupational downgrading among UK graduates. The proportion of graduates who work in abstract occupations has been stable around 80%, over a period when the share of graduates in the workforce more than doubled. In other words, the increase of graduates was quickly absorbed through employment growth in abstract occupations. Again, this is consistent with the model of endogenous task-biased technical change proposed in Chapter 3,

Specifically, I build a multi-sector equilibrium model of demand and supply of occupational labour. The main innovation relative to the polarisation literature is that there is a choice of technology within each industry and this choice depends on task prices.[1] This new feature makes the demand side very elastic. When the skills distribution shifts in favour of a certain task, firms may switch to the technology that's more intensive in that task, thus the resulting impact on prices and wages will be smaller than if technology is exogenous. In other words, the endogenous adoption of technology helps absorb supply-side shocks, so the effects are seen in the relative quantities of tasks rather than prices. Thus, it simultaneously explains the two facts we described earlier: 1) the employment shift from the middle to the top, and 2) the lack of relative wage changes.

[1] In this chapter, I use 'task' as a short hand for occupational labour.

On the supply side, my model implies that the probabilistic mapping from skills to occupation is just a function of task prices. Given the lack of change in task prices, the mapping will not change much either. As long as the skills composition within education groups does not deteriorate too much, the mapping from observed education to occupation won't change much.

In both chapter 2 and chapter 3, I compare the UK to the US and other European countries where possible. Many European countries share the key facts observed in the UK since the 90s, and are more different to the US. Appendix A.9 shows that among the European countries that saw large increases in higher education, the majority did not see a significant impact on graduates' relative wage. In terms of occupations, employment growth has been strongest in high-paid occupations in most European countries, compared to both the middle and the bottom. Finally, wage polarisation has only been observed in the US, and not in other developed countries that saw job polarisation. All these facts observed for non-US countries are consistent with the idea that there are multiple available technology options to choose from and the choice depends on the supply of skills.

Conceptually, the applicability of different models of technical change depends on the country's position relative to the technology frontier and its contribution to the global advancement of technology. A model of endogenous technology adoption is more suitable for smaller economies that are close to the technology frontier, where firms have access to a range of technologies. In rare historical episodes, major revolutionary technologies become available and clearly dominate the pre-existing technology at any relevant factor prices. In those cases, exogenous technical change could be a reasonable assumption. Meanwhile, models of endogenous innovations may be more suitable for a large technological leader like the US.

While chapter 2 and 3 examine consequences of an increasing supply of graduates, chapter 4 investigates the labour supply behaviour of another population: females near retirement age in Chinese cities.

In China, the employment rate among middle-aged and older urban female residents is exceptionally low. For example, 27% of 55-64-year-old urban women were in work in 2013, compared to over 50% in UK. This is worrying in a country with rapid population ageing and makes the public pension system unsustainable[2]. Chapter 4 examines why women in urban China leave the labour force early, and simulates the impact of increasing their pension age.

The first explanation is the relatively low pension age enjoyed by the group. For urban female residents, the formal retirement age is 50 for most workers and 55 for female "cadres" or managers. In chapter 4, I document a striking coincidence in timing between pension receipt and exiting the labour force. About 70% current pensioners aged 60 or above stopped working in the same year that they became eligible for a retirement pension. This is a bit puzzling, because all the pensions are not means-tested or carry any disincentive for work. The effective marginal tax rate on labour income is actually lower for pension recipients, because they don't need to pay into the social security schemes any more. Thus, in a standard life-cycle model without jumps in preferences or wages at the point of pension receipt, there should not be a large concentration of people leaving the labour force at that exact time, unless liquidity constraints are prevalent. However, we also know that the majority of the population of interest have positive saving rates.

The second reason for the low employment rate among older females is financial transfers from their adult children. Subsection 4.3.2 shows that about 70% of women received financial support from their adult children in the 12 months prior to the survey, conditional on having children living outside the home.[3] In response to the question "Who can you most rely on for old-age support?", the most common responses are: 1

[2] It has been projected to have a financing gap equivalent to 95% GDP for the period 2001-2075 (Sin, 2005).

[3] We only observe transfers between parents and non-coresident children in CHARLS.

"their pensions", 2 "their children", and 3 their own savings. Thus, (expected) transfers from children have an important wealth effect. I also show that the transfers from adult children are negatively correlated with the parents' income, suggesting that it may have an insurance role as well.

In section 4.4, I build a life-cycle model of labour supply and saving behaviour and calibrate it using two Chinese household surveys. The model incorporates a search cost to return to work after reaching pension age, uncertainties in pension income, income-contingent transfers from adult children, and liquidity constraints. Together, these features create a large amount of bunching in the timing of labour market exits. Overall, the model generates realistic patterns of labour supply behaviour. I simulate the labour supply response to different pension reform scenarios. In the counterfactual of raising pension age to 60 for all, for example, the employment rate among 50-59 year old women would be higher by 28 percentage points, compared with the current baseline.

The remainder of this thesis is structured as follows. Chapter 2 analyses the impact of higher education expansion on graduate wages in the UK. Chapter 3 examines occupational polarisation, highlighting the supply shift as an important cause. Chapter 4 studies the labour supply behaviour of Chinese women near retirement age. Chapter 5 concludes and sets out directions for future research.

Chapter 2

Higher education expansion and the graduate wage premium

2.1 Introduction

In the period extending from the early 1990s to the present, the UK economy experienced a dramatic transformation in educational attainment. Specifically, in 1993, 11% of the population held a university degree. This percentage has doubled by 2006 and tripled by 2016. In this chapter we examine the impact of that increase on the UK labour market, using our findings as a basis for contributing to the ongoing discussion about the interaction of educational attainment and technological change.

There is a strong consensus among economists that the Information Technology (IT) revolution has played a central role in determining wage and employment outcomes in many economies in the last four decades and that the effects of education should be viewed in conjunction with that revolution (see Acemoglu and Autor (2011) for a comprehensive review of the literature on these topics, which started with Katz and Murphy (1992)). Most famously, changes in the wage distribution have been described as a race between skill-biased demand shifts emanating from IT innovations and increases in skills generated by changes in education levels (Goldin and Katz, 2008). The core idea underlying this consensus is that the new technologies are complementary with skills. Intertwined with the broad literature on the effects of technology on the wage structure in general is a literature on skills, IT, and the organizational structure of the firm (e.g., Bresnahan et al. (2002); Caroli and Van Reenen (2001); Bloom et al. (2014), which build on Becker and Murphy (1992) and Radner (1993) among others). This literature seeks to look inside the 'black box' of the firm to understand how skills complement IT. Its main message is that IT, by altering information flows and communications within firms, implies a shift in the optimal organization of the firm toward a form that is more decentralized and flexible. Decisions, information transfer, and co-ordination of tasks happen throughout the organization instead of through top-down direction as in the previous, Taylorist form - the form in which tasks are broken into small sub-components with central direction. The shift in organizational form is the channel through which more educated workers benefit from the broad technological change since human capital investment gives workers greater ability to deal with increased change and decision making and makes them relatively more productive in the new environment. On the other side, a large literature on polarization argues that IT replaces routine tasks to the detriment of less educated workers (Acemoglu and Autor (2011)). Our approach, both theoretically and empirically, incorporates both the decentralization and routinization elements of the IT revolution and how they intersect with educational change.

Much of the empirical work and the clear majority of the theorising on the interaction between education and technological change has been done on the US economy. However, there are good reasons to believe that the US is the technological leader in this period and, because of that, may exhibit special relationships between technological change and education that do not apply even to other advanced economies. Given this, we view the UK educational expansion as an opportunity to study the relationship between education and technological change in a technological follower, as we believe the UK has been in terms of skill biased technologies and firm organizational forms. We will argue that taking this perspective has an impact on which model of technological change and education one adopts. In particular, we argue for a model in which firms choose among existing technologies rather than one with new invention. Our claim is that a technological choice model provides a natural explanation for a remarkable fact for the UK: that its very substantial increase in education level was accompanied by a complete lack of change in the university-high school wage differential. We present a model that captures this fact but also has further testable implications that we show are supported in the data.

Our key message is that technological change is not one size fits all. Many papers look for evidence of the importance of technological change in common movements in the relation between educational attainment and wage differentials across countries. The argument being that if new technologies are accessible in all developed countries then, conditional on mediation through relative skill supply shifts, it should act as a common force showing up in the same way in all developed countries. In contrast, differential movements in the combination of educational attainment and skill based wage differentials across countries is taken as evidence of the impact of other, non-technological factors (e.g., Caroli and Van Reenen (2001)'s examination of French and English data or Antoncyzk et al. (2010)'s assessment of education and wage movements in the Germany and the US). In contrast, we argue that the same changes in factor supplies interacting with the same technology can dictate quite different wage outcomes for two countries depending on whether they are leaders or followers in the adoption of that technology.

The paper proceeds in six sections including the introduction. In section 2.2, we establish the core patterns for the UK, relying largely on Labour Force Survey (LFS) data between 1993 and 2016. We show that, despite a rapid increase in the proportion of university graduates, the college wage premium is flat across our time period.[1] We

[1] Note that the period investigated in the paper is after the period when the college wage premium in the UK increased substantially (Machin and McNally, 2007).

demonstrate the robustness of the two findings: they cannot be explained as, for example, declines in the actual wage differential that are masked by changes in composition. We consider compositional changes related to increases in the female participation rate, the shift toward more advanced university degrees over time, the difference between the public sector and the private sector, and the substantial increase in immigration. We also consider changes in unobserved abilities, using a bounding approach. None of these exercises alters the core result that the education wage differential was essentially unchanged during a period of rapid educational growth.

The combination of an increase in the supply of education and no change in the educational differential points to an offsetting relative demand shift favouring more educated workers. Such a shift has, of course, been the focus of considerable investigation, with a common conclusion that technological change associated with the IT revolution has been a key driving force. In section 2.3, we investigate competing models of technological change: the canonical model of exogenous skill biased technological change; models in which increases in education induce skill biased invention; and models of technological choice, in which firms choose among existing technologies. To test among the models, we employ wage and relative wage regressions derived from a general production function that nests all three possibilities. Based on estimates of those regressions, we argue that a model with exogenous technological change (either in its classic form or in a task based form) cannot explain the skill and wage patterns in the UK data. In particular, in the context of those models our estimates would imply that skilled and unskilled labour are nearly perfect substitutes and that there has been no exogenous skill biased demand shift, neither of which seems reasonable. This echoes previous papers that conclude that the canonical model also does not fit more recent US data (Beaudry and Green (2005); Card and DiNardo (2002); Acemoglu and Autor (2011)). The implied substitution patterns are also relevant for empirical specifications derived from the endogenous invention model when holding technological change constant. For those specifications as well, the findings of near perfect substitutability and no ongoing skill bias to demand shifts do not match the model. Implications from the endogenous invention model when not holding technological change constant also do not fit with the UK data patterns.

We also present evidence that the US is a strong candidate for being the technological leader where the new skill biased inventions were made. It had both a higher level of education and a higher amount of investment in IT before any other developed economies. The UK, on the other hand was a laggard in educational attainment. We argue that once the UK did start to increase the educational level of its workforce, its firms could choose to pick up the technologies and organisational forms already developed in the US. In that sense, it is more natural to think of the UK in the context of the third type of model: a model of endogenous technological choice.

In section 2.4, we set out a model of endogenous technological choice which has the ability to capture the core data patterns and the results of our estimation. The model is a variant of models in Rosen (1978) and Borghans and ter Weel (2006) which focuses on the role of decentralization of decisions and information. It is also related to the model of endogenous technological choice in Beaudry and Green (2003). Firms use skilled and unskilled labour and choose between an older, centralized mode of operation and a newer, decentralized mode. The model endogenously generates an unchanging college wage premium. This was the point of using this type of model, and so that outcome provides no proof of the model's relevance. However, the model also generates testable added implications about the effects of the skill supply shift on wage levels (not just wage ratios) and about the pattern of employment in manager positions over time.

We examine these empirical implications of our model in section 2.5. In that section, we also investigate further implications by examining the relationship between the educational composition of the workforce and the extent to which workers feel they control how they do their own work using matched worker-workplace survey data from the UK Workplace and Employer (WERS) data. We show that the areas where the increases in the BA proportion were largest had the greatest uptake of decentralized organizational forms. We establish that this is a causal relationship using an IV strategy using a combination of parental education and the population share of the birth cohorts most affected by the educational increase, measured in 1995 (i.e., before the entry of the most affected cohorts into the labour force). We view this as a credible strategy since the validity of the instrument just requires that differences in fertility rates across areas were not driven by changes in firm organizational forms twenty years later. Thus, the data fits with a model in which increased educational attainment induces more and more firms to choose a decentralized organizational form. One interesting implication of the model that is confirmed in this data is that increases in education levels in an area induce larger increases in individual decision making among less educated than among more educated workers. This arises because under the old, centralized technology, more educated workers were disproportionately managers and were already making their own decisions. It is for the less educated that decentralization is a particularly big revolution.

In section 2.5, we also briefly provide evidence that several other developed and developing economies in this period also experienced a combination of a rapid increase in educational attainment with little change in the education wage differential. That is, in our terms, the UK was not the only technological follower. Section 2.6 concludes.

We are not the first researchers to note the substantial increase in degree-holding in the UK. For example, Carpentier (2004) documented the trend in student numbers from 1920 to 2002, showing that it increased sharply around the early 90s. He also showed a reduction in university expenditure per student around the same time. Many other studies have also documented the substantial increase in the share of graduates in the 1990s or across cohorts (OLeary and Sloane, 2005; Walker and Zhu, 2008; Green and Zhu, 2010; Devereux and Fan, 2011).

Previous papers have also noted the lack of a reduction in the college wage premium over time or across recent UK cohorts (Machin and McNally (2007); McIntosh (2006); Walker and Zhu (2008)). However, those papers either appeal to offsetting relative demand shifts stemming from exogenous skill biased technical change or do not attempt to explain the lack of change in the relative wages at all. We add to the previous literature, in part, by providing an explanation that does not rely on exogenous skill biased demand shifts that just happen to be the right size to match the change in educational attainment across a range of years. Instead, we present a model in which this pattern arises endogenously, which has ramifications for how we think about the interactions of technological change, factor supplies, and factor demand. We also differ from earlier studies in our explicit emphasis on the firm organization part of the process - that is where our empirical work focuses. Combined, these give us new insights into how technological change affects economies. Overall, we view studying the UK as an opportunity to examine the impact of education policy on technological adoption and, through it, on wages in the situation that is likely relevant for most countries - being a technological follower.

2.2 Data and core patterns

2.2.1 Data

Our main empirical work is based on the demographic, education, employment, wage, and occupation variables in the UK Labour Force Survey (LFS). The LFS is a representative quarterly survey of approximately 100,000 adults that is the basis for UK labour force statistics. It is similar in nature to the US Current Population Survey (CPS) which we use as a comparison. We make use of UK LFS data running from the first quarter of 1993 to the last quarter of 2016.

Consistent definitions of education levels over time are obviously important in our investigations. The LFS asks respondents about their highest level of educational qualification, with the potential categories changing over time. We take advantage of detailed responses to construct six more aggregate categories that are consistent over time. For our main discussion, we then further aggregate those categories into three broader groups: a university degree level or above; secondary or some tertiary education below a university degree level; and below secondary qualifications. We draw the bottom line of secondary education as Grade C in the General Certificate of Secondary Education (GCSE), which are exams that students take at age 16 after 11 years of formal schooling. The GCSEs mark the first major point of exit from education in England: around one fifth of the working-age population have GCSEs Grade C or above or equivalents as their highest level of qualification in 2016. We consider a grade of at least C to be equivalent to High School graduation (HS) in the US because the proportion of people strictly below the threshold in the UK is close to the proportion of HS drop-outs in the US.[2] Under UNESCO's International Standard Classification of Education (ISCED 2011), both US High School Diploma and UK's GCSE Grade C or above fall into ISCED level 3 "upper secondary education". We have investigated alternative definitions of education groups and they make little difference to our main results.[3] We restrict our samples to people between ages 20 to 59 because the education qualification question was not asked of people over age 60 before 2007 unless they were working at the time of the survey.

Wages are surveyed in the first and fifth quarters an individual is in the survey. We use the hourly wage derived from the weekly wage in the main job and actual weekly hours. Our sample contains 30,000-75,000 wage observations per year. As we are interested in the real cost of labour to firms, we deflate wages by the GDP deflator[4] .

[2]For example, 10.6% of 25-34 year olds in the US are HS drop-outs in 2012. Coincidentally, the proportion of this age group in the UK who do not have qualifications equivalent to or higher than GCSE grade C is also 10.6% ; and 19.8% have qualifications equivalent to GCSE grade C and no higher qualifications.

[3]These are reported in the Online Appendix. One particular alternative we have tried is to define the UK HS group by A-levels instead of GCSEs. A-levels are typically taken at age 18 and are required for university admission.

[4]Source: we use the variable 'GDP: Total implicit price deflator' in the dataset called 'MEI Original Release Data and Revisions' from OECD.stats.

In places, we use the U.S. CPS to form a comparison. We again use individuals aged 20 to 59. The data is from the Outgoing Rotation Group samples. Following Lemieux (2006), we do not use observations with imputed wages when calculating wage statistics. Wages and employment status refer to the week prior to the survey week, and we only use wage and occupation data for individuals who are employed in the reference week. We aggregate the U.S. workers into three education groups: high school drop-outs; high school graduates (which includes workers with some or completed post-secondary education below a Bachelor's degree); and university degree holders (Bachelors and higher).

2.2.2 UK wage and educational attainment movements

2.2.2.1 Changes in educational attainment

We begin with a figure showing the level of university attainment over time for the UK, with the US as a benchmark. We will use the shorthand of calling the group with university degrees BA's, even though it includes other types of Bachelors degrees and more advanced degrees. For both the US and the UK, we summarize the data by plotting year effects from an exercise in which we first calculate the BA proportion for the set of cells defined by year and 5-year wide age ranges then regress those proportions on a complete set of year and age range dummies. We control for age in this way because we are concerned that the movement of the baby boom through the age structure will affect our BA proportion measure.

Figure 2.1 contains plots of the year effects for the BA proportion for both the UK and the US. The figure includes year effects from the General Household Survey (GHS) for the UK for the years before 1993 along with the same proportions from the LFS starting in 1984.[5] The sample sizes for the GHS are small, especially for the more educated, so we don't use it in our main analysis, but it does provide longer term context

[5]The LFS underwent significant changes in 1984 and in 1992. Before 1984, it was a bi-annual survey. From 1984 to 1991 it was annual. From 1992Q2 onwards, it was quarterly.

for the LFS data patterns. For the overlapping years, both of the UK datasets show a gradually increasing trend, although the level differs. As shown in Figure 2.1, the BA proportion in the UK showed a gradual increase in the 1970's and 1980's but it was still only about 0.13 in 1990, half of the value for the US in that year. Beginning around 1993, however, the UK proportion underwent a rapid acceleration. By 2010, it had surpassed the US.[6]

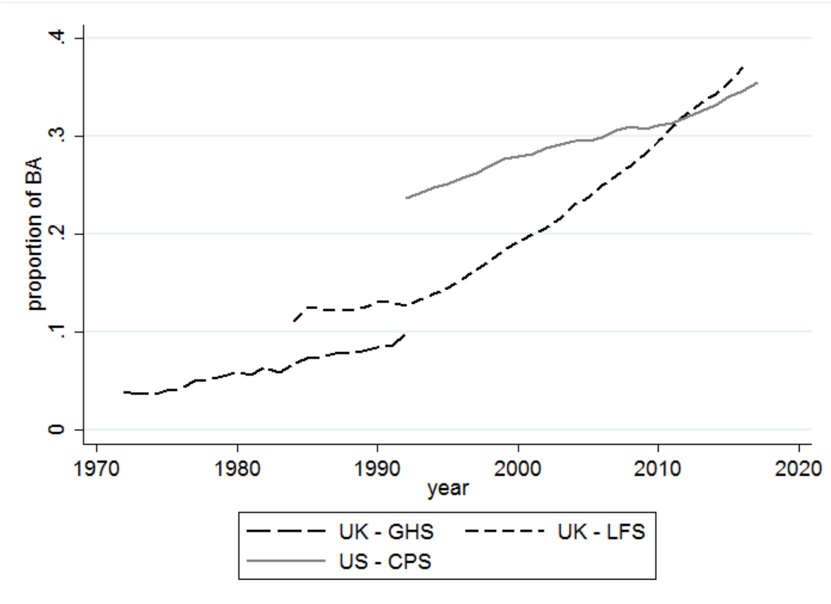


Figure 2.1: Proportion BA for the UK and US

Notes: BA refers to individuals who have a bachelors or higher degree. We aggregate each dataset to the level of year and 5-year age band, and regress the BA proportion on year dummies and age-band dummies. The proportion BA numbers are year effects from these regressions plus the level in 1992 for the 30-34 age band.
Source: Authors' calculation from the UK Labour Force Survey, the UK General Household Survey, and the US Current Population Survey.

The big increase in the UK proportion in the BA group starting in the mid-1990s corresponds to a rapid increase in higher education enrolment from 1988 to 1994. This increase has been documented in many studies (OLeary and Sloane (2005); Carpentier

[6]The rate of increase and the catch-up to the US is even clearer when the data is plotted by birth cohort (Appendix A).

(2006); Walker and Zhu (2008); Green and Zhu (2010); Devereux and Fan (2011)) and has been used as an arguably exogenous source of variation in studies of the causal impact of education (Devereux and Fan, 2011). The expansion of higher education over these decades reflects a sequence of specific policy choices made by the UK government. Further details are provided in the online appendix.

2.2.2.2 Changes in relative wages

The second main pattern relates to wages. In Figure 2.2 we plot the ratio of BA to high school median hourly wage by year for the UK. We will refer to this ratio as the college wage premium. As with the BA proportion, the plot corresponds to year effects from a regression in which age is held constant.[7] The striking point in this figure is its flatness. Over the span of years from 1993 to 2016, the wage ratio shows only minor fluctuations around a flat line. The absence of significant changes to the relative wages is consistent with previous studies which found the UK graduate wage premium to be stable in the 90s and early 2000s (Chevalier et al., 2004; McIntosh, 2006; Machin and Vignoles, 2006; Machin and McNally, 2007; Walker and Zhu, 2008).[8] The flatness of the ratio seems to us to be striking in light of the near tripling of the proportion of the working age population with a BA over this same period. Our goal in this chapter is to provide an explanation for this pair of patterns.

2.2.3 The Effects of composition shifts on the core patterns

One possible explanation for why such substantial increases in educational attainment were associated with little or no change in educational wage differentials is that compo-

[7] In Online Appendix, we present the college wage premium over the life-cycle by birth cohort. The differential is increasing over age in a concave pattern for each cohort. Because of this life-cycle pattern, one would expect the education wage ratio for the economy as a whole to increase as the population in our 20-59 sample ages, due to the baby boomers getting older. Holding age constant allows us to look past these composition related changes to the underlying wage changes.

[8] Two earlier papers OLeary and Sloane (2005); Walker and Zhu (2005), using data up to 2003, found the university premium to have fallen somewhat over the cohorts that experienced the higher education expansion. However, the authors later revised their cohort conclusions with more years of LFS data in Walker and Zhu (2008).

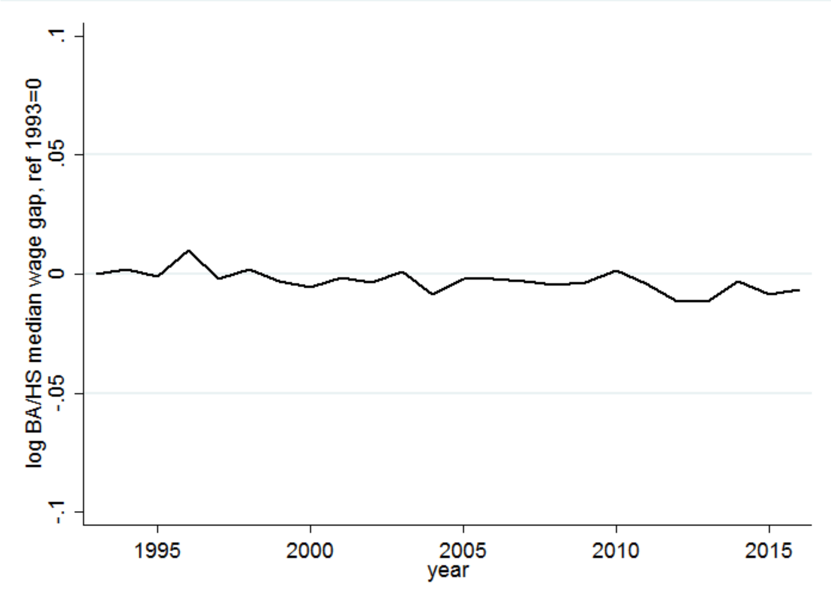


Figure 2.2: Ratio of BA median wage to that of high-school graduates 1993-2016, UK

Notes: Wage is hourly. The sample is 20-59 year olds in LFS 1993-2016. BA refers to individuals who have a bachelors or higher degree. We aggregate LFS to the level of year and 5-year age groups, and regress the log BA to HS median wage ratio on year dummies and age-band dummies. The figure plots the estimated year effects normalized to zero in 1993.

sitional shifts are obscuring the true patterns. To see this, it is helpful to think of workers as bundles of efficiency units of tasks. More able workers supply a larger number of efficiency units per hour worked, and, in a standard neoclassical model, their observed wages will reflect this. As a result, observed average wages can increase either because of increases in the market price per efficiency unit or because the composition of workers shifts in the direction of a higher average number of efficiency units per worker. Since our result is that the observed college wage premium has not fallen as we might expect, the scenario of greatest potential interest is one in which the price differential for BA versus HS tasks declines while the differential in average efficiency units between BA and HS workers increases.

Figure 2.3: UK year effects on BA proportion and college wage premium by gender

Note: The year effects use the same sample selection and regression specification as for Figure 2.2 but here are separated by gender.

2.2.3.1 Observable characteristic composition

Perhaps the most obvious compositional shift in terms of observable worker characteristics is related to the increase in female labour force participation. If the added female entrants with BA's are successively more able (compared to the added HS females) then their entry could hide a decline in the education differential in prices per efficiency unit. However, even the most cursory glance at the data indicates that gender composition shifts are not a source of problems since the wage patterns are the same for males and females. In Figure 2.3 we plot the Proportion of BA's and the college wage premium for males and females separately (again, obtaining year effects from regressions including age polynomials). For both genders, we see the dramatic increase in the BA proportion after 1993, with a faster increase for females. The wage differential remains flat over time for each gender, with each series showing nearly identical values for the differential in 1993 and 2016, and so a change in weighting between men and women would not alter the overall wage picture.

In Online Appendix, we present several further exercises. First, we consider the increase in the proportion of university degree holders with post-graduate degrees. We show that replotting the wage line in figure 2.2 including and not including workers with post-graduate degrees among the BA's does not change the main pattern: both lines show nearly identical values in 1993 and 2016. This is a reflection of the fact that, while the proportion of workers with a postgraduate degree increased rapidly, the proportion of university graduates with these degrees was still small at the end of our period. Second, we consider immigration as another potential source of compositional change since the proportion of UK workers born outside the UK doubled over the past two decades and immigrants' returns to education are lower than those of the native born (Dustmann et al. (2013)). However, the combination of strong increases in education with no accompanying changes in the college wage premium is present even if we look at the UK nationals alone, implying that composition changes related to immigration are not driving our main patterns. We also break the data down into public versus private sector employment and wages. Over the sample period, the public sector's employment share has remained around 25%. Both sectors saw very large increases in the BA proportion, with somewhat faster increases in the private sector. Both sectors again experienced relatively flat movements in the college wage premium, though the private sector trend is slightly more negative (amounting to about a 3% decline over the period from 1994 to 2016 as shown in Figure 8 in the online appendix.)

Overall, we conclude that shifts in composition with respect to observable worker characteristics cannot explain our main pattern of substantial education increases paired with an invariant education wage premium.

2.2.3.2 Unobservable characteristic composition

It is still possible, of course, that changes in the composition of unobservable characteristics has shifted across education groups in a way that could explain the wage patterns. As higher education expands, it draws in pupils from a wider and wider range of prior attainment and perhaps innate ability. The expansion of university education in the UK after 1988 came with a fall in per student resources and was accomplished in part by transforming polytechnic institutions into universities. Both of those changes might also have had a negative impact on the quality of courses and hence of graduates. Thus, it seems possible that the average quality of BA workers has declined across cohorts. It is important to note, however, that this does not necessarily imply that the observed college wage premium is biased one way or the other relative to the composition constant differential. The quality of HS-educated workers is also likely to fall if the more able individuals among those who would have stopped at a HS education level in earlier cohorts now go to university and if some of those who would have been HS dropouts previously now obtain secondary qualifications. Thus, it is theoretically ambiguous whether the ability-composition constant college wage premium is greater or smaller than the observed one.

The idea that BAs have a lower and wider range of quality after the higher education expansion has been advocated in OLeary and Sloane (2005) and Walker and Zhu (2008). Both papers use quantile regressions to estimate the university wage premium across different periods or cohorts, and they report a greater decline in the premium at lower quantiles than at higher quantiles. While it's tempting to interpret such results as evidence of declining quality of BAs at the lower end of the BA wage distribution, examining the wage distributions for BA and HS workers separately suggests a different conclusion. Working with 5-year wide birth cohorts, in Online Appendix we show that the decline in the wage differential at lower quantiles is driven by relative increases in lower end wages for the HS-educated. The 50-10 differential of the BA wage distribution is unchanged across cohorts entering the labour market in our period. Thus, it is difficult to conclude that the fall of the graduate premium at lower quantiles is due to a greater deterioration in the quality of BAs than HS workers at their respective lower ends.

In the Online Appendix, we also present a bounding exercise to examine the limits of the potential impact of shifts in the distribution of unobservable characteristics on the college wage premium. We work at the level of 5-year birth cohorts because any such shifts would be clearest in looking at different cohorts of potential university graduates. Our exercise follows Manski (1994), Blundell et al. (2007) and Lee (2009), and works directly from a bounding approach in a Roy Model context set out in Gottschalk et al. (2014).

We present detailed results from the bounding exercise in the Online Appendix. The nature of the exercise is such that the bounds are defined as movements relative to a base cohort - in our implementation, the 1965-69 cohort. That cohort entered university age just before the major policy generated university expansion that began in 1988 and had a university graduate proportion of 0.16. The following two 5-year birth cohorts (born 1970-74 and 1975-79) represent the main part of the increase in educational attainment. For the 1975-79 cohort, the proportion graduating university reached 0.34. The bounds on the change in the college wage premium between the 1965-69 and 1975-79 cohorts range between an upper bound of 0 and a lower bound of -0.05. That is, even under extreme assumptions, the movements in the relative wage distribution and the proportion of each cohort who graduated university fit with very small changes in the college wage premium. In the following cohorts - ones over which the proportion with a university degree increased at a much slower rate - the bounds move to around -0.15 for the 1985-89 cohort. Re-examining Figure 2.2 in light of this finding, it is possible to see a small (though statistically insignificant) decline in the college premium after 2010. To the extent this is true, it would suggest a decline in the premium that occurs after the main increases in the educational supply. We will return to that possibility later in our discussion. But, our overall conclusion from the bounding exercise is that, under this model of ability, selection on unobservables cannot explain why we do not see a large decline in the education wage differential for the cohorts with the largest increase in their education level.

2.3 Technological leadership and models of technological change

To this point we have established that since the mid-1990s, the UK experienced a substantial upgrading in the education level of its workforce but virtually no change in the wage differential between university and high school educated workers. The obvious implication is that the increase in the relative supply of more educated workers was exactly offset by an increase in the relative demand for more educated workers. That type of skill biased demand shift is, of course, the focus of a very large literature in which much of the attention focuses on the role of technological change. We are convinced by papers such as Bresnahan et al. (2002), Caroli and Van Reenen (2001), and Bloom et al. (2014) which argue that the key technological change in recent decades is broader than just the use of computer hardware and software in specific tasks, taking in changes in organizational form that make use of newly invented IT features. For that reason, we will couch our investigations of the impact of technological change in that wider, organizational context.

One can think of the interaction of increased human capital attainment with technological change in terms of three main models. The first is one in which the technological change is exogenous: a new technology is introduced for an unspecified reason and is so dominant in terms of cost savings over existing technologies that it is adopted on a wide scale. Wage differentials are then determined by the interaction of relative demand shifts arising from this technological change (hinging on the skill bias of the technological change) and shifts in supply. Early versions of this model that focused directly on the college wage premium have generally been shown not to fit the data well (Beaudry and Green (2005); Card and DiNardo (2002); Acemoglu and Autor (2011)) but the more recent literature on polarizing changes in technology also has this broad form (e.g. Autor and Dorn (2013)). In all of these models, wage differentials reflect the classic race between technological change and education, with wages in higher skilled groups (defined by education or occupation) rising less if educational policy generates increases in the supplied labour in that group ((Goldin and Katz, 2008)).

The second model type is one in which the invention of new technologies is a function of movements in the relative factor endowments in an economy. Thus, an increase in the education level in an economy provides an incentive for inventors to create new technologies that are relatively intensive in the use of higher educated labour (Acemoglu (1998), Kiley (1999)). In this case, the relative increase in demand for skills is actually induced by the increase in their supply. Acemoglu (2007) shows that in cases where innovation is created by government funded research or by monopolistic or oligopolistic firms, if the elasticity of substitution between skilled and unskilled labour is high enough then an increase in the relative supply of skilled workers can induce an increase in the relative wage of the skilled workers. In this sense, in the context of this model, attempts to combat inequality by increasing educational attainment could backfire.

The third type of model is one in which a set of technological options already exist and firms choose among them. These endogenous choice models have the structure of a 2 sector by n factor trade model, where the sectors correspond to different technologies, and inherit implications of that model. In particular, if n>2 and all factors are inelastically supplied then these models can yield the same implications as the induced invention models, i.e., that increases in the relative supply of skill can generate increases in the skilled wage differential (Beaudry and Green (2003); Beaudry et al. (2010)). On the other hand if all but two of the factors are perfectly elastically supplied (as one might expect if new organizational capital, for example, requires a one time investment but widely accessible information thereafter) then even large increases in the relative supply of educated labour will leave skill group wages unchanged if the economy remains within a region in which both the new and old technologies are in use (the cone of diversification) (Beaudry and Green (2003)).[9] We believe this class of models fits with the spirit of the literature on decentralization and organizational form which, starting with Milgrom and Roberts (1990)'s seminal contribution, often approaches organizational form as something firms optimally choose given existing options (e.g., Bresnahan et al. (2002); Caroli and Van Reenen (2001)). It also follows a line of reasoning dating back to Griliches (1958) which emphasize endogenous adoption of technologies as the cost of adoption changes (see, for example, Doms et al. (1997) and Borghans and ter Weel (2007).

Deciding which of these models is relevant for an economy is important because, as we have just described, they can have quite different implications for the effect of education policy on inequality. But which model is relevant is potentially context contingent. There may be technologies that are so superior that the exogenous technical change model is clearly relevant (though we suspect those situations are extremely rare). On the other hand, in economies that are technological leaders in time periods when new technological possibilities are opening up, the induced invention model may be more appropriate. However, for other, following economies (and even in the technological leaders in periods after the initial invention is complete) the endogenous technological choice models, with firms choosing from an already invented set of options, may be the most relevant.

Much of the theorizing about these different models has been done with the US economy and US stylized facts in mind. But the US context may be quite unique. In particular, we will argue that there are good reasons to believe that the US has been a technological leader in the development of skill biased technologies and their associated organizational forms in recent decades. The UK - and, potentially, other developed economies - are, then, technological followers. In the remainder of the paper, we inves-

[9]Note that this is different from the first, competitive model in Acemoglu (2007). In that model, firms choose among technologies but those technologies are included as another input in a standard, unitary production function. The technological choice models we are referring to involve choosing among completely different technologies with different substitution elasticities.

tigate the claim that the UK is a technological follower and that, as a result, endogenous technological choice models best describe the functioning of its economy.

2.3.1 Testing among models of technological change

In this section, we use an empirical specification derived from a relatively general production function with UK data to establish the claim that the exogenous technological change and endogenous innovation models do not match patterns in the UK labour market in recent decades. Given that, in subsequent sections, we set out a model of endogenous technological choice and investigate its implications, including for the wage specifications derived and implemented in this section.

To investigate the various models, we derive an empirical specification that nests all three models. Here, we provide a brief description of the derivation, with details in Appendix A.1. We adopt a specification set out in Beaudry and Green (2005) in which there is an aggregate production function given by, $F(\theta_{st}S_t, \theta_{ut}U_t, K_t)$, where S_t is skilled labour used in production, U_t is unskilled labour, K_t is capital, and θ_{st} and θ_{ut} are skilled and unskilled labour enhancing technological change parameters, respectively. Given the focus of the existing literature and to keep the discussion simple, we assume that technological change is labour enhancing, implying that our specification does not nest factor neutral technical change. We discuss the implications of using a form of factor neutral technical change in Online Appendix. We will also assume that F(.,.,.) is constant returns to scale. Apart from that, the production function is left purposefully general so that it can be seen as reflecting any of the three models of technological change. Because we are concerned that there could be age effects arising from the movement of different sized cohorts into the education system, we follow Card and Lemieux(2001) in assuming that both skilled and unskilled labour can be written as CES aggregates of labour supplied by workers of different ages, i.e., $S_t = (\sum_j \Gamma_j S_{jt}^{\frac{\sigma_a-1}{\sigma_a}})^{\frac{\sigma_a}{\sigma_a-1}}$ and $U_t = (\sum_j \Omega_j U_{jt}^{\frac{\sigma_a-1}{\sigma_a}})^{\frac{\sigma_a}{\sigma_a-1}}$, where S_{jt} is the amount of skilled labour from age group j that is employed in period t, U_{jt} is defined analogously, Γ_j and Ω_j are age specific factor augmenting parameters, and σ_a is the elasticity of substitution between age groups within a skill group. In our estimation, we use over time variation within geographic sub-regions in the UK but in our initial exposition we will focus on a single region, suppressing the regional subscript.

Assuming competitive labour markets and employing a log linear approximation, we obtain,

$$\ln w_{ujt} \approx \ln \Omega_j - \frac{1}{\sigma_a} \ln \tilde{U}_{jt} + \ln \theta_{ut} + \alpha_1 \ln(\frac{S_t}{U_t}) + \alpha_1 \ln(\frac{\theta_{st}}{\theta_{ut}}) + \alpha_2 \ln(\frac{K_t}{\theta_{ut} U_t}) \tag{2.1}$$

and,

$$\ln w_{sjt} \approx \ln \Gamma_j - \frac{1}{\sigma_a} \ln \tilde{S}_{jt} + \ln \theta_{st} + \beta_1 \ln(\frac{\theta_{st} S_t}{\theta_{ut} U_t}) + \beta_2 \ln(\frac{K_t}{\theta_{st} S_t}) \tag{2.2}$$

where, $\ln \tilde{S}_{jt} = (\ln S_{jt} - \ln S_t)$ and $\ln \tilde{U}_{jt}$ is defined analogously. Concavity of the production function implies $\beta_1 - \beta_2 \leq 0$ and $\alpha_1 + \alpha_2 \geq 0$.

The difference between the two log wage expressions gives

$$\begin{aligned} \ln \frac{w_{sjt}}{w_{ujt}} \approx & (\ln \Gamma_j - \ln \Omega_j) - \frac{1}{\sigma_a}(\ln \tilde{S}_{jt} - \ln \tilde{U}_{jt}) + (\alpha_2 - \beta_2) \ln \theta_{ut} + (\beta_1 - \beta_2 - \alpha_1) \ln(\frac{S_t}{U_t}) \\ & + (1 + \beta_1 - \beta_2 - \alpha_1) \ln(\frac{\theta_{st}}{\theta_{ut}}) + (\beta_2 - \alpha_2) \ln(\frac{K_t}{U_t}) \end{aligned} \tag{2.3}$$

Equation (2.3) is a generalization of the specification in Card and Lemieux(2001). In that paper, as in most papers in the skill biased technical change literature, only a relative wage equation is estimated. But there is relevant information in the underlying wage equations as well, and we will focus on the skilled wage equation along with the wage ratio equation. With estimates of those two, the unskilled wage equation is redundant.

In order to take the skilled wage equation and the relative wage equation to the data we need to address the fact that the productivity parameter ratio ($\ln(\frac{\theta_{st}}{\theta_{ut}})$) and the θ_{ut} parameter that enter both equations are unobserved. We address these issues using the approach in Beaudry and Green (2005), capturing general productivity increases with measured TFP and allowing for exogenous skill-biased shifts using a quadratic function of time. This allows for a bit more flexibility than the common linear skill biased technical change assumption, which is obviously nested in this specification.

Based on this, we arrive at an estimable specification for the skilled wage equation similar to the one in Beaudry and Green (2005), given by:

$$\ln w_{sgjt} = b_{0j} + b_{0g} + b_0 t + b_1 t^2 + b_2 \ln(\frac{S_{gt}}{U_{gt}}) + b_3 \frac{\ln TFP_t}{(s_t^u + s_t^s)} + b_4 \ln(\frac{K_t}{U_{gt}}) + b_5 \ln \tilde{S}_{gjt} + \varepsilon_{1gjt} \tag{2.4}$$

where ε_{1gjt} is an error that contains approximation error and is assumed to be independent of the right hand side variables. We also obtain a relative wage specification given by:

$$\ln \frac{w_{sgjt}}{w_{ugjt}} = d_{0j} + d_{0g} + d_0 t + d_1 t^2 + d_2 \ln(\frac{S_{gt}}{U_{gt}}) + d_3 \frac{\ln TFP_t}{(s_t^u + s_t^s)} + d_4 \ln(\frac{K_t}{U_{gt}}) + d_5 (\ln \tilde{S}_{gjt} - \ln \tilde{U}_{gjt}) + \varepsilon_{2gjt} \tag{2.5}$$

where, again, ε_{2gjt} corresponds to approximation error. Note that both equations include a complete set of age band effects. In addition, we have introduced a subscript, g, corresponding to geographic region. We include a complete set of region effects (d_{0g}) and, so, are using within-region and age group, over-time variation. We construct the wage and employment variables at the region by age group by time level, but it is important to highlight that neither the TFP_t variable components nor K_t have g subscripts, i.e., the relevant values for both are assumed to be at the national level. For TFP_t, this reflects an assumption that technologies are available equally in all regions of the country. The same assumption underlies the lack of a g subscript on the time trend coefficients. For K_t, the corresponding assumption is that the capital market is national. With capital and technology defined at the national level, we use regional level data to see how regional variation in skill supplies alter sub-national differences in technological adoption and, so, wages. We view differences in regional outcomes within a common capital market as a good scenario in which to examine implications of the relationship between skill supplies and wages. The detailed derivation of these equations in the Appendix provides the direct mapping of the b and d coefficients onto the underlying structural (α,β, and σ_a) parameters. We also include there a discussion of the conditions under which our specification reduces to the Card and Lemieux (2001) version of the canonical specification, which does not include capital or TFP terms.

The exogenous technical change model and the induced innovation model have similar testable implications for the estimated coefficients in our model. In particular, in the canonical exogenous technical change model, the d_2 coefficient equals $-\frac{1}{\sigma}$, where σ is the elasticity of substitution between skilled and unskilled labour, and must be negative (Card and Lemieux (2001)). Further, the coefficients on the time variables in the wage ratio equation should imply a positive and significant trend, representing the exogenous technological shift favouring skilled workers. Given our expanded specification, skill biased technical change could, alternatively, show up as a positive and significant coefficient on $\frac{\ln TFP_t}{(s_t^u+s_t^s)}$ in the wage ratio equation, implying that observed technological change favours skilled workers. In the endogenous innovation model, holding technology constant (as we do using the combination of the time trend and TFP), d_2 is also equal to $-\frac{1}{\sigma}$ and, so, faces the same restrictions as with the exogenous technological change model ((Acemoglu, 2007), equation (18)). Further, if we estimate a specification in which we do not control for technology then the coefficient on $\ln(\frac{S_{gt}}{U_{gt}})$ in the wage ratio equation is an amalgam of the substitution effect and a potentially offsetting innovation effect that would raise the relative wage of skilled workers. The theory implies a connection between the estimated coefficients with and without controls for technology: if the elasticity of substitution estimated when controlling for technology is large then the effect of a shift in relative skill supply on the relative wage should be large and positive. Thus, in order to test the implications of the innovation model, we implement our full specification as well as a specification in which we do not include either time or TFP variables. For comparison to previous estimates, we also estimate the Card and Lemieux (2001) variant of the canonical model for the wage ratio, i.e., a specification that includes all the variables in (2.5) except $\ln TFP_t$ and $\ln(\frac{K_t}{U_t})$. In the appendix, we present further specifications in which we drop $\ln(\frac{K_t}{U_t})$ and replace it with the log price of capital, $\ln r_t$. Our conclusions are robust to these variations.

We use UK LFS data from 1993 to 2016, restricting our sample to 20-59 year olds for whom we observe wages and education. We aggregate to the level of cells defined by 5-year wide age groups and geographic regions, which allows us to control for compositional changes associated with the growing importance of London and other urban centres in our time period. For sample size reasons, we pool the data in 3 year groups.[10]

Within each age x region cell, we obtain the median real log wage for BA and for HS workers. We take the difference of those to form our wage gap dependent variable. We measure S_{gjt} and U_{gjt} as the total number of hours worked by BA and HS workers, respectively, who are in region g, age-band j and year t. We measure S_{gt} and U_{gt} as the simple sums of S_{gjt} and U_{gjt} across age groups within a region.[11] We scale S_{gjt}, U_{gjt}, S_{gt} and U_{gt} so that the aggregate hours supplied each year $\sum_g (S_{gt} + U_{gt})$ matches the na-

[10]The three year groups are 1993-95, 1996-98, etc.. We use the LFS "Regions of Usual Residence" as our definition of geographic regions. There are 19 such regions including, for example, London, Rest of South East, Greater Manchester, and the Western Midlands. These regions are consistently defined over the whole of our sample period. Sample size issues related to the reporting of wages prevents us from using a more detailed geography such as the on used in the organizational forms exercise later in the paper. We treat our production function as being at the level of the region, implying that all of our variables now have a g, for geographic region, subscript. The only exceptions are the capital and TFP variables. We assume that both capital and technological ideas flow freely across the regions in the country, implying that the country-aggregate levels of those variables are relevant.

[11]This deviates from the theory in which the aggregates are functions of $\sigma_a, \Gamma_j, \Omega_j$. We do this for simplicity and transparency so that we aren't forcing this element of our specification on the data. Since our estimates of σ_a imply very high substitutability across age groups, the results change very little when using the CES aggregates with estimated parameters rather than simple sums.

tional time series from the Office of National Statistics (ONS).[12] We obtain aggregate TFP series, capital and aggregate hours from the ONS.[13]

It is worth emphasizing that our estimates are based on variation within region x age cells over time. In figure 2.4, we show the variation we are using by plotting long differences (between 1993-1995 and 2014-2016) in $\ln \frac{w_{sgjt}}{w_{ugjt}}$ against long differences in $\ln \frac{S_{gjt}}{U_{gjt}}$ for all our regions for one of our age groups (30 to 34 year olds). Plots for other age groups show the same pattern. In particular, there is considerable variation in changes in $\ln \frac{S_{gjt}}{U_{gjt}}$ across regions, ranging from just over 1.1 log point increase over the 20 years in Northern Ireland to a high of over 1.5 log points in London and with an even spread in between. Matching that is little change in the within region/age group wage ratio, with most of the long term changes in the ratio being under 10% in absolute value. The correlation between the two series is only 0.15 and is not statistically significantly different from zero. When we put $\ln \frac{S_{gt}}{U_{gt}}$ instead of $\ln \frac{S_{gjt}}{U_{gjt}}$ on the x-axis, we get a similar pattern of weak correlations comparing only between regions. Thus, our data has considerable over-time variation in changes in employment ratios across regions matched with small changes and little variation in the change in the wage ratio. This core moment in the data is what is driving our estimate of the coefficient on $\log S_{gt}/U_{gt}$ in Table 2.1.

We present the results from our specifications in Table 2.1. The first two columns contain estimates of the skilled wage equation and the wage ratio equation by OLS. The second two columns contain 2SLS estimates aimed at addressing the potential endogeneity of the employment levels of the inputs. We instrument for $\ln \frac{S_{gt}}{U_{gt}}$ by using the education reform. In particular, we form a Bartik style instrument in which we interact

[12]The simple sum of hours in our sample every year would deviate from the true aggregate hours because education is missing to varying degrees over time and our sample selects 20-59 year olds only.

[13]The TFP series is the annual series of multi-factor productivity from ONS' release "Multi-factor productivity estimates: Experimental estimates to Q2 2017". Our capital measure is the annual series called "Contribution of capital services to GVA growth (percentage points)" in the same ONS release. Aggregate hours is "labour hours" from the same ONS release. This ONS release can be found here.

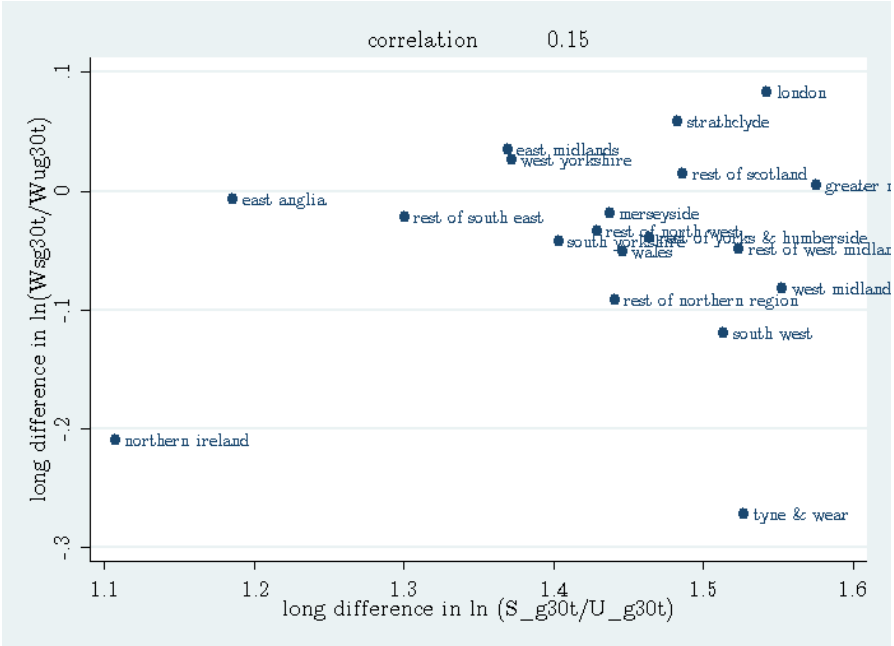


Figure 2.4: Changes in Employment and Wage Ratios by Region, 1995 - 2016

Note: The variables are the difference between 1993-1995 and 2014-2016 in the log ratio of employment (x-axis) and median wages (y-axis) of BA to HS workers for each region. The data is for 30 to 34 year olds in each year.

the proportion of the population in a region in 1993 (the start of our data) who were born in 5 year-wide birth cohorts with the growth in the proportion of that cohort who obtained a BA at the national level. The idea behind this instrument is that regions with a higher proportion in the cohorts that were most directly affected by the education reforms (those born between 1970 and 1974 and between 1975 and 1979) would face a stronger increase in the relative supply of skilled labour for reasons that have to do with historical fertility patterns that are plausibly independent of later education trends. We also construct an instrument as the interaction of the proportion of the parental generation for the 1970-1979 birth cohorts who themselves had a BA with the national growth rate in the proportion of workers with a BA. This is intended to capture the idea that children in locations with more educated parents were more likely to take advantage of the education reforms. Both instruments are strong predictors of $\ln \frac{S_{gt}}{U_{gt}}$ in the first stage and do not suffer from weak instrument issues by any standard test. Using similar logic to the second instrument, for each birth cohort in each region, we construct the proportion of the 'parental' cohort (the one born 25 years earlier) with a BA. We interact that proportion with the growth rate in the proportion with a BA for the specific child's cohort at the national level. Here too, the idea is that the growth in the BA share for an age group in a region will be related to the education level of the parents for that age group combined with the general increase in education level for their cohort. We use this as an instrument for $\ln \frac{S_{gjt}}{U_{gjt}}$ but have to restrict our attention to age 20 to 44 year olds because the first stage is weak when we include older individuals since there is little variation in the proportion of the parents' generations with a BA for the older age groups. Finally, we instrument for $\ln \frac{K_t}{U_t}$ using the interest rate.

