# Essays on Human Capital Accumulation over the Life-Cycle

**David Goll**

A thesis presented for the degree of
Doctor of Philosophy

Department of Economics
UCL

March 14, 2022

# Declaration

I, David Goll, 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 thesis. Chapters 1, 3 and 5 are sole-authored. Chapter 2 is based on joint work with Jack Britton and Monica Costa Dias. Chapter 4 is based on joint work with Richard Blundell, Monica Costa Dias and Costas Meghir.

# Abstract

This thesis contains three papers, each focusing on a different aspect of human capital accumulation after the conclusion of compulsory education.

In the first paper, I focus on higher education. I estimate a novel two-sided matching model, where heterogenous students match to courses with limited capacity. Using the estimated model, I simulate a wide set of policies aimed at boosting intergenerational income mobility. I conclude that achieving substantial improvements in mobility through the higher education system is likely to require reforms to admissions policy, such as percent plans.

In the second and third paper, I focus on human capital accumulation after entering the labour market. There are two alternative models for how individuals accumulate human capital while working: learning-by-doing, where human capital is a by-product of labour, and Ben-Porath, where agents must divide their time between learning and earning.

In the second paper, I estimate models of both types. I find that neither can satisfactorily replicate the qualitative features of women's wages and time use in early working life. I then extend the model to incorporate both mechanisms simultaneously and show that this greatly improves model fit. I consider reforms to the benefit system and find that in my preferred model lowering the withdrawal rate for benefits reduces wage growth when women are young but increases wages for older women.

In the third paper, I use a model incorporating both learning-by-doing and Ben-Porath style human capital accumulation to investigate the role of on-the-job training in reducing the gender wage gap. Women often reduce working hours following childbirth and, as a result, experience slower wage growth than men. I find that training is potentially important in compensating for this loss in work experience, especially for women who left full-time education after completing high school.

# Impact assessment

The research in this thesis has the potential to have an impact both on public policy and within the academic community. My research helps to improve the understanding of the relationship between post-compulsory education and labour market outcomes. Each of the papers develops a model linking education and training to earnings, estimates the model using rich survey or administrative data, and then applies the model to understand the efficacy of alternative policy levers.

In Chapter 2, I develop a novel methodology for estimating a non-transferable utility matching model. I assess the impact of several higher education policies that have been proposed by policy makers, such as abolishing tuition fees, increasing maintenance grants for low socio-economic status children and increasing maintenance grants for high priority fields such as STEM. I find that these policies have a limited impact on inter-generational mobility, as university admissions policies limit access for low socio-economic status students. This research has been partially funded by the Department for Education and the findings have been shared with them.

In Chapter 3, I focus on the interaction between benefit policy and wages. I simulate the effects of a reduction in the benefit withdrawal rate, mirroring a policy that was introduced in November 2021. I show that reductions in the benefit withdrawal rate may reduce wages for young women, even if labour supply increases. Understanding the interaction between benefit policy and wages, via human capital, is crucial for the accurate costing of alterations to the benefit schedule, as well as assessing the social costs and benefits of such policies. I am currently working on an extension of this paper that will explicitly assess the impact of recent changes to the benefit schedule, building on the research contained in this paper. The findings of the extension will feed in to the Institute for Fiscal Studies' Deaton Review, a high profile five-year review into the extent and causes of inequality, chaired by Nobel Laureate Professor Sir Angus Deaton and funded by the Nuffield Foundation.

In Chapter 4, I assess the impact of training subsidies on the gender wage gap. I show that policies that subsidise the training of recent mothers can increase their disposable income (beyond the taxation required to fund it) as well as overall welfare. We also find that a modest subsidy pays for itself by incentivising full-time work both during the eligibility period and after it. Both the gender wage gap and adult learning are areas of interest among policy makers, although there has been limited research considering their relationship. This paper therefore highlights an under-appreciated policy lever that could partially address existing gender disparities in pay. The paper that this chapter is based on has been published in the Journal of Labour Economics, and the findings have been presented widely to both academic and policy audiences.

# Acknowledgements

The completion of this thesis has been aided by financial support from the Economic and Social Research Council, both through an ESRC Studentship and through the Centre for the Microeconomic Analysis of Public Policy (CPP) at the Institute for Fiscal Studies.

Thanks to my supervisors, Eric French and Richard Blundell, for their input and guidance throughout my PhD. I also benefited greatly from working with various co-authors, including Jack Britton, Monica Costa Dias, Rob Joyce, Costas Meghir, and Tom Waters. My psuedo-colleagues at the Institute for Fiscal Studies have helped me more than they know. Particular thanks go to all of the EdSkills team for making me feel so welcome, patiently answering my grumpy Stata questions and tolerating my proprietorial attitude towards office staplers.

I am immensely grateful for the love and support of my family and friends - your generous sympathy, pep talks, and joyful distractions have got me through the last few years. Thank you. Finally, all my love to Mad, who has been with me through every up and down, and without whom I could never have completed this thesis.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

This thesis consists of three self-contained papers. Each paper explores a distinct set of policy instruments - Chapter 2 focuses on higher education policy, Chapter 3 focuses on benefit design and Chapter 4 focuses on training subsidies. In each case, I develop a model linking education and training choices to labour market outcomes, estimate the model using rich survey or administrative data, and then apply the model to assess the extent to which specific policies can address existing inequalities in wages and earnings. In this introduction, I briefly summarise the approach and findings of each of these papers.

In Chapter 2, I examine the capacity of alternative higher education policies to improve intergenerational income mobility. There is a large body of evidence suggesting that whether, where and what people study at university can have a dramatic impact on their subsequent earnings. However, students from poorer backgrounds are less likely to go to university and, if they do go, they are less likely to attend high-earning courses. These socio-economic gaps in higher education outcomes contribute to subsequent inequalities in lifetime earnings.

However, devising concrete policies to address these gaps is not straightforward, as the equilibrium match between students and universities is a complicated function of both the preferences of students for universities and the preferences of universities for students. Policies aimed at incentivising poorer students to go to better universities, such as targeted grants or fee exemptions, may not impact mobility if the lower prior attainment of these students means that those universities are not willing to admit them.

In this paper, I develop and estimate an empirical two-sided matching model of higher education that takes into account preferences on both sides of the market. I then apply this model to assess the impact of various higher education policies.

I first evaluate several policies aimed at encouraging poorer students to go to university (via targeted living cost grants for poorer students and tuition fee exemptions) or to attend higher quality universities (by introducing grants that are conditional on attending such institutions). I find no meaningful impact of these policies on either higher education participation or on social mobility. I also simulate a policy that provides grants to poorer students only if they study higher-earning science, technology, engineering and mathematics (STEM) subjects. I find that a significant proportion of poorer students shift into STEM degrees in response to these grants, and that this has positive implications for income mobility. The effects are small but non-negligible, suggesting policies targeting the demand side should focus more on *what* rather than *where* people study.

I then consider potential supply-side reforms. I first bring in a '10% rule' that forces universities to offer priority admissions to students from the top 10% of their high school class. I find that this policy achieves a comparable increase in income mobility to the STEM grants discussed above. If priority admissions are instead offered to students who are in the top 10% of their class *and* from a low SES background, I find a dramatic improvement in the intergenerational income elasticity. I conclude that policies targeting the supply-side offer greater opportunities for increasing income mobility than comparable demand-side policies.

In Chapter 3, I examine the interaction between human capital accumulation and the benefit system. Two alternative models for human capital accumulation over the life-cycle are common in the economics literature. The first posits that human capital accumulation is a by-product of labour and that there is no trade-off between earning and learning - I refer to this model as *learning-by-doing*. The second instead has human capital accumulation competing with productive labour, so agents must divide their time between learning and earning - I refer to this model as *Ben-Porath*.

The impact of the benefit system on human capital and therefore wages depends crucially on which of these models is applied. Individuals receiving benefits temporarily face high effective tax rates while benefits are withdrawn. The opportunity cost of human capital accumulation within a *Ben-Porath* model is foregone labour and therefore, all else equal, high effective tax rates *decrease* the cost of acquiring human capital for individuals on benefits. The opposite is true in a *learning-by-doing* model, where high effective tax rates may discourage labour supply and therefore discourage accumulation of human capital.

I develop this insight further by estimating several dynamic models of labour supply and skill investment, each nesting an alternative mechanism for human capital accumulation. These models incorporate many of the features necessary to capture realistic trade-offs in benefit design, such as human capital risk, credit constraints, household formation, fertility, partner income and permanent heterogeneity.

I find that, in line with the basic intuition described above, revenue neutral reforms that reduce the rate at which benefits are withdrawn increase wages within a *learning-by-doing* model, whereas the same reform discourages skill investment and therefore decreases wages within a *Ben-Porath* model. However, I also establish that both *Ben-Porath* and *learning-by-doing* models fail to replicate some key features of women's observed profiles of wages and time-use.

I therefore supplement the two pure models with a mixed model, which incorporates both modes of human capital accumulation. I find that this model provides a much better fit to observed wages and time allocation. Behaviourally, the mixed model has similar implications to a *Ben-Porath* model early in working life. Since wages are relatively low and many years of working life remain, women are willing to utilise costly skill investment to accelerate wage growth. As women age, they reduce costly skill investments and rely increasingly on human capital accumulation through work experience. Reducing the rate at which benefits are withdrawn therefore decreases wages early in working life and increases wages at older ages. I conclude that both modes of human capital accumulation are necessary to explain observed behaviour, and that reforms that improve work incentives for low-income workers may also decrease wages, particularly among the young.

In Chapter 4, I examine whether work-related training has a role to play in reducing the gender wage gap. Women generally reduce labour supply in the years following childbirth, both on the extensive and intensive margin. There is no comparable reduction in labour supply for men, and previous research has shown that the resulting gap in work experience is a key contributor to the gender wage gap.

As in Chapter 3, I specify a model of women's labour supply and training over the life-cycle. In the model, women enter the labour market after completing education. Marriage, separation and children arrive exogenously with a probability estimated from the data and depending on prior children, age and marital status. The evolving family structure over the life-cycle is a key feature because it affects the incentives and preferences of women for work and training. While working their human capital grows through experience, at a rate depending on whether work is part time or full time. Job separations imply a loss in human capital and hence earnings. During their working life they may also participate in work-related training, which is paid for by deductions from their earnings but increases human capital and therefore wages in future periods.

I incorporate information on the welfare and tax systems in the UK over many years, which allows me to construct the precise budget constraint that an individual is facing in each year of work. The data includes multiple cohorts, entering the labour market at different times. Each faces a different set of welfare and tax systems. Tax and benefit reforms affect multiple cohorts but at different ages. This generates exogenous variation in the incentives that people face to work and train in different parts of the earnings distribution. This is the key idea that underlies the identification strategy and provides the variation I need to estimate the model.

Our findings point to a potentially important role for training among women who completed high school level education but did not go on to complete university. I show that, for these women, training can have a role in reducing the wage loss that arises from part-time work after childbirth. Moreover, policies that subsidize the training of recent mothers from this group can increase their disposable income (beyond the taxation required to fund it) as well as overall welfare. I also find that a modest subsidy pays for itself by incentivising full-time work both during the eligibility period and after it. Finally, while training can play some role in reducing the labour market costs of children, this cost remains quite large even after systematic training policies. Other policies that would reduce the incidence of part time work, such as better childcare availability, may have a more important role to play.

The remainder of this thesis is structured as follows. Chapters 2 to 4 each contain one self-contained paper. Chapter 5 then concludes by setting out future research questions that build upon the research included here.

# Chapter 2

# Can Higher Education Policy Boost Social Mobility? Evidence From an Empirical Matching Model[1]

## 2.1 Introduction

There is a large body of evidence suggesting that whether, where and what people study at university can make a dramatic difference to their subsequent earnings.[2] However, students from poorer backgrounds are less likely to go to university and, if they do go, they are less likely to attend high-earning courses. Reforms that address these socio-economic gaps in higher education outcomes could therefore have important implications for intergenerational mobility. In a recent paper, Chetty et al. (2020) draw this conclusion, stating that "changing how students are allocated to colleges could substantially increase intergenerational mobility, even without changing colleges' educational programs." This is an appealing argument for policymakers, as it implies well designed policy tweaks could have large payoffs.

However, devising concrete policies to achieve this boost in mobility is not straightforward as, in practice, the equilibrium match between students and universities is a complicated

[1]This chapter is based on a paper co-authored with Jack Britton (University of York & Institute for Fiscal Studies) and Monica Costa-Dias (University of Bristol & Institute for Fiscal Studies)

[2]See, for example R. J. Andrews, Imberman, and Lovenheim (2017), Anelli (2018), Britton et al. (2021), Chetty et al. (2020), Dale and A. B. Krueger (2002), Hastings, Neilson, and Zimmerman (2018), and Kirkeboen, Leuven, and Mogstad (2016) and Zimmerman (2019).

function of both the preferences of students for universities and the preferences of universities for students. For example, Chetty et al. (2020) suggests income mobility in the United States would increase if poorer students were given preferential admission to top universities. But even if places at top universities were available to poorer students, it does not mean they would accept them.[3] On the other hand, policies aimed at incentivising poorer students to go to better universities, such as targeted grants or fee exemptions, might not have much effect on mobility if their lower prior attainment means that those universities do not want to admit them.

In this paper, we develop and estimate an empirical two-sided matching model of higher education that takes into account both the preferences of students for universities and the preferences of universities for students. We use the model to better understand how students sort into higher education programmes in England. We assess the impact of various reforms to the higher education system on intergenerational mobility, while also taking into account any potential efficiency trade-offs.

Our institutional setting is the United Kingdom (UK). Alongside the United States, the UK is one of the worst performing countries in the OECD in terms of the extent to which the earnings of children are predicted by the earnings of their parents (Corak 2013). We use the model to evaluate several policies aimed at encouraging poorer students to go to university (via targeted living cost grants for poorer students and tuition fee exemptions) or to attend higher quality universities (by introducing grants that are conditional on attending such institutions). We find no meaningful impact of these policies on either higher education participation or on social mobility. We also simulate a policy that provides grants to poorer students only if they study higher-earning STEM (science, technology, engineering and mathematics) subjects. We find that a significant proportion of poorer students shift into STEM degrees in response to these grants, and that this has positive implications for income mobility. The effects are small but non-negligible, suggesting policies targeting the demand side should focus more on *what* rather than *where* people study.

We then turn our attention to the supply-side. We first bring in a '10% rule' that forces universities to offer priority admissions to students from the top 10% of their high school class. We find that this policy achieves a comparable increase in income mobility to the STEM grants discussed above. If we instead offer priority admissions to students who are in the top 10% of their class *and* from a low SES background, we find a dramatic improvement

[3]For example, S. E. Black, Cortes, and Lincove (2015) show that poorer students in Texas choose not to attend top universities within the state, even when the are guaranteed admission.

in the intergenerational income elasticity.

Percent plans generally result in low SES students being matched with higher quality courses than they would otherwise be able to attend. These policies could have significant efficiency costs if there is a high degree of complementarity between course quality and students prior attainment. We allow for these complementarities in earnings when estimating the model, and find that, although they exist, they are quantitatively small. We conclude that policies targeting the supply-side offer greater opportunities for increasing income mobility than comparable demand-side policies.

Our model combines a non-transferable utility matching model of the HE market with a life-cycle model of consumption and earnings. In an initial period, potential students either match with a HE course or directly enter the labour market. The HE options available are heterogeneous; we incorporate 150 universities of varying quality levels and courses in three broad subject fields for students to match with.[4] During the period we study, university prices (tuition fees) were fixed externally by the government, and student number controls were tightly regulated and binding. This market structure closely resembles the institutional setting of Agarwal 2015, and our matching model builds on his framework. Importantly, we do not need to identify university cost functions or make assumptions about the profit maximisation behaviour of universities.

After completing their degrees, students enter the labour market where they receive stochastic labour earnings that depend on their socio-economic background, school attainment and higher education match. Our addition of a life-cycle component to the model allows us to incorporate the long-run effects of higher education on earnings. It also allows us to accurately model the English student loan system, which allows students to pay for their university education through income-contingent deductions from their labour earnings throughout their working life.

The primary assumptions of our model are as follows. First, we assume that within field, universities have a common preference ranking of students. This restriction is necessary because allowing for rich preference heterogeneity on both sides of the market creates significant challenges for identification. This assumption seems reasonable within the context of university admissions, where there is a large amount of vertical sorting of students based on test scores. Second, we assume that universities always prefer to fill their places as op-

[4] These are Law Economics and Management (LEM), Science Technology, Engineering and Mathematics (STEM) and Arts, Humanities and Social Sciences (AHSS).

posed to leaving them empty. Empirically, we provide evidence that annual enrolments very closely match student number controls, and that there is no correlation between proportion of places filled and course quality. Third, we assume that prospective students base their salary expectations on the salary outcomes of previous cohorts, and that student loans only affect their utility through their impact on their future net income. This is a strong assumption, but one that does not prevent us from being able to replicate the key patterns in the data, most notably the impact of the major reforms to tuition fees that occurred in 2012. Finally, we assume that the equilibrium match is stable. In this context, this assumption requires that any course that a student prefers to the one that they matched with would not be willing to accept them. We argue that the application process in the UK is sufficient to deliver this stable match, noting the fact that people can apply to multiple places, and can make use of the 'clearing' system if they are unhappy with their outcome.

To estimate the model, we use the Longitudinal Education Outcomes (LEO) dataset which links together administrative school, university and tax records. The dataset includes everyone born between 1986 and 1996 who attended secondary school in England. It has detailed information on examinations taken at ages 11, 16 and 18 including specific grades in specific subjects that we use to construct a two-dimensional skill index for each student that we refer to as 'quantitative' and 'communication' skills. We observe linked tax records for many cohorts of school leavers, including those who did not attend university. These cohorts span the substantial reforms to the higher education system that occurred in 2012, when tuition fees were increased from around £3,000 a year to around £9,000 a year.

The linked school-university records from LEO allow us to characterise the match between students and courses. We also use the school records to generate several instrumental variables that shift student preferences for courses without affecting university preferences. In particular, we can use variation in geographic proximity to different universities and quality of courses available in the local area to shift student choices. We show these instruments to be powerful predictors of both quality of course attended and field of study. These demand-shifters are crucial for separately identifying preferences on both sides of the market. Our panel data on earnings also allows us to account for unobserved permanent heterogeneity in students that can affect both university preferences and students' future earnings.

We estimate the model using a minimum distance estimator that matches data moments to simulated moments from the model. We estimate using four cohorts who left school between 2006 and 2009. This was a benign period for higher education policy, with no changes to tuition fees or student support. We also simulate the 2010 cohort without formally including them in the estimation. This allows us to assess out-of-sample fit. We are able to replicate sorting patterns extremely well, both in and out-of-sample.

As a further test of the out-of-sample predictive power of our model, we show the model is able to replicate the effects of the 2012 higher education reforms on sorting patterns, even though neither of the cohorts immediately before or after these reforms are included in the estimation data. Although we confirm the result from previous work that these reforms had almost no effect on overall participation (for example, Azmat and Simion 2020), we present reduced form evidence that the overall zero effect masks large drops in demand from the highest ability students, with their spots being filled in by lower ability peers. Our model is able replicate these basic patterns. We then use the model to show that these reforms had only very minor implications for income mobility and for the efficiency of the system.

The parameter estimates from the model suggest that students, on average, have a strong preference for AHSS subjects (arts, humanities and social sciences) over LEM subjects (law, economics and management courses). Students have a distaste for attending institutions far from their home, and this distaste is much greater for low SES students. The university utility parameters suggest that universities care a lot about the match between student skill types and the subject they are studying. Universities strongly prefer quantitative skills over communications skills for STEM subjects and vice versa for AHSS subjects. Finally, the wage parameters in the model suggest important differences in returns to university relative to the estimates obtained from OLS regressions. In particular, the model estimates suggest the returns to university quality are only around half as large as the OLS estimates. A key difference between the two approaches is that the model incorporates unobserved heterogeneity, which boosts earnings by around 4.5% and is also an important driver of selection into higher quality institutions.

The rest of the paper is organised as follows. Section 2.2 discusses how our paper connects to the literature. Section 2.3 discusses the LEO data and institutional background before showing some data descriptives, including showing the reduced form effects of the 2012 reforms. Section 2.4 outlines our model and Section 2.5 discusses estimation and identification. Section 2.6 then shows the fit of the model and estimated impact of the 2012 reforms before Section 2.7 runs counterfactual policy experiments. Section 2.8 concludes.

## 2.2 Literature

*Modelling of the higher education market.* Our paper contributes to the growing literature that models higher education choices. Many papers model higher education choices without formally modelling preferences of the supply side (Arcidiacono 2004; Delavande and Zafar 2019; Keane and Wolpin 1997, 2001; Wiswall and Zafar 2014). Counterexamples are Arcidiacono (2005) and Kapor (2020), who focus on affirmative action policies, and Epple, Romano, and Sieg (2006) and Fu (2014), who focus on equilibrium tuition and financial aid polices. These latter papers are all complicated by the fact that in the United States, universities have much more control over their tuition fees, financial aid and student numbers. Because of this, Agarwal (2015), which models the 'medical match' between junior doctors ('residents') and training hospitals ('programs') in the United States, is in fact a much closer institutional setting to ours and is therefore the closest paper to ours methodologically.

We extend Agarwal (2015) in four important ways. First, we increase the dimensionality of both student characteristics and the choices available to students. Students in our model have multidimensional skills and match to a course in a specific subject field at a specific university. This reflects the fact that, in the UK, students make both subject and university choices prior to entry. The fit between student skills and course matters for student preferences, for university preferences, and for subsequent earnings outcomes. Second, we relax the assumption of homogeneous supply-side preferences, allowing courses to rank students differently across different fields. For example, we allow STEM courses to have stronger preferences for quantitative skills than other subjects. Third, we allow for permanent unobserved student heterogeneity that affects both university preferences for students and student earnings. Students can therefore select into courses on characteristics that are unobservable to the researcher, enabling us to correct for potential endogeneity in the earnings equation. Fourth, we incorporate a dynamic lifecycle component to the model, which is important in our context as it facilitates explicit modelling of the English income contingent loans system.