The theory underlying our specifications implies several restrictions. The results reported in Table 2.1 have not imposed these restrictions; imposing them would make little difference to the key estimates and we will show them in Appendix A.

2.3.2 Assessing the exogenous and endogenous SBTC models

The estimates from our wage specifications do not fit with either the canonical exogenous SBTC model or the induced skill biased innovation model. The first strike against these models is the lack of any substantial effects of the skill supplies on the wage ratio. The estimated coefficients on $\log S_t/U_t$ in column (1) (OLS) and column (3) (IV) of Table 2.1 are statistically insignificant and have the wrong sign according to the theory. Results from alternative specifications presented in Appendix A all show this same pattern. Moreover, the lower bound of the confidence interval for the IV estimate in Table 2.1 (-0.05) is very small compared to the earlier use literature (e.g., -0.7 in Katz and Murphy (1992))). Thus, even in a generous interpretation, the coefficient would imply very high and possibly perfect substitutability between skilled and unskilled labour. This is very problematic for both the exogenous and induced innovation models since changes in relative demand created by either exogenous or endogenous technical change cannot move relative wages if the skill groups are perfect substitutes. As a side point, the age-specific skill supply coefficient is also close to zero (-0.038 in the OLS, and the wrong sign in the IV), implying a huge substitution elasticity between age groups (above 25). By comparison, Card and Lemieux (2001) estimated this elasticity to be in the [4,6] range.

Table 2.1: Skilled Wage and Wage Ratio Regressions: UK, 1993-2016

	$\ln \frac{w_{sgjt}}{w_{ugjt}}$	$\ln w_{sgjt}$	$\ln \frac{w_{sgjt}}{w_{ugjt}}$	$\ln w_{sgjt}$	$\ln \frac{w_{sgjt}}{w_{ugjt}}$	$\ln \frac{w_{sgjt}}{w_{ugjt}}$
t	0.006	0.021**	-0.025*	0.000		-0.00330
	(0.008)	(0.007)	(0.012)	(0.012)		(0.005)
t^2	-0.000	-0.001***	0.000	-0.000		
	(0.000)	(0.000)	(0.000)	(0.000)		
$\ln S_{gt}/U_{gt}$	0.047	-0.034	0.256	-0.168	0.002	0.056
	(0.087)	(0.081)	(0.156)	(0.163)	(0.010)	(0.084)
$\ln \frac{TFP_t}{laborshare_t}$	-0.004	0.315**	0.091	0.474***		
	(0.121)	(0.113)	(0.104)	(0.109)		
$\ln K_t/U_t$	-0.104*	-0.073	0.212	0.488**		
	(0.049)	(0.046)	(0.167)	(0.177)		
$\ln \tilde{S_{gjt}}/\tilde{U_{gjt}}$	-0.038		0.015		-0.036	-0.036
	(0.023)		(0.038)		(0.023)	(0.023)
$\ln \tilde{S_{gjt}}$		-0.025		-0.133		
		(0.021)		(0.156)		
IV	no	no	yes	yes	yes	no
N	1208	1208	760	760	760	1208

Notes: standard errors are shown in parentheses. The regression is at the level of 19 regions, 5-year-age-bands and 3-year-periods. The sample without IVs consists of 20-59 year olds. Whenever we use IVs, the sample is restricted to 20-44 year olds. The first 2 columns are the OLS estimation of the two equations. SURE results would be very similar and not shown here. The next 2 columns contain 2SLS estimates. 3SLS estimates would also be very similar to the 2SLS ones. The fifth column does not control time trend or TFP and the sixth is comparable to Card and Lemieux (2001). All specifications include complete sets of age-band and region dummies. *** p<0.01, ** p<0.05, * p<0.1

The second strike against the exogenous SBTC model is found in the coefficients on the time and time squared variables in the wage ratio equations. Recall that these are intended to capture the path of the ongoing skill biased technological changes. The IV estimates imply a negative trend while the OLS estimates imply technological change effects that are small and move from positive in early years to negative in later years. These results are robust to different specifications. Such a pattern, in which technical change is small and either against skilled labour from the outset (IV) or turning against it in later years (OLS) is at odds with our intuition.

These conclusions are reinforced in the last column of the table which contains estimates from the implementation of the the classic Card and Lemieux (2001) specification that includes only the linear time trend, the overall skill supply ratio, the skill supply ratio at the age group level, and a complete set of age and region effects. From this specification, we can see that our estimates of the skill supply and time effects in our main specification are not being determined by the inclusion of the TFP and capital variables. The estimated coefficients on both the time trend and the relative supply variables are statistically insignificant and of the wrong sign. The coefficient on the age group specific relative supplies is also small and statistically insignificant. At best, using the extremes of the confidence intervals, these estimates imply that skilled and unskilled labour are close to perfect substitutes, different age groups are close to perfect substitutes, and there is little or no ongoing skill biased technical change. We view these data patterns as a repudiation of the exogenous skill biased technical change model for the UK in the period after 1992.[14]

The estimated TFP effects provide further evidence against both the exogenous and endogenous skill biased technological change models. The TFP variable has a small and statistically insignificant effect on the wage ratio in column 3 of the table. Combined with the positive and statistically significant effect of TFP on the skilled wage in the estimates in column 4, the implication is that there is technological growth in this period but that skilled and unskilled workers benefit from it to an equal degree. Thus, the data

[14]These patterns are robust to imposing the theoretical restrictions on coefficients in equations (2.4) and (A.6) and to excluding London, out of concerns that it is big enough to be driving the results on its own.

does not fit with technology, as captured by TFP, being skill biased: the core feature of both of the first two technological change models.[15]

All of these implications apply to both the exogenous and endogenous skill biased technological change models, but the endogenous technological change model has added implications. In particular, if we do not control for technological change, then the impact of changes in the skill ratio on the wage ratio no longer has a determinate sign. The negative substitution effect that is estimated when controlling for technological change is combined with an effect on innovation that can generate offsetting, skill biased demand shifts. Under some circumstances, the latter effect dominates and the estimated coefficient in the relative wage regression without technology controls can be positive. In column 5, we present estimates of the wage ratio equation without the TFP and trend variables. We also drop $\ln \frac{K_t}{U_t}$ in order to obtain a specification similar to what is implied in Acemoglu (2007). The estimated coefficient on the relative skill supply variable is close to zero and statistically insignificant. For this to be the case, σ, the elasticity of substitution between skilled and unskilled labour should be near 2 in the endogenous innovation models in (Acemoglu, 2007) and (Acemoglu and Zilibotti, 2001). Instead, our estimates in columns 1 and 3 have the opposite sign and even the lower bound of the estimates indicate much larger σ values. The lower bounds of estimates in columns 1 and 3 would imply the effect of the relative skill supply in column 5 should be much larger. Either way, the data patterns are not consistent with the implications of the endogenous innovation model.

One possible response to our concerns about the model of exogenous SBTC of the type embodied in Card and Lemieux (2001) is that it is an older version of these models which has been supplanted by models of technological change and polarization. This has happened, in part, because other papers have similarly concluded that the exogenous

[15]Problems with the exogenous and endogenous SBTC models can also be demonstrated in a calibration exercise. In Online Appendix, we show that if one assumes typical values from the US literature for σ and σ_a then the combination of the implied path for $\ln \frac{\theta_{st}}{\theta_{ut}}$ of observed TFP implies a very strongly declining path for θ_{ut} that is unrealistic.

skill biased technical change model does not fit even the US data well either (e.g.,(Card and DiNardo (2002); Beaudry and Green (2005); Acemoglu and Autor (2011)). To look further into the role of polarization in the UK wage and employment structure, in Table 2.2 for 30-34 year olds, we present average real wages (in the first column of the first panel) and proportions of employees (in the first column of the second panel) in each of 9 one digit occupations in 1993. The occupations are ranked by their average real wage. In the second columns in each panel we present the change in either wages or proportions between 1993 and 2016.

The second column for employment proportions shows an approximate U-shaped pattern, with growth in employment shares in the top three occupations, declines in the middle (largely routine) occupations and growth in personal services. The relationship is not perfect since the lowest-paid occupation ("elementary") shows a decline, but the pattern is broadly one of polarization. However, when we hold the education composition constant between the cohorts (in the last column), there are small declines in employment in the top three occupation groups and essentially no change in processing and skilled trades in the middle. Thus, it seems that the employment shift in the UK from the middle to the higher-end is largely attributable to the education shifts. That is the hypothesis explicitly examined in Chapter 3. In section 3.5.1, we investigate correlations between occupational wages and employment, and conclude that a task-based version of exogenous technical change model cannot fit UK data, whereas a model of endogenous task-biased technical change can.

Taken together, we view the patterns of changes in wage levels, wage ratios, skill ratios, TFP, and capital for the UK in the last two decades as firmly rejecting both the exogenous skill biased technological change model and the endogenous skill biased innovation model for the UK for this period. Our view is that the endogenous innovation model is better suited to explaining movements in economies that are technological leaders where the innovation is taking place and that this does not describe the UK in

Table 2.2: Changes between 1993 and 2016, by occupations, at age 30-34

occupation	mean real wage		employment shares		
	w_{1993}	%change observed	$share_{1993}$	change observed	change reweighted
Professional occupations	15.00	0.184	0.111	0.070	-0.016
Associate professional and technical	14.00	0.149	0.159	0.029	-0.022
Managers and senior officials	13.06	0.257	0.151	0.010	-0.020
Skilled trades	10.19	0.128	0.133	-0.034	0.013
Administrative and secretarial	9.85	0.227	0.134	-0.046	-0.047
Process, plant and machine operatives	9.10	0.097	0.090	-0.038	-0.005
Personal service	7.89	0.194	0.055	0.032	0.052
Sales and customer service	7.63	0.229	0.060	0.002	0.018
Elementary occupations	7.09	0.184	0.107	-0.026	0.028

Notes: Real wage is in 2012 prices, deflated by GDP deflator. The final column reweights the employment shares of occupations using the education split in 1993 of 30-34 year olds.

this period. We elaborate on this claim in the next section.

2.3.2.1 Induced Technological Change and Technological Leadership

Induced technological innovation models focus on the expansion of the technological frontier. As such, they are about countries which are the technological leaders and would seem to provide a better explanation for movements in leader than follower economies. Working within the induced innovation model, the country that is most likely to be the leader in skill biased technological innovation will be the one with the highest share of skilled workers. A high share provides an incentive for innovator firms to invent machines or forms of organization that complement skills. In 1980, on the cusp of the computer revolution, the US was the leading developed economy in terms of education level. In that year, 22% of the US population aged 25 to 64 had a tertiary education, which was by far the highest in the OECD (Lee and Lee (2016)).[16] Thus, incentives for innovators to generate human capital intensive technologies would have been highest

[16]The next highest were Canada at 18% and Australia and New Zealand at about 15%, with the remainder of the OECD decidedly lower.

in the US. Moreover, the US has had the highest ratio of investment in ICT (Information, Computers, and Technology) capital to total non-residential gross fixed capital throughout the 1985 to 2010 period (OECD(2017)). The idea that the US is the innovation leader is also supported by evidence in Bloom et al. (2012) showing that US multinationals use a more decentralized structure relative to both domestic firms and multinationals from other countries even when all are observed operating in the same economy (the UK).

On the other side, there is also good reason to believe that the UK is a follower in the area of skill biased technologies and their associated organizational forms.[17] Certainly, the UK was well behind the US in educational attainment at the beginning of the computer revolution. This can perhaps be most clearly seen in data organized by birth cohort. For the cohort born between 1955 and 1959 in the UK (and who would have turned 25 in the early 1980s, at the outset of the computer revolution), 12% held a university degree by age 30 compared to 24% for the same cohort in the US.[18] For the cohort born a decade later, the numbers were 16% for the UK and 27% for the US - the UK was still a laggard. Thus, viewed through the lens of the theory of induced invention, we would not expect the UK to have been a leader in skill-biased innovation. However, because of the educational reforms described earlier, by the cohort born between 1975 and 1979 (who turned 25 in the early 2000s), the UK had surpassed the US with 34% attaining a university degree in the UK compared to 32% in the US. That increase in the educational attainment of new labour market entrants in the UK could have provided the conditions for firms to adopt the technologies previously developed

[17] Classifying the UK as a technological follower could imply that we can analyse its wage patterns as the equivalent of a Southern economy in the analysis in Acemoglu and Zilibotti (2001). In their discussion, Northern economies innovate in response to relative skill changes in their workforces as described earlier. Southern countries, in contrast, do not innovate and take the technological level invented in the North as given. However, with no innovation response channel in the South, increases in the relative supply of skill in their workforces will necessarily induce a decline in the skilled-unskilled wage ratio. As we have seen, this does not fit with the wage patterns in the UK in recent decades.

[18] These figures are computed from the UK LFS for the years 1992 to 2015 and the Outgoing Rotation Group sample from the US Current Population Survey for the same years.

in the US. Interestingly, the proportion of investment that was in ICT capital shot up in this decade in the UK, approximately doubling at the same time the proportion of new labour market entrants with a university education also doubled (OECD(2017)).[19] Further, the evidence in Bloom et al. (2012) about use of decentralized organizational forms also suggests that UK firms were following rather than leading. They argue that UK firms were laggards in adopting decentralized structures because of regulation based inflexibilities. We offer an alternative explanation: that at the time of the development of the new IT related structures, the lower education level in the UK implied it was less profitable for UK firms to adopt the new approach. Then, as the UK education level increased, the UK underwent a technological transformation. We think that these patterns fit most naturally with models of technological choice and we turn to a model of this form in the next section.

2.4 Model of technological choice and decentralization

In this section, we set out a model of technological choice in a situation where newly invented technologies involve decentralized organizational forms made possible by IT innovations. We derive implications of the model at the macro level that we compare to our production function estimates and at the micro level that we investigate with workplace data in the following sections.

The general framework we consider is one in which firms can choose to produce a single output either with a centralized (C) technology or a decentralized (D) technology. Having a single output is intended to emphasize the nature of these technologies as general purpose technologies that could be applied to the production of any product. Following Rosen (1978) and Borghans and ter Weel (2006), we will characterize production in engineering terms as having a Leontieff form in which a continuum of

[19]The proportion of total non-residential fixed capital investment in ICT increased by 88% in the UK between 1990 and 2000. Only Finland and South Korea had faster growth in this proportion in this decade. In comparison, the proportion grew by 37% in the US.

tasks, x, defined on the unit interval are required to produce an output.[20] The amount of each task required to produce one unit of output is given by the continuous function, $\alpha(x)$, $x \in [0,1]$. The tasks are performed by two types of workers: U (unskilled) and S (skilled). Total hours of work are inelastically supplied by each type of worker. Workers of each type are described by capacity functions, $\tau_l(x)$, which are continuous functions defined on [0,1] determining the amount of time a worker of type $l = U,S$ needs to produce the amount of task x required for one unit of output. Further, we assume that tasks are ordered from least to most complex and that S workers have comparative advantage in more complex tasks, i.e., $\frac{\tau_S(x)}{\tau_U(x)}$ is decreasing in x.

Rosen (1978) shows that based on such a specification, one can derive a production function defined over n_s and n_u (the number of hours of S and U labour used, respectively) in which the firm allocates a given amount of S and U to each task in order to maximize output. In particular, firms will allocate skill groups according to their comparative advantage in the sense that there will be a task ρ such that all tasks, $0 \leq x \leq \rho$ are assigned to U workers and, conversely, all tasks $\rho < x \leq 1$ are assigned to S workers. Further, ρ is declining in $\frac{n_s}{n_u}$. Thus, if the relative number of S workers is small then they will only be assigned to the most complex tasks and as that relative number grows, they will be moved progressively further down the list of tasks ranked by complexity. The marginal rate of technical substitution between S and U equals $\frac{\tau_S(\rho(n_u,n_s))}{\tau_U(\rho(n_u,n_s))}$, where we have written ρ as a function of n_u and n_s. Thus, profit maximizing firms will hire numbers of hours of U and S labour to equate the marginal rate of technical substitution to the wage ratio, $\frac{w_S}{w_U}$ (where, w_S and w_U are the skilled and unskilled hourly wages), allocating those hours optimally according to comparative advantage over the tasks required to produce. The result is a production function that reflects the efficiencies from

[20]This general form for production has become somewhat common in models of technological change, tasks, and polarization. For example, a variant of it is used in Acemoglu and Zilibotti (2001), and Acemoglu and Autor (2011) use this approach to provide a framework for interpreting existing research on tasks and technological change. Our model differs in the way we introduce decentralization and in our assumption that firms can choose between two such technologies.

taking account of the comparative advantage of the two types of workers and which is, itself, not necessarily Leontieff in form. In this sense, the ultimate production function reflects more than just the engineering 'recipes' since it includes the optimal allocation of workers across the task combinations specified in the recipes.

As Rosen (1978) demonstrates, and as we draw in Diagramme 1, in the case with two types of workers (our case), the unit output isoquant intercepts both axes. The intercept on the N_u (number of unskilled workers) axis equals $\int_0^1 \tau_U(x)dx$. As we move away from that intercept to the left, we begin to introduce S workers, replacing the U workers in the most complex tasks. Thus, the slope of the isoquant is given by $\frac{\tau_S(\rho(n_u,n_s))}{\tau_U(\rho(n_u,n_s))}$ and comparative advantage dictates the standard convex shape. The N_s intercept is given by $\int_0^1 \tau_S(x)dx$.

We will consider an economy with two possible 'recipes' or technological forms. The first is centralized and takes the form as set out above, where we will now write the technological requirements function as $\alpha^C(x)$ and the amount of time a worker needs to complete the number of tasks needed for a unit of output as, $\tau_l^C(x)$. In order to match patterns in the data, we delineate management tasks from other tasks. In the centralized technology, management tasks are necessary in order to co-ordinate the other tasks and the producers of the other tasks just focus on production of their part of the process, leaving communication and co-ordination to the managers. We will arbitrarily denote tasks on the interval $[\theta,1]$ as management tasks. To keep the exposition simple, we will assume that the α and τ functions are continuous from above and below at θ.

The alternative technological form is decentralized. Caroli and Van Reenen (2001) describe modern organizational forms as being 'delayered' with 'some decision-making being transferred downstream.' Multi-tasking is also an important feature of this organizational form with the benefits that the firm becomes more flexible and managers have to spend less time monitoring and co-ordinating workers (Bloom et al. (2014)). Thus, rather than having workers performing physical tasks without regard to others and having a manager who co-ordinates the outcome, in a decentralized form, workers both produce and co-ordinate with other task producers. As a result, less of the pure management task is needed. All of this is made possible by (i.e., is complementary with) IT technological change, which reduced the cost of diffuse information transfer.

We capture the differences in the decentralized form relative to the centralized form, first, by assuming that there is a lower requirement for the pure management tasks in the new form:

$$\alpha^D(x) = \lambda \alpha^C(x), \forall x \geq \theta \tag{2.6}$$

where, the D superscript denotes the decentralized technology, and $\lambda < 1$. For simplicity, we will assume that the requirements for the other tasks remain the same, i.e., $\alpha^D(x) = \alpha^C(x), \forall x < \theta$.

Following much of the literature on technical change and the labour market, we also assume that skilled workers are better at working with the new organizational form (Caroli and Van Reenen (2001); Bresnahan et al. (2002)). We represent this by assuming that skilled workers are perfect multi-taskers and can perform each of the non-management tasks in the same amount of time as before, performing the new, associated communications while they are doing them without extra effort (thanks to IT). For unskilled workers, performing each non-managerial task now requires more time since working with the new IT is more difficult for them. Further, skilled workers are able to take advantage of the new technology in management tasks while unskilled workers are not. Thus,

$$\tau_S^D(x) = \tau_S^C(x), \forall x < \theta \tag{2.7}$$

and

$$\tau_U^D(x) = \gamma \tau_U^C(x), \forall x < \theta \tag{2.8}$$

with $\tau_U^D(x) = \tau_U^C(x), \forall x \geq \theta$, $\tau_S^D(x) = \lambda \tau_S^C(x), \forall x \geq \theta$ and $\gamma > 1$. We view this specification as capturing the notion of decentralization in papers such as Lindbeck and

Snower (1996); Caroli and Van Reenen (2001); Bresnahan et al. (2002), and Bloom et al. (2012): that it is an organizational form in which decision making and communications are spread throughout the firm rather than being done by a small cadre of managers. We could allow for decentralization forms in which communication and decision making are differentially allocated across tasks but elect for the simpler form in which they are essentially allocated evenly across the non-manager tasks for expository clarity.

The literature emphasizes that decentralization has been enabled by the advent of IT. Much of the recent work on IT and the labour market also emphasizes impacts of the new technology in replacing routine tasks that tend to lie in the middle of the wage distribution. Following Borghans and ter Weel (2006) and Acemoglu and Autor (2011), we can model this effect by having the α values in middle tasks substantially reduced under the new (D) technology. Essentially, the idea is that IT capital performs those tasks and, thus, less labour is required in them. As described in Acemoglu and Autor (2011), the result will be a polarization in employment, with relatively more employment in low and high complexity jobs compared to those in the middle. However, this will not alter our main points about movements in educational wage differentials set out below. For that reason, we will not explicitly include the reductions in middle α's in our analysis for simplicity.

Given this setup, if there were only U workers in the economy then all firms would use the C technology since it would be cheaper at any given unskilled wage. Conversely, if there were only S workers in the economy, firms would use only the D technology. But we will start by assuming that the endowment of S and U workers in the economy is such that both technologies are in use (returning to the conditions under which that is true momentarily). We also assume that these are general purpose technologies that can be used for producing any good. Thus, to simplify, we assume both are used to produce a good which is the numeraire. Assuming free entry of firms and that output is the numeraire with a price of 1, that implies two zero profit conditions:

$$1)1 = w_U \int_0^{\rho_C} \tau_U^C(x)dx + w_S \int_{\rho_C}^1 \tau_S^C(x)dx \tag{2.9}$$

$$2)1 = w_U \int_0^{\rho_D} \tau_U^D(x)dx + w_S \int_{\rho_D}^1 \tau_S^D(x)dx \tag{2.10}$$

where, w_U is the unskilled wage, w_S is the skilled wage, ρ_C is the task dividing the U from the S tasks for technology C and ρ_D is the threshold task for the D technology.

Several key points follow from these two equations. First, together they imply a factor price invariance result as in standard trade theory. Because ρ_C and ρ_D are determined by the equality of the wage ratio to the marginal rate of technical substitution (MRTS) in profit maximizing firms and the MRTS is given by $\frac{\tau_S(\rho)}{\tau_U(\rho)}$ (i.e., is technologically determined), everything on the right hand side of both equations can be written as functions of w_U and w_S. That, combined with the assumption that these are general purpose technologies and so are producing the same good with the same price, implies that we have two equations in two unknowns (w_S and w_U). We show the solution diagrammatically in Diagramme 1. The figure shows the unit output isoquants for the two technologies. The isoquant for the centralized technology intersects the number of unskilled workers (N_u) axis at $n_{u0}^C = \int_0^1 \tau_U^C(x)dx$, i.e., the total number of hours to produce one unit if only unskilled workers are being used. Similarly, its N_s axis intercept is $n_{s0}^C = \int_0^1 \tau_S^C(x)dx$. The unit isoquant for the decentralized technology has a larger N_u intercept because of our assumption that unskilled workers take longer to do non-managerial tasks because of the requirement to communicate as well as produce but get no advantage in terms of the time they require to perform management tasks. In contrast, under the decentralized technology, skilled workers require no extra time to do non-production tasks and can take advantage of IT to spend less time on managerial tasks. The result is an isoquant with a larger N_u intercept, a smaller N_s intercept and a lower slope at all values of x than the C isoquant.[21] Given the continuity assumptions and the comparative advantage assumption, the isoquants will cross once. That, in turn, implies that there will be a single unit cost line that is just tangent to the two isoquants, i.e., a single pair of w_S and w_U values at which both technologies are in operation.

Diagramme 1: Wage Setting with Two Technologies

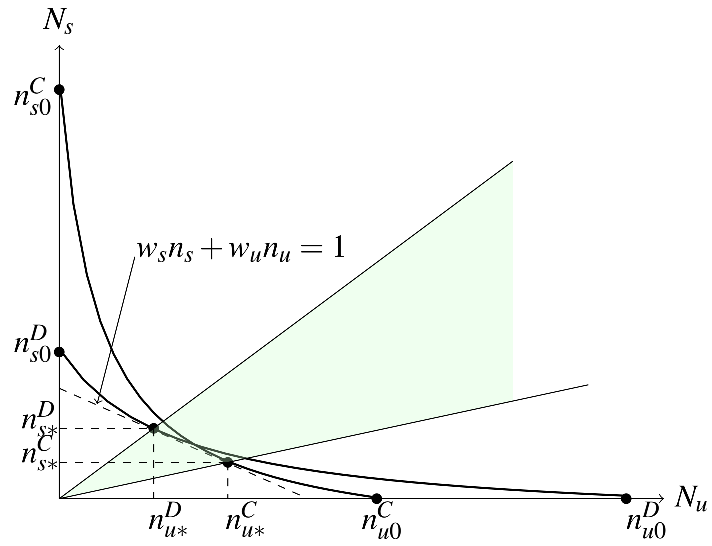


Diagramme 1 is, of course, a standard trade diagramme with two technologies instead of two sectors, and the same conclusions follow here as in the simple trade case. Our assumptions about the two technologies implies that the C technology will be relatively U intensive and in an equilibrium in which both technologies are used, n_{u*}^C and n_{s*}^C hours of unskilled and skilled work, respectively, will be used to produce a unit of the output with this technology. Similarly, n_{u*}^D and n_{s*}^D hours of unskilled and skilled work will be used with the D technology. Rays defined by $\frac{n_{s*}^C}{n_{u*}^C}$ and $\frac{n_{s*}^D}{n_{u*}^D}$ are drawn as diagonal lines in the figure. Those rays form the boundaries of the cone of diversification (the shaded area in Diagramme 1). As long as the ratio of skilled to unskilled hours in the economy falls within that cone, both technologies will be in use. If, instead, $\frac{N_s}{N_u} < \frac{n_{s*}^C}{n_{u*}^C}$

[21] To make the exposition simpler, we assume that $\lambda \cdot \gamma = 1$. This implies that the isoquant is smooth at task θ. Without it, there would be a kink in the isoquant that would complicate the exposition but not the ultimate conclusions.

then only the C technology will be used. This is simple to see in the figure since on rays with lower slope than $\frac{n_{s*}^C}{n_{u*}^C}$, the cost line that is just tangent to the C isoquant will lie below the D isoquant, implying that it is less costly to produce just with C. Conversely, if $\frac{N_s}{N_u} > \frac{n_{s*}^D}{n_{u*}^D}$ then only the D technology will be used.

What is of most interest to us is the implications for wage movements when there are increases in S relative to U. Given that equations (2.9) and (2.10) have a unique wage solution and are not functions of labour quantities, as long as both technologies are in use, changes in the amounts of S and U in the economy do not alter the individual wages or their ratio. This is the standard factor price invariance result from trade theory. Firms in the economy react to larger relative amounts of S labour not by increasing the amount they use with any one technology but by shifting toward the more S intensive technology (D). In fact, it is straightforward to show that a given increase in the ratio $\frac{N_s}{N_u}$ generates a more than proportionate increase in output from the D technology.[22]

Following from this, the empirical implications from the model are as follows. First, if the two technologies are available and the skilled to unskilled labour ratio, $\frac{N_s}{N_u}$, is in the cone of diversification then increases in the ratio of skilled to unskilled labour does not alter the wage ratio, $\frac{w_s}{w_u}$, or the individual wages, w_s and w_u. Second, if $\frac{N_s}{N_u}$ rises enough then eventually all firms will adopt the D technology and then subsequent increases in $\frac{N_s}{N_u}$ will generate decreases in $\frac{w_s}{w_u}$ as in the standard one technology case. Third, assume that there are unskilled managers in the C technology before the increase in the skills in the economy (as is the case in our sample period), i.e., $\rho_C > \theta$. In that case, the ratio of the number of unskilled managers to skilled managers will decline as $\frac{N_s}{N_u}$ increases. This happens because U workers form a larger fraction of managers under the C technology (indeed, they may not be managers at all in the D technology given

[22] To see this note that we can write the ratio of S to U hours employed in the economy as a weighted average of the ratios employed in the two technologies, i.e., $\frac{N_s}{N_u} = \phi_C \frac{n_{s*}^C}{n_{u*}^C} + (1-\phi_C)\frac{n_{s*}^D}{n_{u*}^D}$, where, ϕ_C is the fraction of output generated using the C technology. If the economy is in the cone of diversification, as $\frac{N_s}{N_u}$ increases, the two technology specific ratios do not change but ϕ_C decreases. In fact, ϕ_C must decrease more than proportionally to maintain the equality.

the comparative advantage set up) while S workers form a larger fraction of managers under the D technology. As the number of skilled workers rises, there will be a disproportionate shift toward the D technology that will imply more S than U managers overall even though the proportion of each type of manager will stay the same within each technology. Fourth, as $\frac{N_s}{N_u}$ increases, the proportion of S workers who are managers decreases. This is somewhat surprising given that the economy is shifting toward a more S intensive technology where more of the management positions are held by S workers. However, there is actually a smaller proportion of S workers who are managers with the D technology (since all S workers are managers in the C technology if $\rho_C > \theta$) and so as the economy shifts toward the D technology the proportion of S workers who are managers will fall. This is a reflection of the fact that in the decentralized technology, where S workers can both produce and communicate at the same time, S workers are used farther down into the task structure than in the C technology in equilibrium. Fifth, as $\frac{N_s}{N_u}$ increases and the economy shifts toward the D technology, we should see more workers in all parts of the production structure making decisions and communicating not just to their managers.

It is interesting to compare these implications to those from a more standard model with exogenous technical change. In Online Appendix, we analyse a model in which one technology is in use at a time. The production function is expressed as a function of managerial and production labour with skilled workers having a comparative advantage in managerial tasks. We characterize skill biased technical change as a relative increase in the productivity of skilled workers as managers. This captures both that the technological change favours skilled worker and that it related to managerial tasks. The technological change arrives exogenously, i.e., it alters the production function firms face without their making a choice over whether to adopt it. In this scenario, we show that the ratio of skilled to unskilled wages will remain constant only if the relative supply of skilled workers in managerial tasks increases by enough to offset the increase in their productivity in those tasks. This is the opposite of the implication from our endogenous technological choice model in which the expansion in S is accompanied by a decreasing proportion of S workers who are managers.

2.5 Evidence on model implications

2.5.1 Macro evidence

We begin our investigation of the relevance of our model by examining its implications in relation to the wage and employment patterns documented in the earlier sections of the paper. The first implication of the model is that the substantial increase in the proportion of workers with a university degree should have no impact on either the college premium or skilled and unskilled wages individually. In section 2, we showed that the college premium has not changed since 1992 even as educational attainment has soared and that this pattern cannot be explained as a result of compositional shifts in terms of observed or unobserved worker characteristics. This implication is borne out in our aggregate production function estimation where the coefficient on the relative skill supply variable in the wage ratio equations is small and never statistically significantly different from zero. As described earlier, the endogenous innovation model can also predict this zero effect but Acemoglu's description of the timing of the reaction of an economy to an increase in its relative skill supply involves an initial decline in the wage ratio followed by an increase as the effects of new inventions gradually take hold. In the figures in section 2, instead, the relative wage stays constant throughout the period of greatest education expansion - a pattern that is predicted by the technical choice model.

The technical choice model also has the stronger implication that underlying the lack of movement in the wage ratio should be a lack of response of the skilled and unskilled wages individually to the relative supply shift. As we discussed earlier, the estimated coefficients on $\frac{S_t}{U_t}$ in Table 2.1 imply that movements in the skill ratio have no effect on either skilled or unskilled wages. These zero effects fit with the picture of the isoquant in Diagramme 1. In that figure the isoquant for the economy is formed as the envelope constructed using the Centralized isoquant to the right of n^C_{u*}, the straight line connecting the points, n^C_{u*}, n^C_{s*} and n^D_{u*}, n^D_{s*}, and the Decentralized isoquant to the left of n^D_{u*}. The flat portion of the isoquant matches a flat section of the aggregate production function corresponding to the range of factor employment values in which the economy is operating in the cone of diversification with both technologies in use. That section being flat corresponds to the effect of $\frac{S_t}{U_t}$ on the levels of both wages being zero, which we have just seen is true. It also implies that the determinant of the Hessian of the production function should equal zero. We can construct an estimate of that determinant as either, $(b_2 \cdot d_4 - b_4 \cdot d_2)$ or $(b_2 \cdot -d_3 - (1 - b_3) \cdot d_2)$. These take values of .007 and -.033 from the OLS estimates and -.16 and -.12 from the IV estimates, all of which are not statistically significantly different from zero at the 5% significance level and all but one of which are about the same size in absolute value as their associated standard errors. Thus, our production function based estimates fit with the model implication that the UK economy was operating in a region in which the production function had a flat spot in our time period. This is a very specific implication of our technological choice model. It is worth noting that one could allow ongoing technological changes in both of the technologies in our model, which would be represented by inward shifts in the unit isoquants in Diagramme 1. However, if the rates of technological change were different in the two technologies then the wages and their ratios would change over time and the estimated production function would not have a flat segment. An equal technological change in each would be captured in a TFP measure that did not affect the wage ratio, as is the case in our estimates.

Taken together, we see the evidence from the production function estimates as fitting with technological change affecting the labour market through two channels. The first is a general shift out in the production possibilities frontier that is captured in our TFP measure. The fact that TFP changes induce wage level changes but no change in the wage differential implies that this element of technological change is skill neutral. It may reflect forces affecting productivity other than the IT and skill related changes that we emphasize here. Controlling for movements in TFP, changes in the skill ratio have no impact on wage levels or the wage differential, fitting with our model of endogenous technological choice. Thus, our evidence suggests a non-biased general shift out in the frontier with skill related technological changes corresponding to changes in the choice of the point along a given frontier (i.e., holding TFP constant). The result that the economy is operating on a flat portion of the production function in this period is a key piece of evidence in favour of this view.

The other implications of the model at the aggregate level have to do with occupational composition. In particular, as the relative number of workers with BA's increases, management roles should be increasingly taken over by BA educated workers. Thus, the model predicts that the proportion of managers who have a BA should increase across cohorts. In the left graph of Figure 2.5, we plot the proportion of managers who have a BA over time. The plots are for 30 to 34 year olds in order to hold age composition constant. There is clear evidence of a large shift in the direction predicted by the model: approximately 25% of managers had a BA in the early 90s compared to over 50% after 2010.[23] At the same time, the proportion of the BA educated workforce employed as managers should decline according to the model. In the 2nd graph of Figure 2.5, we plot the proportion of BA workers employed in management jobs, again focusing on age 30-34. We see that 23% of BA workers were managers in the early 90s compared to 19% after 2010. We argued earlier that the pattern depicted in the two panels in Figure 2.5 fits with our model but does not match the predictions of a standard model with an exogenous technological change favouring educated workers.

[23]The data underlying Figure 2.5 is from the LFS with managers corresponding to the first major group in the UK SOC2000, called 'Managers and senior officials'.

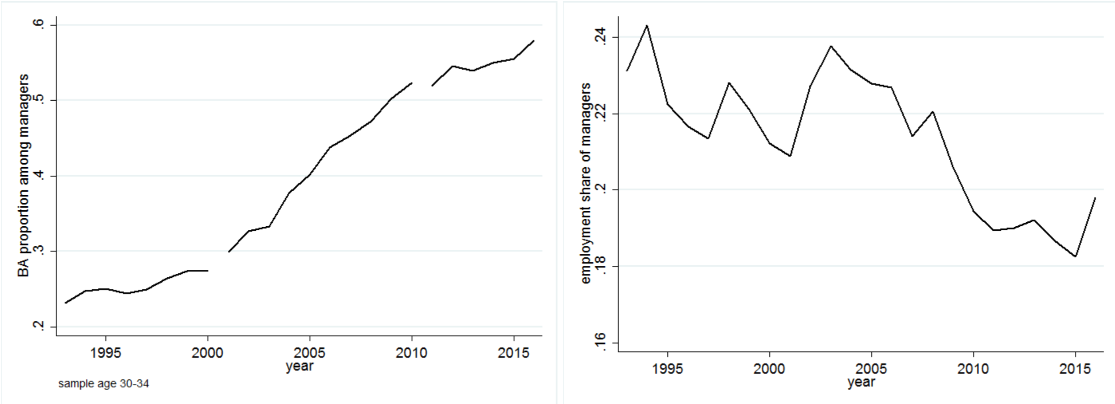


Figure 2.5: BA proportion within managers, and Proportion of BA's Who are Managers

Notes: We define managers as the first major group under UK SOC2000. The occupation classification in the LFS changed from SOC90 in 2000 to SOC2000 in 2001, and then to SOC2010 in 2011. We map the other two classifications to SOC2000 in a probabilistic way, using a matrix from the ONS for the latter period, and a self-constructed matrix based on dual-coded data in 2000-1. The left graph shows the break points in the time series for when the classification changed.

2.5.2 Micro evidence

We turn, next, to using micro data to examine the main implication of the model: that firms in locations with larger increases in the relative number of educated workers make greater use of decentralized organizational forms. Our hypothesis is that in a more decentralized and de-layered organizational structure, workers will be given more autonomy and will report greater influence over their work. We are interested in whether an increase in the relative supply of education skills induces a shift toward a more decentralized organizational form as measured by this marker. We examine this question using the UK Workplace Employment Relations Survey (WERS). The WERS is a survey of workplaces that includes questionnaires both for the manager as well as for a subsample of employees.[24] We focus on employees' responses to three questions:

"How much influence do you have about the following?"

[24] The WERS surveys 25 employees per workplace. When there are fewer than 25 employees at the workplace, they are all given the questionnaire. The WERS is a representative survey and we incorporate its associated weight in all our calculations.

1. “The range of tasks you do in your job”,

2. “the pace at which you work”

3. “how you do your work”.

The responses for each question range from 1 “A lot” to 4 “None”. These questions are included in the cross-sectional WERS surveys for 1998, 2004, and 2011. Rather than use these questions separately we implement principal components analysis to compute an index of the ability of workers to influence their own work. We define the index as 4 minus the first principal component, so that the index is higher where more employees report having more influence. The index accounts for approximately 80% of the total covariance among the three questions. Finally, we normalize the influence index to have mean 0 and standard deviation 1 in the 3-wave-pooled sample. We view the answers of “A lot” to these questions as reflecting a decentralized workplace where decision making on what to do and how fast to do it has been devolved to workers. In this, we follow Bresnahan et al. (2002) and Bloom et al. (2012) who implement surveys of managers to capture organizational practices. Their decentralization measure is based, in part, on ”individual decision authority” which reflects whether workers control their “pace of work” and “method of work”.

Table 2.3 lists the overall mean and standard deviation of the influence index by WERS wave and education of employees. Across all firms, there has been a nearly 0.6 standard deviation increase in the mean influence index value between 1998 and 2011. Thus, there is a clear general trend toward decentralization of decision making. We examine differences between more and less educated workers in the lower panels of the table, presenting weighted averages with the proportion of workers at a firm in the particular education group as the weights. Doing that indicates that the increases in the index value were particularly large for lower educated employees. This makes sense since firms with more low-educated employees are more likely to have used a centralized structure in the past and they would have had the most leeway for adjustment.

Table 2.3: Summary statistics of the influence index

Wave	Number of TTWAs	Number of workplaces	Mean influence index	Standard deviation
Influence index for all employees				
1998	204	1758	-0.32	0.92
2004	230	1657	0.024	0.96
2011	238	1917	0.27	1.025
Influence index for employees with degrees or above				
1998	190	1368	-0.018	0.79
2004	209	1272	0.001	0.83
2011	223	1557	0.13	0.80
Influence index for employees without degrees				
1998	203	1732	-0.34	0.88
2004	229	1620	0.024	0.93
2011	235	1823	0.25	1.047

Notes: for each education group or for 'all employees', we first calculate 4 minus the first principle component of the three influence scores (ranged $1-4$). We then normalize that variable to have mean 0 and standard deviation 1 in the 3 wave pooled sample for the education group or for 'all employees'. Workplaces are weighted by the establishment's employment weight times the proportion of employees of that workplace in that education group. If a workplace has no employees of the labelled education group responding to the influence questions in the employee survey, the workplace is not counted in the sub-table for that labelled education group. *Source:* Authors' analysis of the UK Workplace Employment Relations Survey.

To investigate the role of skill supply in choice of organizational form, we examine the relationship between the local supply of workers with BAs and the influence index at the workplace level. "Local area" here refers to Travel To Work Areas (TTWA), which were developed to capture local labour markets using data on commuting flows in 1991.[25] There were around 300 such areas in the UK in the 1998 through 2011 period. We derive from the LFS the proportion of workers in the TTWA who have a BA or above for the two calendar years up to and including the WERS survey year.[26]

[25]For further information on TTWA, see http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/other/travel-to-work-areas/index.html

[26]For example, for the WERS outcome measured in 2011, the BA proportion is measured from LFS 2010-2011.

Table 2.4 reports the results from OLS regressions of the influence index on the local BA proportion across a range of specifications. In all the specifications, we pool together the data from the three waves and we weight by the size of the workplace. Given that our main variable of interest varies at the TTWA level, we cluster the standard errors at that level. In the first column, we report the results from an OLS regression with the proportion of BA's in the area and year dummy variables as the only regressors. The estimated year effects indicate a secular trend toward organizational forms with greater worker control. This may reflect a response to the general increase in the education level of the workforce but more direct evidence on whether such a relationship exists is found in the estimated effect of the proportion of workers with a BA. We estimate that a 10 percentage point increase in the proportion of BAs in an area is associated with a 0.09 standard deviation increase in the influence index. This result fits with the idea that firms in areas with a higher proportion of educated workers use more decentralized organizational forms.

In the next set of columns, we check the robustness of this result across a series of specifications. In the second column, we condition on the current HS proportion in the area, and the coefficient on the BA proportion changes very little. Thus, what matters for decentralization is the proportion of higher educated workers not more versus fewer high school drop-outs among the less educated. In the third column, we introduce controls for industry, workplace size, and size of the organization.[27] Notably, the size and significance of the BA proportion coefficient remains very similar to what was observed in column 1. This implies that the association between the level of education of the population and the organizational form happens within industries (as one would expect with a General Purpose Technology) rather than through shifts in the industrial struc-

[27]More specifically, industry is measured by the first digit of Standard Industrial Classification 1992; we have 5 categories of workplace size: <25,25-49,50-249,250-999,1000+. Whereas workplace size refers to the number of employees at the specific site, the organization may have multiple sites and hence many more employees. We have 5 categories of organization size: <50,50-249,250-999,1000-9999,10000+.

Table 2.4: Workplace-level regressions of the influence index

	(1)	(2)	(3)	(4)	(5)
Current %BA in the TTWA	0.92***	0.95*	1.022***	1.01***	0.94***
	[0.17]	[0.50]	[0.16]	[0.16]	[0.30]
Wave=2004	0.30***	0.30***	0.28***	0.44***	0.30***
	[0.049]	[0.045]	[0.047]	[0.14]	[0.050]
Wave=2011	0.47***	0.47***	0.42***	0.77***	0.47***
	[0.062]	[0.066]	[0.064]	[0.22]	[0.064]
Current %HS people in TTWA		0.047			
		[0.70]			
Dummies for workplace size, organization size and industry (1digit)	No	No	Yes	Yes	Yes
Full interactions between industry (1digit) and wave	No	No	No	Yes	No
Including London	Yes	Yes	Yes	Yes	No
Observations	5,332	5,332	5,332	5,332	4663
R-squared	0.064	0.064	0.155	0.165	0.156

Notes: All regressions are at the workplace level, with standard errors clustered at the TTWA level. Each workplace is weighted by its employment weight. For both the BA proportion and the HS proportion at TTWA, the variable is the proportion of economically active people (working or unemployed) in that education group from the LFS in the two calendar years prior to the year of the dependent variable. For example, for workplaces observed in 2004, the BA and HS proportions are from the LFS 2003-04.
Source: Authors' analysis of the UK Workplace Employment Relations Survey.

ture. In the fourth column, we further include interactions between industry and wave and the key estimate remains essentially unchanged. Finally, we are concerned that our results are being driven primarily by London as a potential outlier which contains a large number of observations and has both high education and high use of more modern technologies. However, omitting London, in column 5, does not alter our results.

In Table 2.5 we report results with the dependent variable generated either only from the responses of the BA employees or only from the non-BA employees' responses. The specification includes industry, size, and year effects as in column (3) of Table 2.4, and we try both weights based just on establishment size and weights based on employment in the specific education group. The results indicate that the positive correlation between BA proportion and employees' influence at workplace observed in the earlier specifications is not a mechanical result from a combination of BAs having more influence than non-BAs and an increasing proportion of BAs. In fact, the influence over work decisions reported by non-BA employees in their workplace is even more positively correlated with the local supply of BAs than for BA employees. Again, this fits with the idea that under the older, centralized organizational form, BA employees would have had managerial or quasi-managerial roles and, thus, some control over decision making. It is the non-BA's who will experience the greatest change in the shift to a decentralized workplace.

Table 2.5: Influence by education of employees

	Influence Reported By			
	BA employees	non-BA employees	BA employees	non-BA employees
Weighted by	Establishment employment		BA employment*	non-BA employment*
%BA in	0.60***	0.86***	0.24	0.93***
the TTWA	[0.21]	[0.24]	[0.20]	[0.29]
Wave=2004	-0.011	0.28***	0.0088	0.29***
	[0.047]	[0.040]	[0.041]	[0.0429]
Wave=2011	0.021	0.42***	0.092**	0.43***
	[0.052]	[0.052]	[0.045]	[0.059]
Observations	4,197	5,175	4,197	5,175
R-squared	0.08	0.121	0.062	0.143

Notes: The set of controls in each regression is the same as in column (3) in table 2.
* In the first two columns, each workplace is weighted by its employment weight. In the last two, the weight is multiplied by the proportion of employees in that education group. If a workplace does not have any employee in the education group responding to the influence questions in the employees' survey, it is omitted from the respective regression.
Source: Authors' analysis of the UK Workplace Employment Relations Survey.

Whether the estimated association between the local BA proportion and the average influence index value in these regressions represents a causal effect of the level of education is unclear. More educated workers may migrate to areas where firms have more decentralized organizational structures, implying a reverse causality. Alternatively, there could be a third unobserved factor prevalent in some areas that both increases the attractiveness of using a decentralized form and is attractive to more educated workers. We find it difficult to determine what form such a factor would take given that we are already controlling for industrial structure and firm size. In addition, the fact that our results hold up when we drop London (which is a strong candidate as a place where more educated workers migrate to with the aim of working for the most up-to-date firms) is weak evidence against the first endogeneity channel. Nonetheless, we are concerned that there is remaining endogeneity.

To address any remaining endogeneity, we adopt two approaches.[28] The first is to include the value of the dependent variable (the mean value of the influence index) in the first year for which we have it (1998).[29] One can interpret this variable as a parameterization of location fixed effects that uses only the part of the fixed effect that is correlated with the historic mean level of worker control over their workplace.[30] Thus, we compare two regions with the same initial level of use of decentralized organizational forms as a means of holding constant a general proclivity to use such forms for time-invariant reasons and ask whether the region that had a greater increase in the proportion of workers with a BA saw a larger proportion of firms increase the extent of their decentralization. The results without industry and firm size controls are given in column (1) of Table 2.6 and the results including those controls are given in column (2). The estimated effect of the proportion BA is again highly statistically significant and takes a value of about two-thirds of the comparable estimates in the first and third columns of Table 2.4. Thus,

[28]We implement these approaches using data aggregated to the TTWA. Estimation using data at the firm level with clustered standard errors yields very similar results.

[29]Since we have to drop the first year of our data, we are left with firm observations across only two years of data.

[30]Direct fixed effect estimators yield erratic and ill-defined coefficient estimates which we interpret as arising from the shortness of our panel.

the proportion BA variable is picking up longer term differences in the extent of use of decentralized forms to a limited degree and not enough to overturn our conclusion that increases in the proportion BA induces a movement toward those forms. Interestingly, the historical use of decentralized forms itself has only a weak relationship with future use of those forms in a region.

Table 2.6: Influence Index Regressions: Initial Value and IV Estimates

	(1)	(2)	(3) IV
Current %BA	0.62***	0.65*	1.08***
in the TTWA	[0.20]	[0.23]	[0.37]
Wave=2004	-0.21***	-0.17***	0.28***
	[0.049]	[0.048]	[0.057]
Wave=2011			0.41***
			[0.075]
Influence Index in 1998	0.025	0.051	
	[0.052]	[0.047]	
Dummies for workplace size, organization size and industry (1digit)	No	Yes	Yes
Including London	Yes	Yes	Yes
Observations	390	390	672
R-squared	0.119	0.278	
First Stage F-stat			15.35
Associated p-value			0.000

Notes: All regressions are at the TTWA level, weighted by employment, with standard errors clustered at the TTWA level. The instruments in columns (3) are: the population share of the 1970-74 birth cohort; the population share of the 1975-79 birth cohort; the proportion of BA educated individuals in the parents' generation; and the proportion of GCSE/O-level holders in the parents' generation. All the instruments are measured at the TTWA level in 1995-6.
Source: Authors' analysis of the UK Workplace Employment Relations Survey.

Our second approach is to implement an instrumental variables (IV) estimator. In particular, we make use of variation across areas that relates to the expansion of education. As instruments we use the proportion of the population born in the years 1970-74 and the proportion born between 1975 and 1979, measured in 1995-96.[31] The underlying idea is that the proportion of the population with a university degree expanded substantially for the 1970s cohorts. As a result, areas with a high concentration of people of university age at the time of the expansion in the higher education system would be predicted to have a more educated population later to the extent that people have some tendency to stay where they grew up. In addition, we use the educational composition of people in the generation who would likely be the parents of these cohorts (people born between 1945 and 1954). In particular, we construct the proportion of the parental generations who have a BA and the proportion who have a GCSE/O level, again measured in 1995-96. We also include the interaction of these parental education variables with the size of the 1970s birth cohorts in the area. The idea behind the instruments is that areas that one would predict to have a large increase in the proportion of BAs in their workforce between the early 1990s and the early 2000s are ones where there is a local baby boom in those generations and where the parents own education indicates that they would be interested in their children's education. For this set of instruments to be valid, we require that parents in the previous generation - and, in particular, more educated parents - did not have a tendency to have more children in areas which would later turn out to have more decentralized organizational structures. We also require that the parents did not locate in an area because it would undergo a shift toward a more decentralized organizational form several decades later, as part of a shift to a technology that did not even exist at the time at which most of them made their location choice. The fact that we control for industry and firm size effects in these regressions eliminates any concern that their location might have been related to persistent concentration in industries that would ultimately favour decentralization. We view the conditions under which this instrument set fails as very stringent. In particular, we find it hard to come up with

[31] The denominator for the proportions is the population born between 1940 and 1979.

situations in which differences in cohort sizes across areas are determined by the conditions that would affect the adoption of decentralized organizational forms decades later, especially after we control for industrial structure. The set of instruments are highly significant in the first stage, with p-values associated with the F-statistic for the test of their exclusion being effectively zero.

Column (3) in table 2.6 contains the results from our IV specification. The estimated coefficient on the proportion BA is 1.08, which is very similar to the value estimated with OLS in column (3) of table 2.4. This fits with our belief that endogeneity is not a substantial concern once we control for industry and firm size effects.

Our overall conclusion from our estimates is that an increase in the proportion of the working age population with advanced education in a region causes firms in that region to increase their use of decentralized technologies, with the effect being on the order of a 10 percentage point increase in the percentage of the working age population with a BA generating a 0.1 standard deviation increase in the extent to which workers feel they control their own work. This fits with results in Caroli and Van Reenen (2001) where they use UK and French data to show that a relative shortage of educated workers in a local labour market, as reflected in a higher education wage differential, implies that the firms in that market are less likely to implement organizational change. We view their results and ours as corroborating evidence for our model in which the large increase in the education level of new cohorts born after the late-1960s generated a shift in organizational structure toward a more decentralized structure in which workers had more control over their own tasks. As we have seen, in such a model, the technological shift can be accomplished without a change in the wage differential between more and less educated workers. So far our micro evidence looks at worker autonomy, future research may examine a wider-range of measures for skill-biased technology, such as ICT in Draca et al. (2007) and R&D Machin and Van Reenen (1998).

2.5.3 Other technological followers

To this point, we have presented evidence for a claim that the UK's combination of a rapid educational upgrading with no accompanying change in the education-wage differential can best be understood in the context of a model of technological choice in which the UK is a technological follower choosing among technologies developed elsewhere (most likely the US). But the UK is not the only economy to undergo a substantial increase in its education level after the US, and it is worth asking whether other economies experiencing such an increase also have patterns fitting with them being technological followers.