*The returns to higher education.* There is a large literature investigating the impact of attending different universities on earnings outcomes. Dale and A. B. Krueger (2002, 2014) and Mountjoy and Hickman (2020) suggest there are weak returns to course quality, but other papers (Anelli 2018; D. A. Black and Smith 2006; Broecke 2012; Dillon and Smith 2020; Hastings, Neilson, and Zimmerman 2013) suggest otherwise. Many of these papers depend on strong assumptions of selection on observable factors. Kirkeboen, Leuven, and Mogstad (2016) provide one method for circumventing these issues - our paper provides an alternative approach that enables us to identify causal estimates of the returns to quality by explicitly modelling selection on unobserved heterogeneity. Our estimates suggest that the returns to course quality are positive and economically significant, but that OLS overstates them by a factor of 1.5 to 2. We also contribute to the literature investigating returns to field of study (Anelli 2018; Chevalier 2011; Kirkeboen, Leuven, and Mogstad 2016) and our results are consistent with the idea that on average switching from a low to a high-returning field can matter a lot more for earnings than moving to a higher quality institution within the same field. Finally, we contribute to the small set of papers that have estimated match effects in higher education (Dillon and Smith 2020; Mountjoy and Hickman 2020). Consistent with Dillon and Smith (2020), we find evidence of match effects, whereby higher ability students experience greater gains from attending higher quality courses.

*The impact of grant aid support and tuition fees on higher education choices.* Denning, Marx, and Turner (2019) investigate the impact of Pell Grants, finding positive effects on graduation and earnings. Marx and Turner (2018) suggest limited effects of Pell Grants on college enrollment or course quality, but Scott-Clayton and Zafar (2019) find evidence that other types of grants can affect these outcomes. Epple, Romano, and Sieg (2006) also argue that financial aid could be targeted to improve access for poorer students to top colleges. We find that grants can affect participation of poorer students. This aligns with Dearden, Fitzsimons, and Wyness (2014) who, also in the UK context, find that maintenance grants have a positive impact on participation among low SES students. We also find that targeted grants can affect field and course quality for low SES students.

The evidence generally suggests that tuition fees reduce participation. Hübner (2012) exploits a natural experiment in Germany to show that tuition fees reduce enrollment, while Neill (2009) draws a similar conclusion based on evidence from Canada. Azmat and Simion (2020) study the impacts of the 2012 reforms to the English higher education system, concluding that there was a negligible impact on participation overall, but that there was a narrowing of the gap in participation between young people from richer and poorer backgrounds. They suggest that the increase in tuition put people off, but that the (relatively modest) increase in maintenance support boosted participation of poorer students, outweighing the negative fee effects for them. Sá (2019) looks at the effect of these reforms on applications, finding large drops for all groups as a result of the reforms. Our paper contributes significantly to understanding of the effects of these reforms.

*Academic undermatch.* Several recent papers have documented the issue of academic undermatch in the US and UK (Campbell et al. 2021; Griffith and Rothstein 2009; Hoxby and Avery 2013). Our simulations suggest that undermatch could be eliminated via either demand-side (targeted grants) or supply-side (percent plan) policies. This contrasts with S. E. Black, Cortes, and Lincove (2015) and Dillon and Smith (2017), who conclude that student decisions are a key driver of undermatch in the United States. However, it aligns with Campbell et al. (2021), who suggests the drivers of undermatch in the UK might be different. We find that eliminating undermatch on its own would do relatively little to affect intergenerational income mobility; Chetty et al. (2020) draw a similar conclusion, albeit from a much more stylised exercise.

## 2.3 Institutional background and data

### 2.3.1 Higher education policy in England

Students in England most commonly apply for entry into higher education during their final year of secondary school. They apply to up to five subject-university combinations through the Universities and Colleges Admissions Service (UCAS), and demonstrate their ability by including test scores from school and a personal statement. A small number of selective universities also include interviews and additional tests in their entry process. If students receive no offers from the five courses they apply to, or have a place that they are unhappy with, there is a 'clearing' system which matches up any universities with unfilled spots just before the start of the academic year with any students without a place.

Throughout the period we are studying, the system was almost entirely public and was tightly regulated, with both student number controls and price caps. Student number controls were enforced for each university with large reductions in teaching grants if maximum allotted enrollments were exceeded,[5] while tuition fee caps were legally imposed.

There have been several changes to the system, which we summarise briefly in Table 2.1. Our model is estimated using the 2006-2009 higher education cohorts. During this period, tuition fees were around £3,000 per year. Students could either pay this up front, or borrow them

[5]In Appendix A.3 we show that annual enrollments during this period were close to the total student number control and that there is no relationship between university selectivity and how close universities were to their number cap. Exemptions to student number controls were introduced in 2012/13 and 2013/14, but were matched by reductions in student number controls that maintained the same overall tightness of the market. Student number controls were abolished in 2015/16.

from the government using an income contingent loan. They were also eligible to borrow an additional £3,000 to £6,000 per year (depending on their parent's income) in 'maintenance support' to help with living costs during study.

Many graduates are unlikely to repay the full value of their loan. Students who borrowed over this period are required to repay 9% of their income above a threshold of around £15,000. No repayments are due if earnings are below £15,000. Interest rates are low and any outstanding debt will be written off after 25 years. The costs of any write-offs fall on the taxpayer.

In 2012, there was a major reform to the system. This reform included a large increase in the tuition fee cap to £9,000 as well as significant changes to the income contingent loan repayment terms. The repayment rate remained at 9% of income, but the repayment threshold increased to £21,000, and interest rates increased substantially.[6] Finally, the repayment period was extended from 25 to 30 years.[7] Combined, these changes dramatically increased the costs of higher education for borrowers with the highest lifetime earnings, while they actually reduced costs for the lowest-earning borrowers. We discuss this in more detail below.

Table 2.1: Approximate student loan rules by university cohort

| | 1998-2005 | 2006-2011 | 2012-date |
|---|---|---|---|
| Tuition fees | £1,200 | £3,000 | £9,000 |
| Fees borrowable | No | Yes | Yes |
| Maintenance support | Yes | Yes | Yes |
| Repayment rate | 9% | 9% | 9% |
| Repayment threshold | £15,000 | £15,000 | £21,000 |
| Interest rate | RPI | RPI | RPI + 3% |
| Repayment term | 25 years | 25 years | 30 years |

Note: The rules given here are simplified versions of the truth and are written in this way for illustrative purposes.

[6] Specifically, they increased from the minimum of the bank rank rate and RPI to RPI plus up to 3%, depending on income and study status. The RPI is generally considered to overstate true inflation by around 1 percentage point and averaged around 3% in the years following the reform.

[7] There were also some relatively minor changes to student support. Maintenance grants increased by slightly more than the standard incremental increase, while a new 'National Scholarship Programme' (NSP) was introduced. However, the NSP was relatively small, and it was often given to students only after they had started at university (Chowdry et al. 2012)

### 2.3.2 LEO data

We use the Longitudinal Education Outcomes (LEO) dataset, which is a new administrative dataset that links tax records to university and school records. We observe fully linked records for everyone who went to secondary school in England who was born between September 1985 and August 1996. There are approximately 600,000 people in each school-year cohort. We also have partially linked data (tax-university records, but no school records) going back to people born in the mid-1970s. This is summarised in Table 2.2.

Table 2.2: LEO data summary

| Birth Cohort | Final School Year | School Records | University Records | (Usable) Tax Records | Tax Data Age Range |
|---|---|---|---|---|---|
| 1975/76 | 93/94 | x | ✓ | ✓* | 29 – 40 |
| 1976/77 | 94/95 | x | ✓ | ✓* | 28 – 39 |
| 1977/78 | 95/96 | x | ✓ | ✓* | 27 – 38 |
| 1978/79 | 96/97 | x | ✓ | ✓* | 26 – 37 |
| 1979/80 | 97/98 | x | ✓ | ✓* | 25 – 36 |
| 1980/81 | 98/99 | x | ✓ | ✓* | 24 – 35 |
| 1981/82 | 99/00 | x | ✓ | ✓* | 23 – 34 |
| 1982/83 | 00/01 | x | ✓ | ✓* | 22 – 33 |
| 1983/84 | 01/02 | x | ✓ | ✓* | 21 – 32 |
| 1984/85 | 02/03 | x | ✓ | ✓* | 20 – 31 |
| 1985/86 | 03/04 | ✓ | ✓ | ✓ | 19 – 30 |
| 1986/87 | 04/05 | ✓ | ✓ | ✓ | 18 – 29 |
| 1987/88 | 05/06 | ✓ | ✓ | ✓ | 17 – 28 |
| 1988/89 | 06/07 | ✓ | ✓ | ✓ | 16 – 27 |
| 1989/90 | 07/08 | ✓ | ✓ | ✓ | 16 – 26 |
| 1990/91 | 08/09 | ✓ | ✓ | ✓ | 16 – 25 |
| 1991/92 | 09/10 | ✓ | ✓ | x | 16 – 24 |
| 1992/93 | 10/11 | ✓ | ✓ | x | 16 – 23 |
| 1993/94 | 11/12 | ✓ | ✓ | x | 16 – 22 |
| 1994/95 | 12/13 | ✓ | ✓ | x | 16 – 21 |
| 1995/96 | 13/14 | ✓ | ✓ | x | 16 – 20 |

Note: * indicates that we only observe this for people who attended university. Final school year is also the year individuals turn 18.

School records are drawn from the National Pupil Database (NPD). They include detailed exam scores from national examinations taken at age 11, 16 and 18. The NPD data also includes secondary school attended (including whether or not it is a private school), gender and parental socio-economic status (SES), which we discuss in more detail below.

University records are provided by the Higher Education Statistics Agency (HESA). They cover all students attending a UK university and include information on the institution attended and subject studied. Throughout this paper, we define a students higher education outcomes (that is whether, where and what they are studying) based on their HESA record two years after finishing school. We only classify individuals as attending university if they are studying full-time for an undergraduate degree at this point. For simplicity, we categorise people into different cohorts based on their final year of school, and we assume that people compete for spots with other students in their cohort.[8] For our main estimation sample, approximately 35% of each cohort are enrolled in a full-time undergraduate degree two years after finishing school.

Tax records are provided by Her Majesty's Revenue and Customs (HMRC). We observe annual taxable income from employment, self-employment and from partnerships. The tax data we have includes the entire population between 2005/06 and 2016/17. We observe only annual earnings and the individual identifier generated by the DfE. The NPD and HESA records are hard linked based on individual student identifiers. These datasets are linked to HMRC records based on a fuzzy match from address and surname. Around 95% of the NPD are linked to the tax records. We drop individuals who are not linked.

### 2.3.3 Individual skills

The NPD includes information on exam grades in specific subjects from national examinations taken at ages 16 and 18, as well as numerical scores in national Mathematics, English and Science examinations taken at age 11.

For the purposes of our model, we want to reduce the dimensionality of this prior attainment data while maintaining information on the relative subject strengths of each student. We do this by combining information from multiple exam results to construct two skill measures: one that captures 'quantitative' skills and another that captures 'communication' skills. The first variable puts more weight on test scores in subjects such as mathematics and science, while the second puts more weight on scores in English and humanities subjects.[9] Importantly, these two variables are not orthogonal to each other. As seen in Figure 2.1, the

[8]This allows us to avoid having to model the choice of whether to take up a spot immediately or to defer for a year. In practice people typically apply in the final year of school and universities take decisions on offers without knowing if people are going to defer or attend immediately.

[9]See Appendix A.1 for more detail on the construction of these variables.

Figure 2.1: Quantitative and Communication Skills

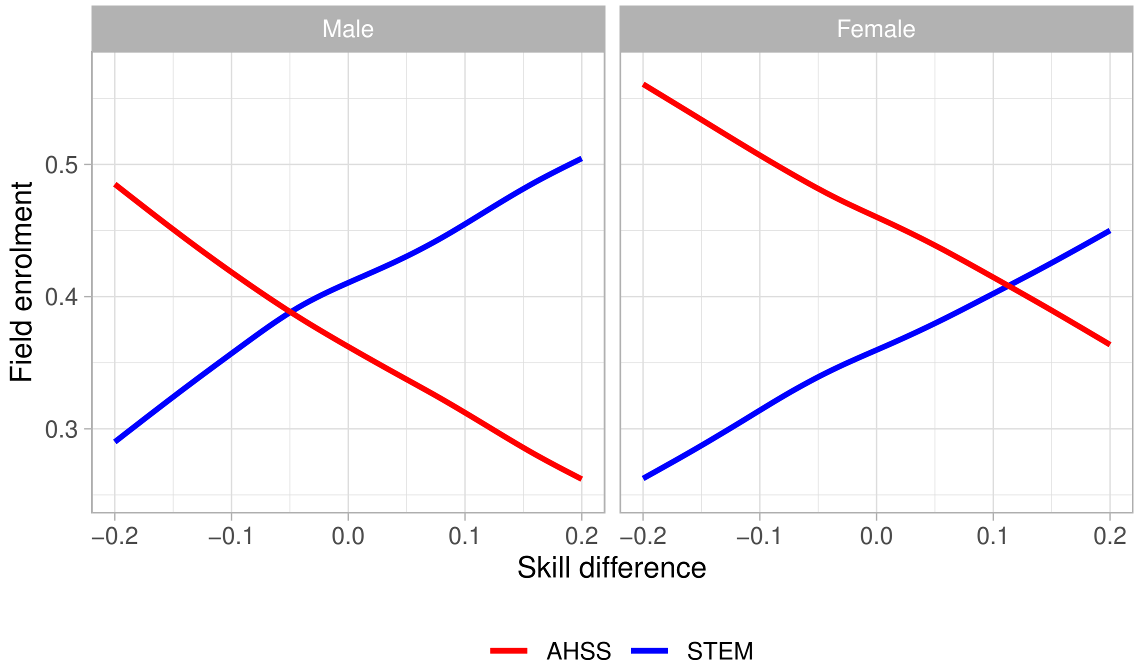

Note: Based on the 2005/06 school leavers. We have added a random jitter to the plot to remove disclosivity risks, without losing meaning.

two skills are highly correlated: students who have high attainment in quantitative subjects are likely to also have high attainment in communication subjects.

These skills measures are strongly related to subject choices at university, the quality of university attended, parental SES, and subsequent earnings. Figure 2.2 shows that despite the high correlation between the two measures of skills, relative advantages in one or the other are highly predictive of subject studied at university. The figure plots the share of students enrolling into STEM and AHSS subjects by the difference between their quantitative and communication skills. Positive differences indicate that the student is relatively stronger in quantitative subjects than in communication subjects. There is a strong positive relationship between STEM enrolment and relative strength in quantitative skills, and a strong negative relationship between AHSS enrolments and relative strength in quantitative skills.

This relative skill strength and subject choice is also reflected in Figure 2.3, where each point is an individual university-subject group combination (e.g. STEM at the University of Manchester). The position of the points indicate the average quantitative and communication skills of enrollees in each course. In the top panel, the enrollees in STEM courses have higher average quantitive skills than the enrollees in AHSS courses, but lower average

Figure 2.2: Relative Skills and Subject Choice




Note: x axis shows quantitative skills minus communications skills. Based on the 2005/06 school leavers. AHSS = Arts, Humanities and Social Sciences STEM = Science, Technology, Engineering and Mathematics. LEM (Law, Economics and Management) subjects are omitted from the plot.

communication skills. LEM courses tend to lie in between. The bottom panel of the figure shows the same plot, but now the colour indicates course quality.[10] This demonstrates the substantial amount of sorting on ability in the UK's higher education system, with the highest quality courses admitting the highest skilled individuals.

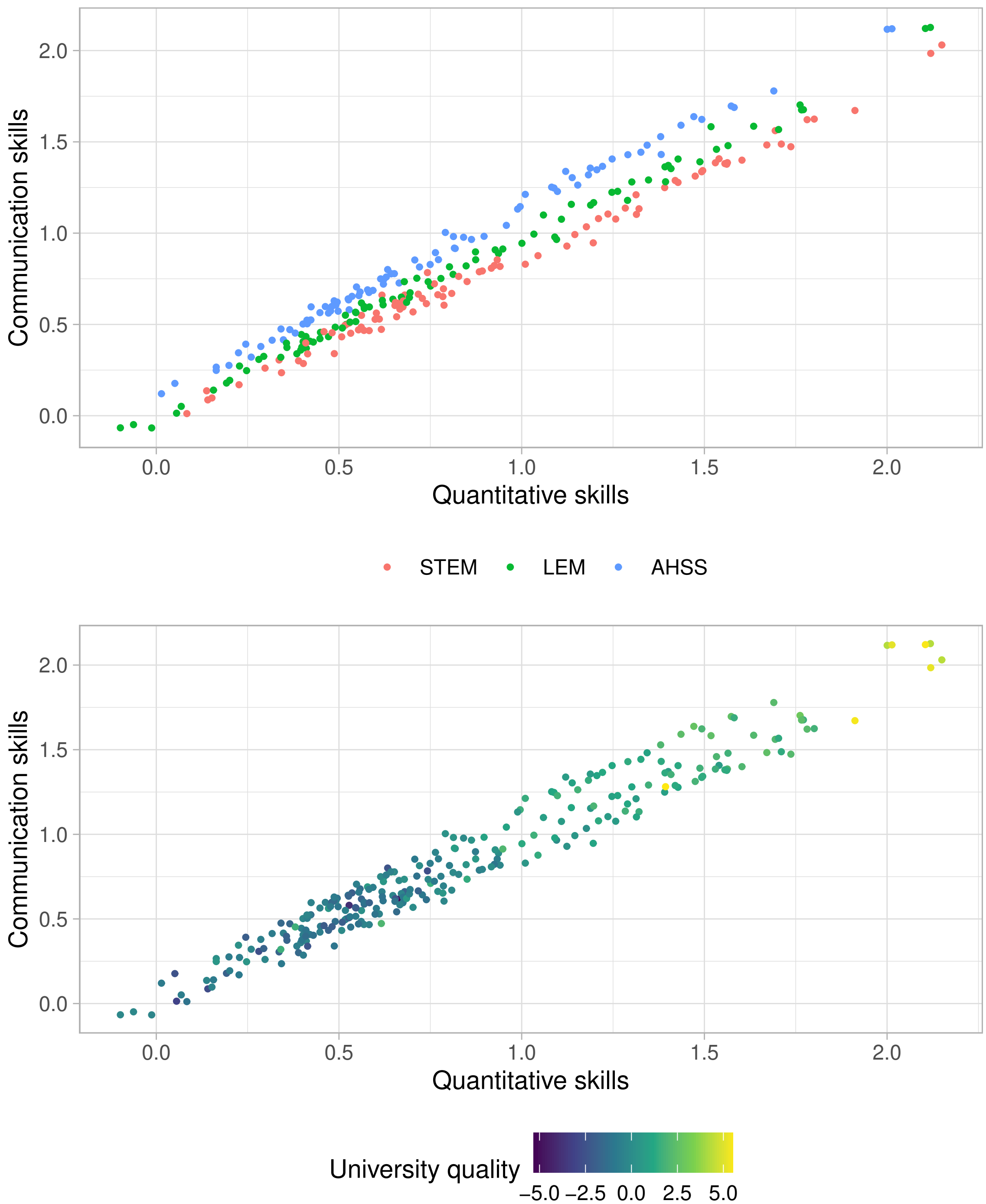
### 2.3.4 Socio-economic background

The NPD also includes information on parental SES, through the 'income deprivation affecting children index' (IDACI), which we base on where each student lived at age 16. For much of our analysis, we divide people into equally sized 'high' and 'low' SES groups based on this index, with the privately educated included in the high SES group.

Figure 2.4 shows that there is a strong relationship between parental SES percentile and child's earnings percentile at age 28. The overall intergenerational income elasticity is around

[10]This is constructed using principal components analysis, extracting the first component from five measures of course quality, including two measures of spending, the student-staff ratio, student satisfaction and research quality. See Appendix A.2 for more details.

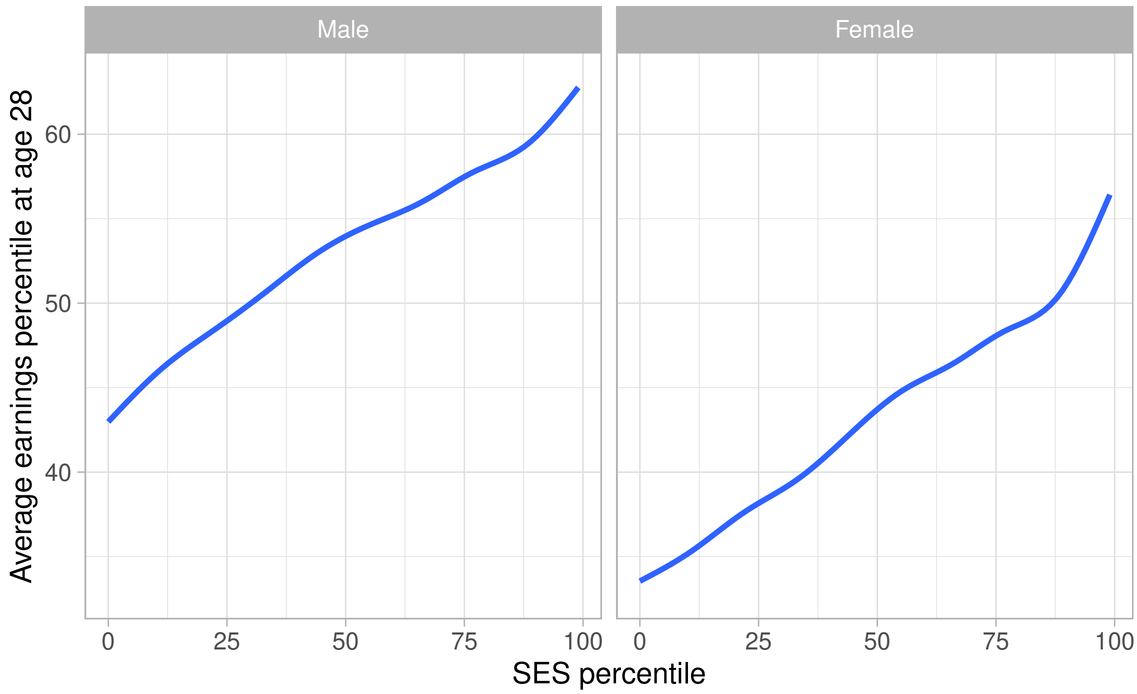
Figure 2.3: Average skill composition by course




Note: Based on 2005/06-2008/09 school leavers (we include more data here because the individual cohorts are noisier).