To address this question, we use data from the OECD on the education levels and education-wage differentials for advanced economies between 1997 and 2010 (OECD (2012)). The data is from the labour force surveys for the member economies and is restricted to 25 to 64 year olds. The period is chosen both because it is one in which we can obtain consistent data and because it roughly matches the period of substantial growth in the UK's education level. That is, it is a period in which other economies also experiencing such growth would face the same set of existing technological choices. In this period, 11 other OECD economies both started the period with a proportion of their population with a tertiary education that was lower than that of the US in 1997 and experienced an increase in that proportion of at least 40%.[32] The lowest increase country meeting these requirements was Belgium (rising from 25% of its population having a tertiary education in 1997 to 35% in 2010) and the highest was Poland (moving from 10% with a tertiary education in 1997 to 23% in 2010). The OECD data indicates a 65% increase for the UK from 1997 to 2010, which is very close to what we obtain using the UK LFS for the same period (71%).

We examine movements in the wage ratio between the mean annual earnings of all workers aged 25 to 64 with a tertiary education and the mean annual earnings of

[32] The countries are: Australia, Belgium, France, Ireland, New Zealand, Norway, Poland, South Korea, Spain, Sweden, and Switzerland.

workers with an upper secondary education being their highest education level. We regress this ratio on a simple linear time trend to summarise the wage differential pattern that coincides with the rapid educational growth in these economies. Out of the 11 OECD economies meeting our education growth criteria[33]: the time trend coefficient is not statistically significantly different from zero at the 10% level or below in 7; two exhibit statistically significant positive trends; and two exhibit statistically significant negative trends.[34] According to the time coefficient regressions, for Poland - the country with the largest percentage increase in tertiary education - the wage ratio fell by 1.8 percentage points per decade (from a base of 170). At the other end, for the country with the smallest educational increase - Belgium - there was a 1.5 percentage points increase per decade in the wage ratio (on a base of 130). The other economies with statistically insignificant time trends for the wage ratio show negative and positive point estimates that are either smaller or somewhat bigger than these two examples. We present the full set of estimated coefficients in Online Appendix. In the online appendix, we also show how our estimates fit with previous results in Crivallero (2016), who estimates very small effects of increases in university attainment on the college wage premium in a pooled sample of 12 European economies, and Chen (2013), who shows that Taiwan also underwent a large increase in educational level with no accompanying change in its college premium.

Taken together, we believe the results in the OECD data and in other papers are consistent with our model for many economies undergoing substantial increases in their education levels. We make no claim that our discussion provides a complete analysis of

[33] This includes the UK. We drop Australia because there are only 3 earnings observations in our period. For the other countries, the wage ratio data is for the years 2000 to 2010, with some missing years in most economies.

[34] The countries with flat wage ratio profiles are: Belgium, France, Ireland, New Zealand, Poland, Switzerland and the UK. The two countries with positive trends are: South Korea and Spain. And the two with negative trends are: Norway and Sweden. Regressions of the wage ratio on the proportion with a tertiary education generate the same pattern of insignificant, significantly positive, and significantly negative coefficients on the education variable.

the determinants of wage movements in these economies. The number of observations for each country in the OECD data is small and we do not investigate factors such as the level of decision making of workers, as we do for the UK. Nonetheless, we think the fact that there are so many economies which both start our period behind the US in their education levels and experience substantial educational growth but do not have statistically significant changes in their wage ratios indicates that it is plausible that other countries could also be described in terms of our model with educational catch-up driving endogenous technological choices. A more complete investigation of this hypothesis for other economies is beyond the scope of this chapter.

2.6 Conclusion

In this chapter, we highlight two empirical patterns: first, the UK underwent a dramatic increase in the proportion of the working age population with a BA since 1993; second, the BA-to-HS wage differential was essentially unchanged over this period. The combination of increased educational supply and a lack of movement in the educational wage differential necessarily implies a skill biased demand shift over time. We consider three models of technological change that imply skill biased demand shifts: the canonical model in which the demand shift is exogenous; a model in which the increase in education induces new skill favouring inventions; and a model in which a variety of technologies already exist, with firms choosing which to implement. We argue that the core patterns in the UK data do not fit with exogenous technological change models, including those that incorporate tasks. Moreover, because the growth in educational attainment varies over time, the exogenous technological change models require that the rate of technological change has to speed up and slow down in just the right way to generate the pattern that we observe of an unchanging college premium throughout the post-1993 period. Of course, we cannot reject a claim that there just happened to be such a variation in the exogenous rate of technological change but we view it as improbable.

Of the remaining, endogenous technological change models, we believe that models of induced invention may be relevant for the US in recent decades since it was in a position to be a technological leader in skill biased technologies by virtue of having a much more educated work force than other developed economies at the dawn of the computer era (Beaudry et al. (2006)). In contrast, the UK underwent its educational expansion much later and, as a result, we believe it is plausible that it was a technological follower for this type of technology - following an induced technological adoption model rather than one of induced innovation.

More explicitly, we argue for a model for the UK in which firms in any sector can choose to produce using a centralized or a decentralized organizational structure as discussed in papers such as Caroli and Van Reenen (2001) and Bloom et al. (2012). In the decentralized structure, workers need to be able to take individual initiative and control their own work - characteristics that we view as fitting more with higher educated workers. The model has a similar construction to a classic trade model in that the economy responds to a shift in the relative supply of more educated workers by shifting toward greater use of the decentralized organizational structure. And, as in the trade model, there is no adjustment in terms of relative wages or wage levels. But the model also has further implications; most notably that the proportion of managers who have a BA should increase but the proportion of BA's who work as managers should decrease as the decentralized technology spreads. The model also implies strong restrictions on the shape of the aggregate production function that we show hold for the UK in this period. In addition, we show that areas in the UK which had more substantial increases in education levels are also areas where workers report having more control over their own work - something we see as a marker of a decentralized workplace. Importantly, this pattern occurs within industries, not because of shifts in the industrial structure, and is robust across a range of specifications. We develop an instrumental variable strategy in which we instrument for area specific educational changes using differences in fertility in previous decades and parental education. We believe these instruments are very likely to be valid, based as they are on an assumption that parental decisions on fertility in the 1960s and 1970s did not arise from predictions of decentralized technologies coming to their areas in the 1990s and after. Again, it is important that we control for industry in all our specifications, implying that parents would have to make their guess about future technology use independently of the local industrial structure for our instrument to be invalid. The IV results indicate that increases in the education level in a local economy have a causal impact on the adoption of decentralized organizational forms by firms in that economy.

The key point we see as arising from this exercise is that the effects of technological change are not one size fits all. There are good reasons to believe the US has been a technological leader and there has been considerable study of the interactions of technological change and educational supply shifts in the US. The question then becomes, can the experience of the US be generalized to other countries? The UK provides an interesting case study to examine this question. Its large expansion in education happened quickly and well after the main expansion for the US. Because of that, we believe that the UK provides evidence on what happens to technological followers as their conditions shift toward favouring the technologies that the leader has developed. We argue that during the transition period for a follower economy, one could observe no real impact on skilled wage differentials even though the economy was being substantially transformed. Our evidence lines up with this interpretation. We believe this calls into question approaches in which technological change effects are identified from commonalities in wage and employment movements across countries, with remaining differences assigned to differences in institutions and differences in supply shifts. This does not mean that there are no commonalities across economies and that we should devolve to studying each economy in isolation. Instead, we view our results as indicating the need for a broader view of the impact of technological change - one which emphasizes the role of differences in movements in relative factor supplies in determining the point in the lifecycle of a technology at which each economy adopts it.

Chapter 3

Occupational polarisation and endogenous task-biased technical change

3.1 Introduction

In many developed countries since the 90s, employment has shifted substantially away from middle-paying occupations towards both the top and the bottom (Goos et al., 2014). This phenomenon - employment polarisation - has been an important factor in rising income inequality. In the polarisation literature, the leading explanation is Routine-Biased Technological Change (RBTC). The basic idea is that technological change (embodied by IT and automation equipment) has displaced workers in carrying out routine tasks, which are important in middle-paying occupations. There is a large literature on this, both theoretical and empirical.[1] Most of it interpret RBTC as a consequence of increasing availability or productivity of capital equipments, or their declining costs. This chapter offers a new perspective, which emphasises the role of increasing skill supply in the diffusion of routine-biased technology.

The idea that firms' choice of production technology depends on the supply of skills is supported by a growing literature (Beaudry et al., 2010; Lewis, 2011; Akerman

[1] Acemoglu and Autor (2011) and Autor (2022) provide a good summary.

et al., 2015).[2] This chapter builds on the RBTC framework by allowing the adoption of technology to respond to skill supply shifts. This allows supply shifts to have different effects on the labour market than alternative theories of RBTC. The UK provides a uniquely-suitable empirical context to examine this, because it has experienced a huge expansion of higher education since the early 90s, and for about two decades student numbers were set by the government. By endogenizing the adoption of technology, my model can explain not just employment polarisation, but three facts about occupations in the UK at the same time.

First, the pattern of employment polarisation in the UK is essentially a shift from middle-paying occupations to high-paying ones, with very little change in low-paying ones. If we broadly classify occupations into 3 groups, the total employment share of the middle group fell by 10.5 percentage points between 1997 and 2015, while that of top group increased by 10.3 percentage points.

Second, there has been no wage polarisation in the UK since the mid-90s.[3] In fact, the movements in wages are uncorrelated with those in occupational employment in the UK. With endogenous adoption of Task-Biased Technical Change (TBTC), my model implies that skill supply shifts will primarily cause shifts in the occupational structure and have less impact on wages.[4] Thus, it is consistent with job polarisation and a lack of wage polarisation at the same time.

The 3rd stylised fact is more striking: the huge increase in educational attainment since the early 90s has led to relatively little occupational downgrading for graduates in the UK. We will see in section 3.2, when the proportion of graduates in the workforce doubled from about 20% in the early 90s to over 40% in the mid-2010s, the share of

[2]Typically, these studies use exogenous geographical variation in the supply of educated workers to prove the causality from skill supply to technology adoption.

[3]Goos and Manning (2007) found substantial growth in both ‘lousy and lovely jobs’ over the 80s and 90s, but wages in lousy jobs were clearly falling relative to those in the middling jobs over their sample period.

[4]I call it TBTC because my model does not presume that technical change is biased against routine tasks per se; the direction of bias will be estimated from the data.

graduates employed in abstract occupations has been stable around $75-80\%$. This phenomenon is consistent with my model, and it would be harder to rationalise in models of exogenous technical change.[5] In fact, each of the three phenomena can be explained by alternative theories, but together they paint a picture consistent with the explanation proposed here. In section 3.5, I will discuss some regression results that reject the hypothesis of exogenous task-biased technical change and support my model.

The model proposed here is an equilibrium model with endogenous adoption of task-biased technologies. In each industry, firms choose between two technologies, which differ in task intensities. The choice depends on task prices. When the skills distribution shifts in a way that favours a certain task, firms may switch towards the technology that's more intensive in that task, thus the resulting impact on prices and wages will be smaller than if technology is fixed. In other words, the endogenous adoption of technology helps absorb supply-side shocks, so the effects are seen in the relative quantities of tasks rather than prices. On the supply side, workers' productivity depends on two dimensions of observable skills and an unobserved general ability, and they choose their occupation based on their comparative advantage and task prices. This means when task prices do not move, the mapping from skills to occupation will not, hence delivering the 3rd fact as a result of the 2nd fact.

The idea that the *adoption* of technology responds to supply-side shifts has not been explicitly investigated in the polarisation literature.[6] Most of the literature has interpreted the pervasiveness of employment polarisation across developed countries as a result of a global technology shock, while attributing the differences in wage trends to unspecified differences in institutions or differences on the supply side(Autor et al.,

[5]It would require the exogenous technical change to increase the demand for abstract tasks at the same time and by the same magnitude as the supply-side shift.

[6]Some papers (Hardy et al., 2018; Salvatori, 2018) have argued for a major role of education increase in the growth of cognitive or high-paying jobs in Europe/UK, by means of shift-share decompositions of over-time changes into between and within components. This chapter uses an equilibrium model to provide a clearer conceptual distinction between supply-side shifts and demand-side shifts.

2003; Goos et al., 2014; Goos and Manning, 2007; Acemoglu and Autor, 2011). As we'll discuss below, many developed countries experienced employment polarisation without wage polarisation (Green and Sand, 2015). By contrast, the chain of events emphasised here starts with a policy-driven positive supply shift, which causes task-biased technical change, and therefore leads to the three aforementioned facts about occupations. This is not a rejection of the hypothesis that technology shocks coming from cheaper machines are routine-biased. Such exogenous technical change is still allowed in my model; but the emphasis here is how the adoption of technology responds to supply-side shifts. This new feature is important because it yields different predictions for how a supply-side policy would affect the labor market. My model's ability to explain all three facts about occupations in the UK gives us confidence that it is a reasonable model for analysing potential policies in the UK, such as skill-based selection of immigrants.

The model also features flexible sectoral shifts. Because industries differ a lot in task intensities[7], differential productivity and demand trends in different industries may affect relative demand for different tasks. Exogenous factors such as population ageing and Chinese imports may lead to rising demand for personal services and falling demand for manufacturing goods. My counterfactual analysis suggests between-industry demand shifts played a big role in occupational polarisation over the period. I allow 7 industries when applying my model to the data. This is a finer disaggregation than most papers in the polarisation literature. For example, Autor and Dorn (2013) distinguishes between low-skill services and the rest. Barany and Siegel (2018) built and calibrated a model of 3 sectors: low-skilled services, manufacturing and high-skilled services. They show that sectoral shifts could explain a large part of the changes in occupational employment shares and in occupational wages in the US since the 1950s.

The idea that technical change is endogenous is not new. Acemoglu (1998, 2002,

[7]For example, finance is intensive in professional task, and construction is intensive in skilled trades.

2003) argued that the extent of skill bias in the new technology is endogenous, which explained the apparent acceleration of skill-biased technical change after an initial increase in the supply of skills in the US. While such models of endogenous innovations are suitable for a big country on the technology frontier, like the US in the last 100 years; for countries that are technological followers, models of endogenous adoption of available technologies may be more suitable. I believe overall the UK belongs to the latter group over my sample period 1997-2015. The UK had a much lower proportion of graduates than the US in the early 90s and have surpassed it by the end of my period. Moreover, models with endogenous innovations imply a downward-sloping short-run demand curve (just like the case with exogenous technology) and a flatter or upward-sloping long-run demand curve; whereas my model with endogenous adoption implies a flatter short-run demand curve. The broad trends observed in the UK support the latter.

This chapter also relates to a growing literature on endogenous adoption of specific technologies and its effects on employment or wages. They usually focus on a tangible technology, such as personal computers (Beaudry et al. (2010), Borghans and ter Weel (2008)), broadband internet (Akerman et al., 2015), automation (Aghion et al., 2020), or industrial robots (Graetz and Michaels (2018), Humlum (2019)). They often find that the adoption of technology was indeed affected by the local supply of skills or local wages. Their research questions centre around the causal effects of adopting that technology on employment, wages, productivity and so on.[8] By contrast, this chapter aims to explain overall patterns in all parts of the economy in a unified framework. So I choose not to focus on one specific technology. In my model, technology boils down to the production function that combines tasks into output.[9] In each industry, there will

[8]Most of these papers did not model general-equilibrium effects. To my knowledge, Humlum (2019) was the first to estimate a general equilibrium model of technology adoption. His model is rich in how manufacturing firms choose whether to adopt robots and parsimonious for the rest of the economy. Specifically, the production function outside manufacturing is Cobb-Douglas and contains no task-biased technical change.

[9]We do not model capital explicitly in this chapter. We can think of the choice of capital equipment as a choice of the function that combines occupational labor into output. For example, adopting robots

be an ‘Old’ technology and a ‘New’ technology. I believe technical changes manifest differently in different firms. It could be automation equipment in manufacturing, some software in financial services, and some sort of organisational restructuring in another services firm. And all those kinds of technical changes may be complementary to each other (Bresnahan et al. (2002), Caroli and Van Reenen (2001)). Empirically, we will use a wide range of proxies to measure the share of the ‘New’ technology at the industry-year level.

The paper is most closely related to Chapter 2 (published as Blundell et al. (2021)). Chapter 2 noted that the rapid growth of graduate numbers in the UK had no noticeable impact on graduate wages, and explained it by an endogenous adoption of skill-biased technical change. This chapter uses the same intuition but in a different context, because the aim here is to explain three facts about occupations and to allow policy analysis. There are three other differences. First, the model in chapter 2 has two labour inputs (graduates and others), whereas the model here is about occupational tasks and it features multiple industries. Second, in this chapter each worker has 2-dimensions of observable skills: analytical and social skills are what matters for productivity, not education per se.[10] This opens up the possibility of modelling changes in skills distribution within education groups over time. Third, chapter 2 did not estimate or calibrate its model, whereas I do. This means I can simulate the effects of counterfactual policies. Overall, this chapter corroborates Chapter 2 with a richer model and more supportive empirical evidence. In addition to chapter 2, Carneiro et al. (2018) and Dustmann and Glitz (2015) also found that production technology responds to changes in the local supply of educated/uneducated workers. Like chapter 2, they differentiate labor by education and have nothing to say about occupations.

Currently, the model is partly calibrated and partly estimated, at the level of 9 oc-

in the production process could mean you would need more technicians and fewer production workers to produce one unit of output.

[10] The model also allows an unobserved general ability, that can vary across individuals freely.

cupations and 7 industries. The 9 occupations are SOC2000 major occupation groups: 1, managers and senior officials, 2 professional, 3 associate professional and technical, 4 administrative and secretarial, 5 skilled trades, 6 personal services, 7 customer services, 8 process, plant and machine operatives, and 9 elementary. The model fits the UK trends pretty well. The good fit is not mechanically guaranteed by the model design, because most of the key parameters do not vary over time. The estimates suggest that technological change in the UK over 1997-2015 was biased against all three routine tasks, favoured managerial and professional tasks, and neither favoured nor biased against the remaining four (3 manual tasks and technicians).[11] Counterfactual analysis suggests that the shift in skills distribution alone can account for between a third and two thirds of the decline of manual routine occupations, and between a third and half of the increase in the three abstract occupations. The shift in the industry demand could account for similar magnitudes of occupational shifts.

The model offers a suitable framework to investigate some counterfactual policy questions. For example, how will further increases in higher education affect the wage structure? If the UK government selects EU immigrants by skills[12], how will it affect the labour market? Such counterfactual scenarios are outside the range of historical observation and are difficult to answer by reduced-form methods. I plan to use the UK Skills for Life Survey to obtain the differences in literacy and numeracy skills between immigrants and natives, and to examine whether the skills distributions within education groups have deteriorated over birth cohorts.

For reasons that will be discussed in Section 3.2 and Conclusion, this chapter also provides a promising framework that can be used to study occupational trends in other European countries. The paper is structured as follows. Section 3.2 documents three

[11]This direction of biases are consistent with previous studies. For example, Humlum (2019) estimated that in Danish manufacturing, robot adoption reduced the productivity of production workers but increased that of tech workers (engineers, researchers and skilled technicians).

[12]It already does for non-EU immigrants, and now after Brexit it will be able to reject EU immigrants on skill-related criteria.

phenomena in the UK labour market, with comparison to other developed countries where possible. Section 3.3 develops the model and explains how to identify the technology trend and the model parameters. Section 3.4 describes the data sources, including the proxies used to impute the share of the 'New' technology. Section 3.5 investigates various correlations in micro data, which are supportive evidence for the model. Section 3.6 explains how various parameters are estimated, discusses some key estimates and the fit of the model, and conducts counterfactual analysis. Section 3.7 concludes.

3.2 Motivating facts

This section documents three phenomena in the UK labour market since the 90s.

1. Employment has shifted significantly away from middle-paying occupations towards primarily the higher end and to a lesser extent the lower end.

2. There is no clear U-shaped pattern in occupational wage changes during the period of employment polarisation.

3. The huge increase in education attainment in the UK has not led to much occupational downgrading, nor decline in the skilled wage premium.

3.2.1 Fact 1: employment polarisation

Employment polarisation refers a 'hollowing out' along the occupation spectrum. This phenomenon has been documented extensively in the literature for the US (Acemoglu and Autor (2011), Autor and Dorn (2013), Hershbein and Kahn (2018)) as well as many other developed countries (Goos et al. (2014), Breemersch et al. (2017), Michaels et al. (2014)). It's been documented since the 1980s for the UK (Goos and Manning, 2007) and for Germany (Kampelmann and Rycx, 2011) and even earlier in the US (Barany and Siegel (2018)). The phenomenon is robust to different ways of classifying and ranking occupations for both the US and the UK. When my model is brought to the UK data, occupation will be at the level of SOC2000 major occupation groups.[13] So in this section I present occupational facts at this level, too. At this level of nine occupations, the three middle-paying occupations are normally considered 'routine': 'Administrative and Secretarial Occupations', 'Skilled Trades Occupations', and 'Process, Plant and Machine Operatives'. The three high-paying ones will be referred to as 'abstract', and the low-paying ones as 'manual'.

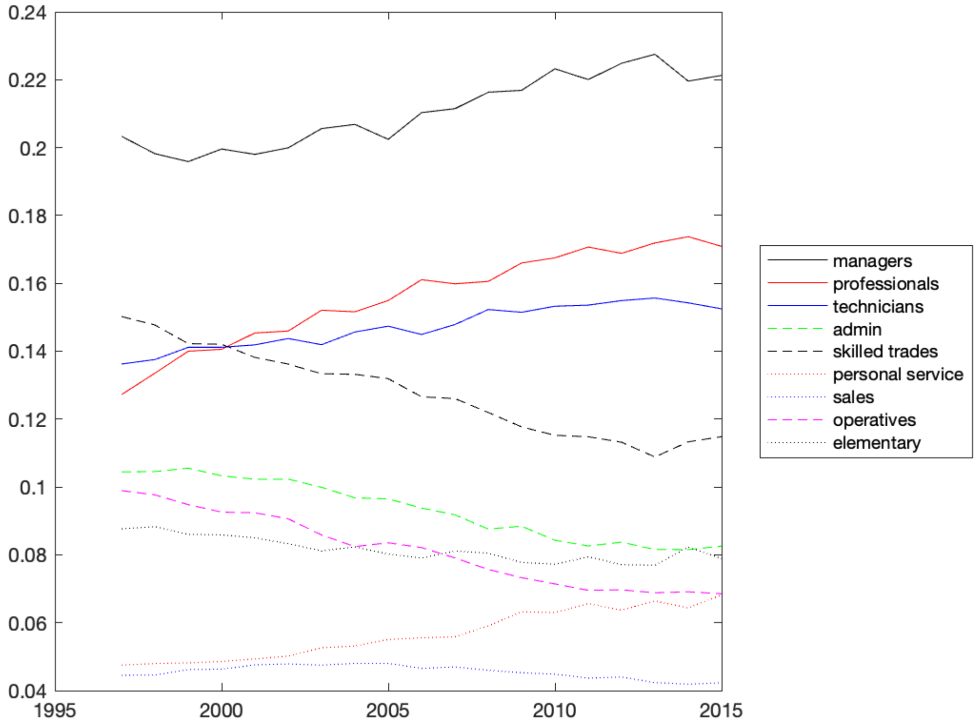
Figure 3.1 shows that each of the three routine occupations saw a very substantial decline in employment share. Over 1997-2015 (the period for which my model will be estimated), the total employment share of the 3 routine occupations fell from 39.1% to 28.5%. Meanwhile, each of the three abstract occupations grew substantially. In particular, professional occupations grew from 9.9% of aggregate employment to 15%. Together, the abstract employment share grew from 39.1% to 49.4% over the sample period. Among the manual occupations, there is some decline in elementary occupations[14], which is more than compensated by the increase in personal services (such as care assistants). Overall, the pattern of employment polarisation in the UK is more of a shift of employment from the middle to the top, with little change in the bottom.

At a similar level of aggregation, Figure B.1 shows a V shape in employment growth across ISCO occupation groups in a number of European countries over 2002-14. This echoes the findings in Goos et al. (2014), which looked at 16 European countries and documented pervasive occupational polarisation over 1992-2010. On the other hand, some more recent studies looking at employment changes in European countries found no polarisation pattern but 'occupation upgrading' - meaning fastest growth in the 'best' jobs and weakest growth in the 'worst' jobs. For example, Fernández-Macías

[13]There are 9 occupations in total : 1, managers and senior officials, 2 professional, 3 associate professional and technical, 4 administrative and secretarial, 5 skilled trades, 6 personal services, 7 customer services, 8 process, plant and machine operatives, and 9 elementary.

[14]which include labourers in agriculture, cleaners, waiters, kitchen assistants, labourers in construction, porters, postal workers and so on.

Figure 3.1: Employment shares by occupation



Note: the 9 occupations are major occupation groups under SOC2000. See section 3.4 for how we adjusted for discontinuities in SOC over 2000-01 and 2010-11.

and Hurley (2017) looked at 23 European countries over 1995-2007 and found polarisation in a handful of countries but the most common pattern is occupational upgrading. Oesch and Piccitto (2019) looked at UK, France, Germany and Spain over 1992-2015 and found job growth was by far the weakest in the 'lowest-quality' jobs using a range of measures of job quality.[15] Murphy and Oesch (2018) looked at Ireland and Switzerland over 1970-2010 and found 'occupational upgrading', and the patterns were consistent with changes on the supply side from women's education and immigration. It's beyond the scope of this chapter to investigate why those studies reach different conclusions. Notably, they all point to strong growth in high-paying occupations. We see in both

[15]The only exception is for the earnings-based indicator in the UK, which suggests a polarising pattern.

Figure 3.1 and Figure B.1 that the professional occupation stands out as having the strongest growth. This is an occupation in which university graduates are likely to have comparative advantage. In the framework proposed here, an increase in the supply of graduates will cause firms to adopt a technology that's more intensive in professional tasks, and therefore the professional employment share will increase. My model does not have a definitive prediction as to whether low-paying occupations should grow or decline relative to the middle. Both 'occupational polarisation' and 'occupational upgrading' could be the consequence of an increase in skills supply. The former follows if the new technology is biased against middle-skilled tasks and in favour of high-skilled tasks; while the latter follows if the new technology is biased in favour of high-skilled tasks and against low-skilled tasks.

3.2.2 Fact 2: no wage polarisation

Meanwhile, apart from the US, there is no such V shape in occupational wage growth in other developed countries that also saw employment polarisation.

Figure 3.2 ranks the 9 occupations from the lowest paid to the highest paid, and plots the occupational wage growth in red markers. The plotted wage changes are net of compositional shifts in education, age and gender.[16] I plan to use another dataset to estimate occupational wages with individual fixed effects in the near future.[17] The three low-skilled occupations have slower wage growth than 5 of the other 6. Skilled trades and operatives have fairly strong wage growth, while admin did have the slowest wage growth. The maximum difference between occupations in log wage changes over 1997-2015 is just under 0.08. This is small relative to the observed changes in employment shares.[18]

[16]In each year, I have regressed log wages on those demographics and occupation dummies. The coefficients on occupational dummies are interpreted as 'composition-adjusted' occupational wages.

[17]This has been delayed due to data access issues during the pandemic.

[18]To give a sense of magnitude, if tasks are neither complements nor substitutes, the response of the log wage ratio to log quantity ratio along the demand curve would be -1. Assuming no demand shift, an increase in the log quantity of professional tasks by 0.5 (its employment share increased from 10% to

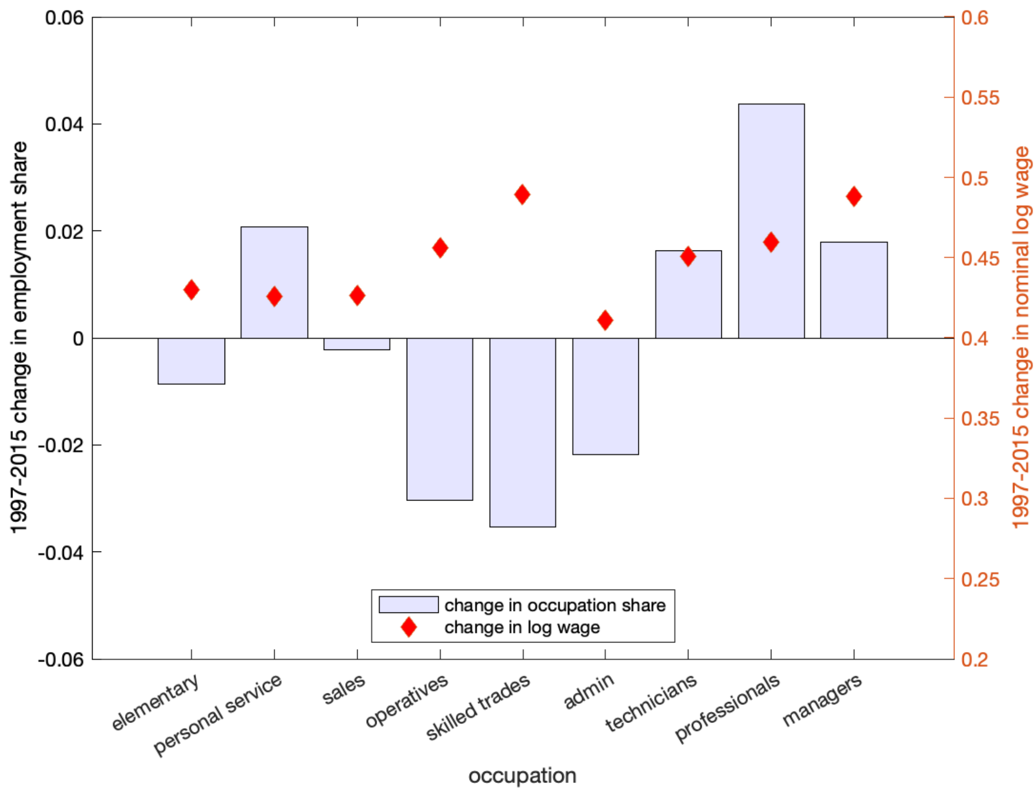
In other European countries, we have not seen wage polarisation since the 80s either. Before 2000, wage inequality increased across the distribution, in the UK during the 80s and 90s (Goos and Manning, 2007), in Germany in the 90s (Dustmann et al., 2009), and in Canada (Green and Sand (2015)). In fact, Green and Sand (2015) summarised that occupational wage polarisation was only observed in the US in the 90s, and not elsewhere or in other decades. After the turn of the century, there was less or no increase in inequality between occupations. In figure B.1, we see in a number of European countries, wage growth over 2002-14 tends to be slightly slower in high-paying occupations such as the professionals. Naticchioni et al. (2014) looked at twelve European countries (subset of EU15) over 1995-2007 and found no evidence of wage polarisation, whether using industry level or individual level data.

The leading explanation for the employment polarisation is routine-biased technical change (RBTC thereafter) (Autor et al. (2003); Acemoglu and Autor (2011), Autor and Dorn (2013), Hershbein and Kahn (2018), Goos and Manning (2007), Goos et al. (2014), Michaels et al. (2014) and many others). Broadly speaking, the hypothesis is that technological changes (such as automation and ICT) were biased against routine tasks, which are important in the semi-skilled occupations around the middle of the distribution. Such a technology-induced demand shock causes polarisation.[19] This has a lot of intuitive appeal, and it fits the polarising trends in employment and wage in the US in the 90s. Guided by the RBTC hypothesis, some papers have asked directly whether occupational wage change correlates negatively with its 'routineness', and the answer is no for Germany and Sweden (Kampelmann and Rycx, 2011; Adermon and Gustavsson, 2015), and yes for the US (Firpo et al., 2011; Böhm, 2020; Acemoglu and Restrepo,

15%) would reduce its log wage by 0.5.

[19]A secondary explanation is the sectoral shift away from manufacturing towards the services. This is also found to contribute to polarisation because manufacturing is more intensive in middle-paying occupations(Autor and Dorn (2013), Barany and Siegel (2018)). But this is also about a shift in the demand curve.

Figure 3.2: Changes in log occupational wages



Note: in each year, we regress log wages on gender-age interactions, detailed education, and occupation dummies. This forms our 'composition-adjusted' occupational wage data $\log Pjt$. Because the occupation classification changed in 2001 and 2011, we then fit each Pjt with a 5th-order polynomial with discrete jumps at 2001 and 2011, and subtract the estimated jump in both pre2001 and post2011 data. Here we show the change in the adjusted $\log Pjt$ between 1997 and 2015.

2021).[20]

The lack of wage polarisation outside the US does not in itself reject the RBTC hypothesis. There are at least four reasons why exogenous RBTC could lead to substantial employment polarisation and no noticeable impact on observed wages, and they could

[20] Adermon and Gustavsson (2015) examined occupational employment and wages in Sweden over 1975-2005, and found that TBTC could explain changes in within-occupation wage differentials but not between-occupation wage differentials. Kampelmann and Rycx (2011) found in Germany, routine jobs have lost employment but there is "no consistent task bias in the evolution of pay rules". By contrast, for the US, Firpo et al. (2011) found that both changes in within- and between-occupation wage differentials in the 90s are consistent with predictions from TBTC. Acemoglu and Restrepo (2021) finds that demographic groups who specialise in tasks that were automated experienced relative wage falls.

be true simultaneously. First, the supply curve could have shifted at the same time in the same direction as the demand due to some exogenous reason. Second, supply could be highly elastic, which would be the case if wage is a key factor in people's selection of occupation and there isn't too high a barrier to switching occupations. Third, wages are sticky for institutional reasons. Fourth, observed wages are confounded by unobserved compositional changes. In particular, the unobserved compositional shift is likely to be negative in expanding occupations and positive in shrinking occupations, and therefore the observed occupational wage changes are attenuated. [21] In the UK case, Cavaglia and Etheridge (2020) has used longitudinal data to estimate task prices for four broad groups of occupations over 1991-2008, and found that the price of abstract tasks has increased by a statistically significant 0.126 log points, while changes in the other task prices are not significantly different from zero.

It's beyond the scope of this chapter to investigate all the plausible explanations of the absense of wage polarization. What this chapter offers is a unified explanation of the facts without deviating from competitive labour markets. It's worth stressing that the changes in occupational wages are not only small and dissimilar to employment changes, they are in fact uncorrelated with the movements in occupational employment in the UK. Section 3.5 further investigates this, comparing my model and a model with exogenous demand shifts. The correlations between occupational wages and employment ratios do not support the latter.

3.2.3 Fact 3: little change in graduates' occupational destinations

Moreover, my framework predicts the 3rd fact, which is harder to explain in alternative model. The fact is the proportion of graduates has increased dramatically in the UK since the early 90s, with no significant deterioration in graduates' relative wage or occupation destinations.

[21] Using German longitudinal data to address the composition shifts, BÃ¶hm et al. (2019) found that the movement in skill price was actually positively correlated with the employment change at detailed occupation level.

This increase in educational attainment was mostly driven by government policy. The vast majority of universities in the UK are publicly-funded: they receive direct grants from the government and tuition fees from students, who can take subsidised loans from the government. The Education Reform Act (ERA) of 1988 changed the funding formula of HE institutions and they responded by increasing their student intake dramatically. Then in 1994, the government introduced student number controls: the number of home students each university could admit every year were capped. The student number controls were raised by the government by a little every year. This resulted in a steady increase in student numbers until the 2010s. In 2012, the cap was abolished for students whose exam grades were above a certain threshold. Since 2015, the cap was abolished for all. Throughout the period, university admission was mostly rationed by prior academic achievement. Figure 3.3 shows that about 20% of workers in the early 90s had higher-education qualifications, and this more than doubled over the next two decades.[22] The pace of increase is much faster than the US.

One might have expected such a big supply-side shift to reduce the relative wage of graduates. In reality, that has not happened. Chapter 2 documents this and explains it in a model of endogenous technology adoption.

One might also expect the huge increase in graduate numbers to lead to 'occupational downgrading', that is, an adverse shift in occupational destinations of graduates over time. However, there has not been much occupational downgrading among graduates in the UK. The right subgraph in Figure 3.3 shows that among graduate workers, the proportion in abstract occupations has been stable over time, at around 80%. There seems to be a little fall after 2010, to around 75% by 2015, which is still very far above the level among high-school workers.

To give a sense of magnitude, I calculate how much the share of abstract occu-

[22] Here 'higher-education' qualifications include both Bachelor's degrees and other tertiary-level qualifications like nursing qualifications. This is a slightly broader definition than 'BA's in chapter 2, because in this chapter we will also need the British Cohort Studies, which only allows this broader definition of graduates.

pations needs to fall within education group if the aggregate abstract share had been constant while the education composition improves.[23] These counterfactual trends are plotted as dashed lines in Figure 3.3: the proportion in abstract occupations conditional on education would need to fall by about a quarter. Thus, the UK story is one where the increase of graduates was quickly absorbed through employment growth in abstract occupations. The model in section 3.3 will formalise this intuition: increasing education leads to an increase in the supply of skills; as those skills are relatively more important in abstract tasks; this would cause firms to switch to the abstract-task-intensive technology and create more abstract jobs.

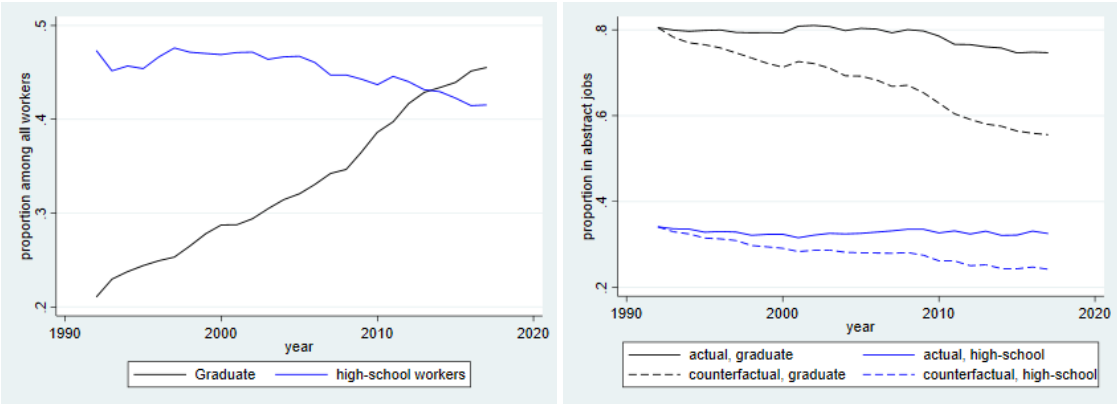
The relatively flat trend documented in Figure 3.3 is robust to the classification of abstract or higher-paid occupations. Consistent with what's shown here, Salvatori (2018) found a 'very small' shift in graduate employment away from top occupations; and Green and Henseke (2016) defined 'graduate jobs' based on skills requirements of detailed occupations and found no increase in over-education following the huge increase of graduates in the UK labour force.

This lack of occupational downgrading among UK graduates is striking when compared against the US. According to Beaudry et al. (2016), in the US the employment rate of cognitive occupations for college graduates fell by nearly 0.1 log point over 2000-2010. The UK trend was basically flat over the 2000s, even though the UK saw a much faster increase of college graduates than the US.

Broadly speaking, most developed countries have seen some increase in tertiary education over the past couple of decades, and the UK is one of the countries with the fastest increase. The US, on the other hand, had the highest level to start with and a slower increase since the 90s compared to most European countries. According to Barro and Lee (2013), the proportion of 15-64 year olds with complete tertiary educa-

[23]In the counterfactual, the education-specific abstract share is proportional to its 1992 level, the aggregate abstract share is at the 1992 level, and the shares of education groups in the workforce are the actual values.

Figure 3.3: Proportion of graduates and their occupation destination



Note: graduates are people with NVQ level 4 qualifications or above. High-school workers refer to those with NVQ level 2 or 3 qualifications. First degrees are NVQ level 4. A-levels and post-16 further education qualifications are NVQ level 3. O-levels and GCSEs (grade C+) are NVQ level 2. ‘Abstract’ refers to the first three occupations in SOC2000: managerial, professional and technicians.

tion was already 24% in the US by 1990, when the proportion in European countries was all below 15%. This supports the view that the US has been the leader of technology in general, with other developed countries closely behind. This means the latter group (including the UK) are in a position to choose between available technologies and this choice would depend on prices and wages. Moreover, Blundell et al. (2021) shows that in 11 OECD countries which experienced substantial increase in tertiary education, there was no significant decline in graduates’ relative wages in 9 of them, like the UK. Finally, it has been documented in Green and Henseke (2021) that in 24 European countries, the share of graduates in non-graduate occupations has increased ‘only modestly’ from 19 to 21 percent over 2005-15.[24] All these similarities suggest that a model of endogenous adoption of technology might be more suitable for these non-US developed countries, whereas the US might need a model of endogenous innovations.

[24]See Figure 3 and the associated description in Green and Henseke (2021). They defined graduate occupations as the top three ISCO-08 major groups, which is very similar to the definition of abstract occupations in this chapter. They looked at 10 Central and Eastern European countries and 14 old EU countries. Only 4 countries saw an increase in the share of graduates in non-graduate jobs by more than 5 percentage points, and they are all in Central and Eastern Europe. And in every one of the 24 countries, the share of graduates in the workforce increased over the period. The UK is around the middle in the distribution of the growth rate of the graduate share among the 24 countries over the decade.

3.3 Model of endogenous adoption of task-biased technology

This section develops an equilibrium model of occupational labour (called 'tasks' for briefness). The model is static because we are interested in long-run comparative statics. On the demand side, there are multiple industries and within each industry firms choose between two technologies that differ in task intensities. This choice of technology means that the demand side is very elastic. On the supply side, workers have two dimensions of observable skills and an unobservable general ability. They sort into occupations based on wages and preferences.

In this chapter I will use 'occupations' and 'tasks' inter-changibly. In reality, the task content within occupations may change continuously as overall demand for tasks change. This is an interesting challenge for future research.[25] In this chapter, 'tasks' should be interpreted as the output of specific occupations. For example, professional tasks are simply the output of labour in professional occupations, whether the actual activity carried out is writing or data analysis or presentation is not studied here.

Each industry produces one good. Denote the goods as $g \in \{1,2,..G\}$. The production of each good is a CES function of tasks $j \in \{1,2,..J\}$, given the technology choice.

To produce any given good g, there are two potential technologies, denoted by $T \in \{O,N\}$. Each firm can choose freely between the 'Old' tech and the 'New' tech. Firms are otherwise identical within the industry. The difference between two technologies is that they have different task intensities α_{gj}^{T}. They also have their own TFP term A_{gt}^{T}, which is neutral with regard to tasks.

[25] Conceptually, what matters in production is tasks, but what workers choose is occupation, The task content within occupation is a choice made by the firm, subject to potentially complex constraints (physical constraints, information constraints, supply constraints and so on). There's also a question of how to organise all the tasks into bundles across individual workers and then to combine them by management.

$$Y_{gt}^{T} = A_{gt}^{T}[\sum_{j} \alpha_{gj}^{T}(y_{gjt}^{T})^{\rho}]^{\frac{1}{\rho}} \quad ,T \in \{O,N\} \tag{3.1}$$

Y_{gt}^{T} is the output produced by technology T in industry g at time t. y_{gjt}^{T} is the amount of task j employed in industry g, using technology T at time t. α_{gj}^{T} is the share parameter of task j in technology T in industry g, note that it does not vary over time. ρ is 1 minus 1 over the elasticity of substitution between tasks. ρ must be below 1. A negative ρ means tasks are complements. A_{gt}^{T} is Total Factor Productivity of technology T in industry g at time t.

Differential productivity trends (A_{gt}^{T}) captures exogenous Routine Biased Technical Change. For example, if the New technology uses robots and the price of robots falls or the productivity of robots increases, then this would be reflected as an increase in A_{gt}^{N}.

Consumers have CES preferences over G goods, with σ being the elasticity of substitution. A good produced by the Old technology is a perfect substitute for the same good produced by the New technology.

$$Q_{gt} = Y_{gt}^{O} + Y_{gt}^{N} \tag{3.2}$$

$$U_{t} = [\sum_{g} B_{gt} Q_{gt}^{\frac{\sigma-1}{\sigma}}]^{\frac{\sigma}{\sigma-1}} \tag{3.3}$$

Q_{gt} is output in industry g at time t. B_{gt} captures time-varying demand for good g. B_{gt} is assumed to be exogenous here. For future research, it would be interesting to allow income growth to differentially affect the demand for goods and services.

Each technology is assumed to have constant returns to scale. We normalise $\sum_{j} \alpha_{gj}^{T} = 1, \forall g, T$.

Because technology O and N differ in task intensities, we can think of a shift between technology O and N as task-biased technological change. This could be caused by changes in TFP in either technology option, industry demand shifts, or changes on the supply side. Ex ante, the model does not prescribe the New technology as routine-biased. It is left for the data to tell us how task intensities differ between the Old and New technologies.

The primary difference between my model and the RBTC literature is the presence of two technologies to choose from. If there's only one technology, then employment shares can only change due to changing task prices or changing parameters in the production function. The latter could be modelled as exogenously evolving share parameters in a CES production function, such as in Johnson and Keane (2013). The downsides are: 1) there is a lot more unobserved parameters to be estimated (one will need α_{gjt} instead of $\alpha^O_{gj}, \alpha^N_{gj}$), and 2) there is one less channel to absorb supply-side shocks, so the result of increasing skills supply will tend to be lower prices of high-skilled tasks. The reality is that the big increase in graduates did not reduce their relative wages, or the relative wage of abstract occupations.[26] In my model, this happens through the endogenous shift towards the New technology, which is more intensive in the tasks that graduates have comparative advantage in. By contrast, in a model with exogenous technology, the technology's parameters would need to shift in favour of the tasks that graduates have comparative advantage in, and at a speed that happens to leave the task prices and the mapping from education to occupation relatively unchanged. In section 3.5.1, I will formally test the hypothesis of exogenous task-biased technical change and reject it in favour of my model.[27] It is worth noting that my model allows for exogenous technical change as well: the TFP trends A^T_{gt} are exogenous, and a sufficiently large increase in the New technology's TFP will induce all firms to switch to it.

The CES formulation is common to the task literature, and many paper make the more restricted assumption of Cobb-Douglas production.[28] One exception is Johnson

[26]The three abstract occupations clearly have much more graduates than other occupations, so I call them 'high-skilled'.

[27]It's a rejection of the hypothesis that all technical change is exogenous. It does not reject the hypothesis that there is some exogenous shock to technology.

[28]For example, Autor (2013) define output as CES over a continuum of tasks; Acemoglu and Autor (2011) models output as Cobb-Douglas over a continuum of tasks; Autor and Dorn (2013) models goods output as Cobb-Douglas over routine task and abstract task, and services output is simply manual labour

and Keane (2013). Johnson and Keane (2013) differentiates labour by occupation, education, gender and age. Their production function is multi-level nested CES.[29] Their formulation is more detailed than my model. To fit the US data over 29 years of data, they found that it's necessary to allow the share parameters to follow 3rd or 4th order polynomials. By contrast, there is no time-variation in the share parameters in my model. Thus, ex ante, it's more challenging for my model to fit occupational trends.

That is the demand side. Now let's specify the supply side.

Suppose each person i is endowed with two dimensions of observable skills: analytical ability a_i and social skill s_i; and an unobserved general ability μ_i. The joint distribution of skills is assumed to be exogenous. Later on we will consider counterfactual policies that shift the skills distribution, through education or immigration. In reality, RBTC may induce workers to undertake more education or training in order to become more productive in abstract tasks (Battisti et al., 2017). Such an endogenous response on the skills distribution is left for future investigation.

In the workplace, only the individuals' skills matter for productivity, not their education per se. Each occupation produces one task. Occupation and task are both denoted by subscript j. The amount of task that worker i in occupation j produces is

$$y(i,j) = k_j e^{\beta_{aj} a_i + \beta_{sj} s_i + \mu_i} \tag{3.4}$$

This formula follows from Autor and Handel (2013), where I specify skills to 2-dimensions. Here k_j is a j-specific scalar. μ_i is worker's general ability which is unobserved. μ_i can be correlated with observed skills freely. The coefficients β_{aj}, β_{sj} are occupation-specific productivities of analytical and social skills. The key assumption

times a scalar; Traiberman (2019) models output in each industry as a Cobb-Douglas function of capital, human capital in each occupation and intermediate inputs produced in other industries.

[29] The bottom three levels are education, gender and age; at the top level, aggregate output is CES between unskilled task and skilled task; unskilled task is 2-level CES of 8 occupations, and skilled task is 2-level CES of capital and 2 occupations.

here is that comparative advantage is captured by 2 dimensions of skills a_i, s_i; conditional on them, there is no omitted factor that makes a person more productive in one task rather than another.

The labour market is competitive. We assume workers do not directly care about the technology chosen by their employer or which industry they are in. Since a worker's task output is the same wherever they work, the task price must equalise between firms that operate with different technologies and across industries. I denote the price of task j at time t as p_{jt}.

Because workers are perfect substitutes in producing any given task (though individuals have different productivities), worker of ability a_i, s_i in occupation j in a firm adopting tech T gets paid the value of their task output

$$W(i, j, t) = y(i, j) p_{jt} \tag{3.5}$$

The utility that worker i gets from occupation j at time t is

$$U_{ijt} = \ln(y(i, j) p_{jt}) + \eta_j + e_{ijt}, \quad j = 1, ...J \tag{3.6}$$

where η_j is occupation-specific amenities; e_{ijt} follows iid Type-1 extreme value distribution, with location parameter at 0 and scale parameter ζ. In future, we may want to allow η_j to vary across demographic groups (e.g. gender, age, family type, and immigration status). Idiosyncratic preference shocks e_{ijt} mean that for any given (a_i, s_i), there is positive probability of the worker going to any occupation j. Omitting the time subscript t for simplicity, the probability of worker i choosing occupation k is simply

$$\pi_k(i,\mathbf{p}) = (y(i,k)p_k e^{\eta_k})^{\frac{1}{\zeta}} / [\sum_j (y(i,j)p_j e^{\eta_j})^{\frac{1}{\zeta}}] \tag{3.7}$$

$$= [e^{\beta_{ak}a_i+\beta_{sk}s_i+\mu_i+\eta_k} k_k p_k]^{\frac{1}{\zeta}} / \sum_j [e^{\beta_{aj}a_i+\beta_{sj}s_i+\mu_i+\eta_j} k_j p_j]^{\frac{1}{\zeta}} \tag{3.8}$$

$$= [e^{\beta_{ak}a_i+\beta_{sk}s_i+\eta_k} k_k p_k]^{\frac{1}{\zeta}} / \sum_j [e^{\beta_{aj}a_i+\beta_{sj}s_i+\eta_j} k_j p_j]^{\frac{1}{\zeta}} \tag{3.9}$$

where $\mathbf{p}$ denotes the price vector of all tasks. Comparative advantage plays a role in the sorting into occupation: a worker with higher a_i is more likely to go to an occupation with higher β_{aj}. A smaller ζ means the preferences are less varied and so the wages are more important in determining the occupation choice. Note that the unobserved heterogeneity term μ_i does not enter into occupational choice. Thus $\pi_k(i,\mathbf{p}) = \pi_k(a_i, s_i, \mathbf{p})$.

Given task prices, the supply of task j in the economy is

$$LS_j(\mathbf{p}) = \sum_i \pi_j(a_i, s_i, \mathbf{p}) y(i,j) \tag{3.10}$$

$$= \int\int \pi_j(a,s,\mathbf{p}) y(a,s,j) f(a,s) dads \tag{3.11}$$

where $f(a,s)$ is the joint density function, and $y(a,s,j)$ is the expected output in task j conditional on observing a,s. The derivation of (B.14) is in Appendix B.2.

Thus, the only relevant unknowns on the supply side are η_j, ζ, $y(a,s,j)$ and $f(a,s)$. As long as we get $y(a,s,j)$, we don't need to estimate the distribution of unobserved heterogeneity μ_i or how it depends on (a_i, s_i), or the returns to skills β_{aj}, β_{sj}.[30]

On the demand side, we do not observe y^O_{gjt}, y^N_{gjt} separately as opposed to $EMP_{gjt} = y^O_{gjt} + y^N_{gjt}$. After some algebraic manipulation, we obtain a demand-side prediction

[30]For counterfactual analysis, we do need assumptions about how $y(a,s,j)$ would change under the counterfactual. For example, I currently assume that it does not depend on the counterfactual policy. This would be the case if the distribution of μ_i conditional on (a,s) does not depend on the counterfactual policy.

about the relationship between task price ratio and the observable task quantity ratio.

$$\ln(\frac{p_{jt}}{p_{1t}}) = (\rho - 1)\ln\frac{EMP_{gjt}}{EMP_{g1t}} + (1-\rho)\ln[(1-w_{gt})r^O_{gj} + w_{gt}r^N_{gj}] \tag{3.12}$$

where w_{gt} is the share of 'New' technology in industry g at time t; $w_{gt} = y^N_{g1t}/(y^O_{g1t} + y^N_{g1t})$. And $r^T_{gj}, T \in \{O,N\}$ are functions of parameters (α^T_{gj}, ρ). See Appendix B.1 for the derivation.