0.2 for women and 0.15 for men. A key motivation of our paper is to better understand how higher education policy can be used to flatten this relationship.

Figure 2.4: Parental SES and child's income



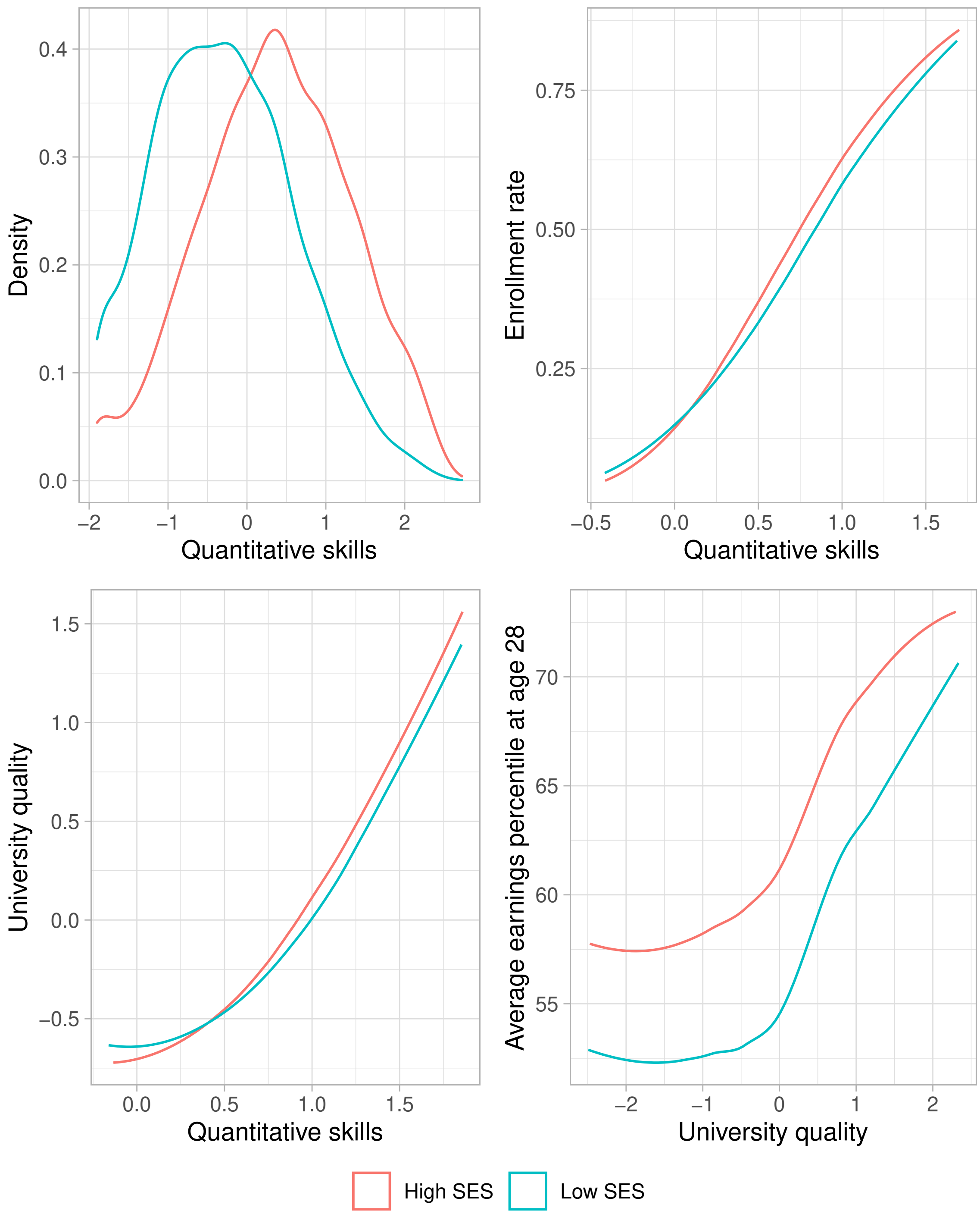

Note: Based on 2005/06 school leavers.

Figure 2.5 documents the relationship between parental SES, skills, higher education outcomes and earnings. The top-left plot shows that there are large differences in the skill distribution for high and low parental SES students. However, the top-right and bottom-left panels also show that for any given level of skills, low parental SES students are both less likely to attend university and, conditional on attending, are likely to attend a lower quality course. Policies that address these gaps could potentially increase intergenerational income mobility.[11]

Finally, the bottom-right plot shows that there is a gap in earnings between the high and low parental SES group, even when conditioning on course quality. This may capture a direct effect of SES on earnings, over and above the impact on higher education outcomes, but may also result from differences in higher education outcomes not captured in course quality. High SES students could, for instance, select into fields with higher returns or attend courses that are a better match for their skills. We explore these alternative explanations further in the next section.

[11]The bottom left plot aligns with the result from Campbell et al. (2021) that poorer students are more likely to undermatch at university than their wealthier peers.

Figure 2.5: Child outcomes by parental SES




Note: Based on 2005/06 school leavers. The figures above focus on quantitative skills, but the results are very similar when we use communication skills instead.

### 2.3.5 Earnings

Table 2.3 shows parameter estimates from the following earnings model, which we estimate by OLS:

$$\ln(y_i) = \alpha_{0F} + \alpha_{1F}X_i + \alpha_{2F}S_i^q + \alpha_{3F}S_i^c + \alpha_{4F}Q_jH_i + \alpha_{5F}Q_jH_iS_i^q + \alpha_{6F}Q_jH_iS_i^c + \epsilon_i \quad (2.1)$$

$y_i$ is the annual earnings of individual $i$. $X$ is a vector of background characteristics that includes gender, SES and private school status. $S_i^q$ and $S_i^c$ are quantitative and communication skills, respectively. $Q_j$ is course quality, which has a direct effect and interacts with individual skills. This allows high skilled individuals to have larger returns to attending high quality courses than low skilled individuals, for instance. $Q_j$ is also interacted with $H_i$, which is a dummy set equal to one if the individual enrolled in higher education, and zero otherwise. All of the parameters are allowed to take different values for individuals who did not attend university ($F = 0$) and, among those who do attend, take different values for each of the three fields that university attendees can study ($F \in \{1, 2, 3\}$). We estimate this equation using the 2005/06 school leaver cohort, with earnings measured in the 2016/17 tax year.

There is a strong relationship between parental SES and children's earnings at age 28, even conditional on university, gender, and skills. Going from the bottom to the top of the SES distribution is associated with an 20-40% higher earnings, depending on field of study. The estimates also show that earnings are higher at age 28 for graduates in each of the three subject areas than they are for individuals who do not go to university. Female graduates earn around 15% less than male graduates, while female non-graduates earn about 35% less.[12] There are large positive returns to quantitative skills and smaller positive returns to communications skills, and strong returns to university quality.[13] There is also some evidence of match effects for quantitative skills, as returns to quality are greater for those with higher quantitative skills. We include these OLS parameter estimates as moments in the estimation of our model.

[12]The coefficients in the table are in log points. The percentage effects are therefore slightly higher than the log-point values presented in the table.

[13]Each of these variables - $S^q, S^c$ and $Q$ - have been normalised to have zero mean and a variance of one

Table 2.3: Wage parameters

| | No Uni | STEM | LEM | AHSS |
|---|---|---|---|---|
| intercept | 9.901 | 10.076 | 10.144 | 9.954 |
| | (0.002) | (0.005) | (0.007) | (0.005) |
| female | -0.320 | -0.154 | -0.168 | -0.119 |
| | (0.007) | (0.005) | (0.002) | (0.004) |
| SES | 0.213 | 0.201 | 0.355 | 0.229 |
| | (0.005) | (0.014) | (0.018) | (0.014) |
| private | -0.065 | 0.009 | 0.007 | 0.041 |
| | (0.007) | (0.006) | (0.009) | (0.006) |
| $S^m$ | 0.157 | 0.100 | 0.138 | 0.105 |
| | (0.003) | (0.007) | (0.01) | (0.006) |
| $S^c$ | 0.058 | 0.047 | 0.044 | 0.024 |
| | (0.003) | (0.006) | (0.009) | (0.006) |
| $Q$ | | 0.032 | 0.044 | 0.029 |
| | | (0.002) | (0.004) | (0.003) |
| $Q * S^m$ | | 0.016 | 0.040 | 0.017 |
| | | (0.004) | (0.006) | (0.004) |
| $Q * S^c$ | | -0.013 | -0.032 | -0.015 |
| | | (0.004) | (0.006) | (0.004) |

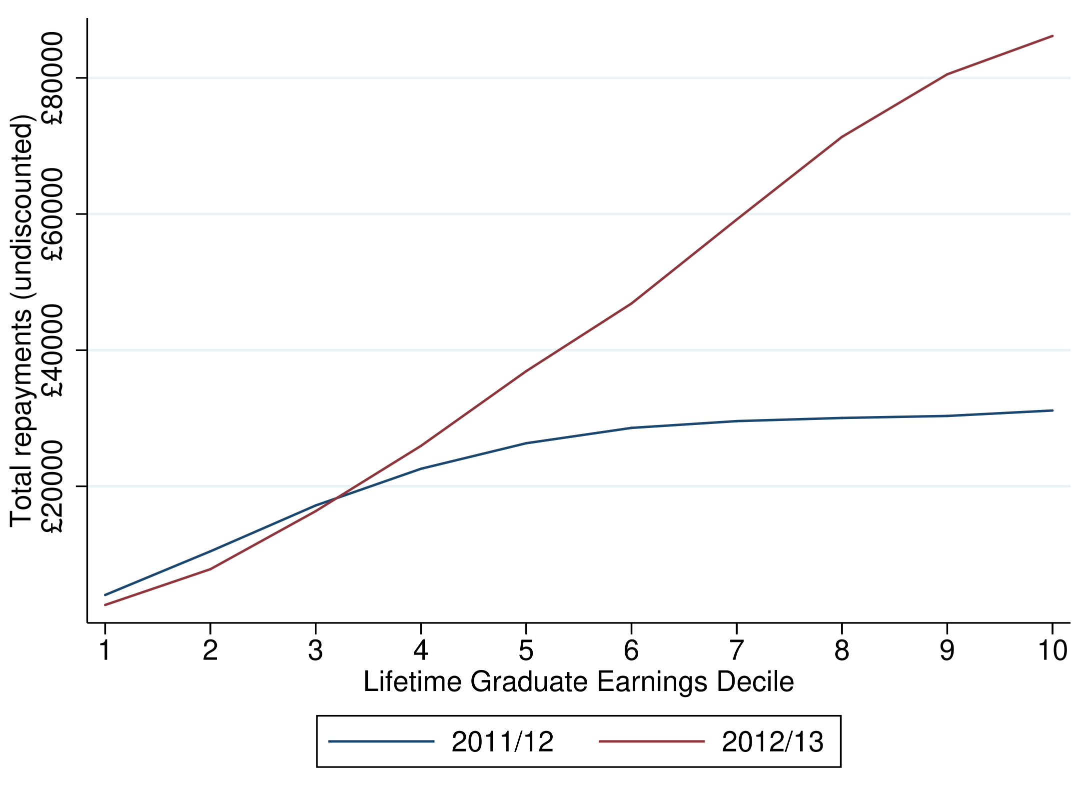
Note: Based on the 2005/06 school leavers, N $\approx$ 450,000. Outcome variable is log of earnings in the 2016/17 tax year (approximately age 28).

### 2.3.6 Participation effects from the 2012 reforms

We use the 2012 tuition fee reforms to validate our model. Rather than including the relevant cohorts in our estimation sample, we estimate the model using earlier cohorts and then simulate the impact of the reform. We then compare the model predicted impact to what we observe in the data. In this subsection, we present the observed effects on how sorting into higher education changed as a result of the tuition fee reforms.

The increase in tuition fees from £3,000 to £9,000 applied to all students. However, the contemporaneous changes to the income contingent loan repayment terms meant that the impact of higher tuition fees on the returns to higher education varied substantially depending on expected earnings.

Figure 2.6: Impact of 2012 reforms on lifetime loan repayments




Source: Taken from Belfield et al. (2017), with permission. Assumes maximum loan uptake and holds lifetime earnings fixed across the two systems.

Figure 2.6, taken from Belfield et al. (2017), shows how total loan repayments vary by graduate earnings decile, for both the pre-2012 and post-2012 financing regimes. The increase in tuition fees had almost no effect on the total cost of higher education for lowest earning graduates. These graduates did not pay back the full balance of their loan in the pre-2012 regime, so increasing the principal of their loan does not increase their repayments. Instead, they slightly benefited from an increase in the repayment thresholds. High earning graduates, on the other hand, faced large increases in total repayments. The repayment threshold and loan principal in the post-2012 regime was sufficiently high that only the top two deciles are likely to repay their loans in full and, as a result, total repayments are increasing across the full earnings distribution. In other words, those with the highest earning potential saw their net returns to higher education fall the most.

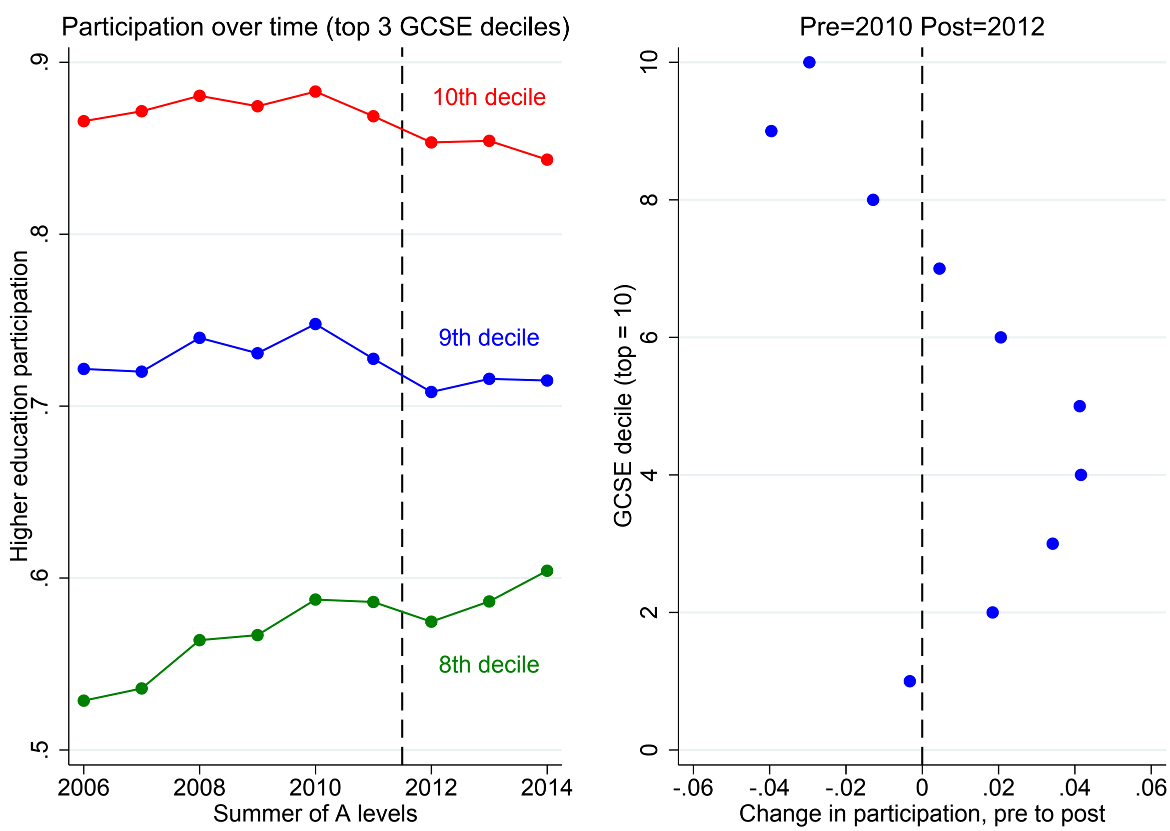
Drawing on the idea that those with the highest earnings potential were most affected by the reform, we investigate how the impact of the tuition fee increase varied by GCSE attainment, which is is highly correlated with later-life earnings, both for graduates and non-graduates. In the left-hand plot of Figure 2.7 we show participation rates by decile of overall GCSE scores, while in the right-hand panel we show the change in participation between the 2010 and 2012 school leaver-cohorts by GCSE decile.[14] There were falls in participation amongst students in the top 20% of the ability distribution of around 4ppts. However, this did not translate into decreased participation among the cohort as a whole. Instead, there were increases in participation further down the ability distribution that offset the lower participation rates of students with high GCSE scores.

One potential explanation for these results is that the government-imposed student numbers caps, coupled with tightly regulated prices, resulted in excess demand for higher education prior to 2012. The increase in tuition fees decreased the returns to higher education for relatively high ability students and discouraged some from enrolling. Their vacated spots were filled by students further down the ability distribution. The biggest participation increases occurred among students in the fourth and fifth attainment deciles.[15] In previous cohorts, these students may not have been able to secure a place on an undergraduate course, or they may only have been admitted to courses in locations or fields that they disliked. The reforms have a limited impact on their returns to higher education, since their expected earnings are significantly lower than their higher attainment peers. As a result, the financing reforms have almost no impact on overall participation, but do result in a fairly big shift in the composition of students going to university. Later we show that our model is able to replicate these patterns. We then apply the model to further understand the effects of the 2012 reforms.

[14] We use entry within 2 years of leaving school as our metric of participation. We therefore consider the 2011 cohort to be partially treated by the 2012 reform, as they would have been affected by the reform if they delayed entry.

[15] We find that this result is extremely robust, including to more sophisticated approaches, such as including decile-specific time trends with post-treatment dummies and including additional control variables.

Figure 2.7: Impact of 2012 reforms on participation by GCSE scores



## 2.4 A model of sorting in higher education

We estimate a two-sided non-transferable utility matching model. Each prospective higher education student has a preference ordering over all of the degree courses (field-university combinations). Similarly, the degree courses have preferences over all of the students. Tuition fees are set exogenously and are uniform across programmes, so preferences and capacity constraints jointly determine the equilibrium match.

Students order degree courses based on a direct non-pecuniary value of attending each course as well the discounted present value of utility from consumption over their working life. The non-pecuniary value is intended to capture the direct costs and benefits of attending the course, relative to entering the labour market. These benefits depend on factors such as the match between the students own skills and the content of the course they are choosing, as well as the students' parental SES and how far away the course is from where they live. The present value of consumption is based on the solution of a lifecycle problem in which individuals choose consumption and savings in each period, depending on their income and assets. Income in the life-cycle model depends on their higher education match. Not attending higher education is one of the options that students face, and they order it amongst their degree preferences as if it is a degree course.

On the other side of the market, universities care about their reputations, and therefore the skills of the students they admit. They are also allowed to care about the composition of their intake, and therefore the gender and parental SES of students. We assume that universities strictly prefer filling available spots to leaving them unfilled. In the rest of this section we provide more detail on the model and how students are matched to courses.

### 2.4.1 Student preferences

Students are characterised by their skills and background characteristics. There are three skills: $S^q$, $S^c$ and $\theta$. $S^q$ and $S^c$ are the quantitative and communication skill measures introduced in the previous section, while $\theta$ is an unobserved skill, known to the agents in the model. $\theta$ affects university preferences for students and earnings in later life. This term could, for example, capture non-cognitive skills that are valued by universities, observable through personal statements and interviews, and positively impact future earnings. Their background characteristics, $X_i$, include students' parental SES ($SES$), an indicator for whether they attended a private secondary schools (*private*) and their gender ($g$). If student $i$ is matched to course $j$, the student receives utility:

$$U_{ij} \quad = \quad u\left(X_i, S_i^q, S_i^c, \text{distance}_{ij}, \text{share}_{ij}, F_j, \eta_{ij}\right) + \text{EV}_{ij} \tag{2.2}$$

where distance$_{ij}$ is the distance from individual $i$'s home at age 16 to course $j$; share$_{ij}$ is the share of the previous cohort from individual $i$'s school who chose to study in the same field; $F_j$ is field (namely STEM, LEM or AHSS); $\eta_{ij}$ is a preference shock for course $j$;[16] and finally $EV_{ij}$ is the expected value of life-cycle consumption for individual $i$ attending course $j$. The specific functional form for $u(\cdot)$ is given by:

$$\begin{aligned} u \quad = \quad & \beta^u_{F_j} + \beta^X_{F_j} X_i + \beta^q_{F_j} S_i^q + \beta^c_{F_j} S_i^c + \beta^d \text{distance}_{ij} + \\ & \beta^{ds}(\text{distance}_{ij} * SES) + \beta^{dp}(\text{distance}_{ij} * private) + \beta^{sh}_{F_j}\text{share}_{ij} + \eta_{ij} \end{aligned} \tag{2.3}$$

There are 150 universities in the model and 3 subject areas, resulting in approximately 450 courses[17]. We allow the parameters of the direct utility function $u(\cdot)$ to vary by field. For example, the relationship between direct utility from attending course $j$ and quantitative skills $S^q$ depends on whether $j$ is a STEM course or an AHSS course.

If student $i$ does not attend university, the student receives utility:

$$U_{i0} = EV_{i0}$$

### 2.4.2 Lifecycle model

After entering the labour market, whether directly from school or via higher education, agents earn labour income ($y$) each period and choose how much to consume ($c$) and how much to save in safe assets ($a$). Earnings and employment are determined by a stochastic

[16] We set course preference shocks to be an additive, separable shock for a specific university and a specific field.