Equation (3.12) looks like a typical demand-side equation from the Skill-Biased Technical Change literature, where the term $\ln[w_{gt}r^O_{gj} + (1-w_{gt})r^N_{gj}]$ would represent technical changes. But it has a particular functional form: it's a weighted average between two technologies, where the weight is at the industry-year level. The standard equation in the SBTC literature would have an exogenous time trend to represent technological progress (for example Katz and Murphy (1992) just had a linear time trend and Johnson and Keane (2013) had 3rd or 4th order polynomial). Those papers categorize labor input by education, whereas here it's across occupation and industry (j,g). In the context of occupation-industry, a standard specification of exogenous technical change would use a j-g-specific time polynomial. I will test such a hypothesis in section 3.5, and show that the evidence does not support it.

3.3.1 Equilibrium characteristics and effect of a supply-side shift

I define the equilibrium as log task prices ($\log \mathbf{p_t} = \{\log p_{1t}, ...p_{Jt}\}$) and technology shares ($\omega_\mathbf{t} = \{\omega_1, ...\omega_{Gt}\}$) such that demand equals supply in each task, and that in each industry, the lower-cost technology is adopted. Both are adopted if their unit costs are equal. Here $\omega_{gt} = Y^N_{gt}/(Y^N_{gt} + Y^O_{gt})$, the share of output produced by the new technology.[31]

In Appendix B.3, we derive the following condition which makes firms indifferent

[31] ω_{gt} is not the same as $w_{gt} = y^N_{g1t}/(y^N_{g1t} + y^O_{g1t})$, the share of new technology in terms of employment in the first occupation. But they are very strongly positively correlated.

between the two technologies in industry g.

$$\sum_j [(\alpha_{gj}^N)^{\frac{1}{1-\rho}} - (\frac{A_{gt}^N}{A_{gt}^O})^{\frac{\rho}{\rho-1}} (\alpha_{gj}^O)^{\frac{1}{1-\rho}}] p_{jt}^{\frac{\rho}{\rho-1}} = 0 \tag{3.13}$$

This equation is linear in $p_{jt}^{\frac{\rho}{\rho-1}}$. Given the alpha parameters, this equation may have no solution in the positive domain if the TFP ratio is very far from 1. In that case, one technology will dominate in that industry. When the TFP ratio is not extreme, there are likely infinitely many points in the $(p_{jt} > 0, 1 \leq j \leq 9)$ space that would equalize the unit costs between the two technologies in all 7 industries.

Denote δ_{jg}^T as unit input, that is, the amount of task j required by tech T to produce one unit of output in industry g. Note it's a function of all task prices $\mathbf{p_t}$.

Given all task prices $\mathbf{p_t}$, the demand for task j is

$$\sum_g [\delta_{jg}^N(\mathbf{p_t}) Q_g(\mathbf{p_t}) \omega_{gt} + \delta_{jg}^O(\mathbf{p_t}) Q_g(\mathbf{p_t})(1-\omega_{gt})] \tag{3.14}$$

$$= \sum_g (\delta_{jg}^N(\mathbf{p_t}) - \delta_{jg}^O(\mathbf{p_t})) Q_g(\mathbf{p_t}) \omega_{gt} + \sum_g \delta_{jg}^O(\mathbf{p_t}) Q_g(\mathbf{p_t}) \tag{3.15}$$

Industry output Q_g is a function of $\mathbf{p_t}$ through industry goods prices. It does not depend on $\mathbf{w_t}$.

Task demand is not uniquely pinned down by task prices. Instead, movements in $0 \leq \omega_{gt} \leq 1$ allows task demand to move within the cone of diversification. The cone of diversification has as many dimensions as the number of industries where the unit costs are equal. For a majority of years in our sample period (1997-2015), it has 7 dimensions.

Recall (B.14), the supply of task j takes this form:

$$LS_j(\mathbf{p_t}) = \int\int \pi_j(a, s, \mathbf{p_t}) y(a, s, j) f(a, s) dads \tag{3.16}$$

where $\mathbf{p_t}$ is the vector of all task prices, $f(a, s)$ is the joint density function, and $y(a, s, j)$

is the amount of task j that workers with skills (a,s) will produce. The latter two do not depend on $\mathbf{p_t}$.

Market clearing requires:

$$\sum_g(\delta^N_{jg}(\mathbf{p_t}) - \delta^O_{jg}(\mathbf{p_t}))Q_g(\mathbf{p_t})\omega_{gt} + \sum_g \delta^O_{jg}(\mathbf{p_t})Q_g(\mathbf{p_t}) - LS_j(\mathbf{p_t}) = 0 \tag{3.17}$$

Given all task prices, these market-clearing constraints are a system of 9 linear equations, linear in the 7-element vector $\omega_{gt}, 1 \leq g \leq 7$.

When some supply-side shock shifts the supply curve (for example if the density $f(a,s)$ in (3.16) changes), it's possible that a change in ω_{gt} will clear the markets without any change in task prices $\mathbf{p_t}$. This requires the shift in $LS_j(.)$ to be in the cone of diversification. In other words, the shift between technologies may absorb supply-side shocks and leave the equilibrium task prices unchanged.[32] Recall that individuals' occupational choice probabilities are functions of their two skills and task prices. When task prices do not change, the occupational employment shares conditional on skills will not change. This is consistent with the UK fact that during a period of rapid increases in higher education, the occupational destinations among graduates did not change much (Figure 3.3). The small amount of occupational downgrading observed within education groups could be interpreted as the education-specific distribution of skills having deteriorated slightly.[33] In short, through technology shifts, an increase in the supply of skills can leave the task prices unchanged, and the occupation destinations conditional on skills unchanged.

[32] If the supply-side shocks are outside the cone of diversification, then price changes will be necessary to return the economy to equilibrium.

[33] In the empirical estimation part of the paper, I will assume that the skills distribution is fixed within education-gender. I hope to relax this assumption in future: when I find skills data for cohorts with different education composition, the data will tell whether the skills distribution of graduates worsens when a larger share of the cohort are graduates.

3.3.2 Identification of technology shares

We don't observe technology share directly, nor do we observe y^O_{gjt}, y^N_{gjt} separately as opposed to $y^O_{gjt} + y^N_{gjt}$. If we knew ρ, we could use observed $y^O_{gjt} + y^N_{gjt}$ to obtain $(1 - w_{gt})r^O_{gj} + w_{gt}r^N_{gj}$ through (3.12). However, r^O_{gj}, r^N_{gj} are also unknown. In fact, the level of w_{gt} is not identified even if we directly observe $(1-w_{gt})r^O_{gj} + w_{gt}r^N_{gj}$. To see why, consider an affine transformation of w_{gt} :

$$\hat{w}_{gt} = kw_{gt} + c, \forall t$$
$$\hat{r}^N_{gj} = r^O_{gj} + \frac{1-c}{k}(r^N_{gj} - r^O_{gj}), \forall j$$
$$\hat{r}^O_{gj} = r^O_{gj} - \frac{c}{k}(r^N_{gj} - r^O_{gj}), \forall j$$

The transformed case is observationally equivalent to the original one:

$$(1-w_{gt})r^O_{gj} + w_{gt}r^N_{gj} = (1-\hat{w}_{gt})\hat{r}^O_{gj} + \hat{w}_{gt}\hat{r}^N_{gj}, \forall j, t$$

Therefore, we will anchor the time series $\{w_{gt}\}$ by assuming $w_{g0} = 0, w_{gT} = 1, \forall g$. This 'normalisation' is not totally innocuous because it assumes that w_{gt} cannot go above w_{gT} or below w_{g0}. This seems true in the UK data, and it allows easy interpretation: we are effectively calling the production function at time 0 the Old technology and the one at time T the New technology.

Empirically, we will estimate w_{gt} from technology proxies. Suppose we have a proxy for new technology called z, such that $z_N > z_O$. The assumption here is that all firms with the New tech have the same level of z, which is higher than the level among old-tech adopters. There is no time variation within z_N or z_O. In practice, we will use several measures of z. We observe z over time and at the industry level.

$$z_{gt} = (1-\tilde{w}_{gt})z_O + \tilde{w}_{gt}z_N \tag{3.18}$$

where $\tilde{w}_{gt}$ is the scale of new technology adopters relative to the entire industry. If z_{gt} comes from employee survey, $\tilde{w}_{gt}$ is the employment share of firms using the new technology in the industry-year. As we anchor $\tilde{w}_{gt}$ to 0 at one point and 1 at another point, we would be setting $z_O = \tilde{z}_{g0}, z_N = \tilde{z}_{gT}$. Thus, we can impute w_{gt} as $\frac{\tilde{z}_{gt}-\tilde{z}_{g0}}{\tilde{z}_{gT}-\tilde{z}_{g0}}$. Thus, w_{gt} is just-identified by one proxy up to an affine transformation. If we have several measures of z, we can allow errors in equation (3.18). In section 3.4.3, we will assume a latent factor model to impute w_{gt}.

3.3.3 Identification of model parameters

The parameters fall into two broad categories: supply-side and demand-side.

On the supply side, the unknowns are: η_j, the utility for working in occupation j; ζ, the scale of preference shocks; $f(a,s)$, the joint distribution of analytical and social skills ; and $y(a,s,j)$, the expected task output conditional on skills (a,s). Note that we don't need to estimate other supply-side parameters such as the returns to skills. The reason was explained around equation (B.14).

η_j is the preference for working in occupation j, and we normalise $\eta_1 = 0$. The higher η_j, the more people will select into occupation j, all else equal. Therefore, η_j can be identified from the occupational employment shares in any given year. If we allow η_j to vary over time without any restriction, we could fit employment shares in every year perfectly. For now, I choose to have fixed η_j, so that no changes in employment will be attributed to preference shifts. Empirically, I search for η_j to match the observed employment shares in LFS 2006 (the mid-point of my sample period).

The smaller ζ is, the more elastic task supply will be with regard to task prices. The identification of ζ relies on movements along the task supply curve. Had there been no changes to the skills distribution, small movements in task prices together with large movements in employment would imply that ζ is small. In future, I could use instruments for demand shocks to estimate the elasticity of task supply.

The joint skill distribution comes from the numeracy score and the literacy score in the British Cohort Studies (BCS), measured at age 34. They are summarised to 7 points of support in each dimension.[34] The skills distribution in the BCS data might be quite different from the aggregate skill distribution in the UK because the BCS only contains the 1970 birth cohort. The aggregate skill distribution might be changing over time due to increasing education as well as immigration. I assume the joint distribution of analytical and social skills is fixed conditional on gender and education.[35] We obtain the distribution from the BCS for each gender-education, get gender-education weights from the Labour Force Survey for each year, and aggregate up. Thus, the shift in skills distribution over time comes from the changing composition of gender and education in the UK workforce.

Because this is a competitive labour market, workers are paid their task output times task price. We can get wages conditional on skills directly from the BCS, and dividing it by p_{jt} gives us the expected task output conditional on skills $y(a,s,j)$.

On the demand side, the unknowns are: ρ, which governs the substitution elasticity between tasks; tasks intensities $\alpha^T_{gj}, T \in \{O,N\}, 1 \leq g \leq G, 1 \leq j \leq J$; TFP trends $A^T_{gt}, T \in \{O,N\}, 1 \leq g \leq G, \forall t$; industry demand B_{gt}; and σ, which is consumers' substitution elasticity.

In principal, ρ is identified from (3.12): the correlation between log wage ratio and log quantity ratio conditional on the technology share w_{gt}. Because we don't observe w_{gt} directly, non-linear estimation of (3.12) is tricky. I hope to estimate ρ directly from (3.12) in the near future. Currently, I calibrate $\rho = -0.1$, which corresponds to Goos et al. (2014)'s estimate of the substitution elasticity between tasks at 0.9.

Given ρ, all the other production parameters are well-identified.

[34] Currently 7 is selected so that each group has at least 10% density. In future, I will experiment with having more or fewer points of support.

[35] In future, I will use other data to test this assumption, by comparing between generations who have very different education composition. This cannot be tested in the BCS because it contains only one birth cohort.

Recall equation (3.12):

$$\ln(\frac{p_{gjt}}{p_{g1t}}) - (\rho - 1)\ln\frac{y^O_{gjt} + y^N_{gjt}}{y^O_{g1t} + y^N_{g1t}} = (1 - w_{gt})r^O_{gj} + w_{gt}r^N_{gj} \tag{3.19}$$

$$= r^O_{gj} + (r^N_{gj} - r^O_{gj})w_{gt} \tag{3.20}$$

Given ρ, we can calculate the LHS of (3.20) directly for all g, j, t. The RHS is a linear function of w_{gt} with unknown parameters. So, regressing the term on w_{gt} by industry and occupation will give us r^O_{gj} as the constant and $r^N_{gj} - r^O_{gj}$ as the slope. Given $r^T_{gj} = (\alpha^T_{gj}/\alpha^T_{gj})^{1/(1-\rho)}$, and that $\sum_j \alpha^T_{j,g} = 1$, we can back out all α^T_{gj} from r^T_{gj}.

A^T_{gt} can be identified using the equation below. This equation assumes $\rho \neq 0$ and uses the F.O.C in firm's profit maximisation. Its derivation is in Appendix B.4.

$$(p_{gt}A^T_{gt})^{\frac{\rho}{\rho-1}} = \sum_j [\frac{p_{jt}}{(\alpha^T_{gj})^{1/\rho}}]^{\frac{\rho}{\rho-1}} \tag{3.21}$$

This equation gives A^T_{gt} as a function of $(\alpha^T_{gj}, p_{jt}, p_{gt}), \forall j$. Once we have identified all the alphas and over-time changes in p_{jt}, we identify the over-time changes in each A^T_{gt} such that $y^T_{gjt} > 0$. For industry g where tech T was not adopted at time t, (3.21) gives the upper bound of A^T_{gt}.[36] We can get the size of A^O_{gt} relative to A^N_{gt}. The absolute scale of A^T_t is meaningless because it's just the inverse of the scale of y^T_{gj}.

Finally, industry demand trends can be identified from observed quantities and prices of all the goods. It doesn't rely on ρ or w_{gt}. Given the CES utility function, the relative trends of B_{gt} are:

$$\ln B_{gt} - \ln B_{1t} = \frac{1}{\sigma}(\ln Q_{gt} - \ln Q_{1t}) + \ln(p_{gt} - p_{1t}) \tag{3.22}$$

[36] In that case, the TFP of the dominated technology is not identified, except that it must be below the upper bound.

σ is unknown. Industry-level prices and outputs can be obtained from the ONS.[37] We can estimate σ by assuming $\ln B_{gt} - \ln B_{1t}$ follows a time polynomial and regressing relative outputs on relative prices. We get $\hat{\sigma} = 0.16$. The absolute level of all B_{gt} is not identified, nor is it necessary because the model features Constant Returns to Scale. To impute $\ln B_{gt}$, we use our own production function to impute industry output rather than directly use the ONS measures. This is because my model does not include capital explicitly, the industry output based on observed employment in my model will be lower than actual output in more capital-intensive industries. To be internally consistent, we calculate industry output from the production function, then combined with observed industry prices and $\hat{\sigma}$, equation (3.22) gives the relative demand trends.

3.4 Sources of moments of data

3.4.1 Occupational employment and wages

The main data source for occupational employment and wages is the UK Labour Force Survey. This a representative quarterly survey of households in the UK, focusing on work-related topics. It is similar in nature to the US Current Population Survey (CPS). I have used the UK LFS data from the first quarter of 1993 to the last quarter of 2017. The main estimation is restricted to the period 1997-2015, because a key dataset for technology proxy is only available over that period.

Occupation in the LFS is based on the Standard Occupational Classification of that decade: SOC1990 until 2000, SOC2000 over 2001-2010, and SOC2010 from 2011 onwards. There are 300+ occupations within each SOC classification. When I bring the model to data, occupations are defined as the 9 major groups under SOC2000. The occupations are: 1, managerial, 2 professional, 3 associate professional and technical, 4 administrative and secretarial, 5 skilled trade, 6 personal services, 7 customer services,

[37] Source: GDP output approach low-level aggregates from the ONS website .

8 process, plant and machine operatives, and 9 elementary.[38] I construct a probabilistic mapping from SOC1990 to SOC2000 on the basis of a subsample of LFS observations linked between LFS2000Q4 and LFS2001Q2, who were in the same job and hence reported SOC1990 and SOC2000 in those two quarters. The mapping takes into account 3-digit SOC1990 and individual's gender and education.[39] On the other hand, SOC2010 is mapped to SOC2000 using the transition matrix from the Office for National Statistics.

Industry is a slight aggregation from SIC80 divisions (in the LFS until 2008) and SIC92 sections (since 2009). To ensure consistency over time and across datasets, I group industries to 7 categories: 1) agriculture, mining, energy and water supply (let's call it natural resources thereafter); 2) manufacturing; 3) construction; 4) wholesale, retail, hotel and catering; 5) transport, storage, and communication; 6) finance, real estate and business activities; 7) all other services including government administration, health, education, social and other services.

For occupational wage bills, I add up all the actual weekly hours in the relevant cell (g, j, t), and multiply it by the mean hourly wage in that cell.[40]

For task price p_{jt}, we run a log wage regression every year on occupation dummies, gender-age interactions and detailed education dummies. We add the observed mean log wage in the reference occupation to the coefficient estimates on occupation dummies. It's possible that the resulting p_{jt} is still contaminated by unobserved compositional changes. In future, we can estimate task prices from the New Earnings Survey Panel Data by allowing individual fixed effect.

The quantity of occupational labor y_{gjt} is simply the wage bill divided by the task price. The change in occupational classification causes discontinuities in the observed

[38] Elementary includes cleaners, waiters, kitchen assistants, labourers in agriculture and in construction, security guards, postal workers and so on.

[39] There are 300+ occupations at the 3 digit level.

[40] This is because wages are not reported for all that report hours.

y_{gjt} and p_{jt}. We remove discontinuities in the time series by the following method. We regress each time series (in log terms) on a 5th order polynomial of time plus a dummy for $t < 2001$ and a dummy for $t \geq 2011$. In other words, we allow the occupation classification change to affect the level of the variable and nothing else. We deduct the estimated jump from the affected period. Figure B.3 in the appendix plots the raw and adjusted p_{jt} for three example occupations. There are clearly jumps in some raw time series at 2001 and 2011, and the adjusted time series are smoother. We use the adjusted data in both descriptive graphs (Figure 3.1, Figure 3.2) and when estimating the model.

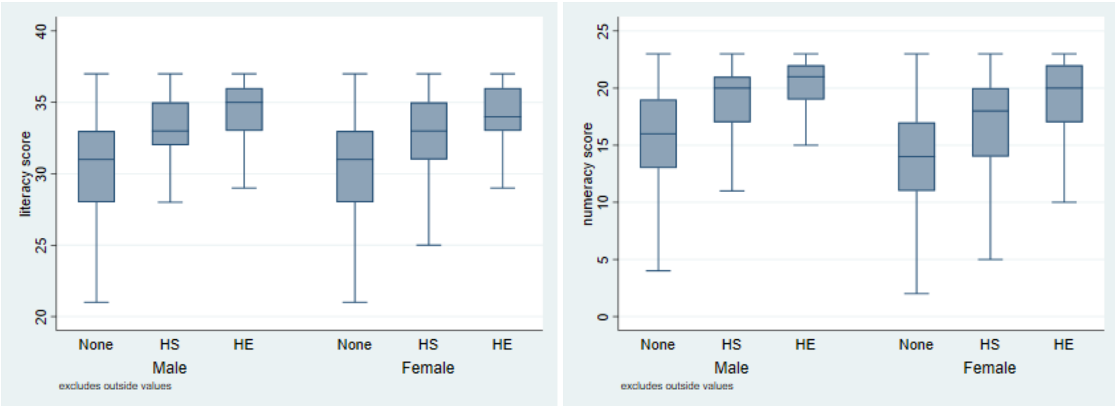
3.4.2 Skills distribution

We use numeracy and literacy skills in the British Cohort Study (BCS). The BCS is a longitudinal survey following around 17,000 people who were born in England in 1970. BCS contains many skill assessments at various ages, sometimes for a subset of the cohort. We are interested in skills measured after the completion of education, because education could have affected skills. We also prefer a larger sample. After age 16, there is only one wave (at age 34) when skills were assessed for the whole sample. Hence, in this chapter we will use literacy and numeracy assessed at age 34. There are about 9500 observations with both skills measured at 34 in the BCS.

Figure 3.4 shows the distributions of two skills by education and gender. For each skill, the mean score clearly increases with education, while the distribution overlaps significantly between education groups. Both skills have scores with 20+ points but the lower range is very sparsely populated. I summarise them to 7 points of support in each dimension.

We pool all the waves together to increase sample size for obtaining wages conditional on skills and occupation. I take age effects out of wages by simply regressing log wages on age dummies, and deducting the age effects from observed log wages. Then for each combination of skills and occupation, I use the mean wage excluding outliers

Figure 3.4: Distribution of literacy and numeracy scores in BCS



Note: from British Cohort Studies. The box edges correspond to the 25th percentile and the 75th percentile within the education and gender group. The line inside the box is the medium skill score. "HE" refers to higher education or above. "HS" refers to secondary school qualifications including A-levels, O-levels, GCSE C+ or equivalents. "None" refers to those without secondary school qualifications.

as the data moment for $E[p_j y(i,j)|a,s,j]$.[41] There are a number of empty (a,s,j) cells (having no individual in the cell or no one reporting wages), and they all have rare combinations of skills where one skill is very high and the other skills is very low. In such cases[42], I use the observed average wage of that occupation.

3.4.3 Technology proxies

When setting out the model, I have not specified what the new technology is or means in practice. This is because I believe its practical manifestation would vary across industries and firms. It could be something tangible such as automation equipment in a manufacturing firm, or high-speed internet in a professional service firm; or it could be something intangible like a decentralized structure of management and decision-making. The literature (Bresnahan et al. (2002), Caroli and Van Reenen (2001)) suggests that the different aspects of changes may be complementary to each other and skill-biased.

[41] Within each (a,s,j) cell, I exclude the top and bottom 5% of wage observations in calculating mean wages.

[42] Such pairs of (a,s) consitute 0.9% of the BCS sample.

Table 3.1: Capital input composition and the graduate proportion

	cap_CT	cap_IT	cap_Soft_DB	cap_pca
propBA	0.0043	0.0301**	-0.0172	4.4629*
	(0.0036)	(0.0093)	(0.0164)	(2.1850)
propDO	-0.0057	0.0069	0.0095	-0.2704
	(0.0048)	(0.0123)	(0.0217)	(2.8928)
Observations	133	133	133	133

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: these regressions are at the level of industry-year, controlling for industry dummies and year dummies. Each dependent variable is the share of overall capital in this type, with the industry-year. 'propBA' is the proportion of people with tertiary qualifications. 'propDO' is the proportion of people without GCSE grade C+ or equivalent.

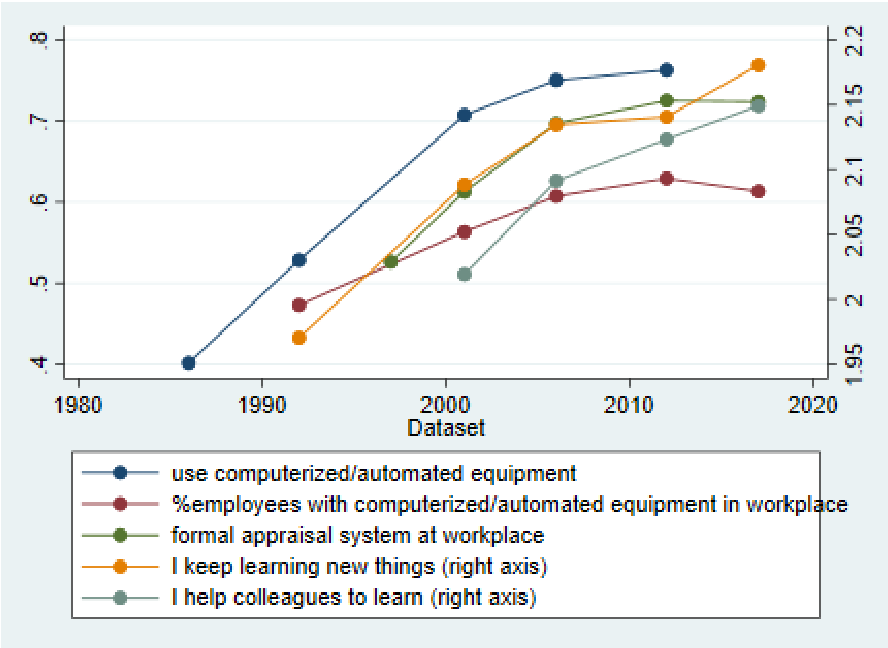
Guided by the literature (Michaels et al., 2014; Machin and Van Reenen, 1998), I consider measures of ICT capital and related tangible technology, as well as measures about intangibles, from two datasets: capital inputs in EU-KLEMS and the British Skills Survey (BSS). The former is available over 1997-2015. The BSS is available for 1986, 1992, 1997, 2001, 2006, 2012, 2017.

In EU-KLEMS, we observe various types of capital by year and across dozens of industries. At the industry-year level, I use the share of overall capital that is in each of the following four areas: Communication Technology, Information Technology, Software&database, and R&D. Most of the variables about capital composition in those areas do show an increasing time trend. I have also verified that the graduate proportion is positively and significantly correlated with IT capital input. Correlations with other capital inputs are mostly positive but insignificant, see table 3.1.

In the BSS, there are dozens of variables about tasks, technology and management that may potentially serve as proxies. My framework implies that the proxy should be positively correlated with the local supply of graduates. So I select proxies from the BSS accordingly.[43] I obtain 5 proxies, which are responses to the following ques-

[43] I summarize the data to the level of industry-region-year. I regress each variable on the graduate

Figure 3.5: Time trends in technology proxies in BSS



Note: the two learning measures take values between 0 to 3, 0 meaning 'strongly disagree' and 3 meaning 'strongly agree'. The other three are valued between 0 and 1.

tions/statements: 'whether job involves use of computerised or automated equipment', 'my job requires that i keep learning new things', 'my job requires that i help my colleagues to learn new things', 'do you have a formal appraisal system at your workplace', and 'In your workplace, what proportion of employees work with computerised or automated equipment?'.

Figure 3.5 shows the aggregate trend in these variables. They are mostly available for 5-6 waves in the Survey. They all increase strongly over time. Table 3.2 shows that all these 5 proxies are very positively and significantly correlated with the local proportion of graduates.

Given a range of proxy measures $z^m_{gt}, 1 \leq m \leq M$, we now impute w_{gt} in a latent

proportion allowing for year dummies, industry dummies, region dummies. I select variables for which the graduate proportion is significant in the first regression, and the raw aggregate trend is also increasing.

Table 3.2: Proxies in BSS, correlation with graduate proportion

	useauto	bnewthin	bhelpoth	E_propcom	E_eapprais
BAprop	0.3276***	0.4234***	0.3081*	0.2733***	0.2000**
	(0.0707)	(0.0929)	(0.1244)	(0.0443)	(0.0702)
Observations	348	390	312	390	389

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: all the outcomes are aggregated to the industry-year-region level. Each regression is at the industry-year-region level, including year dummies, industry dummies, region dummies. 'own use PC' is binary on 'whether job involves use of computerised or automated equipment'. '%PC in workplace' is 'In your workplace, what proportion of employees work with computerised or automated equipment?'. 'appraisal' is binary for 'do you have a formal appraisal system at your workplace'. 'learn new thing' is the reported agreement with the statement 'my job requires that i keep learning new things', it has range 0-3, higher value for more agreement with the statement. 'help others learn' is similar, for the statement 'my job requires that i help my colleagues to learn new things'. 'BAprop' is the proportion of graduates at the industry-year-region level in the BSS.

variable model. Suppose each measure is a linear function of the latent variable w_{gt} plus some measurement error.

$$z_{gt}^m = \zeta_g^m + \psi_g^m w_{gt} + \varepsilon_{gt}^m \tag{3.23}$$

The constant and the slope coefficient is specific to the measure m and the industry g. Because w_{gt} is unobserved, w_{gt} is only identified up to affine transformation. We have 4 measures of capital composition from 1997 to 2015 annually and 5 measures from the BSS available at 4-5 points between 1992 and 2017. Because the different measures have different scales, I standardise each measure within industry so that when I minimise the sum of squared ε_{gt}^m, they are equally important. Then, I do an affine transformation of w_{gt} to equal 0 in 1997 and 1 in 2015. Finally, I smooth each time series with a cubic spline and constrain the value to be in the [0,1] range. Figure B.4 in Appendix B.5 shows the resulting time series for all the industries.

3.5 Corroborative evidence

In this section, we empirically assess 2 implications of the model. First, increasing supply of skills may have little impact on occupational wages. Second, changes in the skill supply predict occupational shifts at the local level.

3.5.1 Do occupational wages respond to supply shifts

The key difference between my model and standard models in the RBTC literature is that the adoption of technology in my model responds to supply shocks. This has different implications for how occupational wages respond to supply-side shocks.

In standard models, the demand curve is downward-sloping. Equation (3.24) below is a typical demand-side equation:

$$\ln(\frac{p_{gjt}}{p_{g1t}}) = (\rho - 1)\ln\frac{EMP_{gjt}}{EMP_{g1t}} + \theta_{gjt} \tag{3.24}$$

$(\rho - 1)$ is the negative reciprocal of the substitution elasticity between tasks in industry g. θ_{gjt} represents technology shift. In the exogenous SBTC literature, the last term θ_{gjt} would be an exogenous trend and it is usually approximated by some polynomial of time. The trend is linear in Katz and Murphy (1992) and Card and Lemieux (2001).[44] Goos et al. (2014) estimated the substitution elasticity between tasks to be 0.9, which would mean a coefficient of -1.1 in front of the log quantity ratio.[45] If tasks are complements in production, we'd expect the coefficient to be below -1. In short, in models with exogenous technology, θ_{gjt} does not respond to supply-side shifts, and so a supply-induced increase in the task quantity ratio will reduce the task price ratio.

By contrast, the demand curve could be flat in my framework. Recall equation (3.12):

$$\ln(\frac{p_{jt}}{p_{1t}}) = (\rho - 1)\ln\frac{EMP_{gjt}}{EMP_{g1t}} + (1-\rho)\ln[(1-w_{gt})r^O_{gj} + w_{gt}r^N_{gj}]$$

This is similar to the exogenous technology formulation (3.24). In both equations, log price ratio equals $\rho - 1$ times log quantity ratio plus a term for technical change. In my

[44] In Katz and Murphy (1992) the coefficient is estimated to be -0.7 (implying an elasticity of 1.4). In Card and Lemieux (2001), the substitution elasticity between college and high-school labour equivalents is estimated to be in the 2-2.5 range. But those estimates are not exactly applicable here because they differentiate labour by education, whereas I do by occupation.

[45] Their estimate did not come from such a regression.

model, the term for technical change is a weighted average between the old and new technologies where the weight w_{gt} is *endogenous*.

How will wages respond to a supply-side shift in my model? As explained in section 3.3.1, if the supply-side shift happens to fall into the cone of diversification, the task prices will stay constant while w_{gt} adjusts to equalise demand and supply. More generally, the endogenous technological shift will tend to offset exogenous shocks on the supply side, so the resulting impact on wages would be smaller than in the case of exogenous technology. To see why, consider a positive supply shock that increases professional employment. There is a direct negative effect on professional relative wage through the first term $(\rho-1)<0$. If the new technology is more intensive in professional task $r^N_{gj} > r^O_{gj}$, the lower professional wage will cause a shift to the new technology: w_{gt} will increase. If instead, we have $r^N_{gj} < r^O_{gj}$, then w_{gt} will fall. In either case, the term $(1-\rho)\log[(1-w_{gt})r^O_{gj} + w_{gt}r^N_{gj}]$ will increase. This will partially offset the negative effect through the first term.

Now let's see how wages have responded to supply-side shifts in the UK data. Specifically, we will regress the log occupational wage ratio on the log occupational employment ratio and a j-g-specific time trend, where the log employment ratio is instrumented by supply-side shifts. The instruments are of shift-share style, using the shift in the demographic composition of the population (defined by education-gender-age) and historical mappings from each demographic group to tasks. The idea is to capture variation that comes from changes in the demographic composition. Thus, the coefficient on the log employment ratio is interpreted as the slope of the demand curve. The specification of time trend is a 5th order polynomial of year, plus two dummies to capture classification discontinuities over 2000-1 and 2010-11. The regression is run separately by industry.

The results are reported in table 3.3. I find that the key estimate $(\rho-1)$ is small and sometimes even positive in some industries. The instruments are reasonably strong:

Table 3.3: Estimating wage response to supply-side shifts, by industry

	Dependent var: $\log wage_{gjt}/wage_{g1t}$			
	natural resources	manufacturing	construction	trade
log emp ratio	0.2773	0.0956	-0.4225	0.1665
	(0.4058)	(0.1323)	(0.2629)	(0.5868)
j-specific trend	yes	yes	yes	yes
Observations	200	200	200	200
	transport, information	finance, business serv	other services	
ln y_gjt/y_g1t	-0.8401	0.0048	0.3706*	
	(1.3341)	(0.2665)	(0.1446)	
j-specific trend	yes	yes	yes	
Observations	200	200	200	

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: The dependent variable is log hourly wage ratio at the industry-occupation-year level. The definition of industry and occupation is the same as the rest of the paper. Occupation 1 is the reference occupation group. The key regressor is log occupational employment ratio $\ln emp_{gjt}/emp_{g1t}$, where $\ln emp_{gjt}$ is the total hours in the g-j-t cell. The instruments for $\ln emp_{gjt}/emp_{g1t}$ are $supply_{gjt}, supply_{g1t}$. $supply_{gjt}$ is a shift-share instrument at the g, j, t level, using contemporary shares of demographic groups and historical mapping from demographic groups to g, j cells. Source: LFS 1993-2017.

the standard errors are small enough to rule out $(\rho - 1) < -1$ in most industries.

In short, the estimates suggest the demand curve is not as downward-sloping as would be expected from standard models. My framework with endogenous technical change offers an explanation as to why it may be flat.

3.5.2 Correlation between education and occupational shares

While there is no geography in the model, we can think of applying it to separate local labor markets. Across regions, we expect a positive correlation between local skill supply and the adoption of new skill-intensive technology. We believe graduates have comparative advantage in abstract tasks and so the increase in graduates would lead to an increase in abstract employment. If the same new technology is also less intensive in routine tasks, then we should also see a negative correlation between local graduate supply and routine occupations' employment share.

In Table 3.4, we look at the correlation between education composition and occupational shares at the region-year level. We include year fixed effects to allow nation-wide task-biased demand shifts. It's clear that all three abstract occupations have their shares very strongly and positively correlated with the local graduate proportion. Two of the three routine occupations are significantly negatively correlated with the graduate proportion. While these correlations are only descriptive, they are supportive of the hypothesis that increasing graduates causes the shift of employment from routine to abstract occupations in the UK.

Table 3.4: Local skill supply and occupational employment shares

	managerial	professional	technician	admin	skilled trades	operative
% Graduate	0.0658**	0.1941***	0.1741***	0.0814***	-0.1699***	-0.2379***
	(0.0215)	(0.0185)	(0.0211)	(0.0201)	(0.0317)	(0.0224)
% HS-dropout	-0.0797**	0.1323***	0.1082***	0.1890***	0.0830*	0.0624*
	(0.0270)	(0.0232)	(0.0265)	(0.0252)	(0.0398)	(0.0281)
Year fixed effect	yes	yes	yes	yes	yes	yes
Observations	491	491	491	491	491	491

Note: each regression is at the region-year level. There are 19 regions in total, and Northern Ireland appeared from 1995 onwards. The dependent variable is the occupation's share of total hours in the region-year. Source: LFS 1992-2017.

3.6 Estimation approach and results

Currently, the model is partly calibrated and partly estimated. I calibrate $\rho = -0.1$ and $\zeta = 0.1$. $\rho = -0.1$ corresponds to Goos et al. (2014)'s 0.9 estimate of the substitution elasticity between tasks. I have experimented with several values of ζ and found $\zeta = 0.1$ yields a good fit of the data overall. It is left for future work to estimate all the parameters by a method of moments.

Given the calibrated ρ, ζ, I estimate all the other parameters according to the methods discussed in section 3.3.3. Given all the parameters, I solve for the equilibrium $(\mathbf{p_t}, \mathbf{w_t})$ in each year. I search for the equilibrium that is closest to the observed and satisfies all the equilibrium constraints within tolerance[46].

[46]Demand minus supply in any occupational employment share is at most 1e-4 in absolute value. The

3.6.1 Parameter estimates and model fit

First, let's compare the estimated task intensities between the two technologies. Figure 3.6 shows the task intensities $\alpha^O_{gj}, \alpha^N_{gj}$ in all 7 industries. Because their identification comes from shifts in occupational employment shares and wage bill shares within industries, the comparison between α^O_{gj} and α^N_{gj} is fairly robust to the calibrated value of ρ.

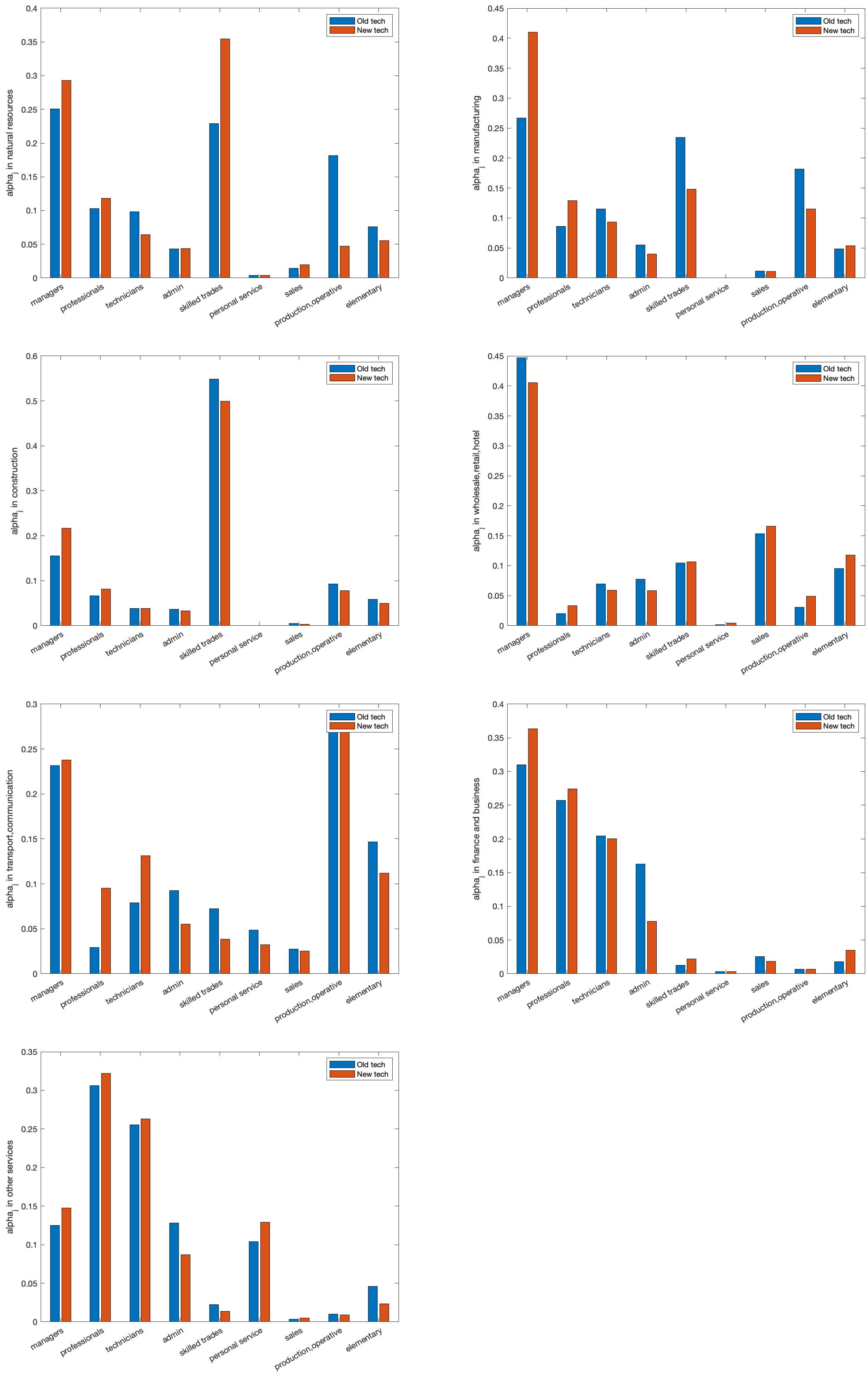
Take the manufacturing industry for example. It is intensive in three tasks: managerial, skilled trades and machine operatives. The New technology is more intensive in managerial task, and less intensive in the other two manual routine tasks. This is what we expect. And this is driven by the data: within manufacturing employment has shifted substantially away from manual routine to managerial. Meanwhile, in non-financial services, the new technology is less intensive in admin and elementary and more intensive in all 3 abstract tasks and personal service task.

Some patterns are common across industries. In all industries, the New technology is more intensive in professional task. In 6 out of 7 industries, the New technology is less intensive in admin task, and more intensive in managerial task. In 5 out of 7 industries, the New technology is more intensive in managerial task. In the natural resources industry, the New technology compared to the Old technology mainly involves a shift from operatives to skilled trades. Other than that, for skilled trades, in the industries where it is sizeable, the new technology is either less or equally intensive in it than the Old technology. The same is true for machine operatives. Among the lower-skilled tasks (personal service, sales and elementary), there is little evidence of the New technology being more or less intensive. While the direction of bias of technological change varies across industries, the overall pattern is that the New technology is biased against the three routine tasks and towards managerial and professional tasks.

---

unit costs of two technologies can differ up to 1%. This is not very big relative to the uncertainties in our parameters estimates. For example, a change of one standard deviation in r^O_{g2} (while holding other r^O_{gj} the same) would change the log unit cost by 0.01-0.04, depending on the industry g.

Figure 3.6: Estimated task intensities in each industry

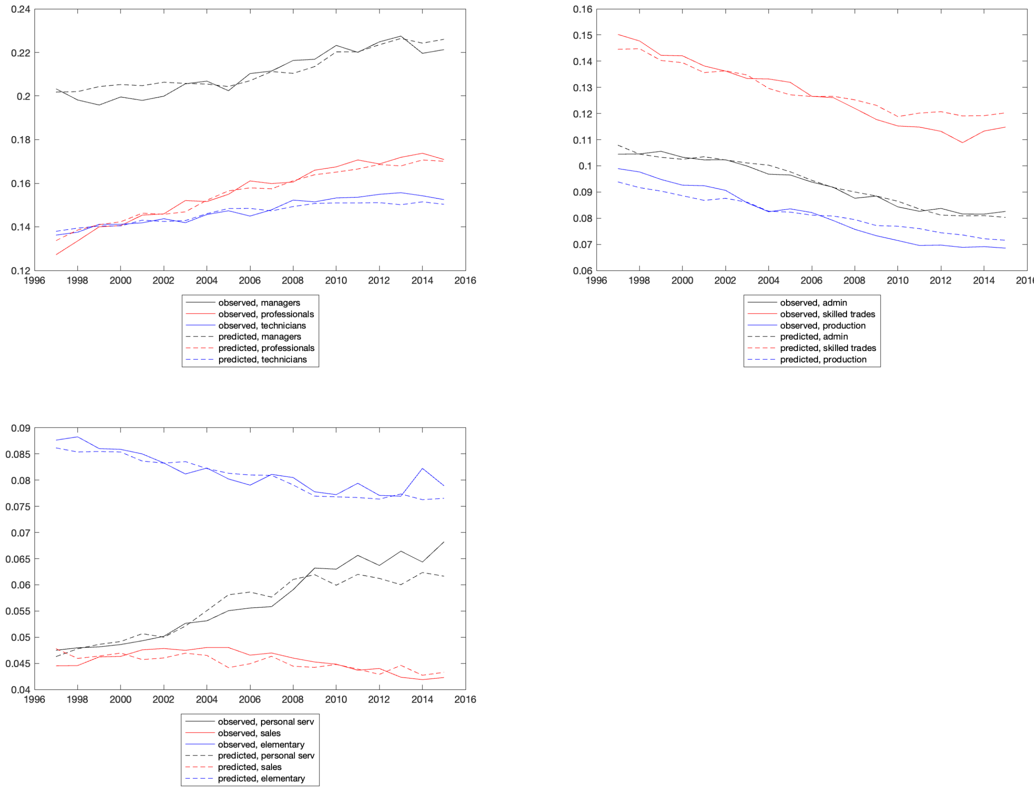


Next, we examine how the key endogenous variables in the model fit the actual trends. Recall that the only time-varying exogenous factors in the model are TFP of both technologies, industry demand, and aggregate skills distribution. And the last is based on education-gender-specific skills distribution and the evolving demographic composition, so there's no free parameter in that aspect. The parameters particularly important for employment shares such as the task intensities α^T_{gj} and the occupational amenity η_j are assumed to be constant. Therefore, the design of the model does not mechanically guarantee a good fit of time trends.

Figure 3.7 shows the observed and predicted trends in occupational employment shares. For all of the 9 occupations, the model fit is quite good. Every occupation with an observed declining [increasing] trend has a predicted declining [increasing] trend. And the difference between observed and predicted employment shares is no greater than 1% of aggregate employment. Figure 3.8 shows similarly a good fit for log task prices. In Appendix B.5, we plot the fit for log industry prices (Figure B.6) and for technology shares w_{gt} (Figure B.5). Note that given the parameters, the endogenous variables are obtained through a search for $\log \mathbf{p_t}, \mathbf{w_t}$ that is closest to the observed and subject to satisfying the equilibrium constraints.[47] This means it is expected that we get a good fit for $(\log P_{jt}, w_{gt}, \forall j, g, t)$. The fact that the model can capture the trends in occupational employment share movements means that the calibrated/estimated parameters are not too bad.

[47]This is because there are multiple points of $\log \mathbf{p_t}, \mathbf{w_t}$ that satisfy the equilibrium constraints within tolerance.

Figure 3.7: Fit of occupation employment share



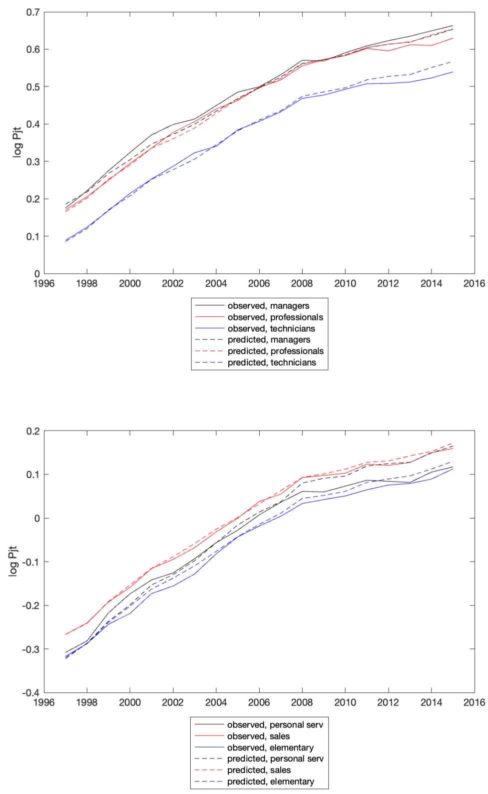
Note: The actual time trends of occupational employment shares are solid lines. The corresponding baseline predictions are dashed lines of the same colour.

3.6.2 Counterfactuals

The model contains three sources of exogenous time-varying factors: TFP of two technologies, industry demand, and the skills distribution. In this section, we will examine how each of them affected occupational prices and employment in the past. In future, I would also like to examine counterfactuals about the Brexit-induced shift in the supply of skills and future education increases.

In each counterfactual, only one exogenous factor changes over time while others stay the same as 1997. Because numerically there are multiple equilibria, for each year, I search for $(\log \mathbf{p_t}, \mathbf{w_t})$ that is closest to some benchmark subject to equilibrium

Figure 3.8: Fit of log task prices P_{jt}

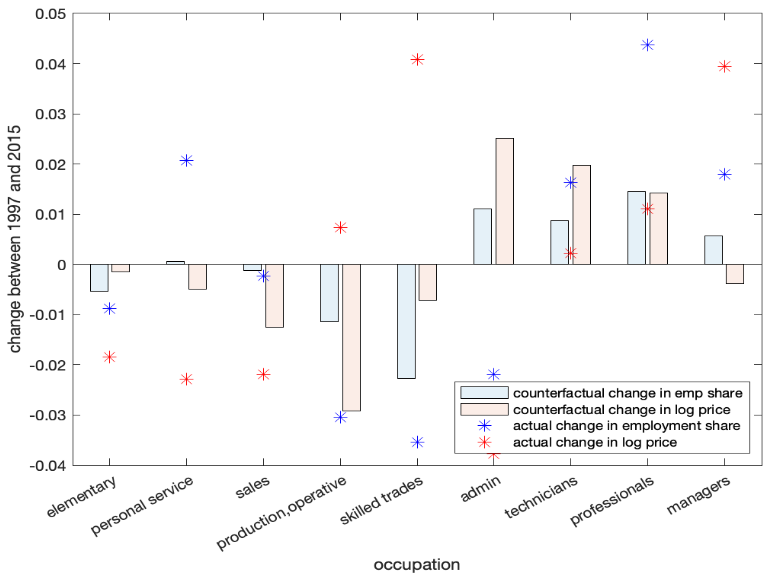


Note: The actual time trends of task prices are solid lines. The corresponding baseline predictions are dashed lines of the same colour.

constraints. The benchmark is the corresponding values in $t-1$ for $t > 1997$, and the observed values for the first year ($t = 1997$). I interpret the result as a lower bound on the effect of shifting that factor.

Figure 3.9 considers the counterfactual where the skills distribution shifts over 1997-2015 while TFP and industry demand are constant. In this counterfactual scenario, the equilibrium task employment would shift significantly. Although I have used the same axis for employment and task price changes, the magnitude of changes should be interpreted differently. An increase of 0.01 in professional employment share is about a 10% increase from its initial employment share, whereas the 0.01 change in its log price is close to a 1% change. For skilled trades, the counterfactual employ-

Figure 3.9: Counterfactual: only skills distribution shifted



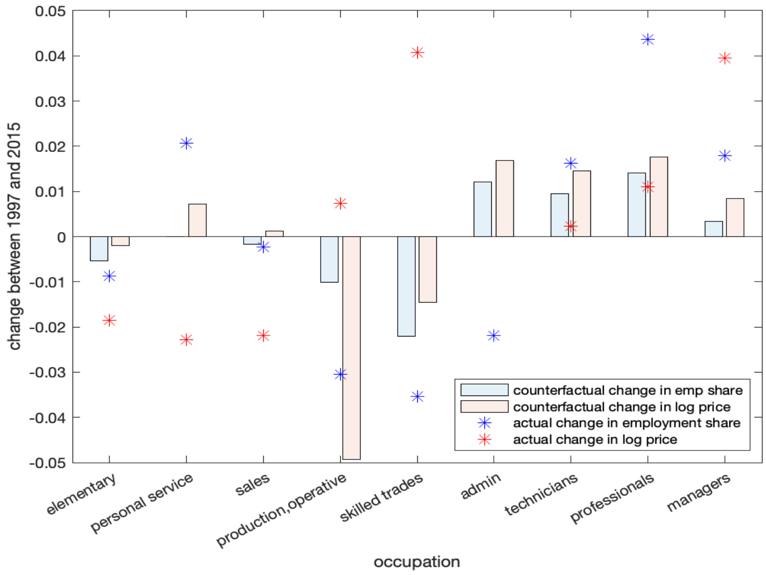
Note: using initial $\log P_{jt}, w_{gt}$ as the benchmark. For log task prices, we normalise the average change across 9 occupations to 0.

ment share falls from 14.5% to 12.2%, that is -0.17 in log terms. Meanwhile, all the counterfactual wage changes are less than 3% in absolute terms.

Figure 3.9 also shows the actual changes as markers, so we can see that the supply shift alone could account for between a third and a half of the actual increase in the three abstract occupations over the 18 year period. It can also account for between one third and two thirds of the actual decline in manual routine occupations. The impact on admin occupational share is in the opposite direction to the observed change, and the impact on manual occupational shares is smaller.