[17] There are slightly fewer than 450 courses because not all universities offer degrees in all three areas

processes that depends on higher education match, skills and background characteristics. Earnings are subject to deductions for income taxes ($I$) and student loan repayments ($P$). The solution to the lifecycle model is governed by the following equations:

$$V_{it}^{j} = \max_{c_{it},a_{it+1}} \left[\ln(c_{it}) + \delta \mathrm{E} V_{it+1}^{j}\right] \tag{2.4}$$

$$\text{st} \quad a_{it+1} = Ra_{it} + d_{it}y_{it} - P\left(d_{it}y_{it}, l_{it}\right) - I\left(d_{it}y_{it}\right) - c_{it} \tag{2.5}$$

where $\delta$ is the discount factor, $R$ is the interest rate on assets, and $d$ is a dummy for working. The probability of being employed is a exogenous function of age and field of study. If the agent is unemployed, pre-tax earnings are set to zero. We assume individuals do not work if they are at university, that they attend for three years and that they consume all of their maintenance loan while they are at university, and as such, no individual has any assets when they leave education.

Pre-tax earnings of employed agent $i$ at time $t$ who studied field $F$ at course $j$ are given by:

$$\begin{aligned} \ln\left(y_{it}^{e}\right) = {} & \alpha_{0F} + \alpha_{1F}X_i + \alpha_{2F}S_i^q + \alpha_{3F}S_i^c + \alpha_{4F}Q_jH_i + \\ & \alpha_{5F}Q_jH_iS_i^q + \alpha_{6F}Q_jH_iS_i^c + \alpha_{7F}\ln(t+1) + \alpha_\theta\theta_i + \epsilon_{Fit} \end{aligned} \tag{2.6}$$

This specification mirrors equation (2.1). As above, $X$ includes gender, ethnicity and parental SES; $S^m$ and $S^c$ are quantitative skills and communications skills respectively; $Q$ is course quality; and $H$ is a dummy for higher education attendance. The specification also includes an interaction between quality and skills to allow for match effects in higher education. All of the coefficients vary by $F \in \{\text{No University, STEM, LEM, AHSS}\}$ whether or not people attended higher education, and if they did, by field.

Whereas equation (2.1) was estimated on one cohort at a specific age, the full model includes earnings growth over the lifecycle. The age-profile of earnings differs across fields and is estimated outside of the model. This model also includes the unobserved skill $\theta$, discussed above. Finally, the error term, $\epsilon_{Fit}$, follows an AR(1) productivity process. Specifically, $\epsilon_{Fit}$ evolves according to $\epsilon_{Fit} = \rho_F\epsilon_{Fi,t-1} + \xi_{Fit}$, where $\xi$ is exogenous iid shock. The AR(1) process is characterised by three parameters: the persistence of productivity shocks ($\rho_f$), the variance of the iid shock ($\sigma^2_{\xi,F}$) and the variance of $\epsilon_{i,t}$ at $t = 0$ ($\sigma^2_{0,F}$). All three parameters are allowed to vary by $F$.

Student loans on graduation are equal to:

$$l_{i0} = \sum_{k=1}^{3}(1 + R_l)^k(T + M(SES))$$

where $T$ is tuition fees and $M$ is borrowing for living costs ("maintenance") that is dependent on parental SES. Student loans accumulate interest during study at the student loan interest rate $R_l$. Subsequently, loans evolve according to the following equation:

$$l_{it+1} = R_l l_{it} - P\left(d_{it} y_{it}, l_{it}\right) \tag{2.7}$$

$$l_{it} \geq 0 \tag{2.8}$$

until any outstanding loan is written off at the end of the repayment period.[18] The student loan repayment function, $P(.)$, which is based on pre-tax income, is given by:

$$P(y_{it}, l_{it}) = \min\{R_l l_{it}, \max\{0, (y_{it} - \phi)\tau\}\}$$

where $\phi$ is the loan repayment threshold and $\tau$ is the repayment rate. Income tax, $I(.)$, is a function of individual gross earnings. Agents are provided with a minimum income floor. Earnings above the income floor are taxed, with step-wise increases in the marginal rate of income tax at specific earnings thresholds. The marginal tax rates, earnings thresholds and income floor are all based on 2019 tax and benefit policy.

[18]To keep the notation tractable, we have omitted the detail that after 2012 student loan interest rates varied depending on whether people were still at university, and on their income once they left university. We do capture this detail in our simulation of the 2012 reforms, however.

### 2.4.3 Course preferences

Courses are characterised by their field, $F_j \in \{\text{STEM, LEM, AHSS}\}$, and their quality, $Q$. If student $i$ is matched to course $j$, the course receives utility:

$$W_{ji} = \gamma^q_{F_j} S^q_i + \gamma^c_{F_j} S^c_i + \gamma^X X_i + \gamma^\theta \theta_i + \eta_{F_j,i} \tag{2.9}$$

Courses are assumed to observe $\theta$, which is unobservable to the researcher. The introduction of $\theta$ therefore allows the model to capture correlations between earnings and sorting outcomes. Importantly, course preferences are allowed to vary by field. This is an extension of Agarwal (2015), which assumes common supply side preferences. Allowing universities to have field-dependent preferences over students is logical as students with different skill sets will match more appropriately to different fields. However, course preferences only vary across fields based on observables; the shock, $\eta_{F_j,i}$, is common to all courses.

Finally, as mentioned above, we assume that courses always strictly prefer filling places to leaving them unfilled. This is consistent with a model of higher education with high fixed costs but low marginal costs of admitting additional students.

### 2.4.4 Solving the matching problem

To describe how we solve the model, we need to introduce some notation. $\mu : N \to J$ is the match function, which assigns students to courses. The inverse of the match function, $\mu^{-1}(j)$, is the set of students matched to course $j$. From above, $W_{ij}$ is the value of student $i$ to course $j$. Let the minimum value of the student assigned to course $j$, $\min_{i' \in \mu^{-1}(j)}\{W_{i'j}\}$, be written as $\bar{W}^k_j$, where $k$ is the algorithm iteration. The algorithm we use to solve the model is:

1. Set $\bar{W}^0_j = -\infty$ for all j
2. Let all students select the course that they prefer given constraint $W_i > \bar{W}^k_j$
3. Count the number of students selecting each course, $tot_j$
4. For all courses with $tot_j > p_j$, where $p_j$ is the capacity constraint:

- Sort students who selected oversubscribed course $j$ by $W_i$
- Set $\bar{W}_j^{k+1}$ to the $W_i$ of the $p_j$th student

5. Repeat steps 2 to 4 until no universities have $tot_j > p_j$

This is a modified Gale-Shapley algorithm. In our setting, the resulting match is the unique stable equilibrium. Here, stability is equivalent to:

$$U_{i,j} > U_{i,\mu(i)} \implies \min_{i' \in \mu^{-1}(j)} \{W_{i'}\} > W_i$$

Stability guarantees that if student $i$ prefers course $j$ to their own course, then course $j$ is at capacity and there is no student on course $j$ that course $j$ ranks lower than student $i$. We assume that the equilibrium match in the data is also stable.

In practice, there is likely to be some strategic application behaviour that violates stability. Over the period studied, students could apply to five universities in the initial round of applications. It is possible that some students failed to apply for a course that would have been willing to accept them and that they preferred to their final match. However, most courses publicise their approximate requirements for accepting applications in advance, and students therefore have a reasonably good understanding of where they are likely to be successful. Five applications is also sufficient for students to include an insurance choice and still apply to four additional places that they would want to attend. Furthermore, students re-apply through the 'clearing' system which allows students a second chance at getting a place they want within the same application cycle. As a result, we think that assuming stability is a reasonable approximation in this market.

## 2.5 Estimation

### 2.5.1 Minimum distance procedure

We use a simulated minimum distance estimator, solving for $\hat{\Theta}$ such that:

$$\hat{\Theta} = min_{\Theta}(M - M(\Theta))'\mathbf{W}(M - M(\Theta))$$

Here $M$ is a vector of data moments and $M(\Theta)$ is the vector of simulated moments from the model. **W** is a k x k weight matrix, where k is the number of moments. We follow several papers (e.g., Blundell, Costa Dias, Meghir, et al. 2016) and use the diagonal weight matrix where each of the elements is equal to the inverse of the variance of each moment in the data (approximated using a bootstrap). We then follow French and Jones (2004) in our computation of asymptotic standard errors.

Each higher education market consists of a single cohort of school leavers. For estimation, we use the four cohorts of school leavers from 2006 to 2009. As described above, this was a relatively benign period for English higher education policy. We estimate the life-cycle earnings parameters outside the model, including parameters defining life-cycle earnings growth ($\alpha_{7F}$ in equation 2.6), the probability of employment, and the parameters defining the persistent earnings shock. We set the discount rate at 0.95.

### 2.5.2 Identification

A key insight from Agarwal (2015) and Diamond and Agarwal (2017) is that, in a many-to-one matching market, data on observed outcomes is sufficient to identify many parameters of interest. Application data is not strictly necessary. Nevertheless, we face several challenges with the identification of our model. First, we need to separate the preferences of the students from those of the programmes. Second, we need to identify returns to programmes in the labour market. [19] The subsections below describe the data features that assist in identification.

#### Excluded variables from earnings and university preferences

We include three types of instrumental variables (which we denote $Z$). Each of these are intended to shift student demand for courses, without directly affecting earnings or the preferences of courses for students. All instruments are defined at the student level. First, for

[19] One point of divergence from Agarwal 2015 is the incorportation of lifetime earnings. We include several earnings moments, including the parameters from the OLS earnings regression in equation 2.1.

each field, we include the average quality ($Q_j$) of courses in the relevant field located within 40km of the student's secondary school. Second, we include the distance from the student's secondary school to the nearest Russell Group university, nearest pre-1992 university nearest and nearest post-1992 university. These university groups are commonly used in the literature as broad proxies for university quality. Third, and again by field, we include the share of students in the previous cohort at each individual's school that studied in that field, conditional on attending university. In the model, we utilise these instruments by including a direct measure of distance to the relevant course and the share of peers selecting the same field directly in the individual's utility function, as outlined in equation 2.2. We also include coefficient estimates from regressions of these instruments on sorting outcomes as moments in estimation.

Table 2.4: Instrument power

| | Study STEM | Study LEM | Study AHSS | Uni. Quality |
|---|---|---|---|---|
| *Average quality of local unis* | | | | |
| STEM | -0.012* | -0.007* | -0.012* | -0.346* |
| | (0.001) | (0.001) | (0.001) | (0.01) |
| LEM | 0.011* | 0.007* | 0.005* | 0.298* |
| | (0.001) | (0.001) | (0.001) | (0.011) |
| AHSS | 0.002* | 0.001 | 0.009* | 0.152* |
| | (0.001) | (0.001) | (0.001) | (0.011) |
| *Distance to nearest* | | | | |
| Russell Group Uni | -0.002* | 0.000* | -0.003* | -0.008* |
| | (0.000) | (0.000) | (0.000) | (0.002) |
| Pre-1992 Uni | -0.003* | -0.001* | -0.003* | -0.034* |
| | (0.000) | (0.000) | (0.000) | (0.002) |
| Post-1992 Uni | -0.004* | -0.002* | -0.005* | -0.015* |
| | (0.000) | (0.000) | (0.000) | (0.002) |
| *Share of peers choosing* | | | | |
| STEM | 0.009* | 0.002* | -0.005* | -0.033* |
| | (0.000) | (0.000) | (0.000) | (0.002) |
| LEM | 0.01* | 0.016* | -0.001* | 0.09* |
| | (0.000) | (0.000) | (0.000) | (0.002) |
| Controls | ✓ | ✓ | ✓ | ✓ |
| F stat | 943.8 | 1707.5 | 280.3 | 1678.3 |
| N | 2,374,368 | 2,374,368 | 2,374,368 | 685,726 |

Note: * indicates $p < 0.05$. The first stages here are estimated on four cohorts of data (the 2005/06-2008/09 school leavers). Controls include SES, a private indicator, gender and our two skill measures.

Table 2.4 shows that these instruments shift educational choices. For the first three columns, the outcome variable is an indicator for studying in a particular field (including those who do not go to university in the regression), while for the final column the outcome is university quality (conditional on attending). In almost all cases the parameters are statistically significant at the 5% level, and the corresponding F statistics are large.

#### Sorting patterns

Sorting patterns reveal whether two different programmes (or two different students) are equally desirable (Diamond and Agarwal 2017). If two different student types are equally desirable, then the distribution of programmes they match to should be similar. These patterns help us identify how programmes value student characteristics, and how students value programme characteristics. We therefore include moments on the covariance between individual level characteristics of enrolees and the characteristics of the matched course.

#### Many to one matching

Agarwal (2015) argue that the comparison of within-course and between-course variation can be informative about the value courses place on particular student characteristics. Comparisons of this type are only possible within many-to-one matching markets. For example, if courses highly value students' quantitative skills then variation in quantitative skill within courses should be small compared to variation between courses. There will be positive assortative matching between quantitative skills and course quality, and as a result little overlap in the skill distribution of students at high quality and low quality courses. By comparison, if universities do not value a characteristic, then the distribution of that characteristic at each course should be similar. In this case, between-course variation of the characteristic will be much smaller than the within-course variation. We therefore include moments describing within and between variation in student characteristics by field of study and the correlations between one's characteristics and those of their peers on their course.

## 2.6 Parameter estimates and model fit

### 2.6.1 Parameter estimates

Tables 2.5, 2.6 and 2.7 report the model parameter estimates. Table 2.5 reports the student direct utility parameters. Direct utility is measured relative to the outside option of entering the labour market. The intercept terms show that AHSS courses give people the highest direct utility while at university, while LEM courses give people the lowest direct utility. Women get less utility from all three subject areas than men do, and they particularly dislike LEM courses. Direct utility increases strongly with parental SES for all subjects (particularly STEM and AHSS), and there is a further jump in utility for the privately educated for all fields. Students have a strong disutility from attending a university that is further away, all else being equal, but the negative effect of distance is much larger for students with low parental SES. Finally, students get more utility from studying in a given subject field if more of their school peers studied in that field.

Table 2.5: Student utility parameters

| | STEM | LEM | AHSS |
|---|---|---|---|
| intercept | 0.023 | -0.787 | 1.827 |
| | (0.0017) | (0.0033) | (0.0022) |
| female | -0.050 | -0.798 | -0.035 |
| | (0.0016) | (0.003) | (0.0018) |
| SES | 0.796 | 0.431 | 0.783 |
| | (0.0021) | (0.0032) | (0.0134) |
| private | 0.378 | 0.360 | 0.371 |
| | (0.0114) | (0.0038) | (0.0026) |
| $S^m$ | 0.010 | -0.013 | -0.007 |
| | (0.0022) | (0.0014) | (0.0018) |
| $S^c$ | 0.011 | -0.002 | -0.027 |
| | (0.0018) | (0.0015) | (0.002) |
| distance* | -1.265 | -1.265 | -1.265 |
| | (0.0008) | (0.0008) | (0.0008) |
| (distance x SES)* | 0.743 | 0.743 | 0.743 |
| | (0.0067) | (0.0067) | (0.0067) |
| (distance x private)* | 0.273 | 0.273 | 0.273 |
| | (0.0013) | (0.0013) | (0.0013) |
| share | 0.099 | 0.214 | |
| | (0.0015) | (0.0012) | |
| field shock* | 1.007 | 1.007 | 1.007 |
| | (0.001) | (0.001) | (0.001) |
| uni shock* | 0.569 | 0.569 | 0.569 |
| | (0.0008) | (0.0008) | (0.0008) |

Note: * indicates parameters that are not allowed to vary by field.

Table 2.6 shows estimated wage parameters from the model, comparing it directly to equivalent estimates from OLS estimation. The model estimates are generally quite similar to OLS but with some important differences. Our model estimates have much lower returns to course quality than those implied by OLS. For example, the model estimates that a one standard deviation increase in quality in STEM boosts earnings by 0.023 log points, compared to 0.032 for OLS. Similarly, a one standard deviation increase of quality in LEM increases earnings by 0.026, compared to 0.044 for OLS. There are similar decreases in AHSS. These are considerable reductions in returns.

The model also finds evidence of match effects. There are stronger match effects between quantitative skills and university quality in STEM degrees (0.030 for the model versus 0.016 for OLS) and also positive, but weaker, match effects between communication skills and university quality in AHSS (0.009 for the model versus -0.015 for OLS). There is no complementarity between communication skills and university quality in STEM.

These differences are likely related to the inclusion of the unobserved skill, $\theta$, in the structural model. $\theta$ has a large effect on wages, with a one standard deviation increase in $\theta$ boosting annual earnings by around 4.5%. Selection into higher quality courses due to high unobserved $\theta$ will generate upward bias in the OLS estimates for the return to quality. Another interesting result from including the unobserved heterogeneity terms is that it flips the sign on the privately educated dummy for those who do not attend university. This means that while OLS suggests that being privately educated is negatively related to earnings amongst non-graduates, our model suggests that this result is driven by unobserved selection.

Finally, Table 2.7 provides the university utility parameter estimates for the model. Importantly, the estimates suggest that STEM courses value quantitative skills much more highly than communication skills (1.57 versus 0.05), while the opposite is true for AHSS (0.108 versus 1.435). LEM courses value the two skill sets similarly. Universities also place a positive weight on unobserved skills that is quite small relative to the weights placed on observed skills. We also find that universities do not have strong preferences by gender or SES, but they do prefer privately educated students. This is perhaps unsurprising, as universities are not allowed to select on gender, while SES is quite hard for them to observe. They will observe whether somebody attended a private school, however. It is also possible that this coefficient is picking up something else that universities like that we are not explicitly modelling: for example, the privately educated might write much better personal statements, or might get better references letters from their teachers, both of which are included in their university applications.

Table 2.6: Wage parameters

| | No Uni | | STEM | | LEM | | AHSS | |
|---|---|---|---|---|---|---|---|---|
| | Model | OLS | Model | OLS | Model | OLS | Model | OLS |
| intercept | 9.897 | 9.901 | 10.081 | 10.076 | 10.109 | 10.144 | 9.989 | 9.954 |
| | (0.0007) | (0.002) | (0.0004) | (0.005) | (0.0005) | (0.007) | (0.0004) | (0.005) |
| female | -0.309 | -0.320 | -0.149 | -0.154 | -0.186 | -0.168 | -0.132 | -0.119 |
| | (0.001) | (0.007) | (0.0005) | (0.005) | (0.0005) | (0.002) | (0.0004) | (0.004) |
| SES | 0.324 | 0.213 | 0.199 | 0.201 | 0.209 | 0.355 | 0.232 | 0.229 |
| | (0.0013) | (0.005) | (0.0008) | (0.014) | (0.0008) | (0.018) | (0.0013) | (0.014) |
| private | 0.051 | -0.065 | -0.026 | 0.009 | -0.005 | 0.007 | 0.000 | 0.041 |
| | (0.0009) | (0.007) | (0.0009) | (0.006) | (0.0006) | (0.009) | (0.0005) | (0.006) |
| $S^m$ | 0.104 | 0.157 | 0.085 | 0.100 | 0.125 | 0.138 | 0.109 | 0.105 |
| | (0.0009) | (0.003) | (0.0006) | (0.007) | (0.0004) | (0.01) | (0.0006) | (0.006) |
| $S^c$ | 0.054 | 0.058 | 0.045 | 0.047 | 0.031 | 0.044 | 0.008 | 0.024 |
| | (0.0011) | (0.003) | (0.0006) | (0.006) | (0.0004) | (0.009) | (0.0005) | (0.006) |
| $Q$ | | | 0.023 | 0.032 | 0.026 | 0.044 | 0.018 | 0.029 |
| | | | (0.0003) | (0.002) | (0.0003) | (0.004) | (0.0003) | (0.003) |
| $Q * S^m$ | | | 0.030 | 0.016 | 0.008 | 0.040 | 0.021 | 0.017 |
| | | | (0.0005) | (0.004) | (0.0004) | (0.006) | (0.0004) | (0.004) |
| $Q * S^c$ | | | -0.002 | -0.013 | 0.008 | -0.032 | 0.009 | -0.015 |
| | | | (0.0005) | (0.004) | (0.0004) | (0.006) | (0.001) | (0.004) |
| $\theta$ | 0.044 | | 0.044 | | 0.044 | | 0.044 | |
| | (0.001) | | (0.001) | | (0.001) | | (0.001) | |

Note: $\alpha_\theta$ is fixed to be constant across all fields. As $\theta$ is unobserved, this is omitted from the OLS regressions. The OLS regressions are based on the 2005/06 school leavers.

Table 2.7: University utility parameters

| | STEM | LEM | AHSS |
|---|---|---|---|
| female* | -0.062 | -0.062 | -0.062 |
| | (0.0007) | (0.0007) | (0.0007) |
| SES* | -0.020 | -0.020 | -0.020 |
| | (0.0009) | (0.0009) | (0.0009) |
| private* | 0.295 | 0.295 | 0.295 |
| | (0.0009) | (0.0009) | (0.0009) |
| $S^m$ | 1.571 | 0.613 | 0.108 |
| | (0.0011) | (0.0012) | (0.0009) |
| $S^c$ | 0.052 | 0.823 | 1.435 |
| | (0.001) | (0.001) | (0.001) |
| $\theta$* | 0.051 | 0.051 | 0.051 |
| | (0.0005) | (0.0005) | (0.0005) |

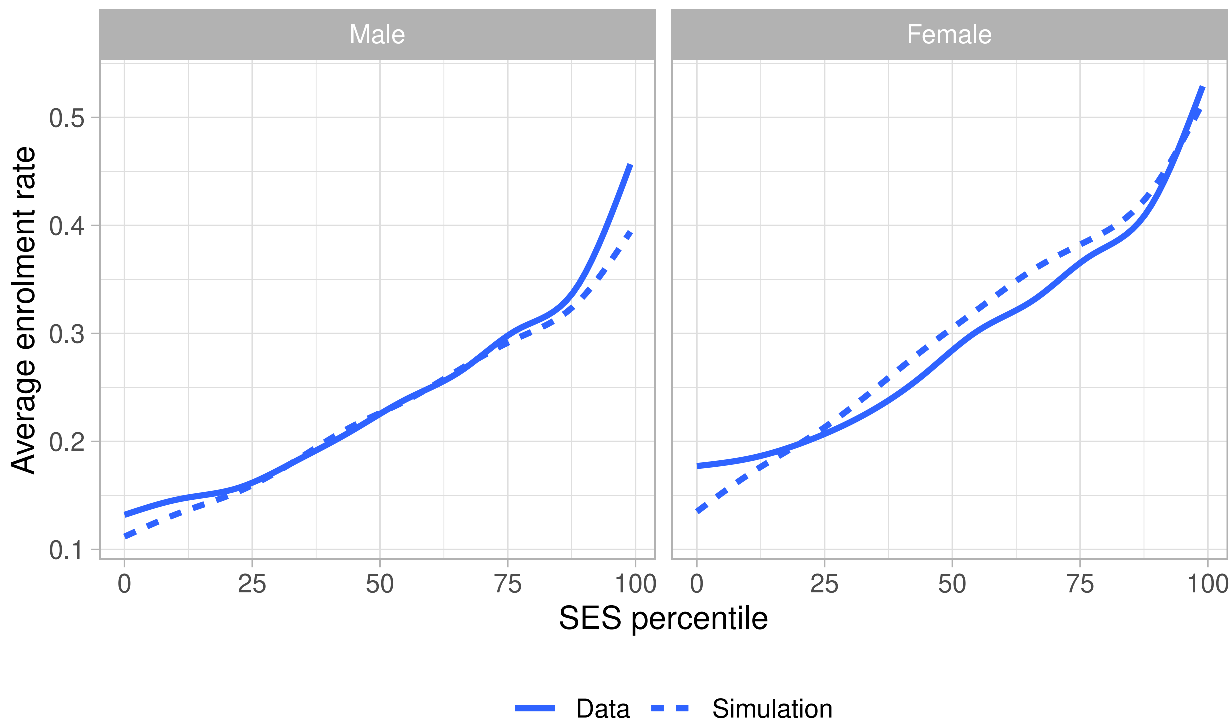
Note: * indicates parameters that are not allowed to vary by field.