Figure 3.10 examines the effect of industry demand shifts. We hold TFP and skills distribution constant, and let B_{gt} follow the actual trend. This counterfactual represents a shift in the demand curve. Because the industry demand shifts were strongly against manufacturing, and the manufacturing industry is very intensive in operatives and skilled trades, we see a large decline in both employment and task price for opera-

Figure 3.10: Counterfactual: only industry demand shifts



Note: using initial $\log P_{jt}, w_{gt}$ as the benchmark.

tives and skilled trades in this counterfactual. Industry demand shifts alone can account for a third of the employment decline in machine operatives, and over half of the decline in skilled trades. It can also account for a third of the increase in professional employment and over half of the increase in technician employment.

3.7 Conclusion

This chapter develops a multi-sector equilibrium model of endogenous task-biased technological change that simultaneously explains three notable phenomena in the UK labour market. First, the UK has seen very strong employment growth in high-paying occupations and an even bigger decline in middle-paying occupations since the 90s. Second, changes in occupational wages are small and uncorrelated with employment changes. Third, there was relatively little occupational downgrading within education groups during a period of rapid increases in education. In addition to these three macro facts, I have provided regression analysis that supports my model and is at odds with the hypothesis of exogenous technical change.

This chapter contributes to the polarisation literature by emphasising the endogenous nature of technology *adoption*. The key driving force in my explanation is a large positive shift in the supply of skills. This supply shift causes firms to adopt a new technology that's biased against routine tasks and in favour of abstract tasks. This technology shift helps to absorb the impact of the supply shock on wages. As a result, we get substantial movements in employment shares, little changes in occupational wages, and little change in the mapping from skills to occupation. To the extent that the skills distribution within graduates are stable, the last outcome means little occupational downgrading within graduates.

The calibrated model can fit UK data well over 1997-2015. While the estimated direction of technical change varies across industries, the overall pattern is that the New technology is less intensive in all three routine tasks and more intensive in managerial and professional tasks, with less difference in other tasks. The shift in skills distribution alone can account for between a third and two thirds of the actual decline in routine manual occupations, and between a third and half of the increase in each of the three abstract occupations. The shift in industry demand can account for similar magnitudes of employment declines in routine manual occupations and increases in professionals and technicians.

While this chapter focuses on the UK, it provides a promising framework to study issues around occupations and education in other advanced economies except the US. These countries often share some of the key facts observed in the UK since the 90s. First, like the UK, employment growth has been strongest in high-paid occupations in most European countries. This is consistent with the New technology being more intensive in abstract tasks. Second, occupational wages did not polarise, except for the 90s in the US. And third, the US had the highest proportion of graduates in 1990 and a slower increase afterwards than many European countries. Among the European countries that saw large increases in higher education, the majority did not see a significant impact on graduates' relative wage; and there has also been relatively little increase in the share of graduates in non-graduates occupations. These empirical differences between the US and the other advanced economies are intriguing, and worth further investigations.

Conceptually, the main point of my proposed framework is that the adoption of technology depends on current prices and skill supply. This is fundamentally different from the scenario where a new technology becomes available and it's unambiguously better than the existing one so that all firms should adopt the new technology immediately in the absence of fixed costs or frictions. That scenario might be a good enough approximation of reality in some cases; but in general, incremental changes of the technology frontier mean that there is often a meaningful choice to be made between relevant technology options. I believe many European countries are close enough to the technology frontier that their firms are in a position to choose between recent technologies, and that the decision depends on prices and skill supply. In principle, the same argument of endogenous adoption should apply to the US as well; but because it's a major innovator and has experienced a smaller increase in education in the past three decades, the role of skill-supply-induced adoption of technology might be much smaller than other factors in the determination of occupational trends.

Finally, the paper offers a data-driven approach to answer a few policy questions about the labour market. By having analytical and social skills (instead of education) as determinants of worker productivity, it allows a lot of heterogeneity within education groups and opens up the possibility of modelling changes in the group-specific skills distribution over time. For example, I plan to use the UK Life for Skills Survey to check whether the education-specific skill distribution has deteriorated between birth cohorts, as higher education has become less selective. Then, the relevant data moments can be fed into the full model to investigate the effects of education expansion. The approach also makes clear that for analysing any policies that shift labor supply, it is important to model potential changes in the distribution of skills that matter for productivity, rather than labels like education.

Another interesting question left for future work is the effects of immigration on the aggregate labour market. Currently, immigrants in the UK are over-represented in both high-paying occupations and low-paying occupations. My next step is to estimate skill differences between British workers and immigrants, and examine to what extent the differences in occupation destinations are explained by skills (not just reported education level), as opposed to preferences or discrimination. Then, we can examine differences in skills distribution between European immigrants and non-EU immigrants, and ask what's the effect of forcing European immigrants to be as skilled as the non-EU ones on the UK labour market.

Chapter 4

Female labour supply in urban China

4.1 Introduction

In 2013, the employment rate among 55-64-year-old urban women in China stood at 27%, well below the rates seen in most other countries at all levels of development. The low employment rate is worrying given the rapid ageing of China's population and the unfunded nature of the public pension system. According to the United Nations' projections, China's total dependency ratio (ratio of population below 15 and 65+ per 100 population 15-64) is estimated to rise from 36.6 in 2015 to 69.7 in 2050, and to peak at 90 in 2085.[1] The biggest public pension scheme (covering enterprise employees) is run mostly on a pay-as-you-go basis and has been projected to have a financing gap equivalent to 95% GDP for the period 2001-2075 (Sin, 2005).[2] The government is already considering a large and gradual increase of the formal pension age from 50 for female workers (55 for female cadres or managers[3]) and 60 for men to 65 for all. To what extent will such a policy increase employment among the elderly is an important

[1] Source: World Population Prospects: The 2015 Revision, File POP/11-A: Total dependency ratio $(< 15\&65+)/(15-64)$ by major area, region and country, 1950-2100 (ratio of population 0-19 and 65+ per 100 population 15-64), Medium fertility variant, 2015 - 2100

[2] Sin (2005) uses administrative data from a few provinces and municipalities. Under the baseline projection, the NPV of sum of the shortfall of revenue relative to expenditure over the period 2001-2075 is estimated to be 95% of the GDP in 2001.

[3] See subsection 4.2 for more details on the current pension system.

empirical question.

The contribution of this chapter is two-fold. First, I offer two broad explanations for the low employment rate among older urban women. I focus on urban women, because the proposed change to retirement age is much greater for women than for men, and because urban women's current employment rates are much lower than their rural counterparts.[4] The two main explanations are the low pension age and the expectation of transfers from and to their adult children. I make the arguments using a wide range of descriptive evidence from household surveys. While similar arguments and the same data source have been explored in Giles et al. (2015) [5], I document some facts and correlations that are new.

Second, this chapter uses a structural life-cycle model to simulate the effects of potential policy changes and the importance of inter-generational transfers on older women's labour supply. To the best of my knowledge, this is the first paper to use a structural model to simulate the impact of increasing the pension age on labour supply in the Chinese context. I found that increasing the female pension age from its current variable levels to 60 for all would increase the employment rate of 50-59 year olds by 28 percentage points.

To understand labour supply behaviour of older women, we must first talk about the Chinese pension system. China's pension system is fragmented and will be described in detail in subsection 4.2. For urban residents, the biggest public pension schemes set the formal retirement age at 50 for female workers in general and 55 for female "cadres" or managers, and 60 for men. However, there are different rules for people with special circumstances such as ill health and compliance is not perfect, so the age at which a woman

[4] The disparity between urban and rural residents is documented in Giles et al. (2015). Meanwhile, Shu (2018) investigates the impact of a new pension program targeted at rural elderly on their labour supply.

[5] It was the first use a newly-available nationally representative survey (CHARLS) to describe labour supply behaviour among older Chinese people. For a description of the CHARLS dataset, see Appendix C.1. It highlights the differences in economic environment and retirement behaviour in rural versus urban residents, and uses descriptive correlations to postulate which factors might be at play.

becomes eligible for a public pension can be as early as 45 or as late as 60. In general, individuals are supposed to complete the retirement process in the month of becoming eligible, not before or after. Here and throughout the paper, "completing the retirement process" and "processed retirement" refer to the act of claiming an employment-based public pension for the first time and shall not be confused with individuals' labour force participation status.[6] According to CHARLS 2011, 90% of current female pensioners completed the retirement process by 55.

The observed timing of exit from the labour force is closely related to the timing of pension incomes. About 70% of current pensioners aged 60 and above had stopped working in the same year as they became eligible for a retirement pension. This coincidence in timing is striking because all the retirement pensions are not means-tested or carry any effective tax rate. In fact, if one has started to receive a retirement pension and continues to work, s(he) and the employer will not need to pay social security contributions any more. In other words, the effective tax rate is actually lower after completing the retirement process.

In the standard life-cycle framework, if individuals are not liquidity constrained and if the net present value of the pension income is known, then how it's distributed over time should not matter. The observed bunching in timing might reflect discontinuous changes in preferences or wages at the time of the first pension receipt. Another plausible explanation is high transition costs: the default is for the individual to leave the employer at the time of the retirement process. Negotiating a new contract with the same employer or searching for a new job might be too costly for some pensioners. It is also possible that the formal retirement age has an anchoring effect on individuals' beliefs about when they should stop working. In section 4.4, my model will incorporate a transition cost (for returning to work), uncertainties in pension income, and liquidity constraints. These features alone are able to create a significant amount of bunching in

[6]More institutional background around retirement can be found in subsection 4.2.

the timing of exit from employment, though not as high as 70%.[7]

The second explanation for the low employment rate among older females is their adult children. First, expected financial transfers from children when one is old and frail reduce the need to accumulate a large stock of assets through working. As we will see in section 4.3.2, about 60% women in their 50s live with their children; and conditional on having non-coresident children, about 70% receive financial support from their non-coresident children in the year prior to the survey.[8] There is some evidence that the financial transfers from adult children respond to parental incomes. Thus, transfers from adult children have a wealth effect as well as an insurance effect, both of which have implications for the older womens' decisions about labour supply and saving.

Moreover, demands from adult children for home production such as caring for grandchildren meant less time is available for paid employment. In urban China, the majority of women have grandchildren before 60. Conditional on having grandchildren, the majority provide regular childcare. And conditional on looking after grandchildren, the typical weekly hours spent on childcare is no less than a full-time job for a majority of grandmothers.

There are reasons to believe that parents cooperate with their adult children when choosing between market and domestic labour supply. First, the prevalence of co-residence and substantial transfers between generations are indicative of strong altruism between generations, which facilitate commitment and cooperation. Second, the older generation often have a comparative advantage in home production. On average, young people today are much more educated than their parents' generation, and they have a higher incentive to accumulate experience as they have a longer potential working life ahead of them. This inter-generational exchange or specialisation could be an important cause of the low employment rate among women in their 50s. Consistent with this

[7] Incorporating discontinuities in preferences and wages at the point of the retirement process will help match the observed amount of bunching. This is left for future work.

[8] We only observe transfers between parents and non-coresident children in CHARLS.

hypothesis, I find that urban female employment is significantly negatively correlated with having grandchildren (conditional on her own age, education and so on), and this correlation is driven by women with more educated children. The latter correlation has not been documented before, and it is more suggestive of comparative advantage rather than altruism as the underlying motivation.

All these inter-related facts about Chinese households call for a rich structural model. Ideally, such a model should endogenize two generations' labour supply, saving, and transfers of money and time in a dynamic way, and it should incorporate uncertainties in wages, pensions, transfers, and health. This chapter takes a first step towards such a model. In section 4.4, I build and calibrate a life-cycle model of labour supply and household saving for women from age 45 to 75. The model includes uncertainties in male income, female pension and female wage, and approximates transfers from adult children as a simple function of parental incomes.

The calibrated model can generate a rapid and realistic fall in female employment rate over the life cycle. It also generates some bunching in the timing of exit from the labor force at the point of pension receipt, without assuming discontinuities in the disutility of work or in wages. Counterfactual simulations show that the female employment rate is sensitive to both the pension system and financial transfers from their children. At each age between 50 and 59, the female employment rate would be 9-36 percentage points higher under the counterfactual of raising the female pension age to 60, compared with the current scenario. If there were no transfers from children at all, the employment rate would be higher by 6-19 percentage points between 50 and 59. Unfortunately, the current model does not generate realistic predictions of household saving. I will discuss the reasons and what model extensions could help improve the fit. These are left for future work.

Section 4.2 describes the institutional background. Section 4.3 documents many descriptive facts and correlations around labour supply and inter-generational transfers. Section 4.4 sets out the model, the parametric specifications and the calibration approach. Section 4.5 discusses the calibrated benchmark case as well as some counterfactual results.

4.2 The institutional background

The population of interest in this chapter are Chinese women aged 45 or above with an urban Hukou. In China, Hukou is a system of household registration that determines a person's rights in the local area. There are two broad categories of Hukou: rural (agricultural) and urban (non-agricultural). One's Hukou status determines one's access to many public services and benefits in the local area, including public pensions, public education, and social insurance for healthcare. Given the large differences in the economic environments faced by urban and rural Hukou-holders, this chapter will focus on the urban ones.

There are multiple public pension schemes covering different populations in China. Among women with urban Hukou who report receiving any kind of pension in CHARLS 2013, 58% received the Basic Pension Program for Firm Employees, and 18% received the Pension Program for Institution Employees (such as schools and hospitals), and 8% received the Pension Program for Government Employees (civil servants and military personnel). In addition, there are commercial pensions, firm-supplementary pensions, and various other non-mandatory public schemes with lower benefit levels and much fewer recipients. This subsection will focus on the Basic Pension Program for Firm Employees, with some discussion of the Pension for Institution Employees and the Pension Program for Government Employees. For the rest of the paper, I will refer to these three as employee pensions.

The Basic Pension Program for Firm Employees was initiated by the 1997 State Council Document No.26 "Decision of the State Council on Establishment of Unified

Basic Old Age Insurance System for Enterprise Staff and Workers".[9] Document No.26 sets out the two-tier pension structure for firm employees that has broadly remained in place today. As with previous pension reforms, the State Council Document outlined the main features of the new pension including the key parameters like pension age. The more detailed policy parameters and rules, and the timing of implementation is left to the discretion of local governments.

The formal retirement age is 60 for men, 50 for female workers generally and 55 for female "cadres".[10] There are exceptions to the formal retirement age, based on factors such as long-term illness and disabilities, whether the occupation is physically harmful or demanding, and length of service. The detailed rules, such as which manual occupations and which illness qualify for earlier retirement, are subject to variable interpretation at the local level. In the 90s, early retirement was commonly granted to employees in bankrupt state-owned-enterprises who were close to their retirement age (West, 1999).

To claim the Basic Pension for Firm Employees, one needs to contribute to the program for at least 15 years.[11] While participation in the program is mandatory for employees in all enterprises, regardless of their ownership, compliance is far from perfect.[12] Moreover, the participation rate varied hugely across localities and types of enterprises. The ones mostly likely to participate were state-owned enterprises, followed by collective-owned ones, privately-owned firms and self-employed individuals. This means, contribution history varied a lot across individuals. Those with insufficient

---

[9] Appendix C.2 gives a brief description of the pension system prior to that.

[10] "Cadre" is an administrative ranking of employees in the public sector. In the private sector, it broadly corresponds to managerial and professional staff. The interpretation of "cadre" is not always clear, so individual women may be unsure of their own pension age.

[11] Continuous employment before the reform counts towards the years of contributions.

[12] As a share of all urban workers, participation rate remained below 50% till the mid-2000s. According to table 4-2 and table 22-44 in China Statistical Yearbook 2010, the number of participants in the Basic Pension Program for Firm employees as a share of the urban working population rises was 50% in 2006. It appears that the nominator does not include workers in government and public institutions, who had separate pension schemes.

contributions upon reaching their retirement age could make a one-off large payment to qualify, or they may have to make contributions for a few more years till they reach 15 years of contributions before qualifying. The rules vary geographically and over time, and some people might have to take back their limited past contributions in a lump sum and they would never qualify for the Basic Pension for Firm Employees.[13]

Upon reaching the formal retirement age and with sufficient years of contribution, one will be notified by their employer to initiate the process of claiming retirement at the local bureau of social security. There is no option of early or late retirement with actuarial adjustment.

In the Basic Pension Program for Firm employees, Tier 1 is a basic pension funded out of a social pool on a PAYG basis. Contributions are a substantial proportion of employees's gross wages subject to a maximum and a minimum base.[14] The benefit level was initially set to 20% of the local wage, but from the mid 2000s onwards it also depends on the number of years of past contributions and the individual's relative-wage index (past wages relative to the local average).[15] Tier 2 is an individual account, designed to be fully-funded but is largely notional in practice. The contribution rate varied across cities and over time, and was unified to 8% by the 2005 State Council Document No.38. The monthly benefit equals the accumulated balance in the individual account divided by a coefficient, which was initially set to 120.[16] For people who worked before the establishment of individual accounts and retired after the reform, there is a transitional supplement which recognises their implicit pre-reform contributions.[17]

[13]Such people could often claim the Urban Resident Pension. This is a non-mandatory scheme and isn't linked to employment. It gives benefits from age 60 onwards. The benefit amount and the amount of annual contribution required before age 60 are both much lower than under the Basic Pension Program for Firm Employees.

[14]The contribution rate varies across municipalities and averaged 19% in 2001.(Sin, 2005)

[15]In Liaoning province and a few cities, the basic pension started to depend on the years of past contributions in 2001. The new formula was stipulated in the 2005 State Council Document No.38, which took effect from the start of 2006.

[16]Since the 2005 State Council Document No.38, the coefficient could depend on the claimant's age and the local life expectancy.

[17]The transitional component depends on current average local wage, the individual's relative-wage

The system does not have fixed rules for benefit indexation over time. Every year, the local government would announce how the entitlements of existing pensioners would increase, which might depend on the individual's age, last year's pension, and number of years worked. In general, the entitlement growth for existing pensioners is similar to local average wage growth.

Meanwhile, public-sector employees are covered by more generous pensions (the Pension Program for Government Employees and the one for Institution Employees). These public-sector pensions have the same retirement age as the Basic Pension Program for Firm employees. What's important for us, the age at which one can apply for retirement from the public sector is also not a choice. Before October 2014, the public-sector pensions were Defined-Benefit: there was no individual contribution into the program. Since October 2014, both workers and employers in the public sector need to pay similar contributions. Before 2014, the amount of pension entitlement under the two public-sector schemes depended on the local average wage and the pensioner's past wages. Since 2014, the entitlement uses a two-tiered formula similar to the Basic Pension Program for Firm employees. For public-sector employees who worked before 2014 and retired after, their work history before 2014 is used to impute notional contributions into their individual account. While the national government proposed to protect workers during the transition, the detailed rules are set by the local government.

In this chapter, we use CHARLS data collected in 2011 and 2013 only. For future research, more recent waves of data may reveal how the reform to public-sector pensions have unfolded over time and across cities.

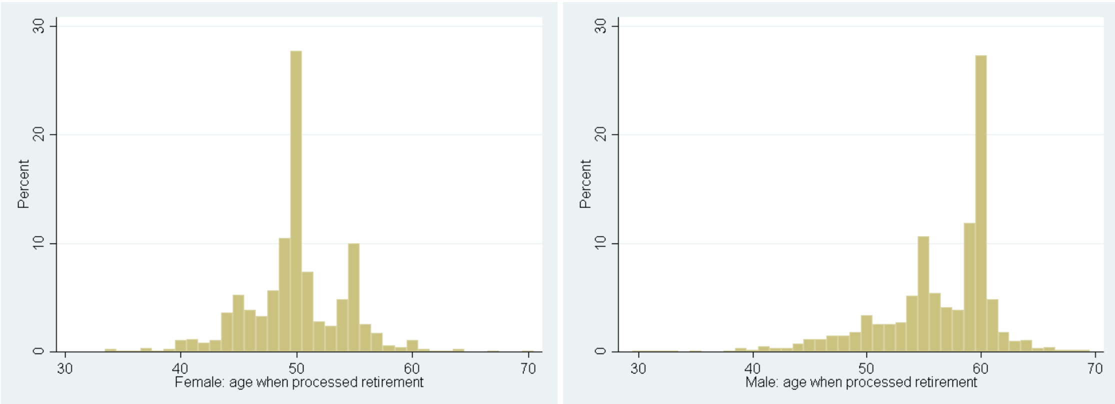
The CHARLS survey asked whether the respondent has completed the retirement process from an enterprise or public-sector employer and when. Figure 4.1 below plots the distribution of the age at which men and women completed the retirement process. The women's distribution has clear spikes at 50 and 55, but more than half of women

---

index and the length of employment before the reform.

completed the retirement process at other ages.

Figure 4.1: Histogram of age when processed retirement by gender



Note: the sample are individuals who have urban Hukou and have completed the process of retirement in CHARLS 2011. Here 'retirement' refers to the claim of Employee Pension, and it does not mean stopping to work.

In addition, a public pension not linked to employment called the Urban Residents' Pension has been gradually rolled out since 2011. In some places it has been merged with the New Rural Pension Scheme (piloted and rolled out since 2009) to become the Urban and Rural Residents' Pension. Contributions to these non-employment-linked pensions are voluntary and are linked to the future entitlement. Pensions are payable from age 60 onwards. 9% of urban pensioners in CHARLS 2013 receive these non-employment-linked pensions. The average amount of these pensions reported in CHARLS is less than half the average value of the employment-based pensions. As we will obtain moments of pension income from CHNS 1989-2011, these non-employment-linked pensions are unlikely to be included.

Finally, 8.6% pensioners in CHARLS 2013 receive a subsidy for "very old" people, with average value below one tenth of that of the Enterprise Employee Basic Pension. Every other type of pension including commercial pension and supplementary pension from one's past employer covers one percent or fewer pensioners in CHARLS.

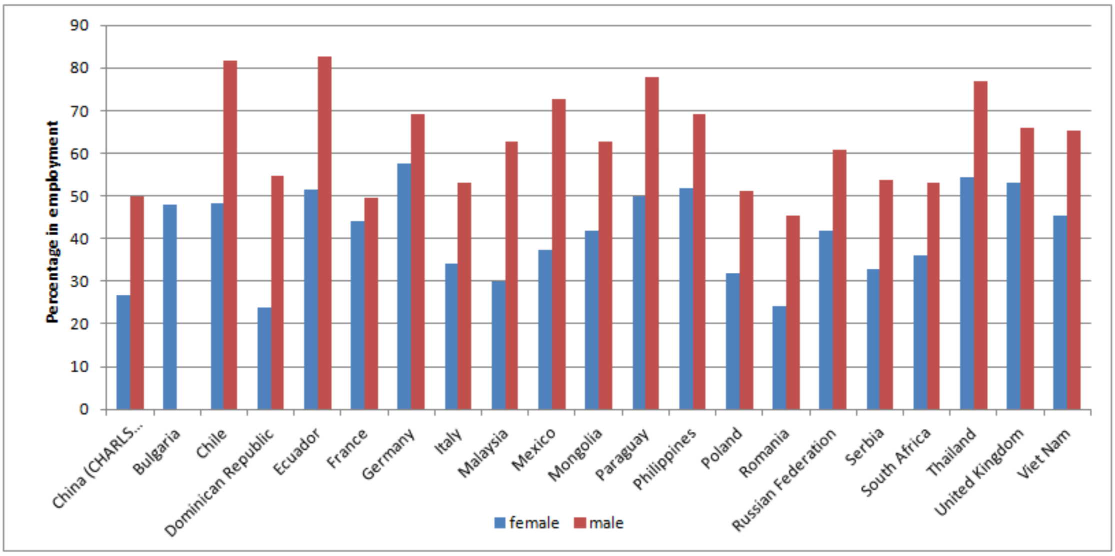
Working after completing the retirement process has no effect on one's pension entitlement. And the time at which one can complete the retirement process is not up to the individual. This raises the question: why don't women work after completing retirement process in their early 50s. This group clearly do not face disincentive to work from their pensions.

4.3 Descriptive evidence

4.3.1 Labour supply of older women

The employment rate of urban females in China is low by international standards. Figure 4.2 shows that the urban female 55-64 employment rate is around 50 – 55% in the UK, Thailand (which has similar GDP per capita to China) and Philippines (which has lower GDP per capita). China's urban male employment rate is also low by international standards.

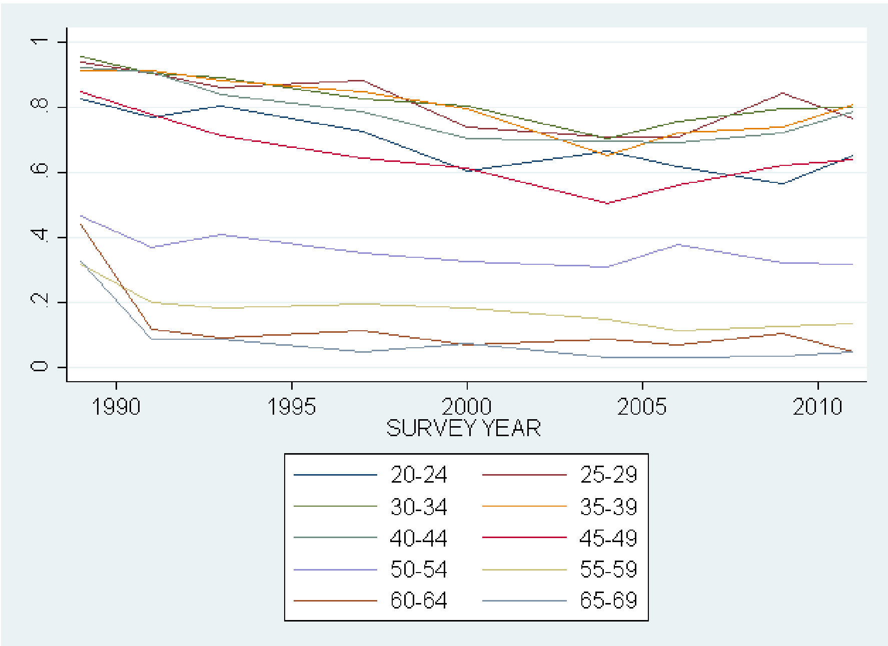
Figure 4.2: Female employment rate in urban areas across countries



Note: The data points for China are based on people with urban Hukou in CHARLS 2013. The CHARLS definition of employment is quite broad, including working at least 10 days in the past year in agricultutral and working at least one hour in the past week in paid work, self-employment or unpaid family business. Alternative definitions would give an even lower employment rate. The statistics for all other countries are from ILOSTAT (the International Labour Organization's website). The statistic is the employment-to-population ratio by sex, age and rural/urban areas. Not all countries have this series. The numbers used here are the 2013 values, or the 2014 or 2015 value if the earlier values are missing.

The employment rate of middle-aged and older urban women has been fluctuating around a low level for many years. Figure 4.3 shows the urban female employment rate by age group, using the China Health and Nutrition Survey.[18] The employment rate declined throughout the 90s and early 2000s among middle-aged women, but the trend is less clear after the early 2000s. The decline in the late 90s is due to restructuring of State Owned Enterprises, when older workers were laid off and those near retirement age were encouraged to take early retirements with formal or informal pension payments (Appleton et al. (2002), Giles et al. (2006)).

Figure 4.3: Female employment rate by age



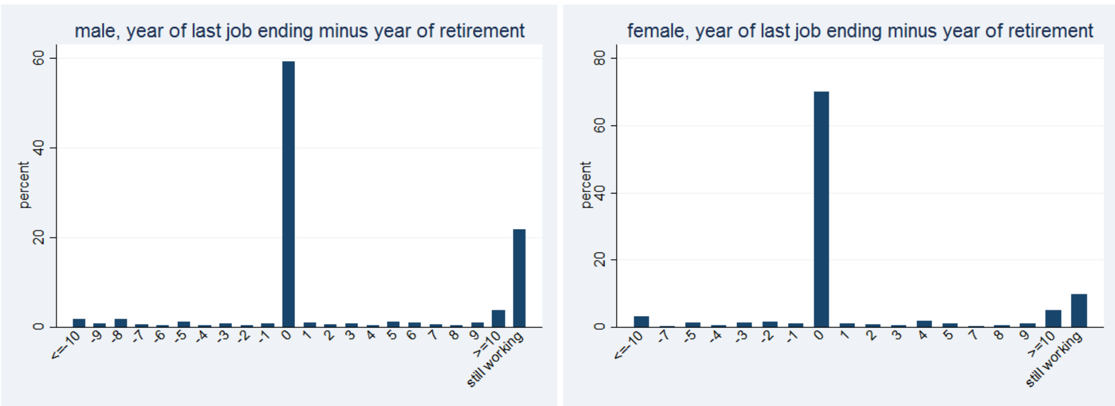
Note: women with urban Hukou in CHNS1989-2011. The level of employment rates measured in CHNS tends to be lower than that in CHARLS. The CHNS questionnaire asks “are you presently working”. Whether respondents’ interpretation of this includes self-employment or family business is doubtful.

I document a large amount of coincidence in timing between stopping work and when one becomes eligible for an employee pension. In Figure 4.4, we see about 70%

[18]Unlike CHARLS, the CHNS is not nationally representative and no weights have been constructed to make it representative. Nonetheless the trend is illustrative.

of current female pensioners above 60 had stopped working in the same year that they became eligible for an employee pension. A small proportion left their last job before retirement and about 10% were still working at the time of the survey.[19] The degree of bunching in Figure 4.4 is much bigger than what's been documented in previous papers. For example, Figure 4 in Giles et al. (2015) shows that 10-20% of women stopped working at age 50 and 55, the theoretical pension ages. This difference arises because I look at the time when women qualify for an employee pension, instead of their age. In the data, many women qualified for pensions at ages other than 50/55 and they stopped work immediately.

Figure 4.4: Distribution of the gap between the year of stopping work and the year of completing the retirement process



Note: the sample are individuals with urban Hukou who have completed the retirement process as last reported in CHARLS 2011 and 2013 and were at least 60 in 2011. It excludes individuals who have missing values in either the year of retirement or missing values in the year when their last job ended if they are currently not working.

The coincidence in timing between pension receipt and labour market exit has been documented for other countries, and a wide range of explanations have been proposed in the literature including specific economic incentives and anchoring effects. For example, in the US there are also pronounced peaks in retirement age at 62 and 65, which are explained by borrowing constraints, actuarial unfairness for delayed retirement and

[19]The sample includes all pensioners aged 60 or above, without missing values in the necessary variables.

lack of health insurance before qualifying for Medicare at 65(Rust and Phelan, 1997). In our empirical context, those explanations are not applicable or less important: individuals cannot choose their pension age with actuarial adjustment; the system contains no disincentive for labour supply after reaching pension age, and the households are not liquidity constrained very often[20]. This bunching phenomenon could be a result of discontinuities in preferences and market wages at the time of pension receipt, both of which would be hard to identify. In section 4.4, I propose a model to rationalise this phenomenon with a search cost for finding a new job after pension receipt, as well as uncertainties in pension entitlement, liquidity constraints, and income-contingent transfers from children. The search cost feature is based on the empirical context: in urban China, exit from the former job is mandatory at the time of processing retirement, although the individual could go back to the formal employer on a temporary or self-employed contractor, or to find other formal or informal jobs.

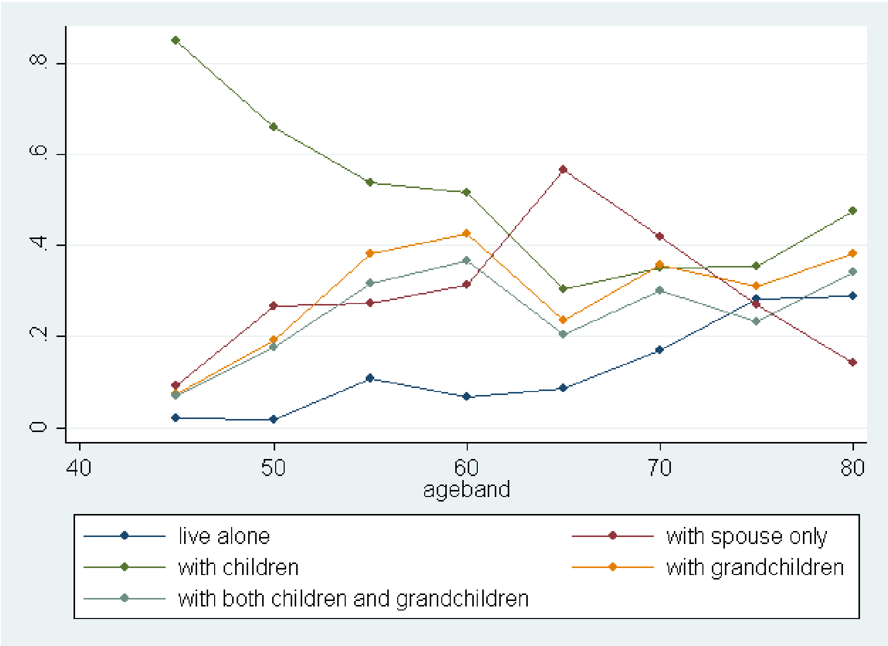
4.3.2 Economic interactions between parents and adult children

In China, the economic relationship between parents and adult children tends to be very close. Supporting and caring for one's elderly parents is not only a long-lasting social norm, but also an obligation of the adult children specified in the Law on the Protection of Rights and Interests of the Aged. When asked "Who can you most rely on for old-age support?", about 60% CHARLS respondents say it's their pensions, 30% say it's their children, and less than 10% say it's their own savings.

Firstly, it is very common for older people to live with or near their adult children, even after the children get married and have children themselves. As shown in Figure 4.5, more than 50% of urban women in their 50s live with children; about a third between 55 and 65 live with both their children and grandchildren. When the children do not live in the same household, they often live nearby. Table 4.1 shows that urban women of any age group (45-79) have on average at least one child living in the same

[20]Chinese households have high saving rates

Figure 4.5: Living arrangements of urban women by age



Note: the sample is all women at or above 45 with urban Hukou in CHARLS2011. The 5-year age bands are 45-49 (labelled 45), 50-54 (labelled 50) and so on. The categories are not mutually exclusive: the 5th category 'with both children and grandchildren' is a subset of both the 3rd and 4th categories.

neighbourhood. For example, 55-59 year olds have on average 1.7 children alive, 1.3 in the same city/county, 1.0 in the same neighbourhood, and 0.8 in the same/adjacent dwelling. For every age group, the proportion of older women who see at least one child weekly ranges between 77% to 93%.

Physical proximity facilitates the transfer of home labour. The older generation may help the younger generation with home production (childcare, cooking, cleaning, etc.), while the adult children may care for the parents when they are ill or frail. Table 4.2 shows that a large share of women in their 50s have grandchildren; conditional on having grandchildren, the majority among those under 60 take care of the grandchildren on a regular basis. The median weekly hours spent on grandchild care (excluding zeros) is 30 for 50-54 year old women, and 40 for 55-59 year olds. This high level of devotion to grandchildren could be based on both altruistic and exchange motives.

Table 4.1: Number of children who are alive and living nearby, and are in regular contact

age band	observations	Mean number of children who				%
		are alive	live in the same			in regular
			dwelling or adjacent	neighborhood	city or county	contact with child
45-49	378	1.4	1.1	1.2	1.3	0.900
50-54	279	1.5	0.9	1.0	1.2	0.796
55-59	362	1.7	0.8	1.0	1.3	0.805
60-64	274	2.3	0.7	1.1	1.9	0.869
65-69	183	2.5	0.5	1.1	2.1	0.769
70-74	162	2.9	0.6	1.3	2.4	0.864
75-79	102	3.9	0.6	1.8	3.2	0.925

Note: the sample is women with urban Hukou in CHARLS 2011. For each non-coresident child, the respondent answers whether the child lives in the same/adjacent dwelling or courtyard, in the same village/neighbourhood, and in the same city/county, and how often they see the child. I define ”Regular contact” in the last column as seeing at least one child at least once a week. Co-resident children are assumed to be in regular contact. All columns about number of children reported above include the number of co-resident children.

Table 4.2: Summary statistics on grandmothers’ provision of childcare

age band	observations	proportion has grandkids< 16	proportion that care for grandkids	median hours per week
45-49	378	0.185	0.605	70
50-54	279	0.440	0.501	30
55-59	362	0.712	0.635	40
60-64	274	0.845	0.297	32

Note: the sample is women with urban Hukou in CHARLS 2011. For each grandchild under 16, the survey asks whether the respondent and their spouse provided childcare in the last year and the usual number of hours per week spent on taking care of the grandchild. The proportion of women providing care is conditional on having at least one grandchild under 16. The For median hours, I have added up the hours spent by the woman on all her grandchildren.

Financial transfers between adult children and parents are common and substantial. CHARLS asks the amount of monetary and in-kind transfers from and to each non-coresident child. Table 4.3 shows that 42% of urban women aged 45-75 have received financial transfers from non-coresident children in the 12 months prior to the survey. As the transfer questions are only asked if she has any non-coresident children and about two-thirds do, that 42% represents about 70% of those with non-coresident children. In terms of magnitude, the median amount of financial transfer from children (excluding zeros) is 3,500 RMB. This is not trivial, given that the median annual parental income is about 35,000 RMB. Conditional on NOT living with any adult children, the transfers are bigger: Table C.1 in appendix C.4 shows that for this group the mean amount of transfers from children (including zeros) is 4,876 RMB, while their mean household income is 47,129 RMB.[21]

On the other hand, 23% of parents have given financial transfers to their non-coresident children in the last year and the average amount given is even bigger (Table 4.3). Overall, parents are a net recipient of transfers from children. For the model in section 4.4, transfers from children will refer to gross rather than net transfers. Conceptually this is treating outward transfer from parents as a form of consumption. There is a great amount of cross-sectional heterogeneity in the amount of transfers between parents and children. Excluding zeros, the 75th percentile of the amount of transfer given to parents is nearly 8 times the 25th percentile. Unsurprisingly, some of that can be explained by the number of children and especially children who are educated or have high income. This can be seen in tables C.5 and C.6 in Appendix C.4, where I regress the incidence and amount of transfers on child characteristics. But most of the variation cannot be explained by the observables.

Meanwhile, there is some evidence that parents are more likely to receive financial

[21]There are many women who have both coresident and non-coresident children. They typically receive smaller amounts of transfers from non-coresident children than women who have only non-coresident children.

Table 4.3: Inward and outward financial transfers

Sources of private transfers						
Givers	% positive	mean inc 0	mean exc 0	median exc 0	25th pct	75th pct
children	42.3	2,856	6,758	3,500	1,000	7,800
other	15.0	861	5,736	2,000	500	8,000
siblings	12.1	479	3,965	1,000	500	3,000
grandchildren	4.2	183	4,380	1,000	400	2,400
parents	3.7	220	5,916	5,000	1,000	8,000

Destinations of private transfers						
Recipients	% positive	mean inc 0	mean exc 0	median exc 0	25th pct	75th pct
children	23.3	2,602	11,159	5,000	1,500	15,000
other	36.3	1,537	4,229	2,000	1,000	5,000
parents	25.8	1,076	4,180	2,300	1,000	4,000
grandchildren	17.6	1,027	5,841	1,920	500	6,000
siblings	16.4	381	2,322	600	500	2,000

Note: the sample is women with urban Hukou and aged 45-75 in CHARLS 2013. Transfers may be reported by the woman or her spouse. Both monetary and in-kind transfers are included. All percentiles are from distributions excluding zeros.

transfers when their incomes are low. In Table 4.4, we run a Linear Probability Model of transfers from children on parental incomes. Current parental incomes, both unearned and earned, are negatively correlated with the incidence of transfers. The correlation between parental unearned income and transfers from children is particularly significant and does not change much when lagged parental incomes are added to the regression. The coefficients on current parental earnings are smaller, and those on lagged parental earnings are more negative. The next two columns run Tobit regressions of the amount of transfers on the same regressors, with the lower bound being zero. Cross-sectionally, an extra RMB in parental unearned income is associated with a 0.07 RMB reduction in transfer from children. The coefficient on female earnings is more negative at -0.13 and significant at 10% level. The coefficient on male earnings is close to zero. All the coefficients on current parental income become less negative or more positive when lagged parental earnings are added. However, if conditioning on the transfer being positive, the amount appears uncorrelated with parental incomes.

Table 4.4: Transfers from children on levels of parental earnings and pooled unearned income

	(1) any in	(2) any in	(3) amount, Tobit	(4) amount, Tobit
parents' unearned incomes	-0.409***	-0.388**	-0.0731*	-0.0445
	(0.113)	(0.141)	(0.0291)	(0.0364)
male earnings	-0.213	-0.0811	0.0136	0.0491
	(0.158)	(0.171)	(0.0410)	(0.0442)
female earnings	-0.551*	-0.219	-0.126	-0.0498
	(0.238)	(0.276)	(0.0699)	(0.0776)
lagged parents' unearned incomes		-0.0534		-0.0456
		(0.129)		(0.0332)
lagged male earnings		-0.317		-0.107*
		(0.202)		(0.0528)
lagged female earnings		-0.732*		-0.214*
		(0.310)		(0.0868)
lagged transfer amount			0.498***	0.518***
			(0.130)	(0.129)
Observations	577	577	577	577
R^2	0.203	0.220		

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: The regressions are at the couple level. The sample is urban couple where the woman is between 45 and 75. It excludes couples who have no non-coresident children, because transfers from co-resident children are not observed. All the income regressors have been scaled down by 100,000. The regressions also condition on each parent's age (linear, squared and 5-year bands) and education (5 categories), home ownership, number of children alive, number of children living away from home, number of sons, number of children who have secondary education, who have higher education, who work, who are managers, who are professionals, and who have income above 50,000RMB a year.

While the negative correlation between parental incomes and transfers from children is intuitive, it's not clear whether it's causal. We do not have a good instrumental variable for parental income in the CHARLS data. However, a credible instrument has been studied. Cai et al. (2006) has exploited the pension arrears caused by the restructuring of state owned enterprises in the 90s as an exogenous shock to parental income.

They found that the net transfer was responsive to parental incomes in a non-linear fashion: the responsiveness is about 0.1-0.16 around the neighbourhood of the poverty line.

For our purpose (to understand parents' saving and labour supply decisions), transfers from children may play multiple roles. First, the Net Present Value of transfers has a wealth effect. As the number of children is declining across cohorts, more recent cohorts of parents might expect a lower NPV of transfers.[22] Second, transfers from children may play an insurance role and therefore reduce precautionary saving. Third, if transfers respond negatively to parental earnings (rather than wage which is unobserved if not in work), that might also reduce the parent's work incentive.

Meanwhile, labour time transfers to children might provide a partial explanation for the low employment rate among middle-aged and older women in urban China. The parents can work either in the formal labour market for a wage or in home production for their adult children. The latter could be for altruistic reasons or for current or future financial support. Due to the rapid improvement in education across cohorts, many parents in their 50s have lower wages than their adult children (in their 20s and 30s).[23] Thus, the parents' comparative advantage might be to take care of the grandchildren and housework, allowing their adult children to focus on paid employment outside the home. This also makes sense for the household collectively in the longer run, because the adult children have more years ahead to benefit from accumulating work experience. From the adult children's perspective, the parents might be a better provider of home-produced services than the market, because their preferences or incentives are more aligned with each other. We showed the prevalence of grandparents' provision of childcare in Table 4.2 earlier in this section. If this inter-generational exchange or cooperation were responsible for the low rate of formal employment of the middle-aged

[22]Realistically though, this wealth effect of transfers is likely to smaller than upcoming changes to pensions, since pensions constitute a much larger share of urban older people's total income.

[23]Table C.7 reports some summary statistics.

or older parents, then better provision of child-care will be important for raising the employment rate among the older generation.

The hypothesis of inter-generational cooperation is consistent with some findings in the literature. For example, Feng and Zhang (2018) found that reaching one's pension age increases women's probability of caring for her grandchildren by 29 percentage points. Maurer-Fazio et al. (2011) has found using 2000 census data that having a coresident parent increased the labour force participation rate of prime-aged urban women (25-50) by 12 percentage points. And most relevant to our hypothesis, Yu et al. (2021) showed that the effect of child birth on materal employment is very negative for those without grandparental support and non-negative for those with grandparental support. Thus, the evidence points to a model where the older woman's labour supply, her provision of childcare, and the younger mother's labour supply are jointly determined. This chapter makes a first-step towards such a model by focusing on the older woman's labour supply.

In this section we'll investigate this hypothesis by looking at correlations between parents' employment and the characteristics of their adult children in CHARLS. These correlations are not causal evidence, but they are interesting suggestive evidence. Table 4.5 reports the results of regressing female employment on her children's characteristics, conditional on her own characteristics. This analysis does not condition on the older woman having any grandchildren.

I find that an extra child with upper secondary education (passing exams at age 18) is roughly associated with a 5 percentage point drop in the older woman's employment. Normally, inter-generational transmission of ability and advantage should create a positive association between parent's employment and children's education. While we have conditioned on the woman's education, there probably remains some unobserved element of ability that is positively correlated with both children's ability and the woman's employment. Other omitted factors like local economic growth are also likely to be

Table 4.5: LPM of women's employment on child characteristics (the number of children who...)

	(1)	(2)	(3)	(4)	(5)
are male	0.0476*	-0.00542	0.0471*	0.0486**	0.0497**
	(0.0246)	(0.0290)	(0.0246)	(0.0247)	(0.0247)
are female	0.0240	-0.0137	0.0229	0.0221	0.0223
	(0.0241)	(0.0267)	(0.0241)	(0.0242)	(0.0242)
finished lower secondary school	-0.0112	-0.0250	-0.0148	-0.0131	-0.0133
	(0.0197)	(0.0199)	(0.0198)	(0.0197)	(0.0198)
finished upper secondary school	-0.0483***	-0.0595***	0.00644	-0.0469***	-0.0458***
	(0.0151)	(0.0163)	(0.0334)	(0.0151)	(0.0151)
are married	-0.0203	-0.0417*	-0.0192	-0.0215	-0.0239
	(0.0221)	(0.0238)	(0.0221)	(0.0221)	(0.0221)
whether have grandkids	-0.0702**	-0.0607*	-0.0186	-0.0352	-0.0328
	(0.0326)	(0.0326)	(0.0430)	(0.0359)	(0.0360)
number of grandkids above 1	0.0367**	0.0297**	0.0375***	0.0390***	0.0400***
	(0.0143)	(0.0143)	(0.0143)	(0.0143)	(0.0143)
work		0.0990***			
		(0.0212)			
has income above 100k		-0.119***			
		(0.0367)			
has income above 50k		0.0593***			
		(0.0204)			
grandkids and num of child w upper 2nd edu			-0.0641*		
			(0.0349)		
if she cares for grandkids				-0.0576**	-0.000588
				(0.0259)	(0.0518)
her weekly hours on grandkids					-0.000267
					(0.000348)
her grandkid care, weeks/year					-0.00107
					(0.00101)
Observations	1689	1689	1689	1689	1689
Adjusted R^2	0.121	0.136	0.122	0.124	0.125

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: sample restricted to 50-64 year old women with urban Hukou. Controls include their age, age dummies (in 5-year bands), education dummies, marital status and Hukou at birth. The 2nd column also includes the number of children living at home, the number of children economically dependent, the number of children working in professional occupations, and the number of children working as managers, which are all insignificant.
Source: CHARLS 2013

positively correlated with both. On the other hand, parents who expect more financial support from higher-ability children have less need to work. Thus, the finding of a negative correlation between female employment and her children's education is supportive evidence of the wealth effect of expected transfers. This estimate on the number of children with upper secondary education does not change much when we include more child characteristics in the 2nd column of the regression table.

Interestingly, the correlation between father's employment and the number of educated children is close to zero (see Table C.4 in the Appendix). It could be that the correlation between father's and children's ability/productivity is stronger than between mother and children; or that the negative wealth effect (the NPV of transfer from children) on labour supply is greater for women than for men.

Meanwhile, having any grandchildren is significantly negatively correlated with the older woman's employment by about 7 percentage points (Column 1 in table 4.5). In the third column of table 4.5, we interact the grandchildren dummy with the number of children that are educated at upper secondary levels. The interaction term is significantly negative, while both the coefficients on the grandchildren dummy and on the number of children with upper secondary education have become close to zero and insignificant. In other words, the negative correlation between female employment and having educated children is coming from those with grandchildren. To the best of my knowledge, this finding is new to the literature on labour supply in urban China. As the presence of grandchildren increases the amount of home production required and having more educated children gives the grandmother a comparative advantage in home production rather than paid employment, this interaction effect is consistent with the hypothesis that older woman's labour supply decision is made jointly with their adult children.

We can look at the relationship between older women's employment status and their provision of childcare. The 4th column in table 4.5 shows that there is indeed a significant and negative correlation. This obviously does not prove causality in either direction, as in theory the grandmother's labour supply and childcare supply could be determined jointly. The last column in table 4.5 suggests that the reported intensity of grandchild care is not significantly correlated with the grandmother's employment.[24] One possibility is that grandchild care requires the carer to be available at certain times of the day (e.g. between school pick-up and the normal end of working hours), which is incompatible with most regular paid jobs in urban China. Thus, what matters is the extensive margin of grandchild care, not the intensive one.

Ideally, we would like to track women before and after they become grandmothers and see whether the timing of exiting the labour force coincides with the arrival of a grandchild. However, there are only about 100 women who have become grandmothers between the first two waves of CHARLS. The survey does not contain the birth dates of one's existing grandchildren. As an alternative, we look at when the woman's child first got married. Having a married child arguably gives the mother a strong signal for when she can expect a grandchild, since the vast majority of babies in China are born to married couples. In table 4.6, we investigate what predicts the year when the woman left her last job. Obviously her birth year and the year that she completed the retirement process are both very significant predictors of when she stopped work. Interestingly, when her child first got married is also positive and significant. This result is robust to the inclusion of more individual characteristics. Overall, the cross-sectional correlations documented in this section supports the idea that older women's labour supply depends on expected financial transfers from her children and expected time transfers to them.

In the next section, I will build and calibrate a model with financial transfers from children, but it will not explicitly contain home production or exchange-motivated intergenerational transfers. These would be interesting extensions, and will require more theoretical work and ultimately some data moments to identify the motivations of trans-

[24]This lack of significant effect remains if we categorise the intensity into dummies.

Table 4.6: When woman stopped working on key events in life

	(1)	(2)	(3)
Year Took Retirement Procedures	0.412***	0.365***	0.356***
	(0.0466)	(0.0462)	(0.0475)
own year of birth	0.176***	0.214***	0.184**
	(0.0527)	(0.0533)	(0.0643)
when child first got married	0.127***	0.0968**	0.0841*
	(0.0365)	(0.0361)	(0.0384)
education==Middle School		1.077	0.962
		(0.572)	(0.592)
education==High school		2.136**	2.022**
		(0.676)	(0.690)
education==Vocational school		3.240***	3.185***
		(0.879)	(0.887)
education==College or above		5.610***	5.560***
		(0.984)	(1.005)
when first got urban or unified hukou			0.000169
			(0.0129)
when oneself got married			0.0487
			(0.0529)
Observations	862	862	862
Adjusted R^2	0.443	0.467	0.465

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: the sample is women with urban Hukou and who are not currently working.

fers. For example, an exchange between time transfers and financial transfers implies a positive correlation between the two. In the current CHARLS data, however, I do not observe any strong positive correlation between older parents providing childcare for their grandchildren and the amount of financial transfer they receive from their adult children, conditional on their characteristics like age and education. This observed (lack of) correlation cannot tell the complete story because parents only report financial transfers from non-coresident children and many grandparents caring for grandchildren live in 3-generation households. It is also likely that the inter-generational exchange of time and money takes place over a long period of time rather than within the same year. In future research I will use more waves of data to check the correlations.

4.4 The model

In this section, we specify a life-cycle model of female labour supply and consumption for urban Chinese households. Each household consists of a woman aged between 45 and 75 and her spouse. In each period, the household observes the realisation of the woman's wage w_t, the woman's pension y_t, and the man's income m_t. They expect a non-negative transfer from children T_t, which is a known function of their own sources of incomes. The couple choose whether the woman should work P_t and how much to consume c_t. We do not differentiate between part-time and full-time work, because a large majority of urban women in work have weekly working hours above 30.[25] The period is indexed by t, which corresponds to the woman's age. T is the final period the household makes any choices. In period $T+1$, the couple collects utility from the final asset. We start the model from $t = 45$ and assume $T = 74$.[26]

The household optimisation problem is formally set out below. In each period $45 \leq t < 75$, the couple maximises expected life-time utility by choosing female labour supply $P_t \in \{0,1\}$ and how much to consume c_t.

$$\max_{c_t,P_t} U(c_t,P_t,S_t) + E_t[\sum_{s=t+1}^{T} \beta^{s-t} U(c_s,P_s,S_s)] + \beta^{T+1-t} V(A_{T+1}) \tag{4.1}$$

s.t.

$$\frac{A_{t+1}}{R_t} = A_t + y_t + m_t + TRAN_t + w_t P_t - c_t \quad t \leq T \tag{4.2}$$

$$TRAN_t = F(y_t, m_t, w_t P_t) \tag{4.3}$$

$$A_t \geq 0 \tag{4.4}$$

$$S_t = S_t(P_{t-1}, P_t, y_{t-1}, y_t) \tag{4.5}$$

[25] Among women with urban Hukou in CHARLS 2013 who report a positive number of hours in non-agricultural work, the 10th percentile is 30 hours a week.