### 2.6.2 In sample fit

This section shows that the model is able to replicate many of the important patterns in the data. Figure 2.8 shows university enrolment rates by SES, which we fit very well. Figure 2.9 shows subject enrollment by relative skills. We accurately replicate the positive relationship between relative advantages in quantitative skills and enrolment in STEM degrees.

Figures 2.10 and 2.11 show that the model is able to replicate broader sorting patterns by subject and by university quality. Finally, Figure 2.12 shows that the model replicates the relationship between parental SES and child's income.

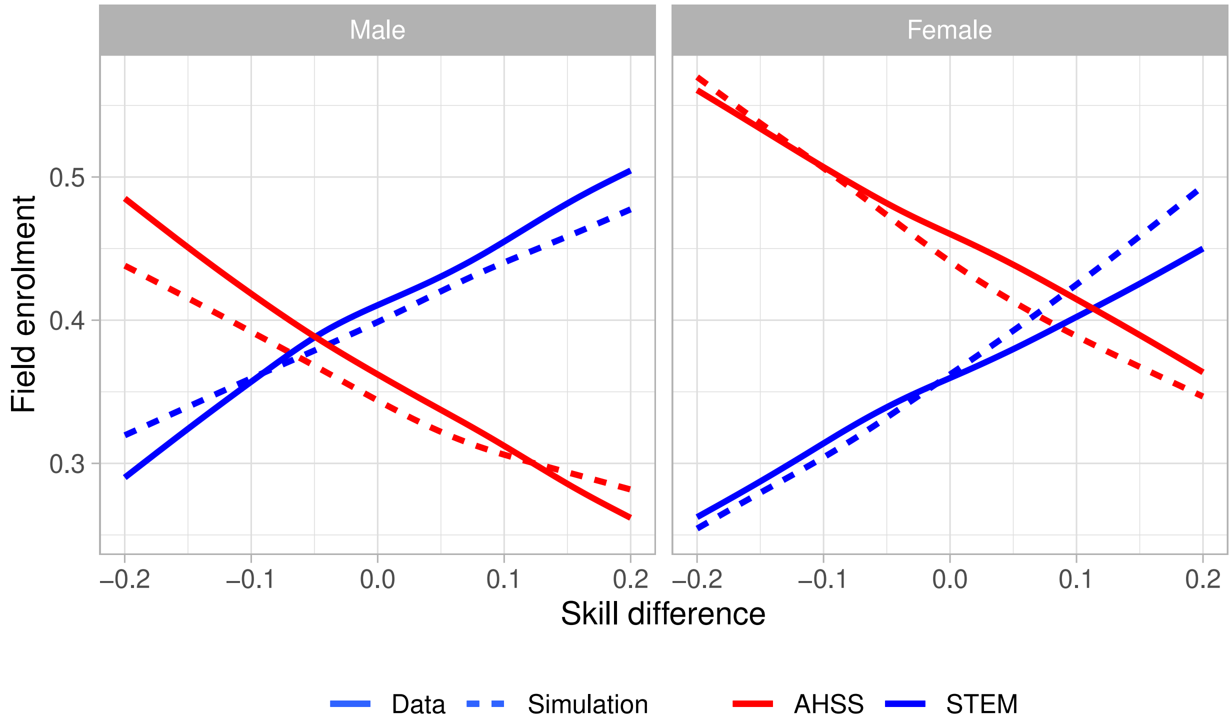
Figure 2.8: Enrolment by SES, by gender




### 2.6.3 Out of sample fit

Figure 2.13 then explores how our model performs out of sample. The points in the figure show the average quantitative skills by subject and university quality decile. The y axis is the model and the x axis is the data, meaning deviations from the identity line reflect differences between the model and the data.

Figure 2.9: Relative skills and subject choice, gender




As described above, the model is estimated on four separate markets, namely the 2005/06-2008/09 school leavers. The circles in the figure reflect in-sample data points from these four cohorts, while the triangles represent the out of sample fit from a market not included in the estimation (2009/10 leavers). Overall we observe a very strong correlation between the model and the data.[20] Furthermore, the out-of sample fit is certainly no worse than the in-sample fit, which means that our model is able to predict the match for markets other than those it is estimated on.

### 2.6.4 Validation of the model through the 2012 reforms

As described earlier in Section 2.3.6, the 2012 tuition fee reforms had no overall effect on participation, but this masked heterogeneity in responses by ability. In this section, we simulate the impact of the reforms using the estimated model. Figure 2.14 compares the change in participation rates through the reforms by quintile of quantitative skills, comparing

[20]This plot closely resembles a plot of model fit produced in Agarwal (2015). This plot and his are visually very similar in terms of fit. We get a very similar plot when we consider average communications skills.

Figure 2.10: Relationship between quant and comms skills within different courses

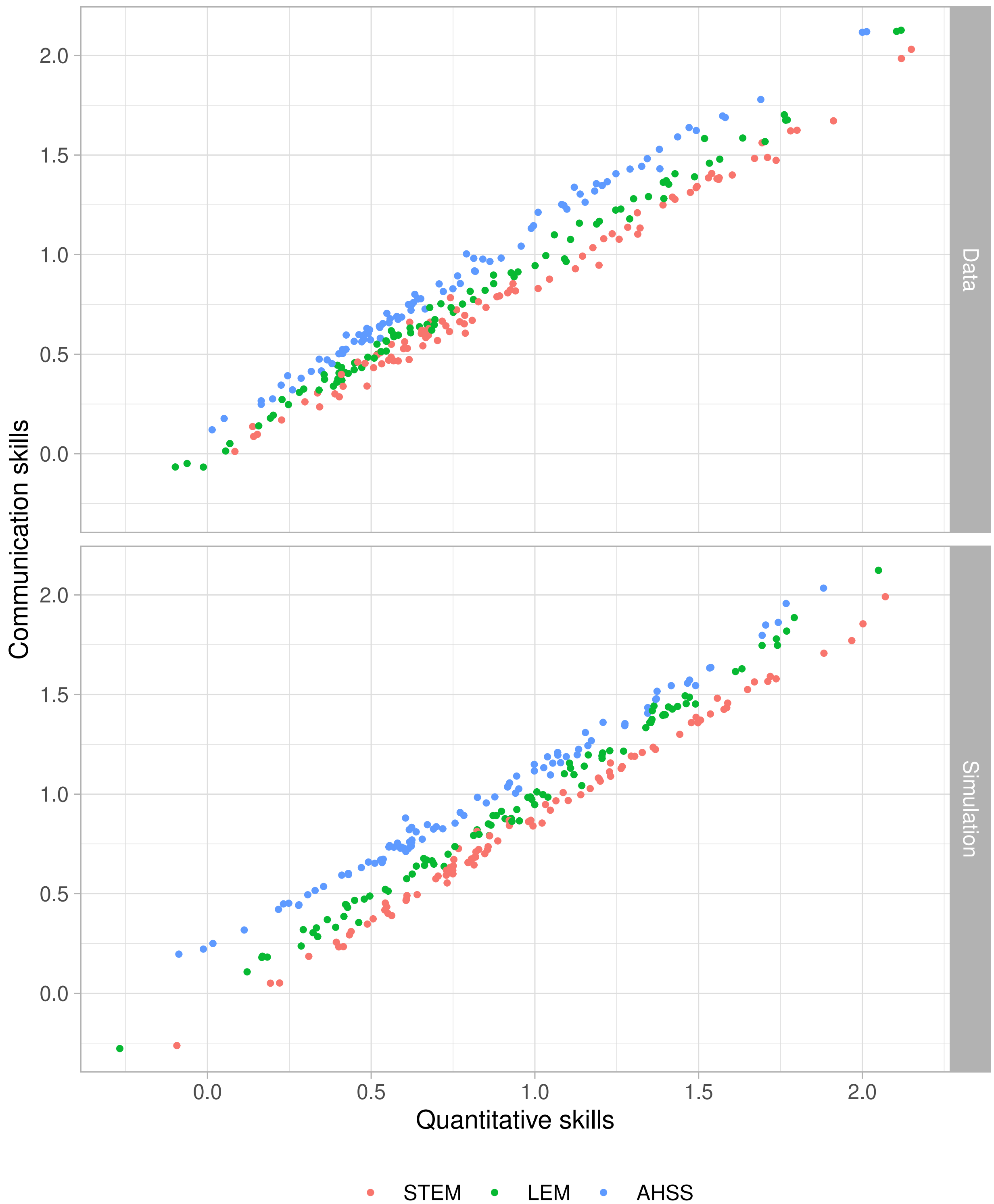


Figure 2.11: Relationship between quant and comms skills within different courses

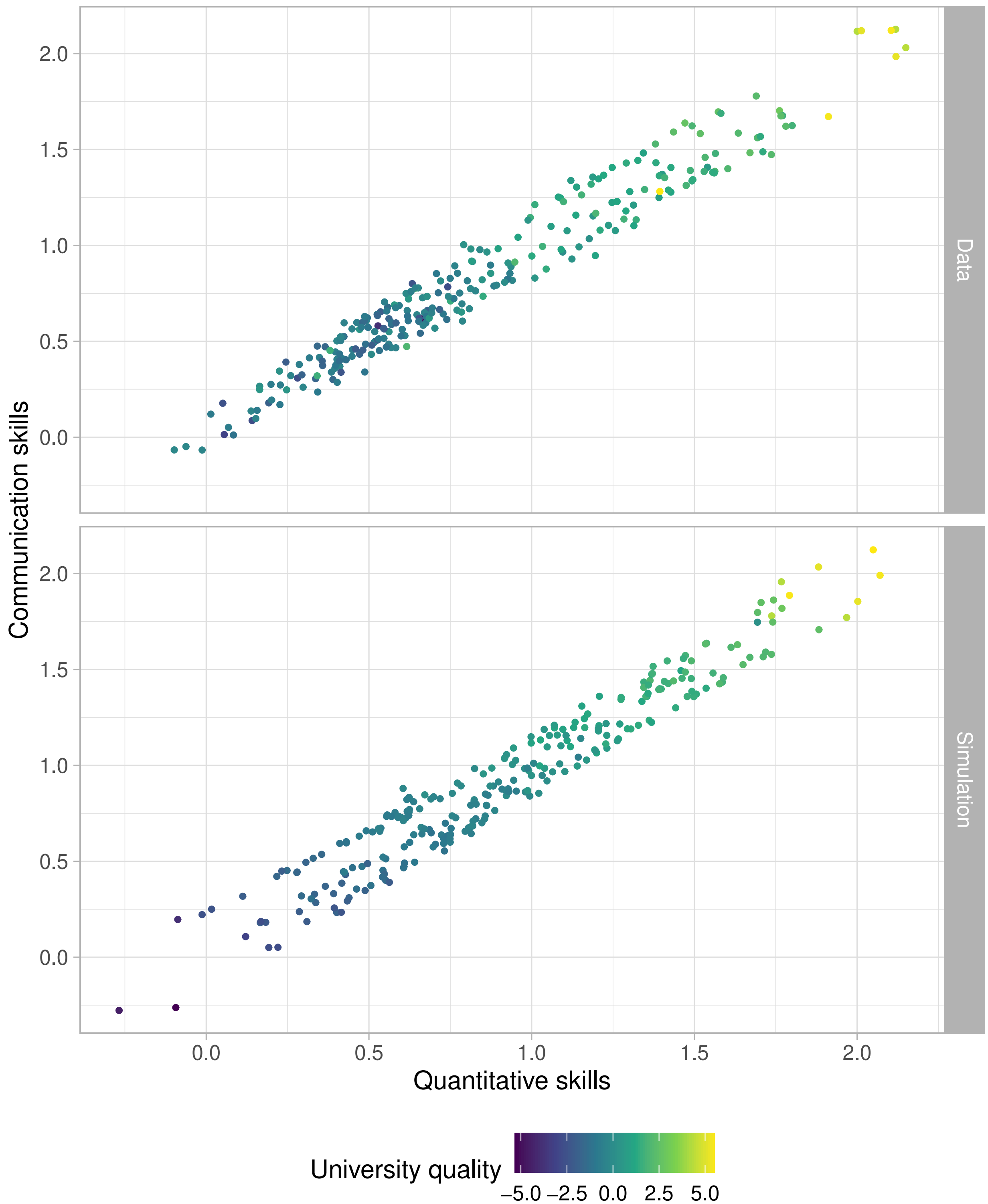
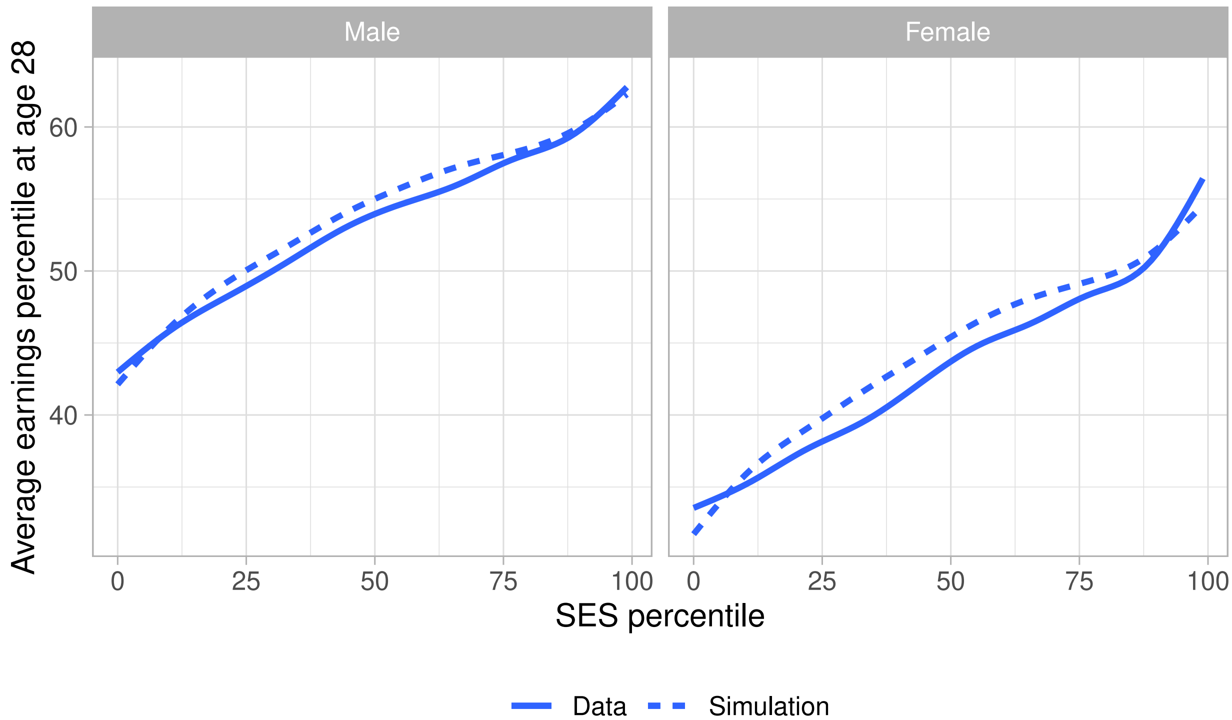


Figure 2.12: Intergenerational mobility, by gender




the data to the simulated effects.[21] Overall, we see that while the model effects are slightly smaller in the middle of the distribution, it is able to replicate the overall pattern of the reform effects extraordinarily well, especially considering that these effects are not formally targeted by the model.

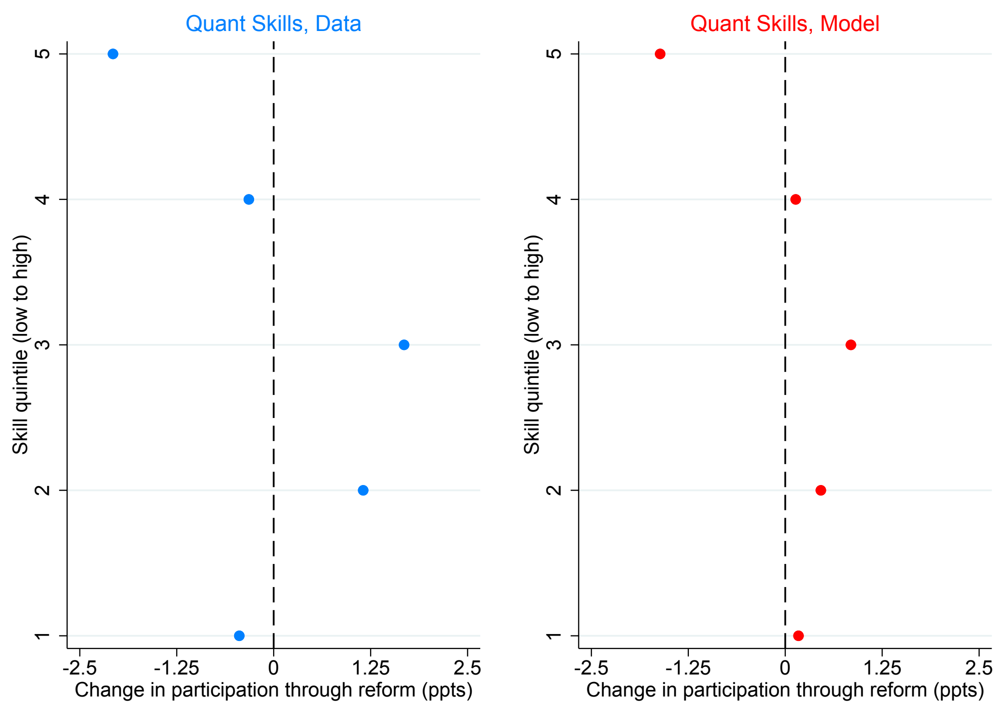
We then turn to unpicking these effects and considering their wider implications in Table 2.8. The top panel summarises the participation effects of the reforms, showing overall participation by parental SES, STEM participation (conditional on attending) by parental SES, and participation of higher ability (top quintile) students by parental SES. The model estimates that the reforms did not change overall participation rates, but did slightly narrow the gaps in participation between higher and lower SES students - this is consistent with previous evidence on the reforms (e.g., Azmat and Simion 2020). Consistent with the above we do see declines in participation amongst higher ability students, however, with the spaces

[21] Note that this differs slightly from Figure 2.8 as here we use quantitative skills rather than GCSE scores, and we demean the effects to allow for overall growth in participation rates. This does not affect the overall pattern of negative participation impacts at the top of the ability distribution and positive effects further down, but does make for a fairer comparison with the model, where overall student numbers and underlying cohort characteristics are held fixed. Note that the equivalent results for communications skills are provided in Appendix Figure A.2, which looks very similar.

Figure 2.13: In and out of sample fit: quantitative skills


Note: Circles show in sample points, triangles show out of sample points.

Figure 2.14: Reform effects by ability: model vs. data




Note: The data shows the raw changes in participation between 2010 and 2012, adjusting for overall growth in participation rates.

they vacate being filled by their lower ability peers.

Table 2.8: Summary of 2012 reform effects

| | Pre-2012 | Post-2012 |
|---|---|---|
| HE Part. (low SES) | 0.240 | 0.243 |
| HE Part. (high SES) | 0.406 | 0.403 |
| High Ab. Part. (low SES) | 0.834 | 0.820 |
| High Ab. Part., (high SES) | 0.847 | 0.830 |
| High Qual (low SES) | 0.180 | 0.181 |
| High Qual (high SES) | 0.299 | 0.300 |
| STEM (low SES) | 0.421 | 0.423 |
| STEM (high SES) | 0.427 | 0.426 |
| Undermatch (low SES) | 1.477 | 1.452 |
| Undermatch (high SES) | 1.265 | 1.235 |
| Intergen. Elasticity | 0.144 | 0.143 |
| High Ability Elasticity | 0.060 | 0.059 |

Note: All estimates are from model simulations. Low SES is based on the bottom 50% of the parental SES index. High ability is based on being in the top 20% of combined skills. The measure of undermatch is explained in the text. Average earnings are based on forecast earnings at age 40.

The second panel considers the effect of the reforms on academic undermatch. Our measure of undermatch here draws upon Campbell et al. (2021). Specifically, in the table we report the average difference between the individual's skill decile and the quality decile of the university they attend, for people in the top three deciles of skill (conditional on attending university). We split this by SES (top and bottom 50%), and consistent with Campbell et al. (2021), we see that low SES students are more likely to undermatch (as undermatch equals 1.477 for low SES students and 1.265 for high SES students). However interestingly, we also see that undermatch of higher ability students declined through the reforms. This suggests that the higher ability students who decided not to go to university as a result of the reform were people who were attending lower quality universities than their abilities suggest they could get into. As there is positive return to attending more selective institutions, this suggests that the people who were put off from university as a result of the reform were people with high earnings potential but relatively low returns to enrolling. This is consistent with Figure 2.6, which shows that it is the highest earning graduates who experience the biggest cuts to their returns to university from the reform.