[26] In CHARLS data, the urban female employment rate falls to around 5% by 75.

The state variables are $(A_t, y_t, w_t, m_t, P_{t-1})$. A_t is the level of asset at the start of period t. All assets are liquid and there is zero borrowing constraint. S_t is the search cost for going back to work. S_t is positive only when the person moves from not working to working, or when the person has just completed the retirement process and works in the same period. Thus, last period's employment status P_{t-1} affects the search cost. Male income m_t is exogenous and uncertain. m_t would be zero if the husband dies. $TRAN_t$ is financial transfer from child in period t, which follows a known function of the parents' current incomes.

We further specify the utility function as:

$$U(c_t, P_t, S_t) = \frac{c_t^{1-\alpha}}{1-\alpha} - \delta_t P_t - \zeta \delta_t P_t max(1 - P_{t-1}, 1[y_t > 0 \& y_{t-1} = 0]) \tag{4.6}$$

It takes the usual CRRA form, with α being the measure of relative risk aversion. δ_t is the disutility of work, which is separable from consumption. I model δ_t as an exponential in t, and this is sufficient to produce a realistic employment rate profile.[27] The last term in equation (4.6) represents the psychic cost of going back to work, and is proportional to the disutility of work. The search cost is payable in the period when one transits from not working to working, and also in the period when one starts receiving a pension if they work. The latter is motivated by the fact that when completing the retirement process, the default is for the pensioner to leave the employer. While continuing to work for the same employer is possible, it requires negotiation for a new contract that could be as difficult as finding a new job.

The transfer from children is assumed to be a linear function of parental incomes in the current period, bounded by zero from below.

$$TRAN_t = max(0, \gamma_t - \gamma_y y_t - \gamma_m m_t - \gamma_w w_t P_t) \tag{4.7}$$

[27] A quadratic function over t would not make much difference.

Because transfers are not well measured in CHNS [28], we use CHARLS to run regressions of transfers on parental incomes to gauge the magnitude of $(\gamma_y, \gamma_w, \gamma_m)$. In the calibration, we set all three to 0.1 in the baseline case. As for the age profile γ_t, the level is based on the observed mean in CHARLS 2013 whereas the slope is estimated from the longer-running CHNS.[29]

Appendix C.3 specifies the processes of other incomes, how we calibrate them and other parameters. The population for calibration is married women with urban Hukou in CHNS 1989-2011, who were 45 years old in 1989 and hence 67 in 2011. Most empirical moments are estimated from CHNS, and the rest are based on CHARLS.

4.5 Results and discussion

The calibrated model generates a good fit of female employment rate over the life cycle. Figure 4.6 shows the key results in the baseline case. The simulated employment shows a small uptick in the last couple of years of life. This is driven by the bequest motive and the relatively high wage at those ages.[30] Also, the assumed disutility of work is not very steep over age (shown in Figure C.4); in reality it could increase much more rapidly towards 75.

The second graph in Figure 4.6 plots the distribution of the gap between the time of exiting labour force and the time of the first pension receipt in the baseline simulation. The density at zero is 42%, significantly higher than the surrounding values, although not as high as the 70% observed in CHARLS. Part of the bunching is due to liquidity

[28] In CHNS, there is a list of income questions about how much the household has received in the last year from different sources, one of them being children. The majority of responses are zeros. By contrast, CHARLS asks the amounts of monetary of in-kind, regular and irregular, transfers from each non-coresident child. The majority of parents in CHARLS report receiving transfers from their non-coresident children.

[29] γ_t is assumed to be linear in t. The constant is set so that the mean of transfer matches that observed in CHARLS 2013, and the linear coefficent on time is estimated from CHNS 2004-2011.

[30] As few urban women are in work at those ages, we have few wage observations. But the data look broadly consistent with a simple linear increase in log wages over age. Figure 4.11 shows that if the effect of final-period asset is smaller [bigger] in the terminal utility function, there would be no uptick [bigger uptick] in the employment rate near the end of the modelled life-cycle.

constraints. Part of it is due to the assumed search cost that would occur if one works in the year of the first pension receipt. Part of it is due to pension uncertainties which become known at the time of the first receipt. But even without the search cost, the model could predict 24.9% bunching at the time of retirement (by liquidity constraints and pension uncertainties); and the extent of bunching would be higher if a higher search cost is assumed (see Figure C.1 in Appendix C.4).

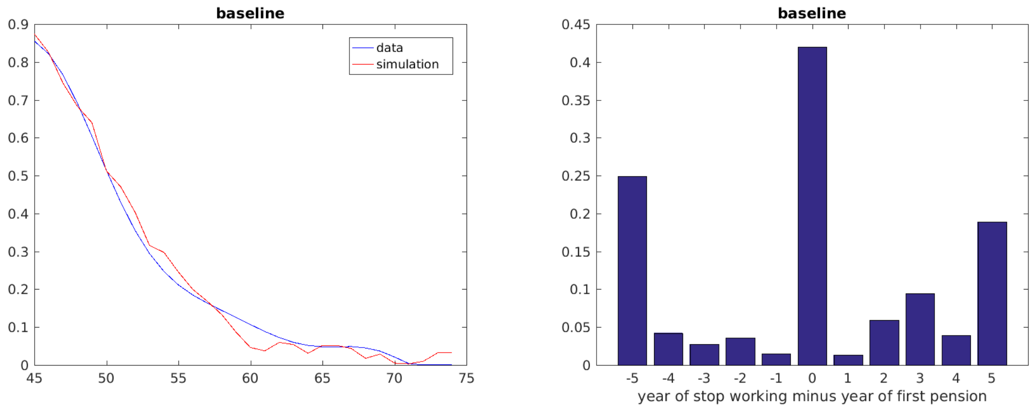
In reality, discontinuities in wages and in the disutility of work might be important causes of the observed bunching phenomenon. The disutility of work might increase discontinuously after the retirement process due to reference-point effects or some desire to conform to the social norm. It is also likely that for many people, their potential wage falls at the point of retirement, because those near retirement age may be paid above their productivity for institutional or contractual reasons. For example, during the restructuring of state-owned-enterprises in the 90s, many women close to retirement age were allowed and indeed encouraged to take early retirement. While they were not forbidden from taking another job afterwards, they were typically low-skilled, had been working for the same state-owned-enterprise for many years, and had much more difficulty finding another job compared to other groups.

To account for the possibility of deteriorating wages at the time of retirement, I calibrate two scenarios where the wage falls discontinuously at the point of retirement by 0.1 and 0.2 log points respectively. Intuitively, a wage drop at the time of retirement induces more people to stop working at the exact time. Figure 4.7 shows the proportion would be 45.5% and 46.0% respectively, compared to 42.0% under the baseline.

In the CHARLS life history data, we see 18% female pensioners returned to work after retirement and the wage change pre and post retirement is usually positive. But they are likely positively selected in the wage change.[31] In the simulations with a 0.2

[31] Because of the small proportion of people staying in work pre and post retirement, there are only 205 men and women whose post-retirement job started in the same or following year of their pre-retirement job and reported both earnings. The median wage change is a 15% increase among the 205. Only 57 of

Figure 4.6: Age profile of employment rate and timing of exit from labour market, baseline



Note: the left panel is the female employment rate over age. The right panel tabulates the gap between the time of exiting labour force and the time of first receiving pensions, where -5 and 5 mean the gap is ≤ -5 and ≥ 5 respectively. The gap graph is for the 86% sub-sample who have ever received a pension before death.

fall in log potential wage, 23% of women continued to work in the year of processing retirement (conditional on working the year before), and among them the median change in log wage pre and post retirement is -0.10. This is much more negative than what is observed among the small sample of people who stayed in work after retirement. Therefore, I doubt that the retirement process is associated with a discontinuous fall in one's wages as large as 20%.

The simulated employment rate is not sensitive to the assumed discount rate. But it would be much lower if households were less risk averse and much higher if they are more risk averse (Figure C.5 in Appendix C.4).

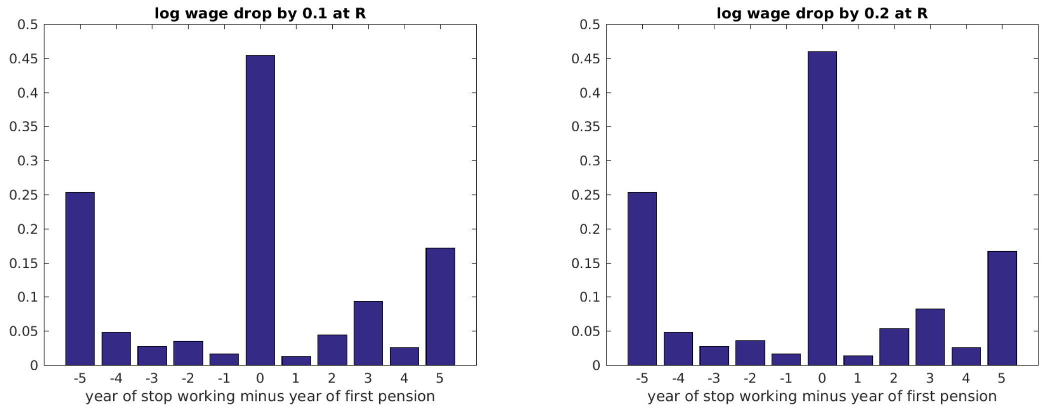
In the baseline, the probabilities of starting to receive pensions at each age are estimated from the data. The age at which women start to receive pensions varies across individuals and has a mean of 50.6 in the simulation.[32] And the average age at which simulated women exit the work force is 53.2.[33] These two are quite strongly correlated

them are women and the median wage change among women is around 60%.

[32]That is conditional on ever receiving a pension by 75. 15.6% simulated individuals never received pensions.

[33]That is conditional on not working at age 74, and 3.2% of simulated individuals work at 74.

Figure 4.7: Timing of labour market exit, by wage change scenario



Note: 5 means 5 or more, -5 means -5 or below. Simulated individuals who never received a pension are not in the graphs. The left graph assumed the log wage falls by 0.1 at the time of retirement, the right assumes a 0.2 log point fall. The baseline assumption is no wage fall.

in the simulations.[34]

For potential policy changes to pension age, we consider four scenarios. First, if all women were certain to become eligible for pension at age 50, not earlier or later; second, if all women were certain to become eligible for pension at age 55; and at age 60 and 65 respectively in the third and fourth counterfactual. As shown in Figure 4.8, the employment rate would fall very sharply at the age of guaranteed pension. Compared to the baseline, the employment rate will be substantially higher before the counterfactual pension age, and substantially lower at that pension age. In the counterfactual of pension age at 60, for example, the employment rate at each age in the 50-59 range would be 9-36 percentage points higher than the baseline. The mean difference over the 10 years is 28 percentage points. In other words, compared to the current scenario of variable pension ages, setting the pension age to 60 for all women would increase the employment rate among 50-59 year old women by 28 percentage points! At 60 however, the employment rate would fall very sharply to zero in the counterfactual, in comparison to 4.6% in the baseline. Similarly, the final counterfactual of retirement age

[34] See Figure C.2 in the appendix for the entire distribution of the age of leaving the work force conditional on the age of the first pension receipt.

at 65 would significantly increase the employment rate before 65: the increase relative to the baseline between age 50 and 64 averages 27.3 percentage points.

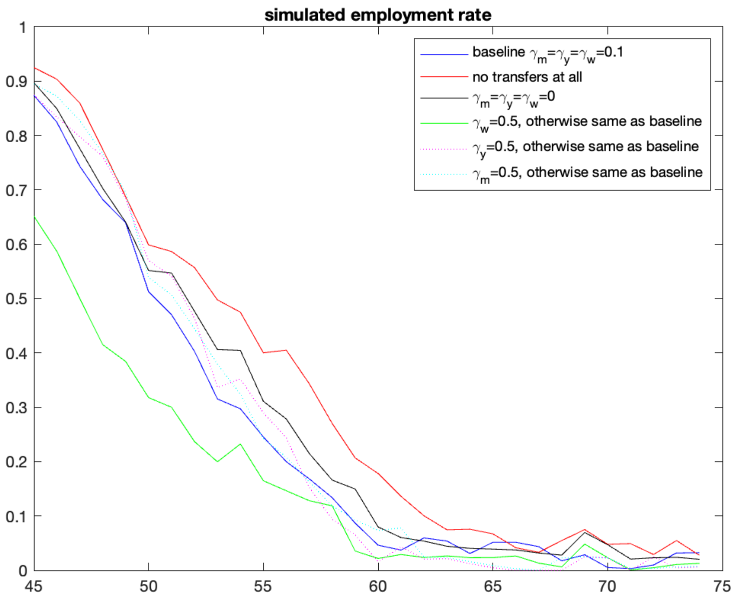
On average, the age at which women exit the workforce would be 51.4, 53.2, 55.7 and 58.1 respectively under the four counterfactuals, compared to 52.9 in the baseline. Thus, increasing the retirement age for all women to 60 would extend the working life by 2.8 years on average from the current scenario.

Figure 4.8: Female employment rate under alternative pension policies

Note: In the baseline, it is uncertain when a woman will qualify for the pension. The baseline pension process follows the specification in (C.3), where the transition probabilities are estimated from CHNS data. In each of the counterfactual, all women are guaranteed the pension at the specific age.

Transfers from children also have significant effects on female labour supply. Figure 4.9 compares a few counterfactual transfer scenarios with the baseline. If there were no transfers at all, the simulated employment rate would be 6-19 percentage points higher than the baseline between age 50 and 59, with a mean difference of 13.4 percentage points over the 10 years. This appears to work through both the income channel and the insurance channel. If transfers were totally irresponsive to parental incomes, that is, having only an income effect, the employment rate would be 4.8 percentage points higher than the baseline on average between age 50 and 59.[35] Thus, even a small degree of responsiveness of transfers to parental income (0.1 in the baseline simulation versus 0 in the last scenario) can have a non-trivial effect on parents' labour supply. If transfers from children were incredibly responsive to parental earnings (assuming it acts as a 50% effective tax rate), the employment rate would be substantially lower than the baseline (by an average of 8.9 percentage points over age 50-59).

Figure 4.9: Female employment rate under alternative transfer functions



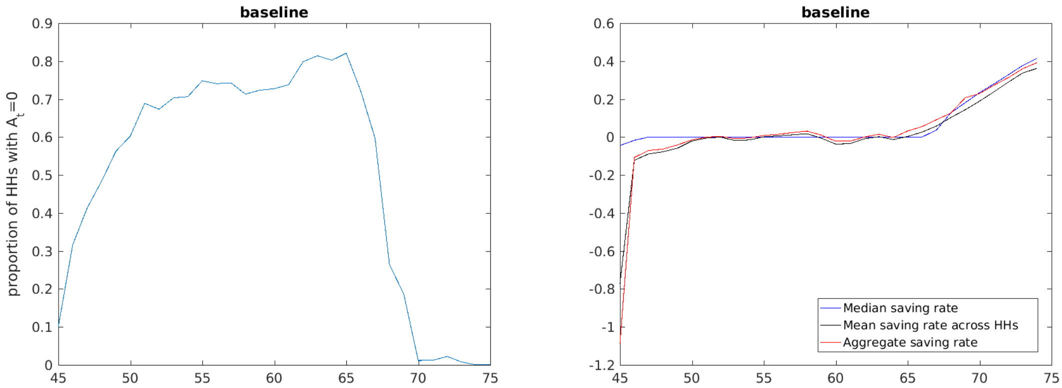
Note: the transfer function is specified in equation (4.7).

Meanwhile, the calibrated model appears to generate too little saving. In the baseline simulation, Figure 4.10 shows the proportion of households with zero asset is in the 60-80% region between age 50 and 65. The aggregate saving rate for the calibrated

[35] In the counterfactual scenarios of different $\gamma_w, \gamma_y, \gamma_m$, the mean profile γ_t is shifted so that the mean level of transfer at age 69 is roughly the same as in the baseline and hence the data.

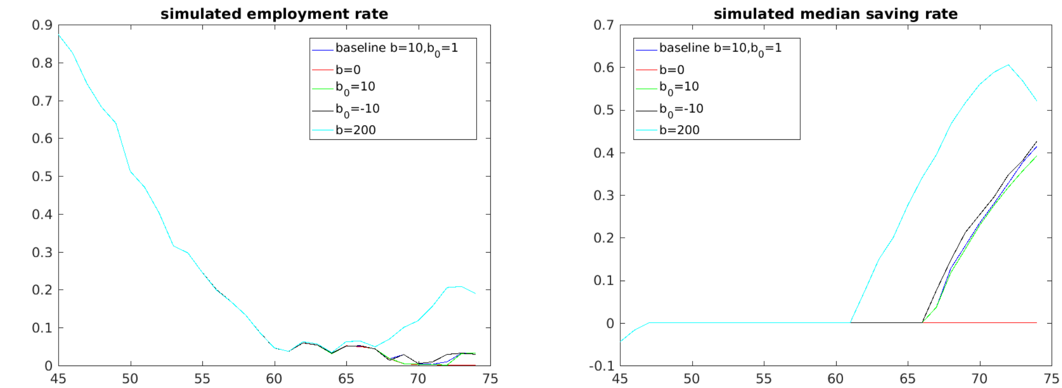
cohort is close to 0 until age 65. This is clearly lower than the reality. The median household saving rate observed in CHARLS is in the range of 20-40% for those households with women aged 50 or above.

Figure 4.10: Age profiles of assets and saving rate in the baseline



Note: In the right graph, household saving rate is defined as 1 minus current consumption divided by current income (including male income, female earnings and pensions, and transfers). The aggregate saving rate equals 1 minus the ratio of aggregate consumption to aggregate income.

Figure 4.11: Age profiles under alternative parameters b, b_0 for terminal utility



Note: the terminal utility function is $b((A_{T+1}+b_0)^{1-\alpha})/(1-\alpha)$.

In the current calibrated model, the main driver of saving is the bequest motive (the role of assets in the final period). In the baseline, we set $b = 10, b_0 = 1$ and $\alpha = 2$. This means the marginal utility of consumption is roughly equal to the marginal utility of bequest when bequest is 3 times the consumption level in the final year. As shown in Figure 4.11, a higher level of b would increase the aggregate saving rate and a lower b would reduce it. If bequests yields no utility, the aggregate saving rate would be around zero throughout life. But all the changes happen towards the end of life, rather than spread throughout the life cycle.

Part of the reason for the low simulated saving rate before 65 is that all wages and incomes have a fast-growing time trend - the growth is around 10% per year. I have tried a counterfactual scenario where all types of incomes are expected to be flat and did remain at the same level in real terms. That would yield median saving rate around 10% before age 60.[36] As standard in the literature, I have assumed expected income growth to be the same as the realised growth. However, in our context, income realisations might have exceeded the expectations throughout the period. The rapid wage growth experienced by this cohort over the latter part of the life cycle was mostly driven by macroeconomic growth rather than their own human capital accumulation. And macroeconomic growth as high as 10% a year and lasting for decades was rather unique, both internationally and historically. If expected income growth is slower than the actual growth, households would save more.

In future, I could add more shocks to the model to induce precautionary saving. The assumed dynamics of pension income means that once a woman starts to receive pensions, as the majority do by 60, there is no uncertainty in her future stream of pension income. As pension income constitutes a large share of household income at old age, increasing uncertainty in pension income will probably improve the prediction of the saving rate. In addition, the current specification of financial transfers means there is effectively a consumption floor, which is relatively high at old age. Thus, to create the possibility of having very low income later in life, it is necessary to add uncertainties to both pension incomes and transfers at all ages.

[36] See figure C.3 in the Appendix.

Moreover, uncertainties about health, mortality and medical expenditures might be important drivers of precautionary saving in reality. To incorporate this into the structural model, we would need two extra state variables, one for health status and one for health care expenditure. The health status can be assumed to follow a first-order Markov process and can be estimated from the transitions of self-reported health status in CHARLS.[37] Health status will affect the marginal utility of consumption and of work. Healthcare expenditure will depend on health status and the amount will be uncertain. The distribution can also be estimated from CHARLS.[38] Finally, uncertainty in longevity would also contribute to pre-cautionary saving.

4.6 Conclusion

This chapter has examined the low employment rate of older women in urban China and proposed two explanations. The primary reason is the early age at which urban women become eligible for pensions. The second reason is their close economic ties with their adult children, in terms of both financial transfers from children and time transfers to children.

These arguments are supported by a wide range of descriptive evidence. Pensions are the biggest source of income for the urban elderly. The timing of leaving the labour market tends to coincide with the timing of the first pension receipt. Transfers from children constitute a non-trivial share of the parents' income. Grandmother provision of child care is common and the average hours are long. Female employment is significantly negatively correlated with their children's education and having grandchildren, and indeed, the interaction of the two. Transfers from children are significantly negatively correlated with the woman's earnings.

[37]CHARLS not only asks individuals to rate their health status into a few categories, but also contains a wide range of more objective measure such as whether has been diagnosed of certain diseases.

[38]For example, I observe in CHARLS 2013 that among those who rated their health as "good" or "very good" or "excellent", 18% had reported out-of-pocket medical expenditure. Among those with worse health ratings, 38% did.

Based on the descriptive evidence and intuition, I have built and calibrated a simple structural model of female labour supply and saving. The model contains uncertainties in male income, female wage and pension income. Transfers are assumed to be an exogenous function of parental incomes. The model yields realistic predictions of female employment rate over the life cycle. It also generates a large amount of bunching in the timing of labour market exit at the point of qualifying for a pension. This bunching comes from the uncertainties in pension entitlement, liquidity constraint, and a search cost for finding new work after the first pension receipt, rather than any discontinuities in wages or preferences.

The simulations suggest that both pension age and transfers have substantial effects on the female employment rate. For example, raising the female pension age from its current level (which varies across individuals and has a mean of 50.6) to 60 would increase the employment rate by 28 percentage points on average over ages 50-59. The average age at which women leave the labour force would also increase from 52.9 in the baseline to 55.7. Had there been no transfers at all, the female employment rate would be 13 percentage points higher over ages 50-59.

In future, I intend to extend the model in a number of directions. First, I would like to improve the fit of saving rates by incorporating more shocks in pensions, intergenerational transfers, and healthcare expenditure. Second, I would like to jointly model the decisions of the older woman and her adult children on labour supply and home production. The descriptive evidence suggests there is substantial cooperation between generations. This means increasing the female pension age is likely to have large spillover effects on the labour supply of the next generation of women, through the need for home production (especially childcare). As the younger generation are more educated and have more to forego in terms of human capital accumulation, taking account of such inter-generational interactions will be important for predicting the overall fiscal and welfare effects of such reforms.

Chapter 5

General Conclusions

Each of the chapters in this thesis contains their own conclusions and discussions about future extensions. Here I will briefly summarise the findings and discuss how I may extend and build upon this body of work.

Chapters 2 and 3 examined the labour market consequences resulting from the large expansion of higher education in the UK. This supply shift has had little impact on graduates' relative wage, or their occupational destinations. This is explained in two models where the adoption of technology depends on the supply of skills. The framework is also supported by various other macro facts and micro evidence: the shift in employment from routine occupations to abstract ones; the absence of wage polarisation across occupations; the correlations between local skill supply and various indicators of the New technology. The broad macro facts are shared by many other European countries; so it would be interesting to explore, in future research, whether the micro predictions of the framework are true in other European countries.

Chapter 3 calibrates the proposed multi-sector model to UK data and simulates counterfactual changes to skills distribution and industry demand shifts. It found that the shift in skills distribution alone can account for between a third and two thirds of the decline of manual routine occupations, and between a third and half of the increase in abstract occupations. Future work could apply the same model to simulate the effects of potential policies. For example, how will selecting immigrants based on their skills affect the future UK labour market? Following Brexit, European immigrants into the UK are now subject to skill-based selection criteria. Therefore the composition of immigrants will differ from the historical range, necessitating a structural model like the one developed in chapter 3.

While Chapters 2 and 3 assumed an exogenous supply of skills and focused on the role of technology on the demand side, Chapter 4 examines labour supply as the main outcome of interest.

Chapter 4 seeks to explain the low employment rate among urban Chinese women in their 50s and 60s. It uses a range of descriptive evidence to argue that there are two main factors. One is the early age at which they qualify for a public pension. The other is the actual and expected financial transfers from their children. The chapter built and calibrated a life-cycle model of labour supply, incorporating pension uncertainties, income-contingent transfers from children, and a search cost of searching for another job after qualifying for a retirement pension. The counterfactual analysis suggests increasing the pension age to 60 for all women would increase employment rate of 50-59-year-olds by 28 percentage points.

Future research can extend the model in a number of directions to answer related questions about Chinese households. One such question is will increasing the pension age of women reduce the labour supply of the younger generation? Chapter 4 has documented some descriptive evidence that point to cooperation between two generations within the household, especially in terms of grandparents' provision of child care. Given the lack of childcare availability outside of the home in today's Chinese cities, it's likely that if grandmothers stayed in work for longer, more of the younger women (mothers) will have to leave the labour force at least temporarily. This will partially offset the fiscal savings from such a pension policy, given that the younger generation of women are much more educated on average.[1] Such inter-generational dynamics calls for a model where at least two women's labour supply are jointly determined.

Another big research question is why do Chinese households save so much? This is a matter of global interest, as it contributes to China's current account surplus. Currently, the model in chapter 4 does not contain enough uncertainties to generate the high level of savings we observe. Future work can incorporate more income uncertainties, and add shocks in health and health-related expenditures. Another probable motivation for saving is for transfers to one's children, both as bequests at death and as a way to improve the child's competitiveness on the marriage market.

[1]Another possible unintended consequence is delayed fertility. Again, this could adversely affect the fiscal balance in the longer run.

Appendix A

Appendix for chapter 2

A.1 Derivation of General Wage Equations

In this Appendix, we provide a more complete derivation of our aggregate skilled wage and wage ratio regressions given in equation (2.4) and (2.5) in the text.

Starting from the production function specification set out in the text and assuming competitive labour markets, we obtain,

$$
\begin{aligned}
\ln w_{ujt} &= \Omega_j - \frac{1}{\sigma_a}(\ln U_{jt} - \ln U_t) + \ln \theta_{ut} + \ln \frac{\partial F}{\partial(\theta_{ut} U_t)} \\
\ln w_{sjt} &= \Gamma_j - \frac{1}{\sigma_a}(\ln S_{jt} - \ln S_t) + \ln \theta_{st} + \ln \frac{\partial F}{\partial(\theta_{st} S_t)}
\end{aligned}
$$

Log linear approximations for these expressions are given in equations (2.1) and (2.2) in the text, and the approximation for the wage ratio is given in (2.3).

In order to take the skilled wage equation and the relative wage equation to the data we need to address the fact that the productivity parameter ratio ($\ln(\frac{\theta_{st}}{\theta_{ut}})$) and the θ_{ut} parameter that enter both equations are unobserved. We address these issues using the approach in Beaudry and Green (2005). In particular, we make use of the fact that given our production function we can write log TFP as,

$$\ln TFP_t = s_t^u \ln\theta_{ut} + s_t^s \ln\theta_{st} \tag{A.1}$$

where, s_t^u is the share of income going to unskilled labour and s_t^s is the share of income going to skilled labour. Rewriting (A.1) slightly, we have:

$$\ln\theta_{ut} = \frac{\ln TFP_t}{(s_t^u + s_t^s)} - \frac{s_t^s}{(s_t^u + s_t^s)} \ln\frac{\theta_{st}}{\theta_{ut}} \tag{A.2}$$

Note that if $\theta_{st} = \theta_{ut}$ then technological change is labour biased but not skill biased and bringing in TFP data alone would be sufficient to get estimable versions of (2.2) and (2.3).

To allow for skill biased technical change, we assume that the log ratio, $\ln(\frac{\theta_{st}}{\theta_{ut}})$ is a quadratic function of t, that is,

$$\ln\theta_{st} - \ln\theta_{ut} = \gamma_0 + \gamma_1 t + \gamma_2 t^2 \tag{A.3}$$

This allows for a bit more flexibility than the common linear skill biased technical change assumption, which is obviously nested in this specification.

Substituting (A.2) and (A.3) into the skilled wage equation yields an estimable specification similar to the one in Beaudry and Green (2005), given by:

$$\begin{aligned}\ln w_{sjt} = {} & \ln\Gamma_j + (\beta_1 - \beta_2 + 1 - (1-\beta_2)\frac{s_t^s}{s_t^u + s_t^s})[\gamma_0 + \gamma_1 t + \gamma_2 t^2] + (\beta_1 - \beta_2)\ln(\frac{S_t}{U_t}) \\ & + (1-\beta_2)\frac{\ln TFP_t}{(s_t^u + s_t^s)} + \beta_2 \ln(\frac{K_t}{U_t}) - \frac{1}{\sigma_a}\ln\tilde{S}_{jt} + \varepsilon_{1jt}\end{aligned} \tag{A.4}$$

Simplifying and gathering terms yields equation (2.4), the skilled wage specification, in the text. To ease comparisons with the existing literature, we estimate the coefficients on the t and t^2 terms as fixed, implicitly pinning the labour share values at average values for our period.

Similarly, we can write:

$$\begin{aligned}\ln w_{ujt} =& \ln \Omega_j + (\alpha_1 - (1-\alpha_2)\frac{s_t^s}{s_t^u + s_t^s})[\gamma_0 + \gamma_1 t + \gamma_2 t^2] + \alpha_1 \ln(\frac{S_t}{U_t}) \\ & + (1-\alpha_2)\frac{\ln TFP_t}{(s_t^u + s_t^s)} + \alpha_2 \ln(\frac{K_t}{U_t}) - \frac{1}{\sigma_a}\ln \tilde{U}_{jt} + \varepsilon_{2jt} \end{aligned} \tag{A.5}$$

and the difference between the two wage equations gives:

$$\begin{aligned}\ln \frac{w_{sjt}}{w_{ujt}} \approx & (\ln \Gamma_j + (1+\beta_1 - \alpha_1 - \beta_2 - (\alpha_2 - \beta_2)\frac{s_t^s}{s_t^u + s_t^s})[\gamma_0 + \gamma_1 t + \gamma_2 t^2] + (\beta_1 - \beta_2 - \alpha_1)\ln \frac{S_t}{U_t} \\ & + (\alpha_2 - \beta_2)\frac{\ln TFP_t}{(s_t^u + s_t^s)} + (\beta_2 - \alpha_2)\ln \frac{K_t}{U_t} - \ln \Omega_j) - \frac{1}{\sigma_a}(\ln \tilde{S}_{jt} - \ln \tilde{U}_{jt}) \end{aligned} \tag{A.6}$$

Simplification and gathering terms yields equation (2.5) in the text.

A.2 IVs for the skill supplies

In this section, we define the instrumental variables (IV's) for the skill supply variables in our aggregate production function estimation. The three IV's we use are:

$$IV1_{gt} = \sum_c \eta_{gc0} * BAgrowth_{ct} \tag{A.7}$$

$$IV2_{gt} = BAprop_{4550,g0} * BAgrowth_t \tag{A.8}$$

$$IV1_{gjt} = BAprop_{t-j-25,g0} * BAgrowth_{t-j,t} \tag{A.9}$$

where η_{gc0} is the share of cohort c in working-age population in region g at time 0 (1993-5), $BAgrowth_{ct}$ is the growth in the BA proportion among cohort c between time 0 and time t, in the UK as a whole, $BAprop_{4550,g0}$ is the BA proportion among the 1945-54 cohorts in region g at time 0, so as to capture the parents' generation of the 1975-84 cohort, $BAgrowth_t$ is the growth in the BA proportion between time 0 and time t, in the UK as a whole,$BAprop_{t-j-25,g0}$ is the BA proportion among the $t-j-25$

cohort in region g at time 0, so as to capture the parents generation of cohort $t-j$, and $BAgrowth_{t-j,t}$ is the growth in the BA proportion for the $t-j$ cohort between time 0 and time t.

The first two IVs are at the region-time level and are included as instruments for $\ln\frac{S_{gt}}{U_{gt}}$ while the third is at the region-age group -time level and serves as an instrument for $\ln\frac{\tilde{S}_{gjt}}{\tilde{U}_{gjt}}$. As seen in the first stage results, the first two IV's are highly significant in the first stage for $\ln\frac{S_{gt}}{U_{gt}}$ while the first and third instruments have substantial significant effects in the age specific skill supply and skill ratio first stages. The results imply easy passing of weak instrument tests.

Table A.1: First stage results

	$\ln S_{gt}/U_{gt}$	$\ln \tilde{S}_{gjt}$	$\ln \tilde{S}_{gjt}/\tilde{U}_{gjt}$
IV_{gjt}	-0.354	3.146***	12.430***
	(0.215)	(0.647)	(0.748)
$IV1_{gt}$	1.772***	-0.677	-0.446
	(0.373)	(1.124)	(1.300)
$IV2_{gt}$	13.809***	-4.165	-20.169***
	(0.951)	(2.863)	(3.310)
t	0.052***	-0.002	-0.004
	(0.004)	(0.011)	(0.013)
t^2	-0.001***	0.000	0.000
	(0.000)	(0.000)	(0.000)
$\ln\frac{TFP_t}{laborshare_t}$	0.016	-0.097	0.045
	(0.071)	(0.213)	(0.247)
N	760	760	760

Notes: standard errors are shown in parentheses. The regression is at the level of 19 regions, 5-year-age-band and 3-year-period. The sample is restricted to 20-44 year olds. All specifications include complete sets of age-band dummies and region dummies. *** p<0.01, ** p<0.05, * p<0.1

A.3 The Role of Capital and Constrained System Estimates

In this appendix, we discuss the role of capital and TFP in wage specifications and present results when we impose theoretically implied cross-equation restrictions. The regression equations (4) and (5) in the paper allow for general productivity growth but also incorporate a flexible skill biased technological trend. Many of the specifications estimated in the micro labour literature on technological change do not include either capital or TFP, and our specification obviously nests such an approach. In particular, if $\alpha_2 = \beta_2$ then neither TFP_t nor K_t appear in the relative wage equation. This would imply that capital is equally complementary with skilled and unskilled labour and would occur, for example, if the production function were multiplicatively separable in K_t and the overall labour component. That is what was assumed in the seminal Katz and Murphy (1992) paper and is one explanation for why most of the Skill Biased Technical Change (SBTC) literature and the polarization literature that followed use specifications that do not include capital. An alternative explanation not including capital comes from the combination of the constant returns to scale assumption and an assumption of a perfectly elastic supply of capital. It is straightforward to derive an expression for the price of capital and use it to substitute out the $\ln(\frac{K_t}{U_t})$ term in our two estimating equations. If we assume that the world price of capital is constant then here, as in the case with multiplicatively separable capital, we end up with the canonical specification for the relative wage equation with only a time trend and the relative skill supply variables on the right hand side.[1] We can, alternatively, allow the price of capital, r_t, to vary over

[1]The two different approaches for eliminating capital from the relative wage equation have different implications for the skilled wage equation. If the production function is multiplicatively separable in capital and a labour aggregate then both TFP_t and K_t enter the skilled wage equation. If, instead, the production function is not multiplicatively separable in capital and labour but capital is perfectly elastically supplied then TFP_t but not K_t is present in the skilled wage equation. As with the relative wage equation, we can include $\ln r_t$ as an added regressor in the perfectly elastic capital supply case with a time varying price of capital.

time, implying an adjusted version of the canonical specification that includes $\ln r_t$ as a regressor. Estimates of this adjusted specification are available upon request. That specification yields very similar results in terms of the estimates of the coefficients of interest to those reported in the text. The theory underlying our specifications implies several restrictions. Weak concavity of the production function implies that $\beta_1 - \beta_2 \leq 0$. From equation (14) in the paper, the coefficient on $\ln(\frac{S_{gt}}{U_{gt}})$ in the skilled wage regression equals $\beta_1 - \beta_2$ and the estimates of that coefficient in both our OLS and IV estimates in Table 1 are negative. Second, concavity implies $\alpha_1 + \alpha_2 \geq 0$. We can construct an estimate of $\alpha_1 + \alpha_2$ as, $b_2 + b_4 - (d_2 + d_4)$, which takes on values that are slightly negative in the OLS and IV estimates (-0.11 and -0.15, respectively) but are not statistically significantly different from zero in either case. Thus, here too, we cannot reject the concavity restriction. The third concavity condition, corresponding to the determinant of the Hessian, is $(b_2 \cdot d_4 - b_4 \cdot d_2) \geq 0$. This terms takes a a value of -.16 with a standard error of 0.10. Thus, from the values estimated for all three conditions, we cannot reject the null of weak concavity of the production function.

The framework implies three equality restrictions on the regression equations (4) and (5) in the paper: $b_3 + b_4 = 1$ and $d_3 + d_4 = 0$ and $b_5 = d_5$. The first restriction is clearly rejected in the OLS case while the other two are not rejected in any specification. In the first four columns of Table A.2, we present SURE and IV estimates in which we impose these restriction and show that they make very little difference to the coefficients of interest, which are the coefficients on the skill supplies and the year effect. Overall, our parameter estimates fit well (albeit not perfectly) with the requirements imposed by our assumption that we are estimating parameters associated with a well-behaved production function. The last 2 columns of Table A.2 contain IV results in which we add a cubic in time, showing that this extra flexibility does not alter our results.

Table A.2: Skilled Wage and Wage Ratio Regressions: UK, 1993-2016

	$\ln \frac{w_{sgjt}}{w_{ugjt}}$	$\ln w_{sgjt}$	$\ln \frac{w_{sgjt}}{w_{ugjt}}$	$\ln w_{sgjt}$	$\ln \frac{w_{sgjt}}{w_{ugjt}}$	$\ln w_{sgjt}$
t	0.002	-0.006	-0.016	-0.001	-0.022*	0.006
	(0.006)	(0.006)	(0.009)	(0.010)	(0.011)	(0.010)
t^2	-0.000	0.000	-0.000	-0.000**	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$\ln S_{gt}/U_{gt}$	0.056	-0.000	0.302*	-0.156	0.307*	-0.075
	(0.086)	(0.081)	(0.144)	(0.157)	(0.147)	(0.136)
$\ln \frac{TFP_t}{laborshare_t}$	0.086*	0.955***	-0.001	0.490***	0.099	0.373***
	(0.044)	(0.042)	(0.076)	(0.083)	(0.110)	(0.102)
$\ln K_t/U_t$	-0.086*	0.045	0.001	0.510***	0.104	0.145
	(0.044)	(0.042)	(0.076)	(0.083)	(0.160)	(0.149)
$\ln \tilde{S_{gjt}}/\tilde{U_{gjt}}$	0.010		0.025		0.021	
	(0.012)		(0.029)		(0.037)	
time cubic					0.000	-0.000
					(0.000)	(0.000)
$\ln \tilde{S_{gjt}}$		0.010		0.025		-0.057
		(0.012)		(0.029)		(0.133)
IVs	no	no	yes	yes	yes	yes
constraints	yes	yes	yes	yes	no	no
N	1208	1208	760	760	760	760

Notes: standard errors are shown in parentheses. The regression is at the level of 19 regions, 5-year-age-band and 3-year-period. The sample without IVs consists of 20-59 year olds. Whenever we use IVs, the sample is restricted to 20-44 year olds. The first 4 columns are the same as the first 4 columns in Table 1 in the paper except that we now impose three constrainst, and estimate using SUER and 3SLS rather than OLS and 2SLS. If we just use SURE and 3SLS and do not impose the constraints, the estimates would be very close to those in Table 1. The last 2 columns here are the 3SLS estimation of the two equations with IVs and a time cubic term; so they are the closest to the middle 2 columns in Table 1 in the paper - the only difference being the time cubic term. All specifications include complete sets of age-band dummies and region dummies. *** p<0.01, ** p<0.05, * p<0.1

A.4 The expansion of high education in the UK and education classification

The expansion of higher education over the past few decades reflects a sequence of specific policy choices made by the UK government. Since the Robbins Report in 1963, policy related to the higher education sector has been moving toward implementation of the principle that university places 'should be available to all who are qualified by ability and attainment to pursue them and who wish to do so'. The 1960s saw the foundation of more than 20 universities and dozens of polytechnics. Polytechnics were a form of higher education institution that taught both degree-level courses and below-degree-level courses, with their degrees certified by a chartered body called the Council for National Academic Awards (CNAA). A CNAA degree from a polytechnic was technically equivalent to a university degree and we treat them as equivalent in our analysis. The Education Reform Act (ERA) of 1988 changed some block grants to tuition fees (paid by Local Education Authorities for each student). In response, polytechnics increased enrolment with lower funding per student. The other major education policy change in 1988 was the replacement of CSEs and O-Levels with GCSEs as the exams that students take at age 16.[2] That reform led to an increase in educational attainment at the secondary level and hence an increase in the proportion of the young with sufficient academic credentials for potential admission to universities. In 1992, polytechnics gained the right to issue degrees and become fully-fledged universities. The reclassification of polytechnics as universities led to a jump in the number of university students in 1992; but the rapid increase in student numbers in higher education started in 1988 and continued until 1994.[3] In 1994, pressures on public expenditures and a desire to protect resources per student led the government to introduce the maximum student number control. This limited the number of full-time undergraduates at individual universities per year. As a result, the growth in student numbers slowed. This acceleration and then deceleration can be seen clearly in the BA proportion across birth cohorts in Figure A.3 in Appendix A.

This paper has focused on the comparison between two education gruops: BA

[2]Certificate of Secondary Education (CSE) and General Certificate of Education Ordinary Level (O-levels) were subject-based qualifications that students in England at the end of secondary school around age 16. CSEs are less academic, and so we count O-Levels in our definition of HS group (equivalent to GCSEs grade C or above), but not CSEs. CSEs are considered equivalent to GCSE below grade C.

[3]This has been clearly shown in Figure 2 in Carpentier (2006)

and HS. Here we show our main result that the BA-HS wage differential has been flat is robust to alternative definitions of education groups. In the paper, we have defined BAs as those whose highest qualification is first degree or higher, and HS as those who obtained Grade C or higher in the General Certificate of Secondary Education exam (GCSE) or equivalent and who did not have any degree-level qualification. We chose these definitions so as to be broadly comparable to college graduates and High School graduates in the US.[4]

The first alternative we investigate is to draw the bottom line of the HS group at A-levels rather than GCSEs. A-levels are subject-based exams taken typically at age 18 and are a pre-requisite for university admission. Under the UNESCO's International Standard Classification of Education (ISCED 2011), both GCSEs and A-levels in the UK are classified as level 3 -"upper secondary education", and so are High School Diploma in the US. The left subgraph in Figure A.1 shows that drawing the line at A-levels instead of GCSEs makes very little difference to the trend in the BA-HS wage gap.

Second, we group people by the age they left full-time education, and look at the wage gap between those who left at age 21-22 and those who left at 17-18. In Figure A.1, we show the estimated trend (net of age effects) alongside the one based on our main definition of education, which was shown in Figure 2 in the paper). Again, both trends are remarkably flat over the sample period. In summary, our main conclusion that the college wage premium has been flat since the early 90s is robust to how it's defined.

Finally, we want to address the concern that the strong increase in the BA proportion observed in the Labour Force Survey may have been over-estimated due to sampling and measurement issues. The LFS is not a compulsory survey and its response rate has been declining over time.[5] If graduates have a differential response rate to less-educated

[4]For example, among 25-29 year olds in 2012, the US proportion of "BA" and "HS" are 35% and 56%. In the UK, the proportions according to our definition are 36% and 53%.

[5]The response rate can be found in the ONS Labour Force Survey Performance and Quality Reports.

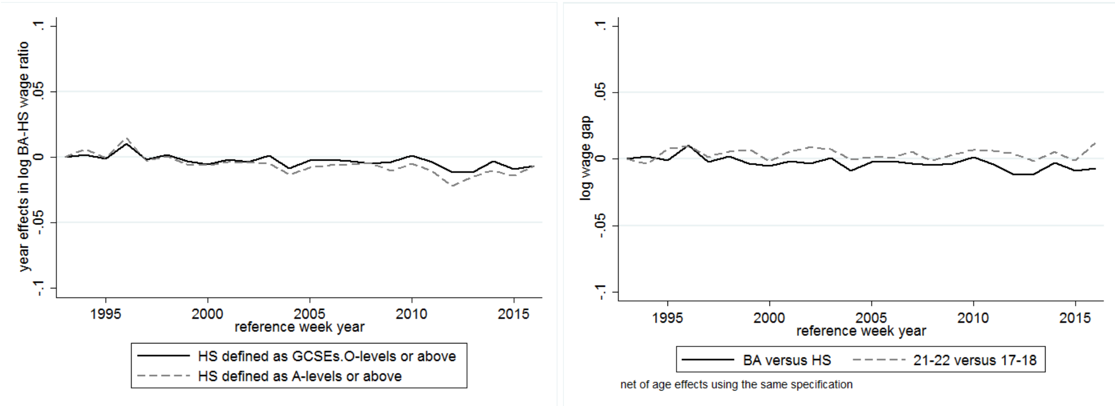


Figure A.1: BA-HS log wage gap under alternative definitions

Notes: Same specification as Figure 2 in the paper. The solid line in each graph is identical to the one in Figure 2.

people, the LFS may yield a biased estimate for the overall BA proportion. As a sensibility check, we obtain the number of graduates from the Higher Education Student Statistics (HESA)[6]. HESA collects student information directly from each university since the early 90s, so the graduate numbers are precise. We use the total number of UK-domicile students obtaining first degrees every academic year. This is plotted as the grey solid line in Figure A.2.

Because information is collected at the time of leaving university, HESA statistics alone cannot tell us how many working-age graduates there are in total in the UK, or anything directly comparable with our Figure 1. So we use the LFS 2016 to derive its implied number of people obtaining first degrees every year. This is also tricky because the LFS doesn't tell us when people obtained each of their qualifications, only when they obtained their highest qualification. Thus, when we plot the number of people over the year they obtained their highest qualification (solid black line in Figure A.2), the number overstates the truth in recent years. This is expected because for those with postgraduate qualifications, they must have obtained first degrees in some earlier

Link
[6]The statistics start in 1994-5 and can be downloaded here.

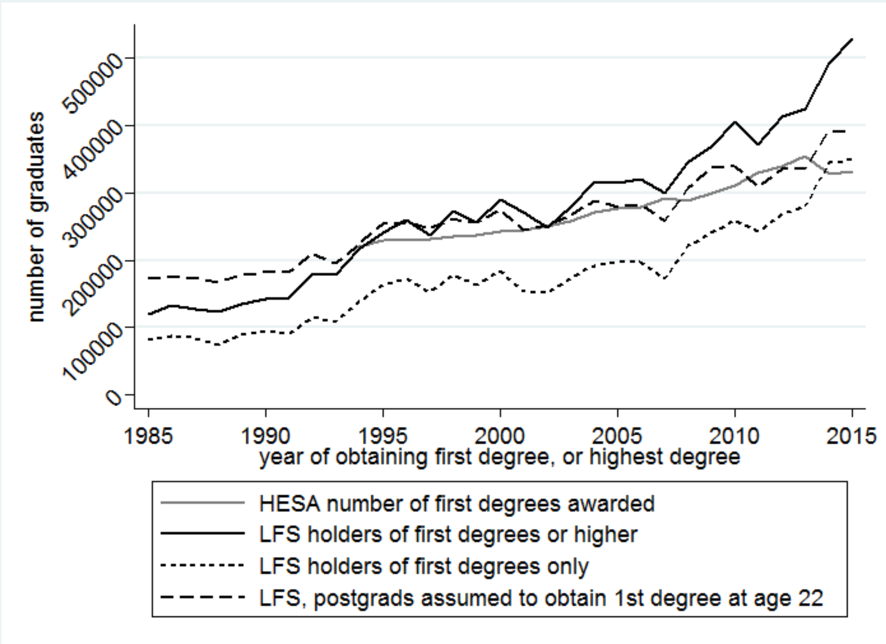


Figure A.2: Number of first-degree graduates over year, HESA and LFS

Notes: the HESA series is the total number of UK-domiciled students obtaining first degrees by academic year, downloaded from here. For all the LFS series, I add up the weight of UK nationals with at least first degrees by the year they obtained their first degree or highest qualification (up to 2015). As I use 4 quarterly LFS datasets in 2016, the weight is divided by 4 to gross up to population totals. The first LFS series counts all those with first degrees or above, by the year they obtained their highest qualification. The second counts those whose highest qualification is a first degree, by when they obtained it. The third assumes that all those with higher degrees obtained their first degree at age 22 and then counts everyone by the year they obtained their first degrees.

unknown years. If we omit all the postgraduates as we do in the short-dashed line in Figure A.2, then obviously we would under-estimate the truth. If we assume that all the postgraduates obtained their first degree at age 22 and add them to the last series, then we get the long-dashed line in Figure A.2. This time series happens to be very similar to the HESA trend. In fact, all three measures from the LFS and the HESA one show a strong increase in the number of new graduates over time. Together with the aging of less-educated older cohorts, this means the overall proportion of graduates in the working-age population increased rapidly since the early 90s.

A.5 Core Patterns by Birth Cohort

In section 2 in the paper, we aggregated the LFS data by 5-year age bands and year to examine time trends. Here we look at trends across birth cohorts. We aggergate the LFS data to the level of age and 5-year birht cohorts. The left subgraph of figure A.3 shows the college wage premium over the life-cycle by cohort. The pattern is striking: the differential is increasing and concave over the life-cycle and there is not much difference across cohorts in either the shape or the level of the differential.

Unsurprisingly, when we regress these wage differentials on an age polynomial of order 5 and a complete set of cohort dummies, we find that the estimated cohort effects are quite flat. This is plotted in the right sub-graph of figure A.3. The same graph also plots the cohort effects in the BA proportion, which is net of age effects in the way. It is clear that the BA proportion is increasing across cohorts and the increase was particularly sharp between the 1965-69 cohort and the 1975-79 cohort. This coincides with the timing of the HE expansion. As the UK Higher Education sector expanded rapidly from 1988 to 1994, the first cohort to be directly affected was born in 1970.

One may suspect that as the BA proportion increased so much, their quality, especially at the lower end of the BA quality distribution, may have fallen. If this is true, one may expect a fall in the wage gap at lower percentiles in the distribution. In Figure A.4, we plot the cohort effects in the wage gap at various percentiles: the 10th, the 25th, the 50th, 75th and 90th.The trend across cohorts is relatively flat for all: the difference from the 1965-69 cohort is 0.1 log terms or less in absolute terms. The 10th percentile of the BA wage relative to the 10th percentile of the HS wage appears to have fallen a bit, by around 0.07 between the 1965-69 and 1975-79 cohorts. However, this decline in the wage gap was driven by a fast increase in the real HS wage at the 10th percentile, rather than a real wage decline among BAs at the 10th percentile. As shown in the 2nd sub- graph of Figure A.4, the 10th percentile of the HS group grew by more than 15% between the 1965 and 1985 cohorts, when that of the BA group was about 10%,

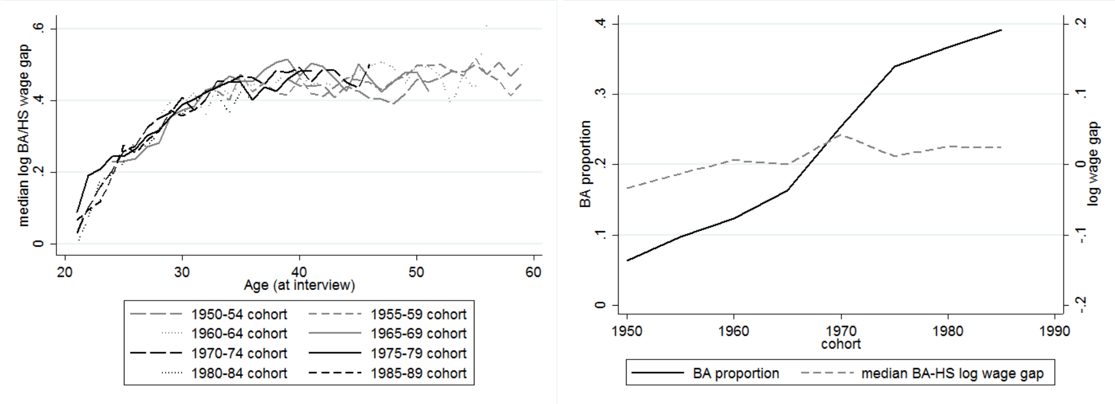


Figure A.3: BA proportion and wage ratio over cohorts

Notes: We aggregate LFS data 1992-2016 up to the level of 5-year-birth-cohorts and age, where age is restricted to 20-59. We look at cohorts 1950-1985 only, so that each chort appears many years in the data. The BA-HS median wage ratio is plotted at this level in the left sub-figure. For the right sub-figure, we regress the BA proportion on cohort dummies and an age polynomial of order 5. For the BA proportion, the cohort effects are scaled to the observed porportion for 1965 cohort at 30 year old. For the wage gap, the cohort effects are normalized to 0 for the 1965 cohort.

and growth was lower at the 25th and 50th percentiles for both groups. This decrease in within-group inequality, particularly for the HS group, looks like a natural consequence of the National Minimum Wage (NMW). The NMW was introduced in 1999 and has been raised at a faster pace than the median wage. Thus, there is no evidence of increasing supply of BAs reduing their relative wage in any part of the distribution.