The third panel then considers the implications for social mobility, showing the overall intergenerational elasticity was unaffected by the reform. This is also true the case when looking at the intergenerational elasticity amongst high ability students only.

## 2.7 Policy experiments

We now turn to our our policy experiments. We simulate six counterfactual policy reforms, with overarching aim of understanding the implications of each of these policies for social mobility:

- **Policy 1:** No student loans for low SES students - tuition fees set to £0 and maintenance loans converted to grants (but held at current levels)
- **Policy 2:** Additional maintenance grants of £4,000 per year, on top of existing loans, for low SES (bottom 50%) students
- **Policy 3:** As for Policy 2, but only if attending a high quality university.
- **Policy 4:** As for Policy 2, but only if studying STEM courses.
- **Policy 5:** A '10% rule' of preferential admission to those who graduated in the top 10% of their high school class.
- **Policy 6:** A 10% rule that applies only to low SES students.

The results of these simulations are summarised in Table 2.9. The three panels of the table contain information about participation, income mobility and average earnings, respectively. As in the previous sub-section, we include information on overall participation, participation of high ability students, participation at high quality universities, participation in STEM and academic undermatch (as defined above), all split by SES. Policy 1 (loan write-offs for low SES students) and Policy 2 (additional maintenance grants for low SES students) shift up the participation of higher ability students from poorer backgrounds by 1.5 and 3.5 ppts respectively, although the effect on overall participation of poorer students is much smaller than this (due to the lack of response from lower ability poorer students). The policies both result in a slight flattening of the relationship between parent and child's earnings at age 40 (see the 'High Ability Elasticity' row) but almost no effect on the overall intergenerational elasticity. One potential reason for this is that undermatch of poorer students increases under both policies, suggesting the students induced to attend university are not going to high quality (and hence high return) courses.

Policy 3, which conditions the grants on attendance at high quality institutions, does not have much effect on participation of poorer students, but it does increase the share of poorer

Table 2.9: Counterfactual policy reform effects

| | Baseline | Policy 1 Loan write-off | Policy 2 Maint. grants | Policy 3 Qual. grants | Policy 4 STEM grants | Policy 5 10% rule | Policy 6 Cond. 10% rule |
|---|---|---|---|---|---|---|---|
| HE Part. (low SES) | 0.232 | 0.235 | 0.240 | 0.232 | 0.237 | 0.246 | 0.247 |
| HE Part. (high SES) | 0.410 | 0.407 | 0.403 | 0.410 | 0.406 | 0.396 | 0.396 |
| High Ab. Part. (low SES) | 0.806 | 0.821 | 0.841 | 0.807 | 0.836 | 0.826 | 0.886 |
| High Ab. Part., (high SES) | 0.839 | 0.837 | 0.834 | 0.839 | 0.838 | 0.819 | 0.811 |
| High Qual (low SES) | 0.168 | 0.171 | 0.163 | 0.187 | 0.173 | 0.204 | 0.314 |
| High Qual (high SES) | 0.298 | 0.297 | 0.303 | 0.287 | 0.297 | 0.280 | 0.211 |
| STEM (low SES) | 0.429 | 0.436 | 0.431 | 0.433 | 0.546 | 0.438 | 0.433 |
| STEM (high SES) | 0.438 | 0.433 | 0.436 | 0.435 | 0.369 | 0.432 | 0.435 |
| Undermatch (low SES) | 1.453 | 1.474 | 1.570 | 1.349 | 1.687 | 1.103 | 0.126 |
| Undermatch (high SES) | 1.189 | 1.193 | 1.168 | 1.239 | 1.177 | 1.192 | 1.861 |
| Intergen. Elasticity | 0.158 | 0.157 | 0.157 | 0.158 | 0.152 | 0.153 | 0.141 |
| High Ability Elasticity | 0.063 | 0.060 | 0.058 | 0.061 | 0.043 | 0.039 | -0.001 |
| Top 20 Share (low SES) | 0.149 | 0.150 | 0.151 | 0.149 | 0.153 | 0.148 | 0.155 |
| Top 20 Share (high SES) | 0.251 | 0.250 | 0.249 | 0.251 | 0.247 | 0.252 | 0.245 |
| Av. Earns. (low SES) | £19,784 | £19,866 | £19,897 | £19,806 | £20,060 | £19,878 | £20,260 |
| Av. Earns. (high SES) | £26,582 | £26,532 | £26,497 | £26,563 | £26,293 | £26,077 | £25,646 |
| Av. Earns Gap (%) | 0.344 | 0.336 | 0.332 | 0.341 | 0.311 | 0.312 | 0.266 |
| Av. Earns Overall | £23,183 | £23,199 | £23,197 | £23,185 | £23,177 | £22,977 | £22,953 |
| $\Delta$ rel. to baseline (%) | - | 0.001 | 0.001 | 0.000 | 0.000 | -0.009 | -0.010 |

Note: All estimates are from model simulations. Low and high SES is based on the top and bottom half of the SES distribution, with the privately educated included in the top half. High Ab. refers to high ability people and includes those in the top 20% of skills (combined). High Qual. is a dummy for attending a high quality (top quartile) university. STEM is a dummy for doing STEM - each of the last two are given as shares conditional on entry. Undermatch is defined in the text in the previous section. All earnings outcomes are based on earnings at age 40.

students attending high quality institutions by 1.9ppts. Again, however, this does not have much effect on the intergenerational earnings elasticity, which is possibly due to the fact that the earnings returns to small changes in university quality are relatively small. Policy 4, which instead targets the grants at those who choose to study STEM courses, has a larger effect. The overall intergenerational elasticity falls from 15.8% to 15.2%, while the high ability elasticity falls by about a third, from 6.3% to 4.3%. These results arise from the fact that the STEM grants do appear to encourage quite a big shift towards STEM amongst poorer students, from 42.9% of enrolees to 54.6%. Interestingly, undermatch of poorer students actually increases with this policy, suggesting that crossing subject margins in potentially more important than moving up the university quality distribution. The policy is also associated with a small drop in the gap of the share of rich and poor students making it to the top 20% of the earnings distribution by age 40 (this is similar to the measure of mobility used in Chetty et al. 2020), and results in a small drop in the average earnings gap between richer and poorer students.

Policy 5, the unconditional 10% rule has a similar effect on mobility to the STEM grants plan. However, it achieves this in a very different way, dramatically reducing the gap in undermatch between richer and poorer students. Unlike the STEM policy, this comes with some efficiency loss, as it reduces assortative matching between high ability students and high quality courses.

To further increase the impact of the 10% plan, Policy 6 gives priority admissions to students who are *both* in the top 10% of their class and low SES. This policy achieves a sizeable drop in the intergenerational elasticity (from 15.8% to 14.1%), a complete removal of any relationship between parental and child's income amongst poorer students (from 6.3%), a 12% reduction in the gap between share of low and high SES students in the top 20% of earners, and a drop in the average earnings gap between rich and poor students of around a quarter (from 34.4% to 26.6%). A key driver of this is the dramatic increase in the share of poorer students attending high quality institutions, from 16.% to 31.4%. The policy is so much more effective that the simple percent plan because the English secondary education system is actually less segregated than one might expect. Surprisingly, the drop in overall average earnings is small at just 1%. This suggests very large equity improvements can be achieved with only very minor efficiency losses.

## 2.8 Conclusion

In this paper we develop and estimate an empirical matching model of sorting in the UK's higher education system. The model is able to replicate sorting patterns in the data extremely well, both for the markets included in estimation and for markets excluded from the estimation. It also does an excellent job of replicating descriptive patterns observed from the 2012 reforms.

The ability of our model to emulate these patterns enables us to confidently simulate counterfactual policies. We do this with the overarching aim of trying use higher education policy levers to boost social mobility. We find, contrary to much of the perceived wisdom on the subject, that policies focused on the demand side (cutting fees or boosting short term financial support for poorer students) do not have much impact on social mobility. On the other hand, we find that an aggressive policy targeting the supply side - namely, forcing universities to give priority admission to poorer students that score in the top 10% of their secondary school class - does have a significant effect on social mobility. Despite this policy resulting in much lower-ability poorer students getting into good universities (and wealthier students going to worse universities), the associated efficiency cost is small.

Importantly, all of these policies take skills as fixed initial conditions, not allowing for the fact that higher education policy might feedback into effort and choices in school. These feedback effects could work in either direction, and developing the model further to account for these feedback effects could be a promising subject for future research. The model could also be applied to study other dimensions of inequality, such as inequality in outcomes by gender or ethnicity.

# Chapter 3

# Human Capital Accumulation and Benefit Design

## 3.1 Introduction

Two alternative models for human capital accumulation over the life-cycle are common in the economics literature. The first posits that human capital accumulation is a by-product of labour and that there is no trade-off between earning and learning - I refer to this model as *learning-by-doing*. The second instead has human capital accumulation competing with productive labour, so workers must divide their time between learning and earning - I refer to this model as *Ben-Porath*.

Several recent papers have explored the interaction of these alternative models with optimal income and capital taxation[1]. However, there has been little work on their contrasting implications for the design of benefits. Individuals receiving benefits temporarily face high effective tax rates while benefits are withdrawn. The opportunity cost of human capital accumulation within a *Ben-Porath* model is foregone labour and therefore, all else equal, high effective tax rates *decrease* the cost of acquiring human capital for individuals on benefits. The opposite is true in a *learning-by-doing* model, where high effective tax rates may discourage labour supply and therefore discourage accumulation of human capital. Reductions in the rate at which benefits are withdrawn could therefore be expected to increase wages in

[1]For instance, Blandin and Peterman (2019) and Stantcheva (2017)

a *learning-by-doing* model, but decrease wages in a *Ben-Porath* model.

In practice, the human capital response to benefit policy depends on many factors, including income effects, borrowing constraints, and expected effective tax rates and labour supply in subsequent periods. I therefore develop this insight further by estimating several dynamic models of labour supply and skill investment, each nesting an alternative mechanism for human capital accumulation. In contrast to earlier work, such as Heckman, Lochner, and Cossa (2003), I incorporate many of the features necessary to capture realistic trade-offs in benefit design, such as human capital risk, credit constraints, household formation, permanent heterogeneity and exogenous changes in family composition and partner income.

I find that, in line with the basic intuition described above, a revenue neutral reform that reduces the effective tax rate for individuals receiving benefits increase wages within a *learning-by-doing* model, whereas the same reform discourages skill investment and therefore decreases wages within a *Ben-Porath* model. I also find that matching the same data with the alternative models generates large differences in implied labour supply elasticities. In a *learning-by-doing* model, earnings from labour are only part of the return to labour supply; women also want to work to increase their human capital. Labour supply is therefore relatively unresponsive to changes in wages, particularly early in working life. By contrast, in a *Ben-Porath* model, women can substitute to and from skill investment in response to wage changes, and labour supply elasticities are therefore higher. The differences in labour supply elasticities between the two models reduce as women approach retirement since human capital is less important when fewer periods of working life remain.

However, I also find that both *Ben-Porath* and *learning-by-doing* models fail to replicate some key features of women's observed profiles of wages and time-use. Neither model captures the growth in wages that are observed in the first few years of working life. The *learning-by-doing* model also over-predicts labour supply early in the life-cycle, whereas the *Ben-Porath* model under-predicts skill investments over the same period.

I therefore supplement the two pure models with a mixed model, which incorporates both modes of human capital accumulation. I find that this model provides a much better fit to observed profiles. Behaviourally, the mixed model has similar implications to a *Ben-Porath* model early in working life. Since wages are relatively low and many periods of working life remain, women are willing to utilise costly skill investment to accelerate wage growth. As women age, they increasingly rely on work experience as the primary source of human capital, and behaviour in the mixed model tends towards a *learning-by-doing* model.

Reducing effective tax rates in the mixed model decreases wages early in working life, but increases wages at older ages.

I conclude that both modes of human capital accumulation are necessary to explain observed behaviour, and that reforms that improve work incentives for low-income workers may decrease wages, particularly among the young.

I estimate all models using UK household panel data. I model the decisions and outcomes of women, since their labour supply decisions are more likely to respond to the design of the benefit system. Women also have greater variation in labour supply across the life-cycle, which is useful for validating alternative human capital models. I observe time spent in training at work alongside more formal educational investments such as university courses. This allows me to characterise time investments in education across the life-cycle, which I match when estimating models with costly skill investment. To the best of my knowledge, the only previous paper to make use of observed time investments in education to assess a *Ben-Porath* type model is Blundell, Costa Dias, Goll, et al. (2021)[2].

The paper is structured as follows. Section 2 reviews the relevant literature. Section 3 outlines the data, and presents descriptive information on wages, work and education. Section 4 describes the structural model, including the alternative human capital functions. Section 5 provides intuition on the interaction between the human capital accumulation function and the impact of benefit policy. Section 6 outlines the estimation methodology. Section 7 presents the estimation results for the *Ben-Porath* and *learning-by-doing* models, discussing both parameter estimates and the model fit. Section 8 presents estimation results for the mixed model, comparing both the parameters and model fit to each of the prior models. Section 9 presents the policy experiments. Section 10 concludes.

## 3.2 Literature Review

A growing literature analyse the interaction between human capital accumulation and taxation. This literature can broadly be divided into two strands. Several recent papers within the dynamic public finance literature adopt a Mirlessian approach, considering the interaction between optimal non-linear tax functions and endogenous human capital accumulation. Examples of these papers include Bohacek and Kapicka (2008), Kapička (2015), Kapička

[2] Included in this thesis as Chapter 4

and Neira (2019), and Stantcheva (2015, 2017). Among these papers, Stantcheva (2015) is most related to the questions analysed in this paper. She derives expressions for the optimal labour wedge and training wedge in a dynamic model with observable training investment and idiosyncratic risk. The degree of substitutability between labour supply and training is parameterized such that pure *learning-by-doing* and *Ben-Porath* models are incorporated as limit cases of the general model. In simulations, she finds that the labour wedge is smaller and grows slower over the life-cycle when human capital accumulation is *Ben-Porath* as compared to *learning-by-doing*. In both pure models, the optimal training subsidy almost completely compensates for the distortion to training investments resulting from labour taxation, such that the net wedge on training is close to zero. This mirrors earlier results in Bovenberg and Jacobs (2005) who found that, in a model without idiosyncratic risk, optimal education subsidies should perfectly compensate for tax distortions on learning. It is unclear how the labour wedge in either of the pure models in Stantcheva (2015) would differ if educational investments were unobservable and, therefore, distortions to training resulting from labour taxation could not be neutralised.

Another strand of the literature considers a more limited set of tax instruments. These papers include Blandin and Peterman (2019), Costa and Santos (2018), Karabarbounis (2016), and Peterman (2016). Of these papers, Blandin and Peterman (2019) has a similar comparative approach to this paper, and consider how the optimal level of capital taxation differs between pure *learning-by-doing* and *Ben-Porath* models. They develop a life-cycle model featuring permanent heterogeneity in learning ability, but without demographic transitions or risk to human capital. They find that optimal capital taxation is significantly higher if human capital is accumulated through *learning-by-doing*. As with the papers listed above, they do not consider the extent to which their estimated model is compatible with data on time-use or educational investments.

Finally, a much smaller literature has considered the interaction between human capital accumulation and benefit design. Early work by Heckman, Lochner, and Cossa (2003) studied the impact of the EITC on incentives to accumulate human capital, considering both *learning-by-doing* and *Ben-Porath* type models. The authors estimate corresponding life-cycle models with no heterogeneity or risk. They find that introducing the EITC decreases earnings in both cases, but has much larger effects within the *learning-by-doing* model.

Blundell, Pistaferri, and Saporta-Eksten (2016) estimate a *learning-by-doing* life-cycle model using a sample of working-age women in the BHPS. They argue that variation in the benefit system is a key source of identification within the model, since different individuals are exposed to a varying history of benefit systems that shift them exogenously into and out of work. Their model incorporates similar demographic transitions and partner earnings processes to those presented in this paper, as well as idiosyncratic productivity risk. Blundell, Costa Dias, Goll, et al. (2021) extends the paper to incorporate costly training investment. While both papers explore some potential reforms to the benefit system, they do not consider how the trade-offs in benefit design differ depending on the assumed human capital accumulation mechanism. In addition, the model estimated only incorporates rough adjustments to labour supply and training; women either work full-time, part-time or not at all, and when in work they have an additional discrete choice on whether to participate in training. As such, the model cannot capture the impact of benefits on the intensive margin.

## 3.3 Data

In estimation, I use the UK Household Longitudinal Study (UKHLS). This survey combines the British Household Panel Survey (BHPS), which was collected between 1991 to 2008, with Understanding Society (USoc), which is an ongoing panel survey collected since 2009. My sample includes all 18 waves of the BHPS and waves 3 to 9 of USoc. I exclude the initial two waves of USoc from my sample, as they coincided with the financial crisis.

5,500 households were included in the initial wave of the BHPS. All individuals within these households remain in the sample in all subsequent waves, except for those lost due to attrition. Data is collected from all adults and children within the household of an original sample member, whether or not they were present within the initial wave. Children who are born to original sample members over the course of the survey are also incorporated within the sample.

USoc follows a similar methodology to BHPS but with a larger initial sample size of 40,000 households. The questionnaire includes many of the same questions. This allows me to construct equivalent measures across both surveys. From wave 2 of USoc, the remaining BHPS sample are also included within the sample. I therefore have up to 25 observations per individual, though the majority of my sample are observed on fewer than 8 occasions.

Both data sources are annual, with rich data on earnings, labour supply, training and educational investments. The longitudinal design allows me to observe fertility, household formation and dissolution. I focus on women between the ages of 22 and 55 who have completed full-time education and entered the labour market. I also incorporate data on their partner's earnings and labour supply where appropriate. The full, working age data set includes 23,585 individuals and 148,066 observations.

Table 3.1: Sample information

| | BHPS | USoc | Combined |
|---|---|---|---|
| Total observations | 61,912 | 86,154 | 148,066 |
| # of individuals | 8,615 | 17,845 | 23,585 |
| Average observations per individual | 7.19 | 4.83 | 6.28 |

Notes: # of individuals indicates the number of distinct women included in the sample. Data collected from additional household members is used in estimation, but these household members are not incorporated into the main sample and therefore not counted in this table. 2,875 women appear in both the BHPS and USoc sample.

### 3.3.1 Time-use measurement

Throughout the paper, I utilise data on time spent on labour supply and time spent on education. I briefly document the construction of these variables below.

All employed individuals are asked both "how many hours are you expected to work in a normal week?" and "how many hours overtime do you usually work in a normal week?". For observations in this category, labour supply hours are the sum of the two measures. I drop all observations where the respondent reports having two or more jobs since labour hours are, in that case, only collected for the "main" job. This affects less than 1% of the employed sample each wave. Self-employed individuals are only asked "how many hours in total do you usually work in your job?", which I take as the relevant measure. The hours of both employed and self-employed are adjusted to account for time in training, as described below.

In both the BHPS and USoc samples, all individuals are asked annually whether they have engaged in any part-time or evening courses, training provided by an employer, day release schemes, apprenticeships or government training schemes since the last interview. If they have, they report the number of such courses they have taken and provide detailed information on the provider, purpose and length of training for up to three training schemes. If they report more than three episodes of training, they are asked to provide detailed information on the three *longest* courses they have taken. 98% of respondents report three or fewer courses.

Total hours of education is measured as the sum of training hours in each of the (up to) three episodes reported. If training is reported as taking place at work, or if the purpose of training is reported as either "to help you get started in your job", "improve your skills in your current job" or "maintain professional status", then it is assumed that the education hours reported are also reported within the estimated number of hours worked per week. I therefore subtract the hours reported from these training episodes from the labour hours measure discussed above.

Finally, individuals also report if they are currently engaged in full-time education. As noted above, I only include women who have left full-time education and entered the labour market within my sample, but I do not exclude those who subsequently return to full-time education later in their career. For these individuals, I set the number of hours of education to 40 per week. Less than two per cent of the working age sample are in full-time education. I make no adjustment to their reported hours of labour supply.

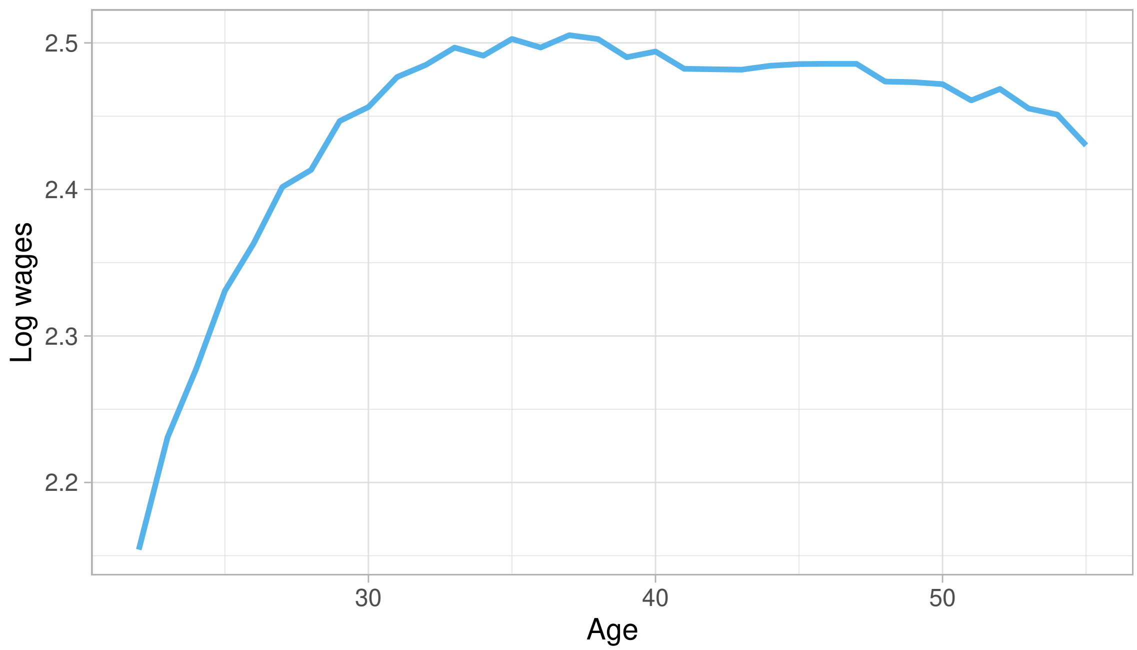
The resulting measure of education hours is closely related to the training hours constructed in Blundell, Costa Dias, Goll, et al. (2021), with two key differences. First, in that paper, individuals who return to full-time education are excluded from the sample. Here, I instead keep them in the sample and include their investments in my measure of education hours. Second, in Blundell, Costa Dias, Goll, et al. (2021), we consider only educational investments that occur at work. We therefore do not include training that is reported as unrelated to current or future jobs within measured training hours. Here, all training is included in our measure of education, regardless of the reported purpose.