A.6 Observable compositional changes

In this appendix, we present added investigations into compositional change effects. The first relates to the expansion of post-graduate degree holding.

The dark, solid line in Figure A.5 plots the proportion of people with a postgraduate degree conditional on having a university degree. Similarly to what Lindley and Machin(2006) show for the US, the importance of postgraduate degrees increases for the UK in our period. Nonetheless, the proportion of postgraduates among university degree holders remains relatively low and so its change is unlikely to be a major driver

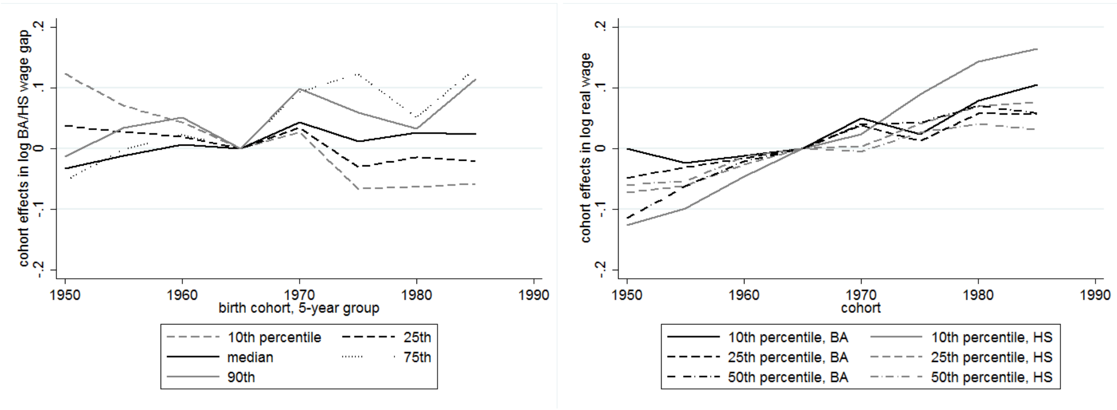


Figure A.4: BA-HS wage ratio at different percentiles

Notes: We aggregate LFS data 1992-2016 up to the level of 5-year-birth-cohorts and age, where age is restricted to 20-59. We look at cohorts 1950-1985 only, so that each chort appears many years in the data. For each percentile shown in the left graph, we regress the BA-HS log wage gap on cohort dummies and an age polynomial of order 5. The cohort effects are normalized to 0 for the 1965 cohort. For the right graph, the dependent variable is the real log wage for each of the shown percentile of the education group.

of relative wage patterns. This is, in fact, what we see in the two wage gap lines in the figure. One line is a replotting of the line in figure 2 in the paper, which includes post-graduate degree holders among the university graduates,t re while the other line shows the wage gap relative to high school educated workers when we include only those with exactly a bachelor's degree and no higher. The two lines are very similar, with both showing nearly identical values in 1993 and 2016.

The second compositional shift we consider relates to immigration. The proportion of UK workers without UK nationality has more than doubled over the past two decades, from under 5% to above 10%. As immigrants are more likely to have university degrees (as confirmed in Figure A.6), the large flows of immigrants contribute directly to the aggregate increase in the share of BAs in the workforce. But it is not clear whether we should count every immigrant with a university education as the equivalent of a university educated native born worker. As demonstrated in Dustmann et al. (2013), immigrants often work in jobs that do not match their observed skills or qualifications,

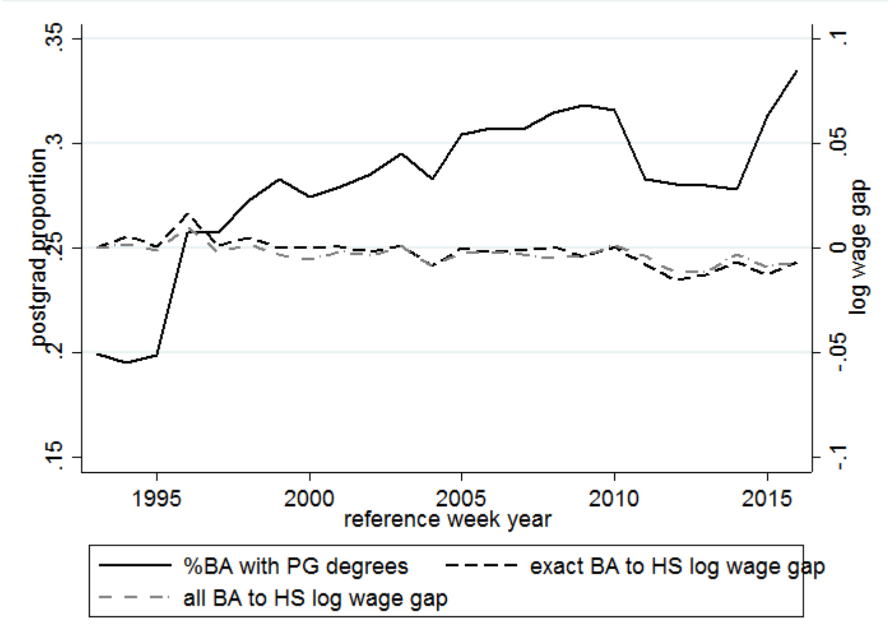


Figure A.5: Year Effects for the Proportion of University Graduates with Advanced Degrees and the BA to HS Wage Ratio

Note: The year effects use the same sample selection and regression specification as for Figure 2 in the paper.

implying that a simple count of the number of immigrants with a university education may over-state the contribution of immigration to the effective supply of highly educated labour. Given the size of the increase in the immigrant proportion in the past 20 years, the positive bias in the measured supply of university labour may become substantial. To address this concern, we can look at the BA-HS wage ratio among UK nationals only. Figure A.6 shows that the BA-HS log wage gap is essentially flat and very similar to the trend including immigrants.

The second observable composition dimension we investigated was between public and private sectors. Public sector employees are, on average, better educated and, with wages largely protected from direct market forces, we might expect wage differentials within the public sector to be more rigid. Given that, an expansion in the public sector

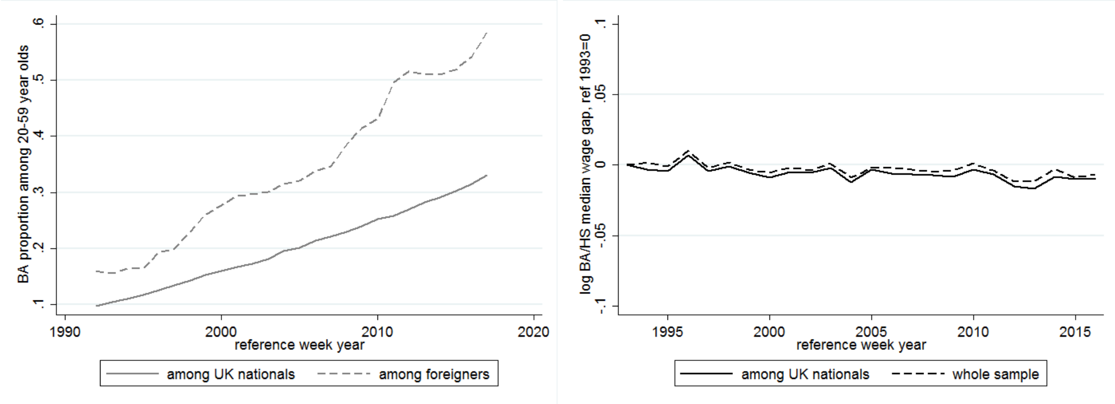


Figure A.6: BA proportion and wage ratio over year, among those born in the UK

Notes: The BA proportions are not normalized. The year effects in wages are normalized to 0 in 1993. The whole sample series is the same as in Figure 2 in the paper.

might partly explain the patterns we have described. That possibility, though, falls short in two ways with respect to employment numbers. First, the proportion of workers in the public sector does not change substantially over our data period. Second, the growth in the proportion of workers with a BA is very similar between the private and public sectors.

The public-private sector dimension of movements in wage differentials is a bit more nuanced. In Figure A.7, we regress the wage differential at the year-age-band level on age dummies and year dummies and plot the year effects. The trend is slightly declining, but relatively flat. Compared to the whole economy (Figure 2 in the paper), the private sector trend is slightly more decreasing: the change in the log wage gap over 22 years is about 0.03, rather than aobut 0.01 for the whole. This is still very small compared to what you might expect from an increase in the relative quantity of BA-to-HS (which is more than 1 full log point over the period).

One place we might look for a compositional shift is at the extensive margin: if the large increase in the relative supply of BAs combined with their constant relative wages induced a relative decline in the employment rate of BAs then this could imply changes

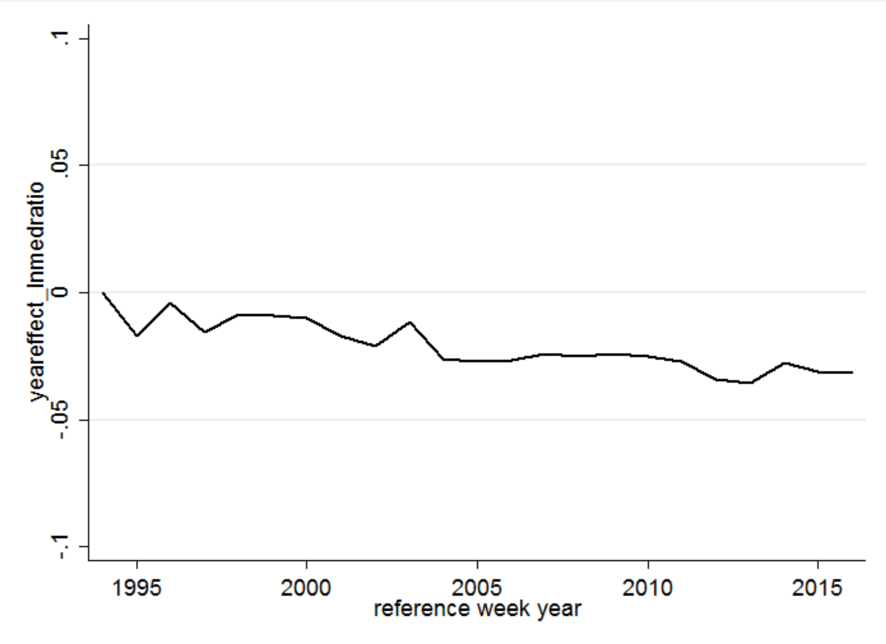


Figure A.7: Time effect in BA-to-HS wage differentials, UK private-sector only

Notes: The time effect is normalized to 0 in 1994, becaue the variable on public versus private sector is available since 1994 only.

in the relative "quality" of BA versus HS workers. In Figure A.8, we plot the estimated year effects in the employment rate of BA's and that of the HS population. The two series move very closely together over time. Thus, the lack of a relative wage response to the educational supply shift was not offset by a relative decline in employment. The change in relative employment rates is also small in the context of a near-tripling in the BA proportion over the period. Thus, we believe compositional shifts based on changes at the extensive margin are not a key driver of the main patterns.

A.7 Unobservable compositional changes: bounds

Implementation of a bounding approach rests on some (preferably minimal) assumptions about the model of wage determination. We will consider a simple but very stan-

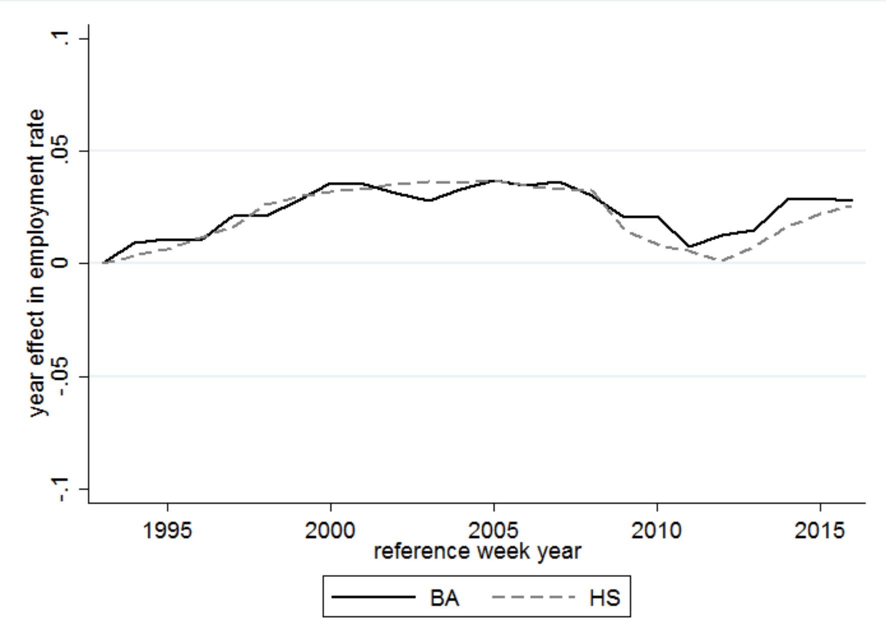


Figure A.8: Time effects in employment rates among BAs and HS workers

Notes: The sample is LFS 1993-2016. The data is collapased to the level of year and 5-year age bands and education. We then regress the employment rate on a complete set of year dummies and age-band dummies. The time effect is normalized to 0 in 1993.

dard model in which the wage for person i in education group j is given by:

$$\ln w_{ict} = \sum_{j=1}^{3} D_{ijt}\beta_{cj} + \sum_{j=1}^{3} D_{ij} f_{cj}(age_{it}) + \sum_{j=1}^{3} D_{ij}\lambda_j \eta_i + \varepsilon_{ict} \tag{A.10}$$

where c indexes the person's birth cohort, D_{ij} equals 1 if person i is in education group j, and zero otherwise, f_{cj} is a cohort-and-education-group-specific age profile of wages, normalized to 0 for age 30 and ε_{ict} is an idiosyncratic error that is independent across time and people and of all other right hand side components in the regression. The specification incorporates a person-specific ability factor, η_i, the effects of which differ across education groups according to loading factors, λ_j. Importantly, both the distribution of η_i and its factor loadings are stationary across cohorts. This model is extreme in its assumption of only one ability factor, but it is also very standard and allows us to see clearly the effects of selection.

We are interested in the price per efficiency unit of workers with a given type of education ($\beta_{cj} + f_{cj}(age_{it})$ in (A.10)). This is unobservable because we do not observe the median wage for a composition constant group, Below we will adopt some assumptions and bounds on the composition-constant median wage for each education group.

We shall assume that the values of the λ's and other parameters are such that for each cohort, the three education groups correspond to three contiguous, non-overlapping ranges of ability. In particular, the groups are defined by two cohort-specific thresholds A_{uhc}, A_{hdc}. University graduates are those with $\eta > A_{uhc}$; high-school grads have $A_{hdc} < \eta \leq A_{uhc}$; and high-school dropouts have $\eta \leq A_{hdc}$. In theory, such a hierachical model of selection could be rationalized by a Roy model where individuals choose education levels by comparing their expected net present value of wages and of costs, and assuming $\lambda_u > \lambda_h > \lambda_d$ and that the costs of obtaining education are weakly decreasing in ability. In addition, the hierachical model fits the idea that university admission is largely rationed by prior attainment.

Consider a situation in which the university proportion increases between cohorts c and c+1, because there is less rationing. This corresponds to a decline in the value of A_{uhc}. Importantly, some individuals who would not get a university degree if they were born with their respective ability in cohort c will get a degree if they belong to cohort c+1 but no one is induced to make the opposite switch. That is, there will be flows in only one direction. Let's call the set of individuals who would get a degree if they face the conditions in cohort c+1 but not if they were in cohort c, "joiners". Their ability distribution has a range with a top value of A_{uhc} and so it lies entirely below that of the rest of university graduates in cohort c+1. The latter group have abilities that are high enough for them to enter university even when the costs were higher (as they were for cohort c). We will call them "stayers".[7] Obviously, the joiners' ability distribution lies

[7] Calling them stayers and joiners is a slight abuse of terminology since we are considering different cohorts and so there are no individuals actually staying or joining. Instead, these groups correspond to

above that of those who remain in the HS group in cohort c+1.

The observed wage distribution of BAs in cohort c+1 is a combination of that of the joiners and that of the stayers. Under our assumptions, if the number of BA's increases across cohorts then that must reflect an inflow of joiners but no outflow. That means we can use the observed median wage for BA's in the first cohort as corresponding to the median wage of the stayers. In the second cohort, we can form two extreme bounds based on what we assume about the joiners. In the first, we could assume that all the joiners have lower ability than the median stayer. We could then form one extreme estimate of the median wage for stayers by first trimming a number of observations equal to the number of joiners from the bottom of the observed wage distribution for the second cohort and then getting the median of the remaining observations. For example, if the size of the BA group increases from 20 to 30 percentage points of the population between cohort c and cohort c+1 at a given age, then we trim the bottom one third of the BA wage distribution of cohort c+1 and the median of the remaining distribution is the upper bound of the median of the stayers. Another extreme bound could be formed by similarly trimming the top third of the cohort c+1 distribution and getting the median for the remaining sample. However, under an hierarchical model of the kind we are discussing, the best the joiners could be is as good as the stayers (if they were better than the stayers, they would be in the sector already). If they are as good as the stayers then the observed median wage for BA's in cohort c+1 would be the same as the median wage for the stayers. Thus, the observed median forms the other bound on the cohort c+1 median wage for the stayers. The next two pages explain mathematically why the trimming method and the observed median are THE upper and lower bounds under the hierachical model. Differencing these bounds for the stayers' median wage in cohort c+1 from the observed median wage for cohort c then gives us bounds on the movements in the price for BA labour for a composition constant group.

---

different ranges in the stationary η distribution.

Because people (or, more properly, ability values) can be induced to switch into or out of higher education but not both at the same time, we can decompose the distribution function for BA wages in cohort c+1 into a component related to the distribution function for the "stayers" and a component for the "joiners":

$$\Pr(\ln W_{uc+1} < w|\eta > A_{uhc+1}) = p_{uc+1}\Pr(\ln W_{uc+1} < w|\eta > A_{uhc}) \\ + (1-p_{uc+1})\Pr(\ln W_{uc+1} < w|A_{uhc} \geq \eta > A_{uhc+1}), \forall w \qquad \text{(A.11)}$$

where, p_{uc+1} is the proportion of the university educated in cohort c+1 who are stayers. Equation (A.11) holds for any wage level w, but we are interested in a particular level: the median wage in cohort c+1 for the university sector stayers, denoted as $\tilde{w}_{uc+1}$.

We can write $\tilde{w}_{uc+1}$ as,

$$\tilde{w}_{uc+1} = \beta_{c+1u} + f_{c+1u}(age_{it+1}) + \lambda_u med(\eta_i + \varepsilon_{ic+1t+1}|\eta_i > A_{uhc}) \qquad \text{(A.12)}$$

Assuming stationarity of the η and ε distributions across cohorts, differencing this relative to the median conditional university wage in cohort c at the same age, age^* would yield,

$$\tilde{w}_{uc+1} - med(\ln W_{uct}|\eta_i > A_{uhc}) = \beta_{c+1u} + f_{c+1u}(age^*) - \beta_{cu} - f_{cu}(age^*) \qquad \text{(A.13)}$$

That is, by comparing wage movements for people with the same set of $\eta's$ (the ones corresponding to choosing to get a university degree under either set of costs), we could obtain an estimate of the change in the actual wage profile across cohorts.

We cannot observe $\tilde{w}_{uc+1}$ because we are comparing across cohorts and so cannot see who has ability levels that would result in their choosing the university degree in the different rationing situations. But we can obtain bounds for it. Returning to equation (A.11), we can obtain an estimate of p_{uc+1} based on changes in the size of the u group between cohort c and c+1 combined with the argument that people (or, rather, ability levels) either enter or leave the group but not both. We know that the second term on the right hand side of (A.11) ($\Pr(\ln W_{uc+1} < \tilde{w}_{uc+1}|\eta > A_{uhc})$) equals 0.5 by the definition of $\tilde{w}_{uc+1}$, and the left hand side corresponds to a quantile of the conditional distribution of wages for the u group in the c+1 cohort, and so is calculable from the data. That only leaves the last term ($\Pr(\ln W_{uc+1} < \tilde{w}_{uc+1}|A_{uhc} \geq \eta > A_{uhc+1})$) unknown and unknowable. However, since it is a probability, we can bound it on one side as $\Pr(\ln W_{uc+1} < \tilde{w}_{uc+1}|A_{uhc} \geq \eta > A_{uhc+1}) = 1$, which corresponds to the marginal people who obtain a degree in cohort c+1 but would not have done so in cohort c having wages that place them below the median wage for the group who would get a degree in either cohort. Based on this, we can get an upper bound on $\tilde{w}_{uc+1}$ by solving,

$$\Pr(\ln W_{uc+1} < \tilde{w}_{uc+1}|\eta > A_{uhc+1}) = \frac{1}{2}p_{uc+1} + (1 - p_{uc+1}), \tag{A.14}$$

This is equivalent to trimming the bottom $(1 - p_{uc+1})$ proportion of observations from the c+1 university wage distribution and obtaining the median of the remaining sample.

Since the abilities of university "joiners" between cohort c and c+1 are assumed to be entirely below the abilities of the "stayers", a joiner's wage can be higher than a stayer's only when the joiner has a particularly positive shock ε_{it} or the stayer has a particularly negative shock. As the idiosyncratic shock is assumed to be independent of ability, it follows that the joiners' wage distribution is first order stochastically dominated by that of the stayers. Mathematically,

$$\Pr(\ln W_{uc+1} < \tilde{w}_{uc+1}|A_{uhc} \geq \eta > A_{uhc+1}) \geq \Pr(\ln W_{uc+1} < \tilde{w}_{uc+1}|\eta > A_{uhc}) \tag{A.15}$$

Using the right side of this expression as the lower bound on $\Pr(\ln W_{uc+1} < \tilde{w}_{uc+1}|A_{uhc} \geq \eta > A_{uhc+1})$ in (A.11) implies that the right hand side of (A.11) just equals 0.5. That is, the other bound is the c+1 median itself.

Meanwhile, we can implement a similar exercise for the HS group. In this case, though, if the BA group grows between cohort c and c+1 this must be directly matched with an emigration of individuals from the top of the HS ability distribution between those cohorts. In trimming terms, this means that one bound can be obtained by appending a number of workers equivalent to the increase in size of the BA group to the top of the cohort c+1 wage distribution for HS workers. At the same time, if the Drop-out group shrinks then, under the single factor Roy model, they must have moved to the bottom of the ability distribution in HS and we would trim a number of workers equivalent to the decrease in size of the Drop-out sector from the bottom of the cohort c+1 HS distribution. Doing both the BA and Drop-out related trimming and appending yields a new adjusted HS sample in cohort c+1 that corresponds to one bound on the wages for the HS group stayers. Taking the difference between the median wage in that sample and the actual median wage for HS workers in cohort c yields an upper bound on the change in the log wage profile at a given age for HS workers. Consider the benchmark case where the upper bound scenarios for the BA and HS workers correspond to one another (i.e., the movements out of the top of the HS distribution become the movements into the bottom of the BA distribution). We can then obtain one bound on the movement in the university - high school wage differential by taking the difference between the upper bound on the movement in the university median and the upper bound on the movement in the high school median. The other bound is the actual change in the median wage ratios shown in Figure A.3.

We repeat the sample trimming exercise for each cohort using the 1965-69 cohort as the base of comparison (cohort c in our example). The resulting quality-adjusted wage differentials are reported in the left panel of Figure A.9. The second panel shows cohort effects derived in the same manner as in the earlier figures. The cohort effects show an increase in the adjusted upper bound differential between the 1965-69 and 1970-74 cohorts. Given that the other bound is the actual change in the median wage

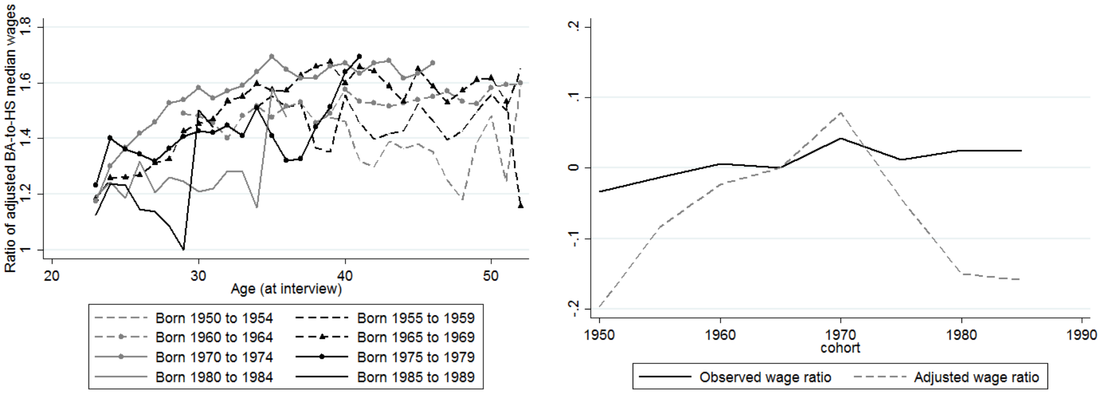


Figure A.9: UK Median BA-to-HS wage ratio, adjusted to the education split of 1965 cohort

Notes: For each age and cohort, we adjust the wage distribution by using the proportions observed for 1965 cohort as reference points. For example, if the observed proportion of BAs is higher than that for the 1965 cohort at the same age, we would trim the bottom of the observed BA distribution.

ratio, the implication is that under this ability model, one cannot argue that selection on unobservables obscured what was actually a decline in the true wage differential. For the difference between the 1965-69 and 1975-79 differential, one bound shows a near zero change and the other shows a 4 percent decline. Thus, here there is some room to argue that selection is hiding a true decline in the ratio, but that decline is still very small compared to a doubling of the proportion of the population with a BA. For the post-1980 cohorts, the bounds include larger declines - about 15% relative to the 1965-69 cohort. However, a glance at the profiles in the left panel suggests the need for some caution in interpreting the cohort coefficients. The age profiles for the various cohorts no longer look parallel once the extreme bound trimming is implemented, implying that the age at which we evaluate the cohort differences can alter our conclusions. But, overall, our conclusion from this exercise is that, under this model of ability, selection on unobservables cannot explain why we do not see a large decline in the education wage differential for the cohorts with the largest increase in their education level.

A.8 Implications of Exogenous Skill Biased Technological Change with Managerial Tasks

In this appendix, we examine the implications of an exogenous skill biased technological change in the context of a standard production function that incorporates two skill levels and two broad types of tasks. The model exposition is similar in nature to that used in the Borghans and ter Weel (2008) paper on technology diffusion and the labour market. In particular, we consider a model in which one technology is in use at a time. Output, Y, is produced according to the Cobb-Douglas production function:

$$Y = M^{\alpha} L^{1-\alpha} \tag{A.16}$$

where, M is hours of managerial labour, L is hours of production labour, and α is a parameter. Each task is performed by a combination of skilled and unskilled labour, with the labour aggregated through CES functions:

$$M = [aS_M^{\sigma} + (1-a)U_M^{\sigma}]^{1/\sigma} \tag{A.17}$$

and

$$L = [bS_L^{\rho} + (1-b)U_L^{\rho}]^{1/\rho} \tag{A.18}$$

where, $\frac{1}{1-\sigma}$ is the elasticity of substitution between skilled and unskilled labour in managerial tasks;$\frac{1}{1-\rho}$ is the elasticity in labouring tasks; a and b are parameters; S_M is the amount of skilled labour in the managerial task; and U_L is the amount of unskilled labour in the basic labouring task. We assume that skilled labour is relatively more productive in the managerial tasks (i.e., $a > b$) and that skilled and unskilled labour are more substitutable in the labouring task (i.e., that $\rho > \sigma$).

We assume that the numbers of unskilled and skilled workers in the economy are given exogenously in any period and that each worker supplies a fixed endowment of labour inelastically. Market clearing in the labour market corresponds to the total number of workers with each skill level in the economy being equal to the sum of the numbers employed in the various occupations and technologies:

$$S = S_L + S_M$$

and,

$$U = U_L + U_M$$

Workers of each skill type can choose freely whether to work as a manager or a labourer and so there will be one skilled wage, w_s and one unskilled wage w_u.

In this framework, a skill-biased technological change can be represented as an increase in a, i.e., an increase in the productivity of S workers as managers. This captures both that the technological change favours S workers and that it is related to management tasks. Note that we are assuming that the technological change arrives exogenously and alters the production function of firms without them choosing whether or not to adopt the new technology.

To understand the impact of this change note that, working from the firm's first order conditions, it is straightforward to show that the wage skill ratio is,

$$\begin{aligned} \frac{w_s}{w_u} &= \frac{a}{1-a}(\frac{S_M}{U_M})^{\sigma-1} \qquad \text{(A.19)} \\ &= \frac{b}{1-b}(\frac{S_L}{U_L})^{\rho-1} \end{aligned}$$

Rearranging these expressions slightly, we get:

$$\frac{a}{1-a}\frac{S_M^{\sigma-1}}{S_L^{\rho-1}} = \frac{w_s}{w_u} = \frac{b}{1-b}\frac{U_M^{\sigma-1}}{U_L^{\rho-1}} \qquad \text{(A.20)}$$

In the context of this model, in order to match the main data pattern of an increase in S accompanied by no change in $\frac{w_s}{w_u}$, equation (A.19) shows that we need an increase in a of just the right size so that the skill biased demand increase just balances the relative supply shift. We view it as somewhat implausible that there were an exogenous set of technological changes that just balanced the supply shifts over an extended period of time, but we cannot reject that this could have occurred. Instead, we ask about the further implications of such changes if this were the mechanism driving our main data patterns. Examining (A.20), note that if a increases then the ratio of the number of skilled workers who are managers to the number who are labourers must also increase in order to match the unchanging wage ratio. This is the opposite of the implication from our endogenous technological choice model in which the expansion in S is accompanied by a decreasing proportion of S workers who are managers.

A.9 Results on education expansion and wages in other countries

Our analysis fits with results in Crivallero (2016). She uses two European surveys to examine wage and education patterns in 12 European countries between 1994 and 2009. Many of the economies in her data are in our sample of countries with substantial educational growth in this period.[8] and she shows that the proportion of the population who are tertiary education graduates for all of these countries pooled together goes up by 50% across the birth cohorts she studies. The dependent variable in the main exercise in the paper is the wage premium to having a tertiary or other post-secondary education relative to a high school diploma. This is regressed on a relative educational supply variable, a variable intended to capture skill biased demand shifts, and a complete set of country, year, and birth cohort effects. The results indicate statistically significant but very small relative supply effects with a 10% increase in the relative number of post-

[8]The countries in her data are: Austria, Belgium, Germany, Denmark, Spain, Finland, France, Greece, Ireland, Italy, Portugal, and the UK.

secondary to secondary graduates being associated with a 1.2% decline in the log wage ratio for the two groups in their OLS estimates. In addition, the relative demand effect is very small and not statistically significant from zero. Thus, Crivellaro's results with a set of 12 European economies matches closely with our results for the UK: substantial increases in education have little effect on the wage ratio and there is also little evidence of an ongoing skill biased demand shift.

Our results also fit with findings in some other papers examining wage differentials and education increases in other economies. Chen (2013) examines these patterns for Taiwan, which underwent a dramatic boom in creating new post-secondary institutions between 1990 and 2000. As a result of that boom, between 1990 and 2010, the number of post-secondary graduates increased by a 600%. Yet over that same period, the difference between the mean log hourly wage for university graduates and workers with less than a university education was quite flat. That university wage premium was approximately 0.6 in 1980, 1990, 2000, and 2010. Chen interprets this outcome within an exogenous skill biased technological change model. As we argued earlier, for such a model to generate a flat premium trend requires a lucky, exact balance of relative supply and exogenous demand shifts. We believe that our model, in which the flat profile provides a more natural explanation. Choi and Jeong (2005) and Choi (2015) examined a similarly large increase in education levels driven by policy changes in South Korea in the 1980s and 1990s. Between 1990 and 2005, the proportion of high school graduates who enrolled in a post-secondary programme increased from approximately 30% to 80%. Choi and Jeong (2005) show that the post-secondary wage premium declined in the 1980s but was flat during the substantial educational expansion that started in the mid-1990s. They show that the latter patterns coincided with an increase in expenditures on IT and conclude that the flat premium reflected an endogenous technological change model. These trends could fit with our model, with the initial decline in the wage premium in the 1980s corresponding to a period before the economy entered the cone of diversification. The post-1994 period is then the period of transition to taking up more skill-biased technologies, as evidenced by the coinciding increase in IT expenditures. Finally, **?** examine wage impacts of an earlier large increase in post-secondary attainment in Norway, taking advantage of regional variation in the creation of universities in the 1970s. They show that the regions where new universities were added had a significant jump in the education level of their workforce but that the wage differential between university and high school educated workers either stayed flat or increased. They interpret this within the context of an endogenous technological change model and show evidence that the productivity of skilled workers increased in relative terms in the regions with new universities.

Finally, we present the results from a simple exercise based on the data on educational attainment and wage differentials from OECD (2012). As mentioned in the text, we focus on the set of OECD economies that have a lower proportion of their population than the US with a tertiary education in the initial year of the data (1997) and experience a growth in that proportion by at least 40% by 2010. In the table, below, we present estimates for this set of countries from a regression on a constant and a linear trend of the wage ratio between the mean annual earnings of all workers aged 25 to 64 with a tertiary education and the mean annual earnings of workers with an upper secondary education being their highest education level. Of the 11 countries meeting our criterion, 7 have trend coefficients that are not statistically significantly different from zero, 2 have positive and significant coefficients, and 2 have negative and significant coefficients.

Table A.3: Regressions of Wage Differential on a Time Trend by Country

Country	Const	t
Belgium	30.02**	0.15
	(2.19)	(0.23)
France	47.25**	0.036
	(4.24)	(0.46)
Ireland	47.84**	1.47
	(12.2)	(1.2)
Korea	31.82**	1.86*
	(7.26)	(0.73)
New Zealand	22.26**	-0.28
	(2.9)	(0.3)
Norway	31.98**	-0.33**
	(0.91)	(0.11)
Poland	72.66**	-0.18
	(6.58)	(0.68)
Spain	19.63**	1.72**
	(2.88)	(0.3)
Sweden	32.57**	-0.58**
	(1.06)	(0.11)
Switzerland	57.76**	-0.21
	(2.12)	(0.21)
UK	58.59**	0.045
	(2.81)	(0.29)

Authors's calculations based on data from OECD (2012). Standard errors in parentheses.
*, ** statistically significant at the 1% and 5% significance levels, respectively.

Appendix B

Appendix for chapter 3

B.1 Derivation of a demand-side equation

In this section, we will derive a prediction about the relationship between task price ratio and task quantity ratio. That is equation (3.12) in chapter 3.

The F.O.C. with regard to task j for a firm using technology T is:

$$p_{jt} = p_{gt}\frac{\partial Y_{gt}^{T}}{\partial y_{gjt}^{T}} = p_{gt}\alpha_{gj}^{T}(y_{gjt}^{T}/Y_{gt}^{T})^{\rho-1} \quad \forall j,g,t,T \in \{O,N\} \tag{B.1}$$

Apply $j=1$ to (B.1) and take the ratio of the same equation between j and 1, we get

$$\frac{p_{jt}}{p_{1t}} = \frac{\alpha_{gj}^{T}}{\alpha_{g1}^{T}}(\frac{y_{gjt}^{T}}{y_{g1t}^{T}})^{\rho-1} \quad \forall j,g,t,T \in \{O,N\} \tag{B.2}$$

$$\frac{y_{gjt}^{T}}{y_{g1t}^{T}} = (\frac{p_{jt}\alpha_{g1}^{T}}{p_{1t}\alpha_{gj}^{T}})^{\frac{1}{\rho-1}} \quad \forall j,g,t,T \in \{O,N\} \tag{B.3}$$

Because we don't directly observe technology, we don't observe y_{gjt}^{T}. What we can observe is industry-level occupational employment $EMP_{gjt} = y_{gjt}^{O} + y_{gjt}^{N}$.

$$\frac{EMP_{gjt}}{EMP_{g1t}} = \frac{y^O_{gjt}}{y^O_{g1t}+y^N_{g1t}} + \frac{y^N_{gjt}}{y^O_{g1t}+y^N_{g1t}} \tag{B.4}$$

$$= \frac{y^O_{g1t}}{y^O_{g1t}+y^N_{g1t}}\frac{y^O_{gjt}}{y^O_{g1t}} + \frac{y^N_{g1t}}{y^O_{g1t}+y^N_{g1t}}\frac{y^N_{gjt}}{y^N_{g1t}} \tag{B.5}$$

$$= \frac{y^O_{g1t}}{y^O_{g1t}+y^N_{g1t}}\left(\frac{p_{jt}\alpha^O_{g1}}{p_{1t}\alpha^O_{gj}}\right)^{\frac{1}{\rho-1}} + \frac{y^N_{g1t}}{y^O_{g1t}+y^N_{g1t}}\left(\frac{p_{jt}\alpha^N_{g1}}{p_{1t}\alpha^N_{gj}}\right)^{\frac{1}{\rho-1}} \tag{B.6}$$

Denote $w_{gt} = y^N_{g1t}/(y^O_{g1t}+y^N_{g1t})$. We can interpret w_{gt} as the share of 'New' technology in industry g at time t. Denote

$$r^O_{gj} = (\alpha^O_{gj}/\alpha^O_{g1})^{1/(1-\rho)} \tag{B.7}$$

$$r^N_{gj} = (\alpha^N_{gj}/\alpha^N_{g1})^{1/(1-\rho)} \tag{B.8}$$

Equation (B.6) simplifies to

$$\frac{EMP_{gjt}}{EMP_{g1t}} = \left(\frac{p_{jt}}{p_{1t}}\right)^{\frac{1}{\rho-1}}[(1-w_{gt})r^O_{gj} + w_{gt}r^N_{gj}] \tag{B.9}$$

The last term is a weighted average between two technologies, where the weight w_{gt} is endogenous.

Flipping the task price ratio to the left hand side, we get

$$\ln\left(\frac{p_{jt}}{p_{1t}}\right) = (\rho-1)\ln\frac{EMP_{gjt}}{EMP_{g1t}} + (1-\rho)\ln[(1-w_{gt})r^O_{gj} + w_{gt}r^N_{gj}] \tag{B.10}$$

B.2 Derivation of task supply equation

Let's denote expected task output conditional on observed skills as

$$y(a,s,j) = E[y(i,j)|a_i = a, s_i = s] \tag{B.11}$$
$$= k_j e^{\beta_{aj}a+\beta_{sj}s} E[e^{\mu_i}|a_i = a, s_i = s] \tag{B.12}$$

Note that $y(a,s,j)$ does not condition on the actual occupational choices, which would be endogenous.

Going back to (3.9) and using (B.12) to substitute for $k_j e^{\beta_{aj}a+\beta_{sj}s}$, we get

$$\pi_j(a,s,\mathbf{p}) = [e^{\beta_{ak}a+\beta_{sk}s+\eta_k} k_k p_k]^{\frac{1}{\zeta}} / \sum_j [e^{\beta_{aj}a+\beta_{sj}s+\eta_j} k_j p_j]^{\frac{1}{\zeta}}$$
$$= [e^{\eta_k} p_k y(a,s,k)/E[e^{\mu_i}|a_i = a, s_i = s]]^{\frac{1}{\zeta}} / \sum_j [e^{\eta_j} p_j y(a,s,j)/E[e^{\mu_i}|a_i = a, s_i = s]]^{\frac{1}{\zeta}}$$
$$= [e^{\eta_k} p_k y(a,s,k)]^{\frac{1}{\zeta}} / \sum_j [e^{\eta_j} p_j y(a,s,j)]^{\frac{1}{\zeta}}$$

This last equation says occupation choice depends on task prices, ζ, occupation amenities η_j, and $y(a,s,j)$ for all j.

Given task prices, the supply of task j is

$$LS_j(\mathbf{p}) = \sum_i \pi_j(a_i, s_i, \mathbf{p}) y(i,j) \tag{B.13}$$
$$= \int\int \pi_j(a,s,\mathbf{p}) y(a,s,j) f(a,s) dads \tag{B.14}$$

where $f(a,s)$ is the joint density function.

B.3 Derivation of when will firms be indifferent between two technologies

This section derives equation (3.13) in chapter 3.

Given the CES production function, the cost of using technology T to produce one unit of output in industry g is

$$unitcost^T_{gt} = [\sum_j (\alpha^T_{gj})^{\frac{1}{1-\rho}} p_{jt}^{\frac{\rho}{\rho-1}}]^{1-1/\rho} / A^T_{gt} \tag{B.15}$$

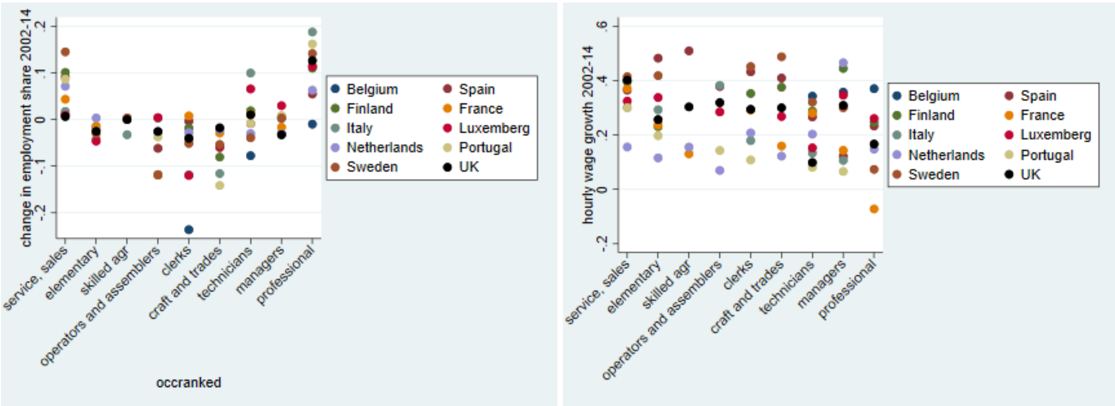
The ratio of unit costs between the two technologies is:

$$\frac{unitcost^N_{gt}}{unitcost^O_{gt}} = \frac{A^O_{gt}}{A^N_{gt}} [\frac{\sum_j (\alpha^N_{gj})^{\frac{1}{1-\rho}} p_{jt}^{\frac{\rho}{\rho-1}}}{\sum_j (\alpha^O_{gj})^{\frac{1}{1-\rho}} p_{jt}^{\frac{\rho}{\rho-1}}}]^{1-1/\rho} \tag{B.16}$$

When the two technologies in industry g have exactly the same unit cost, we have

$$\begin{aligned}
&[\frac{\sum_j (\alpha^N_{gj})^{\frac{1}{1-\rho}} p_{jt}^{\frac{\rho}{\rho-1}}}{\sum_j (\alpha^O_{gj})^{\frac{1}{1-\rho}} p_{jt}^{\frac{\rho}{\rho-1}}}]^{1-1/\rho} = \frac{A^N_{gt}}{A^O_{gt}} \\
\Rightarrow &\frac{\sum_j (\alpha^N_{gj})^{\frac{1}{1-\rho}} p_{jt}^{\frac{\rho}{\rho-1}}}{\sum_j (\alpha^O_{gj})^{\frac{1}{1-\rho}} p_{jt}^{\frac{\rho}{\rho-1}}} = (\frac{A^N_{gt}}{A^O_{gt}})^{\frac{\rho}{\rho-1}} \\
\Rightarrow &\sum_j (\alpha^N_{gj})^{\frac{1}{1-\rho}} p_{jt}^{\frac{\rho}{\rho-1}} - (\frac{A^N_{gt}}{A^O_{gt}})^{\frac{\rho}{\rho-1}} \sum_j (\alpha^O_{gj})^{\frac{1}{1-\rho}} p_{jt}^{\frac{\rho}{\rho-1}} = 0 \\
\Rightarrow &\sum_j [(\alpha^N_{gj})^{\frac{1}{1-\rho}} - (\frac{A^N_{gt}}{A^O_{gt}})^{\frac{\rho}{\rho-1}} (\alpha^O_{gj})^{\frac{1}{1-\rho}}] p_{jt}^{\frac{\rho}{\rho-1}} = 0
\end{aligned}$$

Figure B.1: Employment and wage growth by ISCO major group



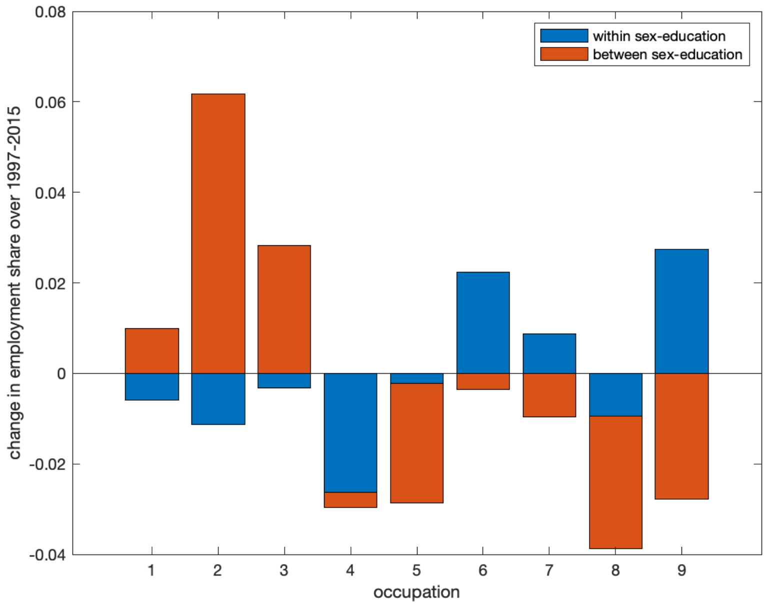
Source: SES 2002 and 2014. To compute the change in hourly wages, we exclude cells where the occupation's employment share has more than tripled or halved because those cases may involve large compositional changes.

B.4 Derivation of an equation to identify TFP terms

We can get A^T_{gt} as an analytical function of $(\alpha^T_{gj}, p_{jt}, p_{gt}, \rho)$, assuming $\rho \neq 0$. This is because the profit maximisation gives a FOC:

$$
\begin{aligned}
& p_{gt}A^T_{gt}[\sum_j \alpha^T_{gj}(y^T_{gjt})^\rho]^{\frac{1-\rho}{\rho}} \alpha^T_{gj}(y^T_{gjt})^{\rho-1} = p_{jt} \\
\Rightarrow & p_{gt}A^T_{gt}[\sum_j \alpha^T_{gj}(y^T_{gjt})^\rho]^{\frac{1-\rho}{\rho}} [\alpha^T_{gj}(y^T_{gjt})^\rho]^{\frac{\rho-1}{\rho}} (\alpha^T_{gj})^{1/\rho} = p_{jt} \\
\Rightarrow & [\frac{1}{[\sum_j \alpha^T_{gj}(y^T_{gjt})^\rho]}]^{\frac{\rho-1}{\rho}} [\alpha^T_{gj}(y^T_{gjt})^\rho]^{\frac{\rho-1}{\rho}} = \frac{p_{jt}}{p_{gt}A^T_{gt}(\alpha^T_{gj})^{1/\rho}} \\
\Rightarrow & \frac{\alpha^T_{gj}(y^T_{gjt})^\rho}{[\sum_j \alpha^T_{gj}(y^T_{gjt})^\rho]} = [\frac{p_{jt}}{p_{gt}A^T_{gt}(\alpha^T_{gj})^{1/\rho}}]^{\frac{\rho}{\rho-1}} \\
\Rightarrow & 1 = \sum_j [\frac{p_{jt}}{p_{gt}A^T_{gt}(\alpha^T_{gj})^{1/\rho}}]^{\frac{\rho}{\rho-1}} \\
\Rightarrow & (p_{gt}A^T_{gt})^{\frac{\rho}{\rho-1}} = \sum_j [\frac{p_{jt}}{(\alpha^T_{gj})^{1/\rho}}]^{\frac{\rho}{\rho-1}} \qquad \text{(B.17)}
\end{aligned}
$$

Figure B.2: Within-between decomposition of the change in occupational employment shares

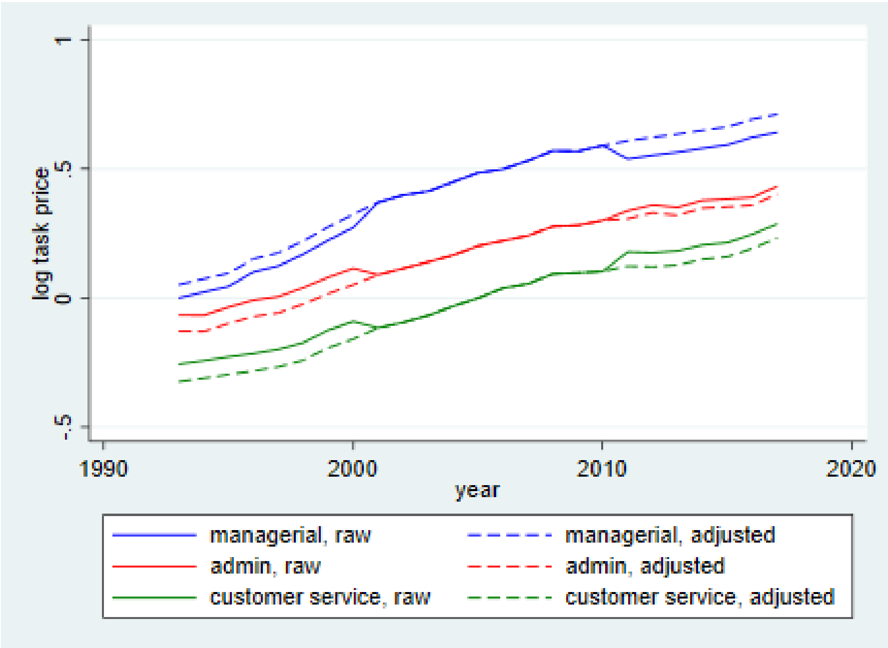


Source: UK Labour Force Survey

B.5 Additional figures

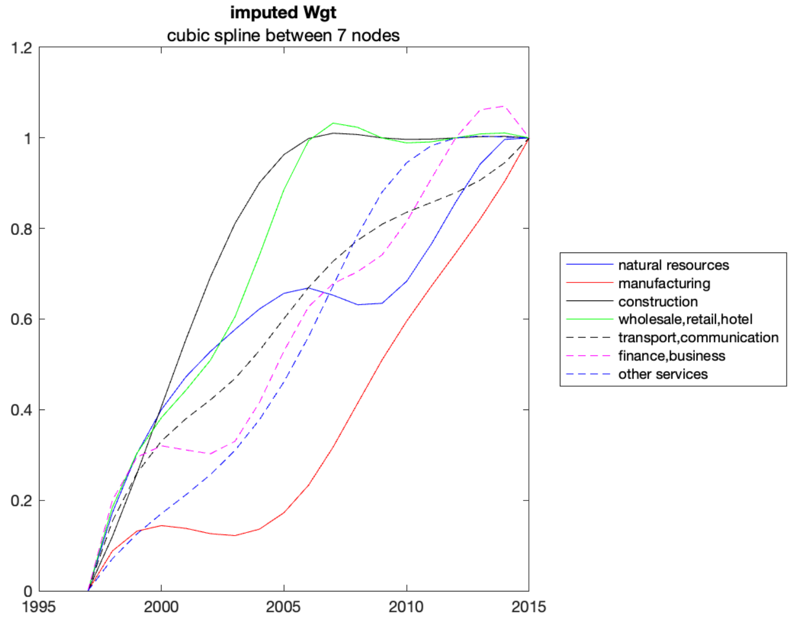
The flip side of the flat proportion of abstract occupations within graduates is that almost all of the aggregate increase in abstract occupations' share can attributed to the increase in education. Using the LFS (1997-2015), I decompose the change in occupational employment shares into within-gender-education-group component and between-group component. Figure B.2 suggests that all of the increase in abstract employment is between-group, and almost all of the decline in skilled trades and operative employment is between-group.