### 3.3.2 Life-cycle wages, labour supply and human capital investments

Figure 3.1 shows the life-cycle log-wage profile of women in our sample. Wages increase rapidly for the first ten years of working life, before plateauing at approximately 34. Wages then decline gradually for the remainder of our sample.

Figure 3.2 shows the life-cycle labour and education hours. Labour hours start low and increase rapidly, peaking at age 25. They then decline between ages 25 and 35, which coincides with years in which many individuals within our sample have a young child in the household. Labour hours subsequently recover, before declining again as they approach

Figure 3.1: Women's average log wage by age



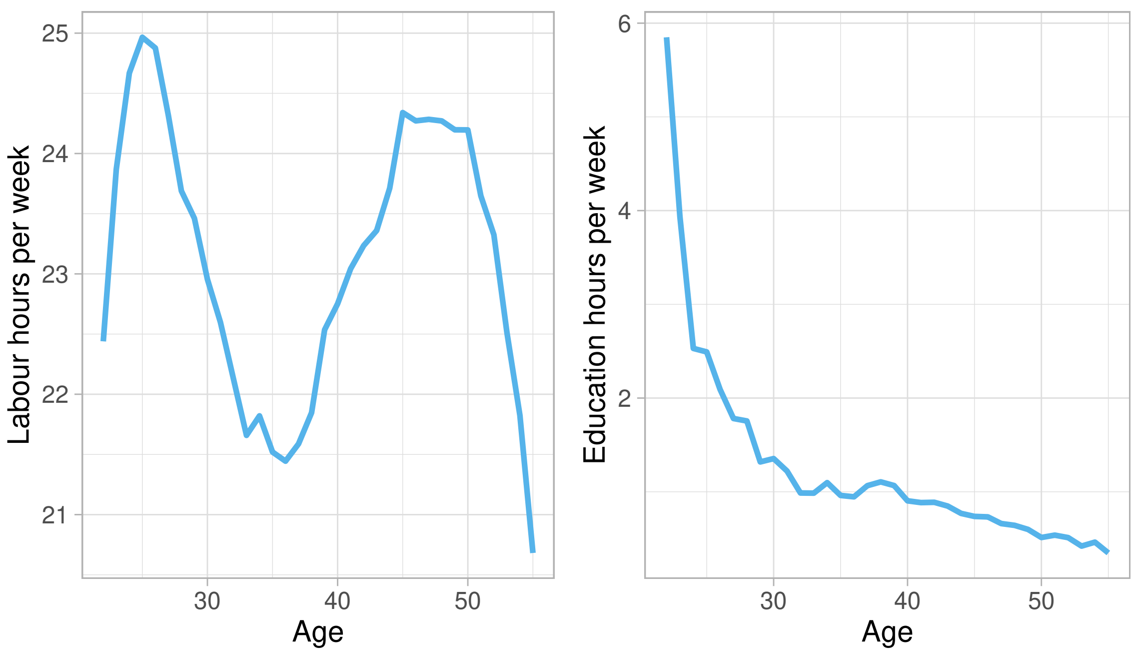

Source: UKHLS. Line depicts unweighted average of log-wages conditional on each age. Average log-wage is calculated for sub-sample with positive work hours at each age. Wages are corrected for inflation using CPI, such that the measured wage is in 2021 prices.

retirement. This contrasts with the labour supply patterns of men, which generally follow an inverse-U shape across the life-cycle. The additional variation in the working hours of women over the life-cycle offers a more interesting setting for testing the performance of alternative models of human capital accumulation, particularly *learning-by-doing* models that imply a direct link between recent work experience and wages.

Average education hours, on the other hand, declines monotonically throughout our sample. Initially the decline in investment is rapid, falling from 10 hours per week to around 1 hour per week by age 30. Investments then plateau for several years, before slowly declining in the final two decades of working life. Unlike labour hours, education hours do not appear to respond strongly to life-cycle changes in demographics.

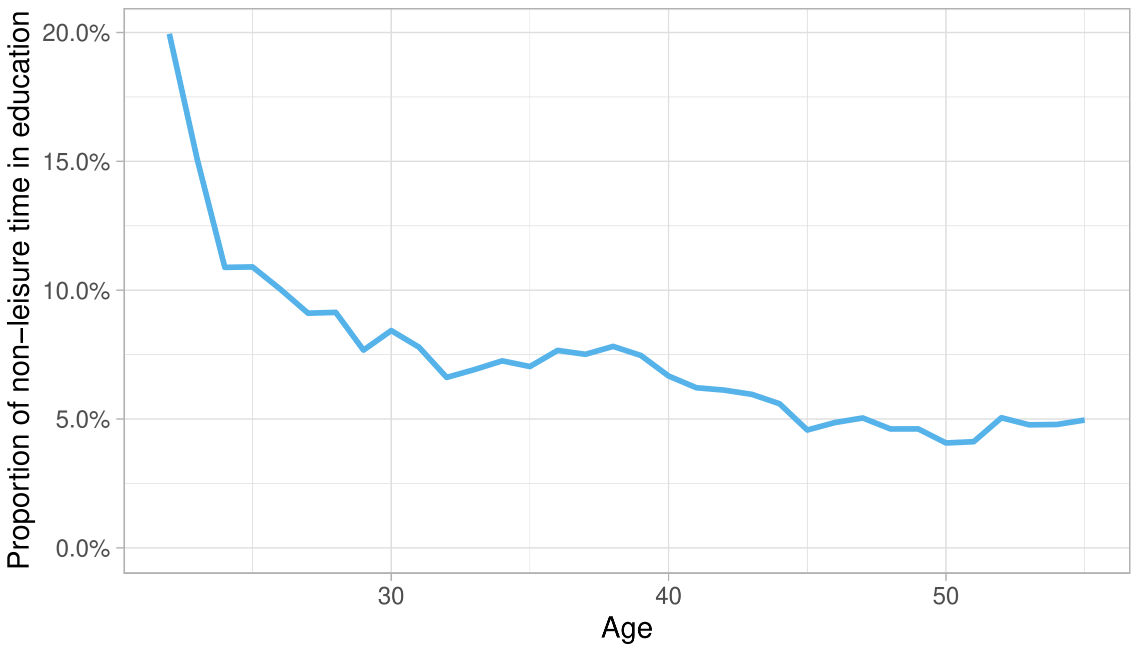
Although these investments are small in absolute terms, they still constitute a reasonable proportion of non-leisure time, which we define as the sum of labour hours and education hours. Figure 3.3 shows proportion of non-leisure time spent in education for the sub-sample with positive non-leisure time at each age. At age 22, women spend around one-fifth of their non-leisure time in education. This declines to around 10% by age 25, and then gradually falls to 5% by age 55.

Figure 3.2: Women's average labour and education hours by age




Source: UKHLS. Line depicts unweighted average of labour and education hours conditional on each age. The definition of labour hours and education hours is discussed in Section 3.3.1. Averages are taken over the full sample, including individuals who report no work or education hours.

Figure 3.3: Education hours as a proportion of total non-leisure hours




Source: UKHLS. Non-leisure hours are defined as the sum of labour hours and education hours. Proportion of non-leisure hours in education at each age is calculated for the sub-sample with positive non-leisure hours at that age. This differs from the sub-sample in Figure 3.2, where women with zero labour and education hours are included.

These moments constitute the major life-cycle patterns that I will attempt to replicate with each model. Equivalent information will be included in the moments matched in estimation, as discussed in Section 3.6, alongside additional moments that capture the variance and co-variance of wages, labour hours and education hours. The following section describes the model in detail.

## 3.4 Model

### 3.4.1 Model outline

Women enter the model at age 22. At this age, they differ only in their initial human capital ($h_0$), whether they are in a couple and/or have children ($f_0$) and their permanent learning ability ($a$). All women start with no financial assets ($k_0 = 0$). Each year, they determine the time they dedicate to labour ($n_t$) and education ($e_t$). Their total income depends on the amount of time they spend working, their human capital, their financial assets and, if they are in a couple, the income that their partner earns in the labour market. Both partners labour earnings are taxed individually, and then benefit eligibility is assessed at the household level based on total post-tax income and family demographics. They decide how much of their net income to consume ($c_t$), with the remainder saved in safe financial assets ($k_{t+1}$).

Each period, they accumulate human capital depending on their existing stock of human capital, their time use and the human capital production technology. Between periods, their human capital stock is hit by a shock ($z_t$). The average value of the shock is negative ($\mu_z < 0$) ensuring that, on average, their human capital will depreciate between periods.

Demographic transitions also occur between periods. Family type indicates whether there is an additional adult in the household and the age of the youngest child. Transitions occur stochastically across the life-cycle according to a exogenous process that depends on age and current family type. If the woman has a partner, their partner's income is also modelled as a persistent, stochastic process that depends on the woman's age and partner's income last period.

Women retire exogenously at age 65. Partners (if present) are assumed to retire at the same age. From this point on, no members of the household earn additional labour income. Households are all childless, but the number of adults within the household can change. These transitions implicitly model mortality of partners as well as separation and formation of new partnerships. For the remaining periods, the household receives a public pension (or two public pensions, if two adults are present) and consume their remaining financial assets.

The following subsections provide a full description of the model. In Appendix B.1, I discuss details of the model solution.

### 3.4.2 Per-period utility

I adopt an additively separable per-period utility function. Utility depends on consumption ($c_t$), time spent working ($n_t$), time spent in education ($e_t$) and family type ($f_t$):

$$u(c_t, n_t, e_t, f_t) = \frac{(c_t/\psi(f_t))^{1-\gamma_1}}{1-\gamma_1} - \chi(f_t)\frac{(n_t + s_t)^{1+\frac{1}{\gamma_2}}}{1+\frac{1}{\gamma_2}}$$

$\gamma_1$ and $\gamma_2$ parameterise the curvature of utility with respect to consumption and non-leisure time, respectively. $\psi(f_t)$ is the consumption equivalence scale, which depends on the number of adults and children in the household and is based on the OECD scale. Finally, $\chi(f_t)$ scales the cost of non-leisure time, depending on the family type. This allows the dis-utility of labour or education to increase if, for instance, there is a young child in the household. The utility function remains the same in retirement, although $n_t$ and $e_t$ are both zero once the individual is retired.

### 3.4.3 Demographics

Household demographics evolve according to an exogenous, age-dependent Markov chain.

$$Pr(f_{t+1}|\Omega_t) = g_t(f_t)$$

For tractability, I include six family types. These family types capture whether there are one or two adults in the household, and whether there is a child aged 5 or under in the household, a child under 12 or under in the household or no children aged 12 or under in the household. I find that these family types are sufficient for the model to provide a reasonable fit to labour supply patterns over the life-cycle. Including an additional group for households whose youngest child is aged between 13 and 18, for instance, has almost no impact on the fit of the model.

### 3.4.4 Budget constraint

I assume that markets are competitive. Women are paid a constant rental rate per unit of human capital provided per hour. I normalise this rental rate to one, and use $h_t$ to denote the money-metric measure of human capital. The data analogue to $h_t$ is therefore real wages. Women's gross earnings from labour are a function of hours spent working ($n_t$) and human capital ($h_t$):

$$y_{f,t} = h_t \times n_t$$

If there is an additional adult in the household, their wage follows a stochastic process with a polynomial trend in women's age and a persistent shock:

$$\begin{aligned} \ln h_{m,t} &= \beta_0 + \beta_1 t^{\frac{1}{2}} + \beta_2 t + \epsilon_t \\ \epsilon_t &= \rho \epsilon_{t-1} + \xi_t \end{aligned}$$

$\rho$ captures the persistence of partner income. $\xi_t$ are random innovations to partner productivity and are distributed $\mathcal{N}(0, \sigma_\xi)$. The polynomial is a function of women's age, rather than age of the partner, to economise on variables in the state space.

In addition, a partner may be unemployed. The probability of being unemployed is an exogenous function of women's age. Employed partners are assumed to work forty hours a week.

The full budget constraint is given by:

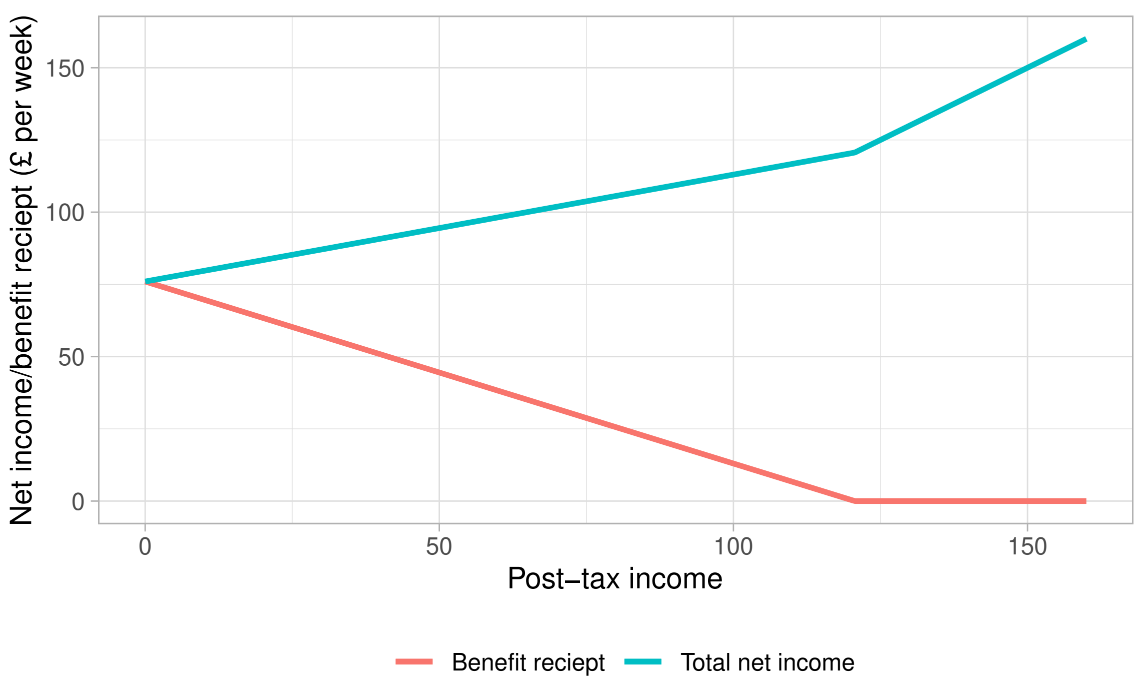
$$c_t + k_{t+1} = T_l(y_{f,t}) + T(y_{m,t}) + B\left(T(y_{f,t}) + T_l(y_{m,t}), f_t\right) + Rk_t$$

where $R$ is the risk-free return to financial assets, $T(.)$ is the tax function for individual labour earnings and $B(.)$ is the benefit function.

Figure 3.4 shows a benefit schedule, based on the universal credit allowance for a single adult household in the UK. Benefits are assessed at the household level and therefore benefit receipt is a function of total post-tax household labour income. The benefit schedule facing a household depends on the intercept, which I denote as $\iota$, and on the slope, which I denote $\kappa$. The intercept determines the award level when household post-tax income is zero, whereas the slope determines the rate at which benefits are withdrawn as post-tax earnings increase. Similarly to universal credit, I allow the intercept to vary based on family type. The slope is identical for all households. Benefit receipt is given by:

$$B(x, f_t) = \max\left\{\iota(f_t) - \kappa x, 0\right\}$$

Figure 3.4: Benefits and net income




Notes: Benefit receipt and total net income (following receipt of benefits) for a household composed of a single adult, based on the UK universal credit schedule.

I abstract from several features of the benefit system in the UK. In practice, some families qualify for a work allowance which allows a limited amount of earnings to be disregarded when calculating benefit eligibility - this policy effectively creates an initial flat section of the benefit schedule. Additional cost factors, such as rent, can also shift the minimum income guarantee. Otherwise, I base the parameters defining the benefit schedule on universal credit in the 2020/2021 tax year.

During retirement, neither partner earns any labour income, so the budget constraint is simply:

$$c_t + k_{t+1} = PP(f_t) + Rk_t$$

where $PP(f_t)$ is the public pension function, which depends only on the number of adults in the household.

### 3.4.5 Human capital

The human capital accumulation function is given by:

$$\tilde{h}_{t+1} = h_t + a\left(\alpha e_t^{\phi_1} + (1-\alpha) n_t^{\phi_2}\right)$$
$$h_{t+1} = \exp(z_t) \times \tilde{h}_{t+1}$$

where $\tilde{h}_{t+1}$ denotes pre-shock human capital next period, $h_{t+1}$ denotes post-shock human capital next period, $z_t$ is the human capital shock, $a$ is permanent learning ability, $e_t$ is education hours and $n_t$ is labour hours.

This formulation encompasses both *learning-by-doing* and *Ben-Porath* technologies. When $\alpha = 1$, human capital accumulation depends only on costly skill investment. When $\alpha = 0$, human capital accumulation depends only on labour hours. Intermediate values of $\alpha$ indicate some contribution from both labour and education hours.

I assume $\ln(a)$ and $\ln(h_0)$ (e.g. initial human capital) have a bivariate normal joint distribution. This allows wage growth across the life-cycle to correlate with wages at the beginning of the working life, and offers a reduced-form approach to capturing prior educational attainment and other unobserved, permanent heterogeneity.

The human capital shock ($z_t$) occurs between periods and the value is unknown at the point where women decide on consumption and time allocation. I assume $z_t$ is distributed normally. $E[\exp(z_t)] < 1$, which implies that, on average, human capital depreciates each period.

## 3.5 Benefit design and human capital accumulation

### 3.5.1 Impact of temporary labour tax

To build intuition for the interaction between benefit design and human capital accumulation, I first briefly consider the impact of a temporary labour tax. For ease of exposition, I consider the impact of a tax in an economy with no taxes, no benefits and with only one family type (e.g. $f_t = 1 \; \forall \; t$). Since there is only one family type, I denote the scale for the utility of leisure as $\chi$ rather than $\chi(f_t)$. In addition, I assume that no individuals are credit constrained.

In period $t$, the government introduces a flat labour tax $\tau_{temp} > 0$. The introduction of this tax is fully compensated through lump sum redistribution, such that the marginal utility of wealth ($\lambda_t$) is unchanged. I consider the impact within a *learning-by-doing* model (e.g. $\alpha = 0$) and a *Ben-Porath* model (e.g. $\alpha = 1$) in turn. I show that in a pure *learning-by-doing* model, the tax will discourage human capital accumulation. The effect is reversed in the *Ben-Porath* model, where the tax will encourage human capital accumulation.

#### Learning-by-doing

The first order condition for $n_t$ in the pure *learning-by-doing* model is given by:

$$\underbrace{\chi n_t^{\frac{1}{\gamma_2}}}_{\text{MC of labour}} = \underbrace{\lambda_t(1-\tau_{temp})h_t}_{\text{MB of earnings}} + \underbrace{\beta \frac{\partial \tilde{h}_{t+1}}{\partial n_t} \frac{\partial E_t V_{t+1}}{\partial \tilde{h}_{t+1}}}_{\text{MB of human capital}} \tag{3.1}$$

The marginal cost of supplying additional labour, due to lost leisure time, is equated to the marginal benefit of increased (net) income plus the marginal benefit of increased human capital. The introduction of the temporary tax reduces earnings from labour, but has no direct impact on either the marginal cost of labour or the marginal benefit of human capital. The direct effect of the tax is therefore to reduce labour supply and, by extension, human capital next period.

There could be, in addition, indirect effects on labour supply through the marginal benefit of human capital. Less human capital next period reduces wages. If lower wages change the expected future path of labour supply, this will feedback into the marginal benefit of human capital accumulation this period. For instance, if the agent expects to work less in future periods, any human capital acquired this period will have lower utilisation and the marginal benefit of human capital will therefore fall. In general, we would expect these indirect effects to reinforce the direct negative effect of the tax on labour supply as long as the labour supply response to a permanent increase in wages is positive.

### Ben-Porath

The first-order condition for $n_t$ has a simpler form in the pure *Ben-Porath* model:

$$\underbrace{\chi(n_t + e_t)^{\frac{1}{\gamma_2}}}_{\text{MC of labour}} = \underbrace{\lambda_t(1 - \tau_{temp})h_t}_{\text{MB of earnings}} \tag{3.2}$$

The agent equates the marginal cost of labour, due to lost leisure time, to the marginal benefit of increased (net) income. This is identical to Equation 3.1, but without the additional benefit from human capital. However, there is also a first-order condition for $e_t$:

$$\underbrace{\chi(n_t + e_t)^{\frac{1}{\gamma_2}}}_{\text{MC of educational investment}} = \underbrace{\beta \frac{\partial \tilde{h}_{t+1}}{\partial e_t} \frac{\partial E_t V_{t+1}}{\partial \tilde{h}_{t+1}}}_{\text{MB of human capital}} \tag{3.3}$$

which implies that the agent equates the marginal cost of educational investment, again due to lost leisure time, to the marginal benefit of additional human capital. Since the left-hand side of Equation 3.2 and 3.3 are identical, the two first-order conditions jointly imply:

$$\underbrace{\lambda_t(1-\tau_{temp})h_t}_{\text{MB of earnings/MC of educational investment}} = \underbrace{\beta\frac{\partial\tilde{h}_{t+1}}{\partial e_t}\frac{\partial E_t V_{t+1}}{\partial\tilde{h}_{t+1}}}_{\text{MB of human capital}} \tag{3.4}$$

We can think of agents as first choosing total non-leisure time, $(n_t+e_t)$, and then determining the proportion of non-leisure time to spend in educational investment. With this framing, the cost of additional educational investment is less labour time. The agent equates the marginal cost of educational investment, which is lower earnings due to less labour time, with the benefit of additional human capital.