Figure B.3: Adjusting occupational wage for classification changes



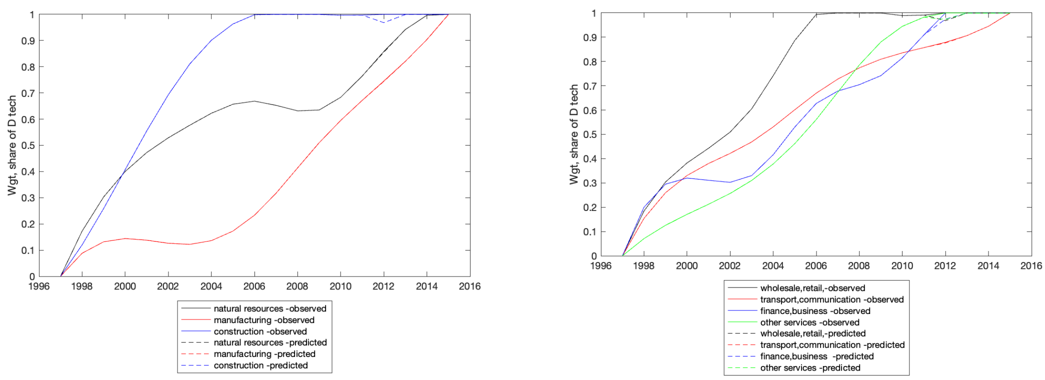
Source: UK Labour Force Survey 1993-2017.

Figure B.4: Estimated w_{gt} from 9 proxies measures



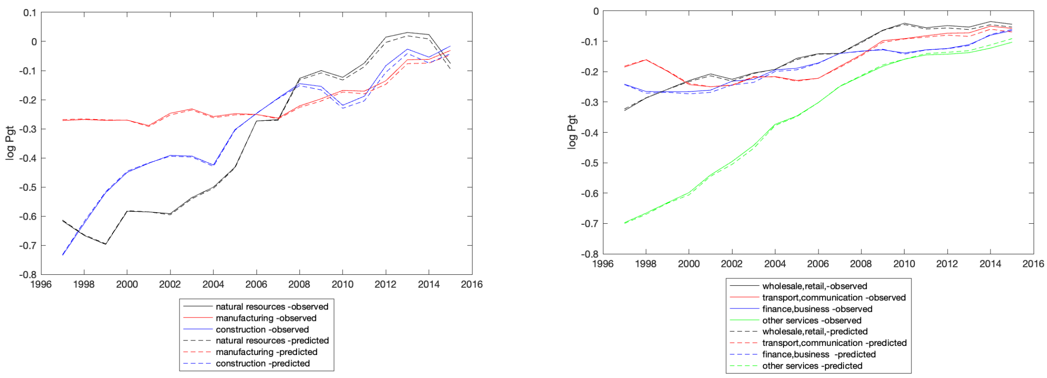
Note: estimated w_{gt} under the assumption that it follows a cubic spline in between each pair of nodes, nodes are 3 years apart from 1997 to 2015, the value in 1997 is constrained to be 0 and the 2015 value is constrained to be 1. Note that w_{gt} is not really comparable between industries, because of the affine transformation within industry.

Figure B.5: Fit of New technology's share w_{gt}



Note: The actual time trends of technology shares are solid lines. The corresponding baseline predictions are dashed lines of the same colours.

Figure B.6: Fit of log industry prices P_{gt}



Note: The actual time trends of industry prices are solid lines. The corresponding baseline predictions are dashed lines of the same colour.

Appendix C

Appendix to chapter 4

C.1 Chinese datasets

Two household survey data have been used repeatedly in this paper.

The China Health and Retirement Longitudinal Study (CHARLS) surveys a nationally-representative sample of Chinese people aged 45 or older. The first two waves of CHARLS were collected in 2011 and 2013, and those waves of data are used in the paper. The third wave of CHARLS has been collected and will be only available in 2017. The CHARLS wave 1 data contains about 10,000 households and more than 17,000 adults aged 45 or above. As this paper only considers those with urban hukou, the effective number of adults are around 3,800.

Like most household surveys, CHARLS contains the usual information such as age, education, employment and incomes. As its focus is on the elderly, it contains a great amount of detailed information on pension incomes and participation, health insurance, past diagnoses and health expenditures. CHARLS is also unique in the amount of information that is available about one's children, who may or may not live in the same household. For each child, it contains age, gender, education, employment, broad occupation group and bands of family income. Moreover, for each non-coresident child, both waves of CHARLS contain the amount of financial transfer in both directions. In this paper, I have used CHARLS to look at income composition, pensions, parental employment, and inter-generational co-residency and transfers.

The China Health and Nutrition Survey (CHNS) is the longest-running panel survey of Chinese households. There have been 9 waves of data, unevenly spaced between 1989 and 2011. There are a bit over 2000 adults aged 45 or above per wave. All waves of the CHNS contains some information on earnings and pensions, but it's not always consistently defined over time and it's less detailed than CHARLS. For example, pension income is reported as one figure at the household level for the first 5 waves and at the adult level afterwards, whereas CHARLS 2013 explicitly asks about 9 types of pensions. In this paper, the CHNS is mainly used to estimate income profiles for model calibration, because the CHARLS panel is too short and too recent for this purpose.

C.2 Historical pension system in China

Before 1986, enterprises were solely responsible for the pensions of their employees. At that time, state-owned enterprises (SOEs) and collective-owned enterprises (SOEs) accounted for most of the employment in urban areas.[1] Most enterprise pensions were Defined Benefit, with the mandatory retirement age set at 60 for men, and 50 for female workers generally and 55 for female ”cadres”. ”cadre” is an adminisitrative ranking of employees, and it usually corresponds to managerial or professional positions.

Since 1986, Social Insurance Agencies were established at the county and city level to take contributions from SOEs and their employees. Two thirds of workers in SOEs were covered by the end of 1991 and Collective-owned enterprises (COEs) became covered in 1992 (Salditt et al., 2007).[2] But the system remain decentralized, with diffferent programmes being piloted in different cities.

[1] According to the 2014 China Statistical Yearbook table 4-2, there are 128 million employerd persons in 1985, of which 90 million were at state-owned units, 33 were at collective-owned units and 4.5 million were self-employed.

[2] According to the 1996 China Statistical Yearbook table 4-4, in 1991 there were 153 million urban employees, of which 103 million were at State-owned units and 35 million at urban collective-owned units.

The 90s saw a series of State Council documents attempting to integrate the system. For example, the 1991 State Council Resolution on the Reform of the Pension System for Enterprise Workers recommended a three-tier pension, where the first tier would be jointly funded by the government, employer and individual and funding of the first tier would be pooled at the province level. However, some provinces adopted two-tier pensions which were funded by employer and employee and did not have a government-backed socially-pooled tier, because it's more popular with private and joint-ownership enterprises (Salditt et al., 2007). In 1995, the State Council Document no. 6 recommended two models of social pension and required the city and prefecture governments to select one or design a third model.[3] In practice, the design of the Enterprise Workers' pension (e.g. entitlement formula and indexation rules) continued to vary substantially across local governments.[4]

C.3 Parametric specifications and calibration

The terminal utility function V takes the following form, with parameters b, b_0 and CRRA parameter α.

$$V(A_{T+1}) = b\frac{(A_{T+1} + b_0)^{1-\alpha}}{1-\alpha} \tag{C.1}$$

The woman's pension income y_t is uncertain at first and deterministic after the first positive amount was received. If the woman received a pension last year $y_{t-1} > 0$, she will continue to receive a pension and the amount will be uprated in line with the age profile μ_t^y. If the woman did not receive a pension last year, the likelihood that she will start receiving a pension at age t is π_t. If she becomes eligible for pension at age t, the replacement ratio y_t/w_{t-1} is drawn randomly. In theory, the amount of pension payable depends on the given worker's complete work history before the retirement process. In the data, however, the replacement ratio at retirement is almost uncorrelated with

[3]There are 300+ of prefecture-level cities in China.
[4]Some examples are provided in (West, 1999).

the number of years of eligible work for the population of interest.[5] Given this weak correlation and the fact that post-retirement earnings do not interact with retirement pensions at all, I will assume that pension incomes are independent of past labour supply in the calibrated model.

$$\texttt{If}\quad y_{t-1} > 0,\ \texttt{then}\quad \log y_t = \log y_{t-1} + \mu_t^y - \mu_{t-1}^y \tag{C.2}$$

$$\texttt{If}\quad y_{t-1} = 0,\ \texttt{then} \begin{cases} \texttt{with probability } 1-\pi_t, & y_t = 0 \\ \texttt{with probability } \pi_t, & \frac{y_t}{w_{t-1}} \in N(\mu_r, \sigma_r^2) \end{cases} \tag{C.3}$$

The age profile of pension income μ_t^y is estimated from CHNS. We regress log female pension income on age and cohort.[6] μ_t is the predicted log pension amount for the 1944 birth cohort at age t. Also in CHNS we can track individuals over waves. We select waves that are two years apart, and infer the annual transition probabilities π_t from the observed between-wave transition probabilities using a minimum distance estimator. For the distribution of the initial replacement ratio, we look into CHARLS life-history survey for the reported net monthly earnings at the end of the last pre-retirement job and the first monthly pension amount. Among individuals who underwent the retirement process in the same or following year of ending their last pre-retirement job, the median replacement ratio is 1 and the distribution is very wide.[7] I will set $\mu_r = 1, \sigma_r = 0.2$ and approximate the normal distribution with 5 points.

These assumptions about pension dynamics are designed to capture two stylised facts. First, there is substantial uncertainty in pension entitlements because the exact

[5]For urban women born in the 1940s who reported their first pension amount and pre-retirement wage in CHARLS 2011, I regress their replacement ratio on a number of characteristics and found the coefficient on the years of eligible work to be essentially zero. The regression results are in table C.2 in the Appendix. The coefficient on the log pre-retirement wage is -0.03 and significant at 10% level only.

[6]Higher orders of age are found to be insignificant, so age is only included linearly.

[7]The 90th percentile is above 3. After restricting the range of the replacement ratio to be between 0.1 and 2, the standard deviation is still 0.34. This seems likely a result of reporting errors.

formula and eligibility rules are determined by local governments and change over time. Second, existing retirees' pension entitlements are usually uprated by the local government every year, based mainly on their last pension amount and their age. Thus, there is little uncertainty about future pension entitlements once an individual starts to receive it.

The structure of pension dynamics assumed here means there can be quite a lot of uncertainty in lifetime total pension wealth due to the uncertainty in when one will qualify and the initial pension amount, and that there is no uncertainty in pension income after the first receipt. In other words, all the uncertainty in pension wealth is revealed at the time of the first pension receipt. This may help explain why so many people stop working at the same time as when they start to receive pensions.

Woman's market wage w_t is assumed to follow a simple process in equation (C.4). This assumption is made so that only one state variable is needed to model the wage process. Future work can relax this. [8]

$$\log w_t = \log w_{t-1} + \mu_t^w - \mu_{t-1}^w + u_t, \quad u_t \in N(0, \sigma_w^2) \tag{C.4}$$

The age profile in female log wage μ_t^w is estimated from CHNS in the same way as that for log female pension.

Male income m_t is uncertain and exogenous. We observe quite a lot of zeros in the data (due to the man not working or dying), so we will model the transitions between zeros and positive values. Conditional on $m_{t-1} > 0, m_t > 0$, we assume the value of m_t is deterministic. This is again a simplifying assumption that may be relaxed in future

[8] In reality, wage shocks are probably not as persistent as assumed here. I have tried estimating the log wage process as a deterministic age trend plus an AR(1) persistent shock plus a transitory shock, and found the correlation coefficient in the persistent shock to be around 0.35.

work.

$$\text{If } m_{t-1} > 0, \text{ then} \begin{cases} \text{with probability } q_t, & m_t = 0 \\ \text{with probability } 1-q_t, & \log m_t = \log m_{t-1} + \mu_t^m - \mu_{t-1}^m \end{cases}$$

$$\text{If } m_{t-1} = 0, \text{ then} \begin{cases} \text{with probability } p_t, & m_t = 0 \\ \text{with probability } 1-p_t & \log m_t = \mu_t^m + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_m^2) \end{cases}$$

The male income profile μ_t^m is estimated from CHNS in the same way as μ_t^y. Likewise, the annual transition probabilities p_t, q_t are estimated from the observed between-wave transition probabilities in CHNS using minimum distance estimation.

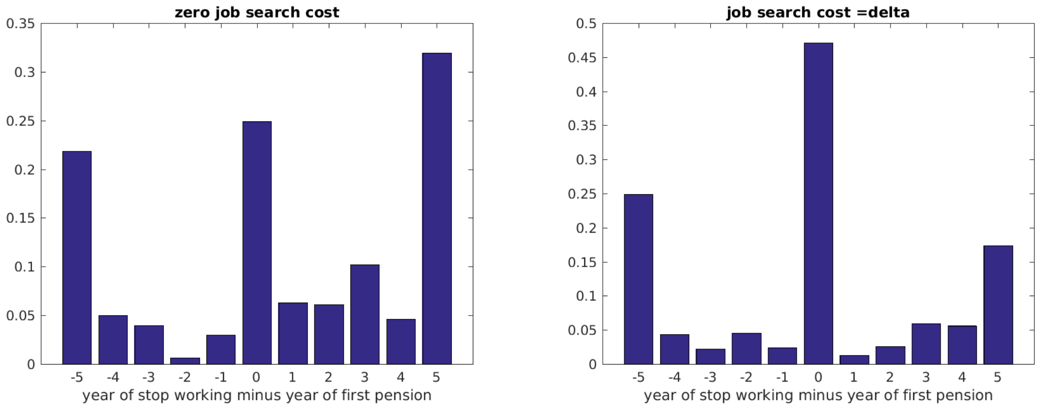
In summary, the model has four continuous state variables A_t, y_t, m_t, w_t and one binary state variable P_{t-1}.

In the baseline case, I set the following parameters.

$$\alpha = 2$$
$$\beta = 0.97, R = 1/\beta$$
$$\zeta = 0.5$$
$$b = 10, b_0 = 1$$
$$\sigma_w = 0.05, \sigma_r = 0.2, \sigma_m = 0.7$$
$$\gamma_y = 0.1, \gamma_w = 0.1, \gamma_m = 0.1$$

The baseline choice of α comes from Chen (2014). The baseline assumption of δ_t is exponential in t and it increases about three-fold between 45 and 75 (Figure C.4 in the appendix.) I have tried alternative assumptions of δ_t and plotted the resulting simulated employment rates in Figure C.4 as well. $\zeta = 0.5$ means the disutility of moving back to work increases the disutility of working by half. I have tried alternative values of ζ in the next section.

Figure C.1: Year when stop working minus year when first received pension, by different assumptions of search cost



Note: 5 means 5 or more, -5 means -5 or below. Simulated individuals who never received a pension are not in the graphs. The left graph assumed $\zeta = 0$, the right assumes $\zeta = 1$. The baseline assumption is $\zeta = 0.5$.

For simulation, we have one simulated household for each point of the initial state variables. The distribution of the initial values of the state variables are mostly based on the CHNS data. All women are assumed to have zero pension at 44. 87% were in work at age 44 (the rate observed in CHNS). 20% of them have zero male income (as in CHNS). Because the CHNS does not contain much information about assets (other than durable goods and housing), we look at the asset distribution in CHARLS 2013, and assume the initial 1989 distribution has the same density at zero and the log level is lower by 2.4. That is 0.1 increase in log level per year, similar to the annual increase in mean incomes observed in CHNS.

C.4 Additional tables and figures on Chinese households

Table C.1: Summary statistics on household incomes

		Nuclear households		
income	percent positive	mean inc. 0	mean exc. 0	median exc. 0
pensions	77.3	29,068	37,612	30,480
transfer from children	59.1	4,876	8,245	4,500
other private transfer	27.6	2,162	7,823	1,200
earnings	25.9	8,820	34,039	31,200
public transfers	17.4	650	3,732	1,012
self emp	5.3	1,553	29,233	20,000
total income		47,129	50,920	41,760
		Non-nuclear households		
income	percent positive	mean inc. 0	mean exc. 0	median exc. 0
pensions	61.4	18,181	29,626	24,000
other HH members	50.5	20,977	41,553	24,000
transfer from children	36.6	2,538	6,931	2,500
earnings	34.4	11,916	34,651	30,000
other private transfer	27.1	2,380	8,782	2,000
public transfers	20.3	768	3,783	1,000
self emp	11.5	2,724	23,614	24,000
total income		59,485	68,431	51,600

Note: the first panel is based on 775 households in CHARLS2013 where the female is aged 45-75 and has urban hukou and the household contains no person other than her spouse or dependent child (below 16), and where the household has completed the individual income and household income modules. The second panel is based on all other households in CHARLS2013 where the female is 45-75 and has urban hukou. There are 1518 such households who have completed the individual income and household income modules. The self-employment income includes incomes from agriculture, livestock and family business. For the first column, missing values are treated as zeros. The 2nd column, the mean including zero, is the product of the first and third columns. The "total" of "mean inc.0" is simply the sum of all the income types in the same column. For "total income" in the last two columns, I add up incomes at the household level, drop those with some missing earnings, pensions or public transfers, and then find the mean and median of the resulting non-zero distribution.

Table C.2: OLS of the replacement ratio

	(1)	(2)	(3)	(4)
VARIABLES	female 1940-49	male 1940-49	female 1950-59	male 1950-59
log pre-retire wage	-0.0287*	-0.0648***	-0.0898***	-0.161***
	(0.0154)	(0.0169)	(0.0207)	(0.0379)
birth year	0.00709	0.0114***	0.000120	0.00516
	(0.00490)	(0.00405)	(0.00539)	(0.0102)
age at retirement	0.00329	0.00519	0.000669	0.00432
	(0.00390)	(0.00339)	(0.00463)	(0.00663)
years of elegible work	-0.000571	0.00166	0.00535**	0.00853
	(0.00204)	(0.00194)	(0.00247)	(0.00528)
GongLin_ms	-0.00687	0.129**	0.0281	0.0250
	(0.158)	(0.0596)	(0.129)	(0.0793)
cadre	0.0208	0.0324	0.1000**	0.127**
	(0.0330)	(0.0238)	(0.0435)	(0.0494)
Constant	-12.78	-21.03***	1.155	-8.510
	(9.547)	(7.878)	(10.52)	(19.95)
Observations	152	247	177	82
R-squared	0.032	0.087	0.142	0.239

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Note: the replacement ratio is the first pension amount divided by the pre-retirement wage. Extreme observations with the replacement ratio below 0.5 or above 1.5 are excluded. The sample is urban retirees in the specified birth cohorts in CHARLS 2011.

Table C.3: Summary statistics on household saving rate by total number of children alive

number of children	observations	25the percentel	median	75th percentile
0	27	0.355	0.373	0.373
1	419	-0.361	0.303	0.600
2	373	-0.320	0.236	0.515
3	307	-0.424	0.280	0.572

Note: the sample is households in CHARLS 2011 where the woman is aged between 45 and 75 and has urban hukou. Observations with income or expenditure more extreme than the first and 99th percentiles are treated as missing values. The household saving rate is defined as one minus household expenditure last year divided by household income (including all earnings, pensions and public transfers).

Table C.4: LPM of men's employment on child characteristics (the number of children who...)

	(1)	(2)	(3)	(4)	(5)	(6)
are male	0.00276	-0.0134	-0.0476	-0.0409	-0.0355	-0.0310
	(0.0153)	(0.0211)	(0.0287)	(0.0287)	(0.0295)	(0.0295)
are female	0.0246	0.00949	-0.0161	-0.0104	-0.00511	-0.00292
	(0.0148)	(0.0210)	(0.0272)	(0.0272)	(0.0281)	(0.0282)
if wife is in work	0.211***	0.212***	0.207***	0.215***	0.216***	0.215***
	(0.0224)	(0.0226)	(0.0227)	(0.0227)	(0.0227)	(0.0227)
finished lower secondary school		0.0170	0.0145	0.0111	0.00824	-0.00809
		(0.0196)	(0.0198)	(0.0197)	(0.0199)	(0.0452)
finished upper secondary school		-0.00351	-0.0108	-0.0255	-0.0233	0.0358
		(0.0151)	(0.0153)	(0.0162)	(0.0166)	(0.0430)
are married			-0.0148	-0.00932	-0.0247	-0.0232
			(0.0185)	(0.0187)	(0.0244)	(0.0244)
work			0.0539**	0.0385	0.0391	0.0349
			(0.0199)	(0.0203)	(0.0203)	(0.0204)
work as professionals				0.0539**	0.0539**	0.0536**
				(0.0195)	(0.0195)	(0.0195)
work as managers				0.0671*	0.0664*	0.0697**
				(0.0264)	(0.0265)	(0.0265)
has income above 100k				0.120**	0.117**	0.119**
				(0.0407)	(0.0408)	(0.0408)
has income above 50k				-0.0431*	-0.0425*	-0.0402
				(0.0212)	(0.0212)	(0.0212)
number of grandkids					0.00376	0.00422
					(0.0138)	(0.0140)
whether have grandkids					0.0371	0.0835
					(0.0330)	(0.0474)
interaction term, lower 2nd						0.0113
						(0.0464)
interaction term, upper 2nd						-0.0722
						(0.0454)
Observations	1728	1728	1728	1728	1728	1728
Adjusted R^2	0.249	0.250	0.252	0.259	0.259	0.260

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: sample restricted to 50-64 year old married men with urban hukou. Controls include for each spouse, their age, age dummies (in 5-year bands), education dummies and hukou at birth; and whether the wife is in work. Source: CHARLS

Table C.5: OLS whether receive any transfers from children on child characteristics

	(1)	(2)	(3)	(4)
are co-resident	0.0279	-0.0305	-0.0442	-0.0510
	(0.0254)	(0.0293)	(0.0306)	(0.0325)
are not co-resident	0.0490**	-0.0352	-0.0511	-0.0641*
	(0.0171)	(0.0276)	(0.0291)	(0.0311)
are male	-0.0140	-0.0200	-0.0203	-0.0267
	(0.0196)	(0.0195)	(0.0195)	(0.0199)
are married		0.0965***	0.0920***	0.0752*
		(0.0249)	(0.0250)	(0.0300)
have finished middle school			0.0281	0.0150
			(0.0177)	(0.0185)
have bachelor degrees			0.0174	0.00132
			(0.0244)	(0.0273)
work				0.0302
				(0.0243)
work as managers				-0.0121
				(0.0297)
work as professionals				0.000657
				(0.0234)
has income above 20k				0.0504**
				(0.0157)
has income above 50k				-0.000571
				(0.0225)
Observations	818	818	818	818
Adjusted R^2	0.079	0.095	0.097	0.109

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C.6: OLS log amount of transfers from children on child characteristics

	(1)	(2)	(3)	(4)
are co-resident	0.0700	0.148	0.0585	0.0189
	(0.0985)	(0.121)	(0.122)	(0.130)
are not co-resident	0.145*	0.248*	0.124	0.0756
	(0.0620)	(0.111)	(0.115)	(0.124)
are male	-0.0382	-0.0355	-0.0315	-0.0614
	(0.0717)	(0.0717)	(0.0697)	(0.0703)
are married		-0.114	-0.120	-0.134
		(0.102)	(0.0993)	(0.113)
have finished middle school			0.158*	0.0857
			(0.0631)	(0.0656)
have bachelor degrees			0.442***	0.245*
			(0.0883)	(0.0994)
work				0.141
				(0.0901)
work as managers				0.146
				(0.105)
work as professionals				0.146
				(0.0858)
has income above 20k				0.0657
				(0.0551)
has income above 50k				0.177*
				(0.0763)
Observations	615	615	615	615
Adjusted R^2	0.072	0.073	0.124	0.151

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C.7: Monthly wage by age band, urban hukou

age band	number of observations	% in work	wage observations	mean wage	median
20-24	228	0.62	110	2,250	2,000
25-29	374	0.83	279	2,695	2,000
30-34	450	0.85	338	2,796	2,100
35-39	524	0.85	367	2,825	2,000
40-44	642	0.85	443	2,734	2,000
45-49	763	0.74	475	2,707	2,000
50-54	689	0.52	304	2,513	2,000
55-59	888	0.35	255	2,663	2,000
60-64	659	0.12	61	2,154	2,000

Note: the sample is individuals with urban hukou in 2011 CHNS. Zero wage observations and exteremly large (above the 99th percentile) wage observations are excluded for the last three columns.

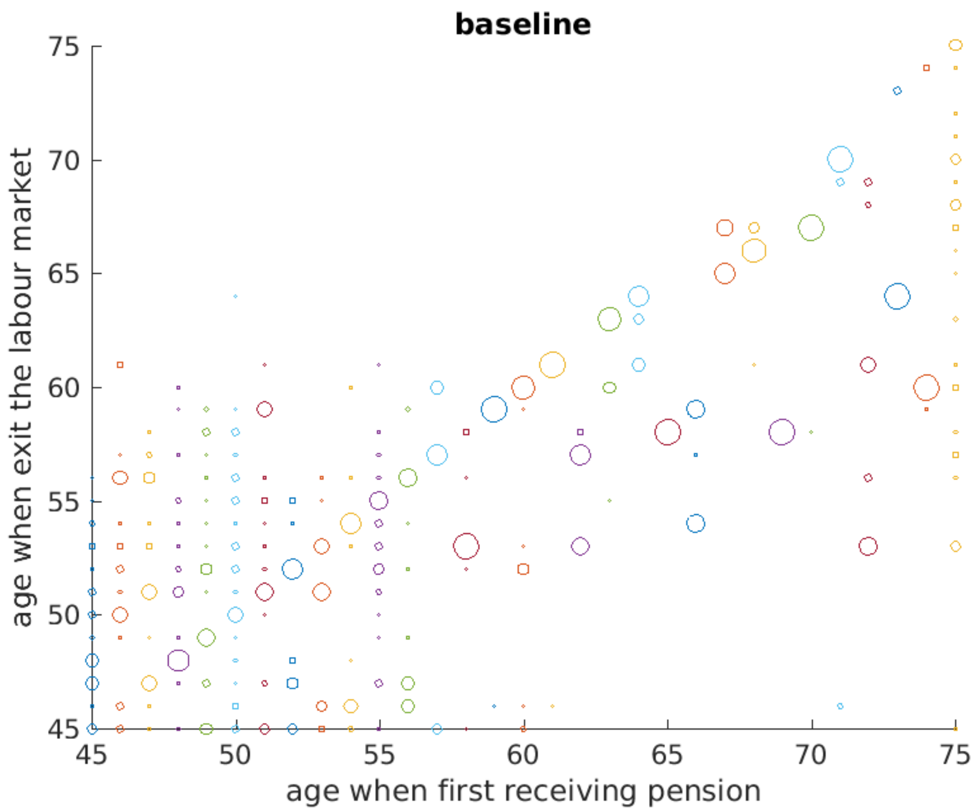
Table C.8: Healthcare cost by age band, urban hukou

out-of-pocket cost for out-patient treatments in the last month							
age band	% w cost	num observed	mean cost	25th pct	50th pct	75th pct	% w high cost
45-49	0.191	91	775	100	200	500	0.039
50-54	0.200	91	686	100	300	900	0.032
55-59	0.219	127	827	100	400	1,000	0.041
60-64	0.236	130	1,462	90	200	800	0.025
65-69	0.262	100	444	100	200	525	0.016
70-74	0.288	76	1,826	100	300	650	0.013
75+	0.279	84	967	115	353	1,000	0.017
45+	0.234	699	990	100	270	700	0.027

out-of-pocket cost for hospitalization in the last year							
age band	% w cost	num observed	mean cost	25th pct	50th pct	75th pct	% w high cost
45-49	0.084	46	6,278	1,000	2,850	6,000	0.022
50-54	0.105	58	7,233	1,800	3,000	6,700	0.012
55-59	0.133	84	7,136	1,200	3,000	6,000	0.013
60-64	0.173	105	7,003	1,500	3,030	6,000	0.013
65-69	0.155	74	8,258	1,150	3,000	6,000	0.012
70-74	0.207	69	6,776	2,000	3,000	6,000	0.023
75+	0.246	88	9,056	1,550	3,000	7,750	0.020
45+	0.151	524	7,478	1,455	3,000	6,000	0.016

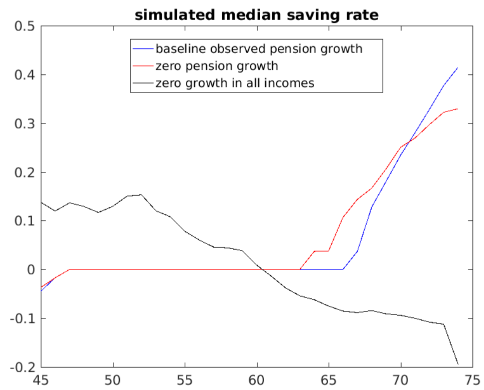
Note: the sample is individuals with urban hukou in 2013 CHARLS. "num observed" is the number of observations with positive costs. Observations are weighted by the individual weight in CHARLS 2013. The means and percentiles are based on the non-zero distribution. For the last column in the upper panel, "high cost" is defined as whether the cost of medical treatment last month exceeds one twelveth of last year's household income (the sum of both spouses' earnings, pensions and public transfers). For the last column of the lower panel, "high cost" is defined as whether the cost of hospitialization last year exceeds last year's household income.

Figure C.2: Age when qualifying for the retirement and age when leaving the labour market



Note: The size of the bubbles are proportional to the share of women who left the labour market at the specified age conditional on receving pension at the specified age. A person who is still in work at 74 is labelled as leaving the work force at 75, and a person who has never received any pension from 45 to 74 is labelled as 75 on the x axis.

Figure C.3: Median household saving rate under counterfactual income profiles



Note: in each scenario, growth expectations are assumed to be consistent with the mean growth rate of incomes.

Figure C.4: Age profiles with alternative paths of disutility of work

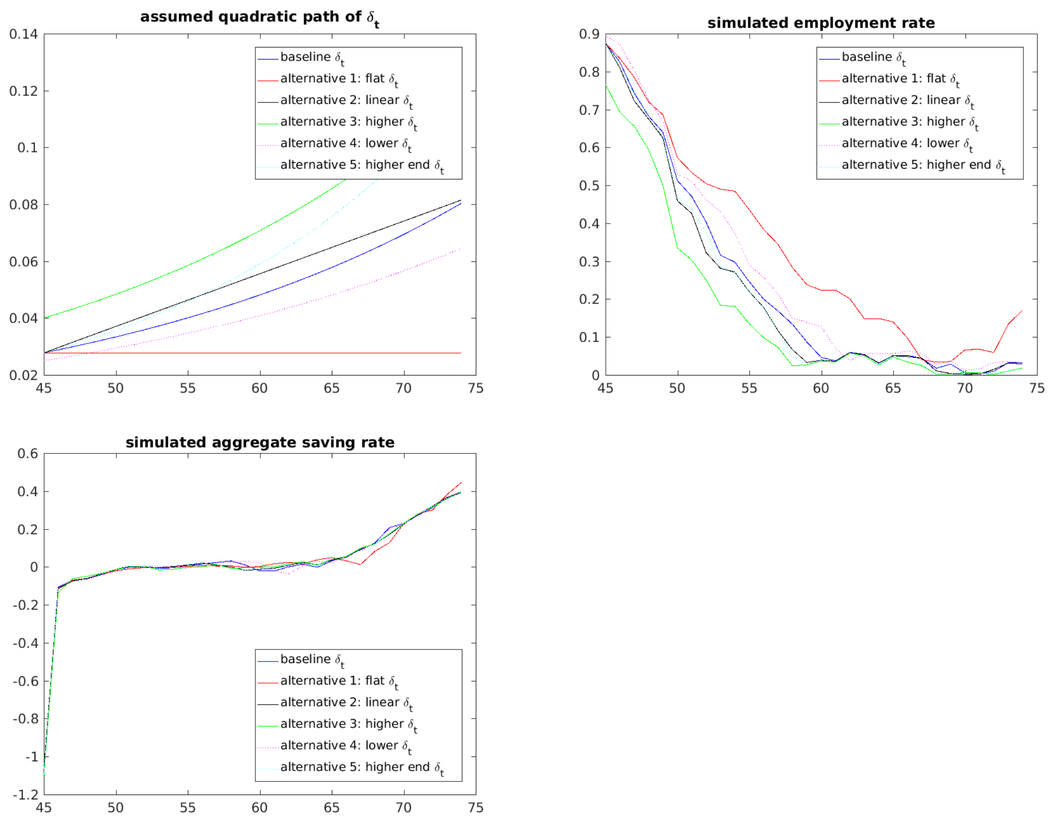
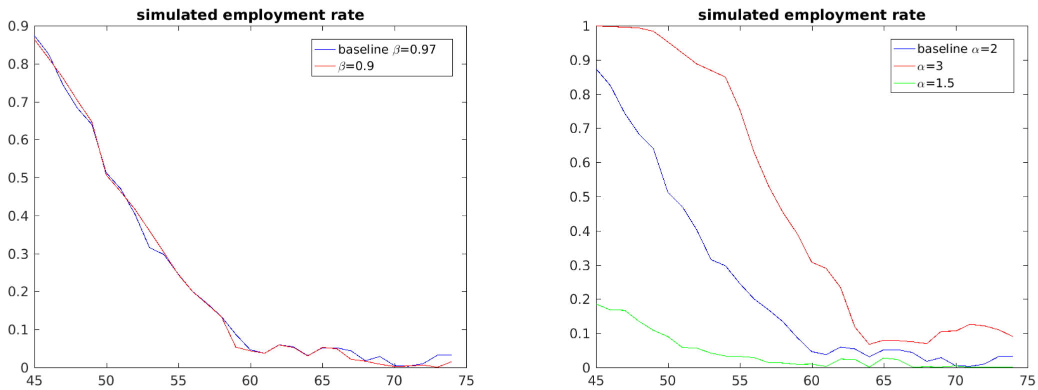


Figure C.5: Employment rate under alternative preference paramters



Note: consumption at age t generates utility $c_t^{(1-\alpha)}/(1-\alpha)$ in the same period. For alternative α scenarios, δ_t are set such that $(\delta_t * (\alpha-1))^{-1/\alpha}$ is the same as in the baseline.

Bibliography

Acemoglu, Daron, "Why do new technologies complement skills? Directed technical change and wage inequality," *Quarterly journal of economics*, 1998, *113* (4), 1055–1089.

_ , "Directed Technical Change," *Review of Economic Studies*, 2002, *69* (4), 781–809.

_ , "Patterns of skill premia," *The Review of Economic Studies*, 2003, *70* (2), 199–230.

_ , "Equilibrium Bias of Technology," *Econometrica*, 2007, *75* (5), 1371–1409.

_ **and David Autor**, "Skills, tasks and technologies: Implications for employment and earnings," *Handbook of labor economics*, 2011, *4*, 1043–1171.

_ **and Fabrizio Zilibotti**, "Productivity Differences," *The Quarterly Journal of Economics*, 2001, *116* (2), 563–606.

_ **and Pascual Restrepo**, "Tasks, Automation, and the Rise in US Wage Inequality," Technical Report, National Bureau of Economic Research 2021.

Adermon, Adrian and Magnus Gustavsson, "Job polarization and task-biased technological change: Evidence from Sweden, 1975–2005," *The Scandinavian Journal of Economics*, 2015, *117* (3), 878–917.

Aghion, Philippe, Céline Antonin, Simon Bunel, and Xavier Jaravel, "What are the labor and product market effects of automation? New Evidence from France," Technical Report 2020.

Akerman, Anders, Ingvil Gaarder, and Magne Mogstad, "The skill complementarity of broadband internet," *The Quarterly Journal of Economics*, 2015, *130* (4), 1781–1824.

Antoncyzk, Dirk, Thomas DeLeire, and Bernd Fitzenberger, "Polarization and Rising Wage Inequality: Comparing the US and Germany," Technical Report, ZEW 2010.

Appleton, Simon, John Knight, Lina Song, and Qingjie Xia, "Labor retrenchment in China: Determinants and consequences," *China Economic Review*, 2002, *13* (2), 252–275.

Autor, David, "The Labor Market Impacts of Technological Change: From Unbridled Enthusiasm to Qualified Optimism to Vast Uncertainty," Technical Report, National Bureau of Economic Research 2022.

Autor, David H., "The 'Task Approach' to Labor Markets: An Overview," Technical Report 13-04, MIT Department of Economics Working Paper 2013.

Autor, David. H and David Dorn, "The growth of low-skill service jobs and the polarization of the US labor market," *American economic review*, 2013, *103* (5), 1553–97.

Autor, David H. and Micheal J. Handel, "Putting Tasks to the Test: Human Capital, Job Tasks, and Wages," *Journal of Labor Economics*, 2013, *31* (S1), S59–S96.

Autor, David H, Frank Levy, and Richard J Murnane, "The skill content of recent technological change: An empirical exploration," *The Quarterly journal of economics*, 2003, *118* (4), 1279–1333.

Barany, Zsofia L. and Christian Siegel, "Job Polarization and Structural Change," *American Economic Journal: Macroeconomics*, January 2018, *10* (1), 57–89.

Barro, Robert and Jong-Wha Lee, "A New Data Set of Educational Attainment in the World, 1950-2010," *Journal of Development Economics*, 2013, *104*, 184–198.

Battisti, Michele, Christian Dustmann, and Uta Schonberg, "Technological and Organizational Change and the Careers of Workers," Technical Report, Working Paper 2017.

Beaudry, Paul and David A Green, "Wages and employment in the United States and Germany: What explains the differences?," *The American Economic Review*, 2003, *93* (3), 573–602.

_ and _ , "Changes in US Wages, 19762000: Ongoing Skill Bias or Major Technological Change?," *Journal of Labor Economics*, 2005, *23* (3), 609–648.

_ , _ , and Benjamin M Sand, "The great reversal in the demand for skill and cognitive tasks," *Journal of Labor Economics*, 2016, *34* (S1), 199–247.

_ , Mark Doms, and Ethan Lewis, "Endogenous skill bias in technology adoption: City-level evidence from the IT revolution," Technical Report, National Bureau of Economic Research 2006.

_ , _ , and _ , "Should the Personal Computer be Considered a Technological Revolution? Evidence from US Metropolitan Areas," *Journal of Political Economy*, 2010, *118* (5), 988–1036.

Becker, Gary S. and Kevin M. Murphy, "The Division of Labor, Coordination Costs, and Knowledge," *The Quarterly Journal of Economics*, 1992, *107* (4), 1137–1160.

Bloom, Nicholas, Raffaella Sadun, and John Van Reenan, "Americans do IT better: US multinationals and the productivity miracle," *American Economic Review*, 2012, *102* (1), 167–201.

_ **, Renata Lemos, Raffaella Sadun, Daniela Scur, and John Van Reenan**, "The New Empirical Economics of Management," Technical Report, National Bureau of Economic Research 2014.

Blundell, Richard, Amanda Gosling, Hidehiko Ichimura, and Costas Meghir, "Changes in the distribution of male and female wages accounting for employment composition using bounds," *Econometrica*, 2007, *75* (2), 323–363.

_ **, David A Green, and Wenchao Jin**, "The U.K. as a Technological Follower: Higher Education Expansion and the College Wage Premium," *The Review of Economic Studies*, 07 2021, *0* (0), 1–39. rdab034.

Böhm, Michael J, "The price of polarization: Estimating task prices under routine-biased technical change," *Quantitative Economics*, 2020, *11* (2), 761–799.

Borghans, Lex and Bas ter Weel, "The Divison of Labor, Work Organization, and Technological Change," *Economic Journal*, 2006, *116* (509), F45–F72.

_ **and** _ , "The Diffusion of Computers and the Distribution of Wages," *European Economic Review*, 2007, *51*, 715–748.

_ **and** _ , "Understanding the Technology of Computer Technology Diffusion: Explaining Computer Adoption Patterns and Implicaitons for the Wage Structure," *Journal of Income Distribution*, 2008, *17* (3), 37–70.

Breemersch, Koen, Joze P. Damijan, and Jozef Konings, "Labour Market Polarization in Advanced Countries," Technical Report 197, OECD 2017.

Bresnahan, Timothy F, Erik Brynjolfsson, and Lorin M Hitt, “Information technology, workplace organization and the demand for skilled labor: Firm-level evidence,” *Quarterly Journal of Economics*, 02 2002, *117* (1), 339–376.

BÃ¶hm, Michael J., Hans-Martin von Gaudecker, and Felix Schran, “Occupation, Growth, Skill Prices, and Wage Inequality,” *CESifo Working Paper*, 2019, (7877).

Cai, Fang, John Giles, and Xin Meng, “How well do children insure parents against low retirement income? An analysis using survey data from urban China,” *Journal of public Economics*, 2006, *90* (12), 2229–2255.

Card, David and John E DiNardo, “Skill biased technological change and rising wage inequality: some problems and puzzles,” Technical Report, National Bureau of Economic Research 2002.

_ **and Thomas Lemieux**, “Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis,” *The Quarterly Journal of Economics*, 2001, *116* (2), 705–746.

Carneiro, Pedro Manuel, Kai Liu, and Kjell G Salvanes, “The supply of skill and endogenous technical change: evidence from a college expansion reform,” *NHH Dept. of Economics Discussion Paper*, 2018, (16).

Caroli, Eve and John Van Reenen, “Skill-biased organizational change? Evidence from a panel of British and French establishments,” *Quarterly journal of economics*, 2001, pp. 1449–1492.

Carpentier, Vincent, “Historical Statistics on the Funding and Development of the UK University System, 1920-2002,” 2004.

_ , "Funding in Higher Education and Economic Growth in France and the United Kingdom, 1921-2003," *Higher Education Management and Policy*, 2006, *18* (3), 1–26.

Cavaglia, Chiara and Ben Etheridge, "Job polarization and the declining quality of knowledge workers: Evidence from the UK and Germany," *Labour Economics*, 2020, *66*, 101884.

Chen, Uniko, "The Impact of the Expansion in Post-Secondary Education on the University Wage Premium in Taiwan," Technical Report, University of Michigan 2013.

Chen, Yi, "Never Too Old to Save & Explaining the High Saving Rates of the Chinese Elderly," Technical Report, Working paper 2014.

Chevalier, Arnaud, Colm Harmon, Ian Walker, and Yu Zhu, "Does Education Raise Productivity, or Just Reflect it?*," *The Economic Journal*, 2004, *114* (499), F499–F517.

Choi, Kang-Shik and Jinook Jeong, "Technological Change and Wage Premium in as Small Open Economy: The Case of Korea," *Applied Economics*, 2005, *37* (1), 119–131.

Choi, Seongsoo, "When Everyone Goes to College:The Causal Effect of College Expansion on Earnings," *Social Science Research*, 2015, *50*, 229–245.

Crivallero, Elena, "The College Wage Premium Over Time: Trends in Europe in the Last 15 Years," *Research in Labor Economics*, 2016, *43*, 287–328.

Devereux, Paul J and Wen Fan, "Earnings returns to the British education expansion," *Economics of Education Review*, 2011, *30* (6), 1153–1166.

Doms, Mark, Timothy Dunne, and Kenneth R. Troske, "Workers, Wages, and Technology," *Quarterly Journal of Economics*, 1997, pp. 253–290.

Draca, Mirko, Raffaella Sadun, and John Van Reenen, "Productivity and ICTs: A review of the evidence," *The Oxford handbook of information and communication technologies*, 2007.

Dustmann, Christian and Albrecht Glitz, "How do industries and firms respond to changes in local labor supply?," *Journal of Labor Economics*, 2015, *33* (3 Part 1), 711–750.

_, Johannes Ludsteck, and Uta SchÃ¶nberg, "Revisiting the German Wage Structure*," *The Quarterly Journal of Economics*, 05 2009, *124* (2), 843–881.

_, Tommaso Frattini, and Ian P Preston, "The effect of immigration along the distribution of wages," *The Review of Economic Studies*, 2013, *80* (1), 145–173.

Feng, Jin and Xiaohan Zhang, "Retirement and Grandchild Care in Urban China," *Feminist Economics*, 2018, *24* (2), 240–264.

Fernández-Macías, Enrique and John Hurley, "Routine-biased technical change and job polarization in Europe," *Socio-Economic Review*, 2017, *15* (3), 563–585.

Firpo, Sergio, Nicole M Fortin, and Thomas Lemieux, "Occupational tasks and changes in the wage structure," Technical Report 2011.

Giles, John, Albert Park, and Fang Cai, "How has economic restructuring affected China's urban workers?," *The China Quarterly*, 2006, *185*, 61–95.

_, Xiaoyan Lei, Yafeng Wang, and Yaohui Zhao, "One Country, Two Systems: Evidence on Retirement Patterns in China," *World Bank, Washington, DC, and Peking University, Beijing*, 2015.

Goldin, Claudia and Lawrence F. Katz, *The Race Between Education and Technology*, Belknap Harvard University Press, 2008.

Goos, Maarten, Alan Manning, and Anna Salomons, "Explaining Job Polarization: Routine-Biased Technological Change and Offshoring," *American Economic Review*, August 2014, *104* (8), 2509–26.

_ **and** _ , "Lousy and Lovely Jobs: The Rising Polarization of Work in Britain," *The Review of Economics and Statistics*, 02 2007, *89* (1), 118–133.

Gottschalk, Peter, David A. Green, and Benjamin M. Sand, "Taking Selection to Task: Bounds on Occupational Task Prices for the US: 1984-2013," Technical Report, University of British Columbia 2014.

Graetz, Georg and Guy Michaels, "Robots at Work," *The Review of Economics and Statistics*, 12 2018, *100* (5), 753–768.

Green, David A. and Benjamin M. Sand, "Has the Canadian labour market polarized?," *Canadian Journal of Economics*, 2015, *48* (2), 612–646.

Green, Francis and Golo Henseke, "The changing graduate labour market: analysis using a new indicator of graduate jobs," *IZA Journal of Labor Policy*, 2016, *5* (1), 1–25.

_ **and** _ , "Europe's evolving graduate labour markets: supply, demand, underemployment and pay," *Journal for Labour Market Research*, 2021, *55* (1), 1–13.

_ **and Yu Zhu**, "Overqualification, job dissatisfaction, and increasing dispersion in the returns to graduate education," *Oxford Economic Papers*, 2010, *62* (4), 740–763.

Griliches, Zvi, "Research Costs and Social Returns: Hybrid Corn and Related Innovations," *Journal of Political Economy*, 1958, *66*, 419–431.

Hardy, Wojciech, Roma Keister, and Piotr Lewandowski, "Educational upgrading, structural change and the task composition of jobs in Europe," *Economics of Transition*, 2018, *26* (2), 201–231.

Hershbein, Brad and Lisa B. Kahn, "Do Recessions Accelerate Routine-Biased Technological Change? Evidence from Vacancy Postings," *American Economic Review*, July 2018, *108* (7), 1737–72.

Humlum, Anders, "Robot adoption and labor market dynamics," Technical Report, Princeton University 2019.

Johnson, Matthew and Michael P. Keane, "A Dynamic Equilibrium Model of the US Wage Structure, 1968?1996," *Journal of Labor Economics*, 2013, *31* (1), 1–49.

Kampelmann, Stephan and François Rycx, "Task-Biased Changes of Employment and Remuneration: The Case of Occupations," 2011.

Katz, Lawrence F and Kevin M Murphy, "Changes in Relative Wages, 1963-1987: Supply and Demand Factors," *The Quarterly Journal of Economics*, 1992, *107* (1), 35–78.

Kiley, Michael T., "The Supply of Skilled Labour and Skill-Biased Technological Progress," *The Economic Journal*, 1999, *109* (458), 708–724.

Lee, David S, "Training, wages, and sample selection: Estimating sharp bounds on treatment effects," *The Review of Economic Studies*, 2009, *76* (3), 1071–1102.

Lee, Jong-Wha and Hanol Lee, "Human capital in the long run," *Journal of Development Economics*, 2016, *122*, 147–169.

Lemieux, Thomas, "Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill?," *The American Economic Review*, 2006, pp. 461–498.

Lewis, Ethan, "Immigration, skill mix, and capital skill complementarity," *The Quarterly Journal of Economics*, 2011, *126* (2), 1029–1069.

Lindbeck, Assar and Denis J. Snower, "Reorganization of Firms and Labor Market Inequality," *The American Economic Review*, 1996, *86* (2), 315–21.

Machin, Stephen and Anna Vignoles, *Education Policy in the UK. CEE DP 57.*, ERIC, 2006.

_ **and John Van Reenen**, "Technology and changes in skill structure: evidence from seven OECD countries," *Quarterly journal of economics*, 1998, pp. 1215–1244.

_ **and Sandra McNally**, "Tertiary education systems and labour markets," *Tertiary Review*, 2007, pp. 1–57.

Manski, Charles F, "The selection problem," in "Advances in Econometrics, Sixth World Congress," Vol. 1 1994, pp. 143–70.

Maurer-Fazio, Margaret, Rachel Connelly, Lan Chen, and Lixin Tang, "Childcare, eldercare, and labor force participation of married women in urban China, 1982–2000," *Journal of Human Resources*, 2011, *46* (2), 261–294.

McIntosh, Steven, "Further Analysis of the Returns to Academic and Vocational Qualifications*," *Oxford bulletin of economics and statistics*, 2006, *68* (2), 225–251.

Michaels, Guy, Ashwini Natraj, and John Van Reenen, "Has ICT Polarized Skill Demand? Evidence from Eleven Countries over Twenty-Five Years," *The Review of Economics and Statistics*, 03 2014, *96* (1), 60–77.

Milgrom, Paul and John Roberts, "The economics of modern manufacturing: Technology, strategy, and organization," *The American Economic Review*, 1990, pp. 511–528.

Murphy, Emily C and Daniel Oesch, "Is employment polarisation inevitable? Occupational change in Ireland and Switzerland, 1970–2010," *Work, employment and society*, 2018, *32* (6), 1099–1117.

Naticchioni, Paolo, Giuseppe Ragusa, and Riccardo Massari, "Unconditional and conditional wage polarization in Europe," 2014.

OECD, "Education at a Glance: OECD Indicators," Technical Report, OECD 2012.

Oesch, Daniel and Giorgio Piccitto, "The polarization myth: Occupational upgrading in Germany, Spain, Sweden, and the UK, 1992–2015," *Work and Occupations*, 2019, *46* (4), 441–469.

OLeary, Nigel C and Peter J Sloane, "The return to a university education in Great Britain," *National Institute Economic Review*, 2005, *193* (1), 75–89.

Radner, Roy, "The Organization of Decentralized Information Processing," *Econometrica*, 1993, *62*, 1109–1146.

Rosen, Sherwin, "Substitution and Division of Labor," *Economica*, 1978, *45* (179), 235–250.

Rust, John and Christopher Phelan, "How social security and medicare affect retirement behavior in a world of incomplete markets," *Econometrica: Journal of the Econometric Society*, 1997, pp. 781–831.

Salditt, Felix, Peter Whiteford, and Willem Adema, "Pension Reform in China," 2007.

Salvatori, Andrea, "The anatomy of job polarisation in the UK," *Journal for labour market research*, 2018, *52* (1), 1–15.

Shu, Lei, "The effect of the New Rural Social Pension Insurance program on the retirement and labor supply decision in China," *The Journal of the Economics of Ageing*, 2018, *12*, 135–150.

Sin, Yvonne, "Pension liabilities and reform options for old age insurance," *World Bank working paper*, 2005, *1.*

Traiberman, Sharon, "Occupations and Import Competition: Evidence from Denmark," *American Economic Review*, December 2019, *109* (12), 4260–4301.

Walker, Ian and Yu Zhu, "The College Wage Premium, Overeducation, and the Expansion of Higher Education in the UK," *IZA Discussion Paper*, 2005, *No. 1627.*

_ **and** _ , "The College Wage Premium and the Expansion of Higher Education in the UK," *The Scandinavian Journal of Economics*, 2008, *110* (4), 695–709.

West, Loraine A, "Pension reform in China: Preparing for the future," *The Journal of Development Studies*, 1999, *35* (3), 153–183.

Yu, Haiyue, Jin Cao, and Shulong Kang, "Who Cares: Deciphering China's Female Employment Paradox," *BOFIT Discussion Paper*, 2021, (7).