This framing clarifies the impact of the temporary labour tax. The introduction of a tax will reduce the marginal loss of income from lower $n_t$, and reduce the cost of human capital investment. Educational investment and human capital next period will therefore increase. The tax has no direct impact on the marginal benefit of human capital but, as in Section 3.5.1, there may be indirect effects resulting from the change in human capital next period. If higher wages next period result in higher labour supply, this will increase human capital utilisation and the marginal benefit from acquiring additional human capital this period. As above, we would expect the indirect effects to reinforce the direct effects provided the labour supply response to a permanent increase in wages is positive.

### 3.5.2 Benefit interaction with human capital

The model presented in Section 3.4 features both idiosyncratic risk and incomplete markets. In such a model, benefits provide valuable insurance against shocks to human capital and partner income. However, there is a trade-off in the design of the benefit system: for a given budget, increasing insurance through larger minimum income guarantees will generally require higher withdrawal rates, and increases in the withdrawal rates distort time allocation.

In the UK, the benefit withdrawal rate in the 2020/21 tax year was 63% and benefits were unified to ensure that this withdrawal rate was constant. Historically, withdrawal rates have varied for different types of benefit, creating complex marginal effective tax schedules with withdrawal rates of 100% for some earnings levels. In this paper I focus on simple linear benefit schedules such as that shown in Figure 3.4, both for tractability and because this forms a reasonable approximation to current policy.

If individuals are temporarily forced on to benefits following an adverse shock, the impact of benefit withdrawal on human capital accumulation may align with the discussion in Section 3.5.1. Consider the incentives facing a woman within a two adult household that, in the prior period, didn't qualify for benefits due to high post-tax earnings. A temporary shock to partner income reduces household post-tax earnings such that, if the woman kept labour supply constant, the household would receive benefits this period. Higher withdrawal rates will, in a *learning-by-doing* model, create larger disincentives to invest in human capital this period and lower wages, labour supply, and therefore tax income, in subsequent periods. By contrast, higher withdrawal rates increase incentives to invest in human capital in a *Ben-Porath* model, increasing wages and tax income in future periods.

However, it is also necessary to consider the direct impact of changes in the withdrawal rate on labour supply. Alternative human capital mechanisms will imply different labour supply elasticities, even when matching the same data. For instance, Blandin and Peterman (2019) find that the elasticity of labour supply is substantially higher in a *Ben-Porath* model than in an equivalent *learning-by-doing* model, even when holding constant the relevant utility parameters. This is because, within a *learning-by-doing* model, earnings are only one component of the total return to labour supply, with the other being the gain in human capital. In a *Ben-Porath* model, by contrast, time can be productively re-allocated from labour supply to education, increasing the responsiveness of labour supply to changes in wage.

Reforms to the benefit system are therefore likely to have different implications depending on the human capital accumulation mechanism. To explore these mechanisms more thoroughly requires a fully specified model. In the next section, I explain my methodology for bringing the model to the data, which will allow me to estimate each of the models and quantitatively explore the implications of revenue neutral reforms to the benefit system.

## 3.6 Estimation

Estimation of the model proceeds in three steps. First, I estimate the exogenous demographic, partner income and partner employment processes. I discuss the methodology for estimating each of these exogenous processes in Appendix B.2, and present evidence on their fit to the data. Second, I calibrate the value of several parameters. Finally, given the exogenous inputs and calibrated parameters, I estimate the remaining parameters of the model using method of simulated moments.

### 3.6.1 Calibrated parameters

Table 3.2 reports the calibrated parameters.

Table 3.2: Calibrated model parameters

| Parameter | Description | Value |
|---|---|---|
| $\beta$ | Discount rate | 0.98 |
| $R$ | Risk-free return | $\frac{1}{0.98}$ |
| $\gamma_1$ | Risk aversion | 2.0 |
| $\mu_z$ | Expected value of human capital shock | $-0.066$ |
| $\sigma_z$ | Human capital risk | 0.2 |
| $\tau$ | Benefit taper rate | 0.63 |
| $\iota(1)$ | Allowance for single adult | 75 |
| $\iota(2), \iota(3)$ | Allowance for single adult with children | 139 |
| $\iota(4)$ | Allowance for couple | 114 |
| $\iota(5), \iota(6)$ | Allowance for couple with children | 177 |

Notes: See Section 3.4 for precise definitions of each variable. The units of the benefit parameters are pounds per week.

We set $\gamma_1$ to 2. This generates a standard degree of risk aversion; the same value is used in, for instance, Conesa, Kitao, and D. Krueger (2009) and Costa and Santos (2018). Similarly, we set $\beta$ to 0.98 following, for instance, Attanasio, Low, and Sánchez-Marcos (2008). The risk-free return is set to the inverse of the discount rate. The benefit schedule, as discussed above, is calibrated to mimic the universal credit schedule as of the 2020/21 tax year. Since we do not track number of children in the state space, the allowance for a single adult or a couple with children is calibrated as if they had exactly one child.

The expected value of the human capital shock is calibrated such that, on average, ten percent of human capital stock depreciates each period. This approximates the depreciation rate estimated in Blundell, Costa Dias, Meghir, et al. (2016). The standard deviation of the human capital shock can be included in the estimation exercise, in which case each model returns values within $\pm 0.02$ of the calibrated value. However, holding the degree of human capital risk constant across models simplifies interpretation of the policy experiments.

The consumption equivalence scale $\psi(f_t)$ is set to mimic the OECD-modified equivalence scale. For single adult households, $e(f_t) = 1$. The equivalence scale is increased by 0.5 if an additional adult is present, and increased by 0.3 if a child is present.

### 3.6.2 Estimated parameters

The remaining model parameters are estimated to match data moments. The moments are constructed from the working-life sample, including only women who have entered the labour market and are aged between 22 and 55.

I create a set of life-cycle moments, calculating the mean and/or variance of specific variables conditional on age. The variables included and the moments calculated using each variable are specified in Table 3.3. In each case, the moments are constructed as either the mean or variance of the specific variable at age 22, 25, 30, and continuing in five year increments to age 55. Labour hours and education hours are observed for the full sample, but wages are only observed for employed individuals. As a result, log wage moments are calculated for the employed sub-sample and differenced log wages are calculated for the sub-sample who are employed in both the current and the previous period. These sample selection criteria are replicated when re-constructing these moments within the simulated data.

Table 3.3: Life-cycle moments

| Variable | Mean by age? | Variance by age? |
|---|---|---|
| Labour hours, $n_t$ | ✓ | ✓ |
| Education hours, $e_t$ | ✓ | ✓ |
| Log wage, $\ln(w_t)$ | ✓ | ✓ |
| Differenced log wage, $\Delta \ln(w_t)$ | × | ✓ |

Notes: For the log wage moments, only women employed at that age are included in the sample. For the differenced log wage moments, only women employed at that age and the year before are included in the sample.

I also include average labour and education hours of women conditional on each of the six family types. Finally, I match an auxiliary regression:

$$\ln w_t = \beta_0 + \beta_1 \ln(w_{t-1}) + \beta_2 \ln(n_{t-1}) + \epsilon_t$$

where $w_t$ is wages at age $t$ and $n_{t-1}$ is labour hours at age $t-1$. Note that wages are the data analogue to human capital. I estimate this regression on the sub-sample of women who are employed in both the current and previous period, since I do not otherwise observe their wages. I include the estimated parameter values and the variance of the predicted residual as moments. I estimate the variance-covariance matrix of the moments by bootstrapping.

Table 3.4 provides a list of the estimated parameters. I also include information on which moments are used in identifying each parameter.

Table 3.4: Estimated model parameters

| Parameter | Description | Identifying moments |
|---|---|---|
| $\chi(f_t)$ | Scale for utility of $n_t + e_t$ | $E[n_t]$ by age & family type |
| $\phi_1$ | Curvature of $h_{t+1}$ wrt. $e_t$ | $\ln(w_t)$ regression & $E[e_t]$ by age |
| $\phi_2$ | Curvature of $h_{t+1}$ wrt. $n_t$ | $\ln(w_t)$ regression & $E[n_t]$ by age |
| $\gamma_2$ | Curvature of $u$ wrt. $n_t + e_t$ | $E[n_t]$ & $\text{Var}[n_t]$ by age |
| $\mu_{h_0}$ | Mean of $\ln(h_0)$ | $E[\ln(w_t)]$ by age |
| $\sigma_{h_0}$ | Standard deviation of $\ln(h_0)$ | $\text{Var}[\ln(w_t)]$ by age |
| $\mu_a$ | Mean of $\ln(a)$ | $\ln(w_t)$ regression & $E[\ln(w_t)]$ by age |
| $\sigma_a$ | Standard deviation of $\ln(a)$ | $\text{Var}[\Delta \ln(w_t)]$ by age |
| $\rho_{h_0,a}$ | Correlation, $\ln(h_0)$ & $\ln(a)$ | $\text{Var}[\ln(w_t)]$ by age |

The estimates of the model parameters are the set of parameter values $\Theta$ that minimise the following expression

$$\sum_{\kappa=1,\ldots,K} \frac{\left(M^d_{\kappa,N} - M^s_{\kappa,S}\left(\Theta\right)\right)^2}{\text{Var}\left(M^d_{\kappa,N}\right)} \tag{3.5}$$

where $K$ is the total number of moments used in estimation, $M^d_{\kappa,N}$ is the estimate of moment $\kappa$ from $N$ observations of observed data and $M^s_{\kappa,S}$ is the corresponding moment calculated on $S$ model simulations for parameter values $\Theta$. I simulate each model five times per parameter set e.g. $S = 5$. As implied by 3.5, I weight according to the inverse variances of the moments rather than the asymptotically efficient weighting matrix, following Altonji and Segal (1996).

## 3.7 Ben-Porath and learning-by-doing estimation results

Table 3.5 reports the remaining parameters for both pure models. All parameters are estimated with the exception of parameter (1), $\alpha$, which controls the weight placed on educational investments relative to learning-by-doing in the human capital accumulation function. This parameter is exogenously set to 1.0 in the *Ben-Porath* model, excluding working hours from human capital accumulation, and similarly set to 0.0 in the *learning-by-doing* model. These two extreme assumptions will be relaxed in Section 3.8.

Table 3.5: Estimated model parameters

| # | Parameter | Description | *BP* | *LBD* |
|---|---|---|---|---|
| (1) | $\alpha$ | Weight in HC on BP vs. LBD | 1.000 | 0.000 |
| (2) | $\phi_1$ | Curvature of $\tilde{h}_{t+1}$ wrt. $n_t$ | - | 0.769 |
| (3) | $\phi_2$ | Curvature of $\tilde{h}_{t+1}$ wrt. $e_t$ | 0.056 | - |
| (4) | $\gamma_2$ | Curvature of $u$ wrt. $n_t + e_t$ | 0.571 | 0.453 |
| (5) | $\chi(1)$ | Util. of $n_t + e_t$: single, no child | 68.109 | 17.422 |
| (6) | $\chi(2)$ | Util. of $n_t + e_t$: single, child under 5 | 113.674 | 75.331 |
| (7) | $\chi(3)$ | Util. of $n_t + e_t$: single, child 6 and 12 | 70.031 | 44.194 |
| (8) | $\chi(4)$ | Util. of $n_t + e_t$: couple, no child | 36.696 | 18.348 |
| (9) | $\chi(5)$ | Util. of $n_t + e_t$: couple, child under 5 | 79.507 | 33.430 |
| (10) | $\chi(6)$ | Util. of $n_t + e_t$: couple, child 6 and 12 | 54.827 | 25.174 |
| (11) | $\mu_{h_0}$ | Mean of $\ln(h_0)$ | 2.352 | 2.351 |
| (12) | $\sigma_{h_0}$ | Standard deviation of $\ln(h_0)$ | 0.272 | 0.252 |
| (13) | $\mu_a$ | Mean of $\ln(a)$ | 0.418 | $-1.970$ |
| (14) | $\sigma_a$ | Standard deviation of $\ln(a)$ | 0.314 | 0.148 |
| (15) | $\rho_{h_0,a}$ | Correlation, $\ln(h_0)$ & $\ln(a)$ | $-0.754$ | $-0.561$ |

Notes: BP = Ben-Porath, LBD = Learning-by-doing. Parameter (1) is not estimated, but is included here to clarify the distinction between each model. Parameter (2) is not estimated in the Ben-Porath model, since $n_t$ is excluded from the human capital accumulation. Similarly, parameter (3) is not estimated in the learning-by-doing model.

Parameters (1) and (2) determine the concavity of human capital accumulation function with respect to educational investments or labour hours this period. Values less than 1 indicate that human capital increases less than proportionally as hours increase, e.g. doubling the relevant hours measure will less than double the pre-shock increase in human capital. Educational investments have no impact on human capital in the *learning-by-doing* model, so parameter (2) is not defined for that model. Similarly, parameter (1) is not defined for the *Ben-Porath* model.

In both models, the increase in human capital is less than proportionate to an increase in hours invested. In the *learning-by-doing* model, working for 40 hours per week generates approximately 70% more human capital than working for 20 hours per week. This contrasts with Blundell, Costa Dias, Goll, et al. (2021) and Blundell, Costa Dias, Meghir, et al. (2016). Both of those papers find that working full-time increases human capital more than twice as much as working part-time, although labour supply decisions within the models that they estimate are limited to working full-time, part-time or not at all. In the *Ben-Porath* model, the concavity of the human capital accumulation function is more extreme. Investing 8 hours per week in education yields only 4% more human capital than investing 4 hours per week. I discuss this result in more detail below, and in Section 3.8 show that when I allow for human capital accumulation through both *learning-by-doing* and *Ben-Porath* mechanisms I estimate parameters more aligned with the existing literature.

Parameter (4) determines the curvature of utility with respect to non-leisure time. In a model without human capital, this parameter would be equal to the Frisch elasticity of labour supply. The estimated values indicate that total non-leisure time should be more responsive to changes in wages in the *Ben-Porath* model than in the *learning-by-doing* model, all else equal.

Parameters (5) through (10) shift the dis-utility of non-leisure time, varying depending on the family type. The overall scale of these parameters cannot be directly compared, since the models differ in the curvature parameter discussed above. However, it is clear that the patterns across demographic groups are similar in both models. Single mothers have a higher dis-utility of non-leisure time than mothers in couples. For both single mothers and mothers in couples, the dis-utility of non-leisure time is highest when their youngest child is under 5.

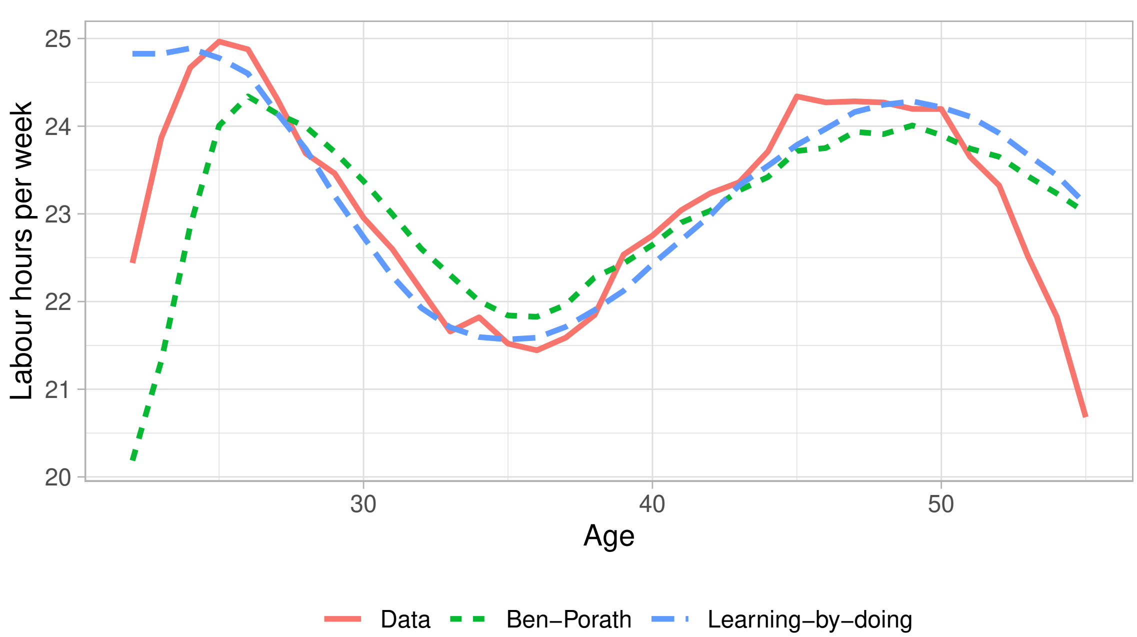
Parameters (11) through (15) define the initial distribution of human capital and ability. Parameter (11) and (12), the mean and variance of initial human capital respectively, are identified by the mean and variance of wages in the first few periods of our model. The values are essentially identical in both models. Parameter (13), the mean of ability, scales human capital accumulation to match average wage growth given the the curvature parameters discussed above and observed investments. The lower average ability in the *learning-by-doing* model is therefore necessary to compensate for greater linearity in the return to human capital investment in *learning-by-doing* as compared to *Ben-Porath*, as well as the higher average labour hours as compared to educational investments. Similarly, the variance of ability must be lower in the *learning-by-doing* to match the same dispersion of growth rates, given the higher variance of labour hours as compared to educational investments.

Finally, parameter (15) is the correlation between initial human capital and ability. Both models estimate a negative relationship between these two initial conditions, implying that individuals with low wages at the start of working life have higher wage growth, all else equal. In the initial period of our model, women are 22 years old. The negative relationship may therefore be capturing heterogeneity in educational qualifications that is not explicitly modelled. Individuals who have just left higher education in the initial periods of the model are likely to have lower wages at age 22 than individuals who have already been in the labour market for several years, but they will benefit from faster wage growth.

Figure 3.5 presents the fit of both models to the average life-cycle profiles of labour hours. Both models successfully replicate the rapid decline in labour hours that begins in the mid 20s, as well as the recovery in labour hours that begins in the late 30s. They have relatively less success in replicating the qualitative features of labour hours very early and late in working life.

Over the first five years of working life, we observe average labour hours increase from 22.5 hours per week to 25 hours per week. Average labour hours in the *learning-by-doing* model, by contrast, are already at 25 hours per week at age 22. The *Ben-Porath* model suffers from the opposite problem, with labour hours too low early in working life and subsequently increasing more rapidly than we observe in the data. Neither model captures the decline in labour hours as individuals approach retirement.

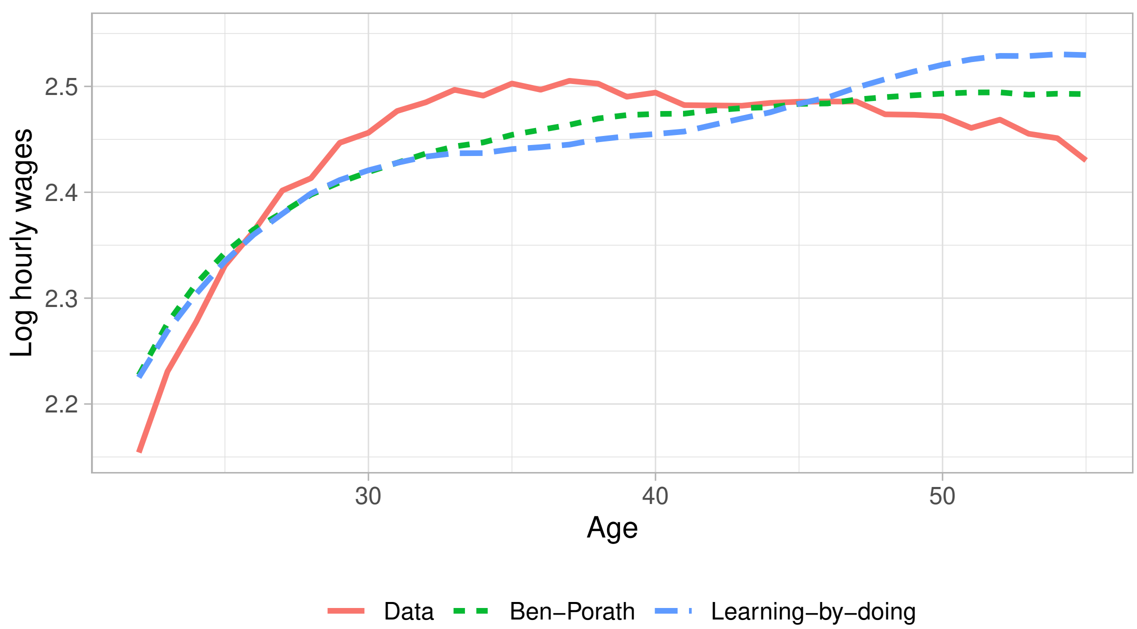
Figure 3.5: Labour hours fit for Ben-Porath and learning-by-doing models




Source: UKHLS and model simulations. Line depicts unweighted average of labour hours conditional on each age.The definition of labour hours is discussed in Section 3.3.1. All averages are taken over the full sample, including individuals who report no labour hours.

Figure 3.6 presents the fit of both models to average life-cycle profiles of wages. Again, both models replicate some general features of the data, but are less successful towards the beginning and end of working life. Both models generate a flatter profile than the data, with wages starting too high and increasing too slowly. Neither model replicates the decline in wages we observe late in working life. The *learning-by-doing* model also generates an increase in average wages from the late 30s to around age 50, which has no analogue in the data. This increase in wages coincides with the recovery of labour hours following child-birth.

Figure 3.6: Log wage fit for Ben-Porath and learning-by-doing models

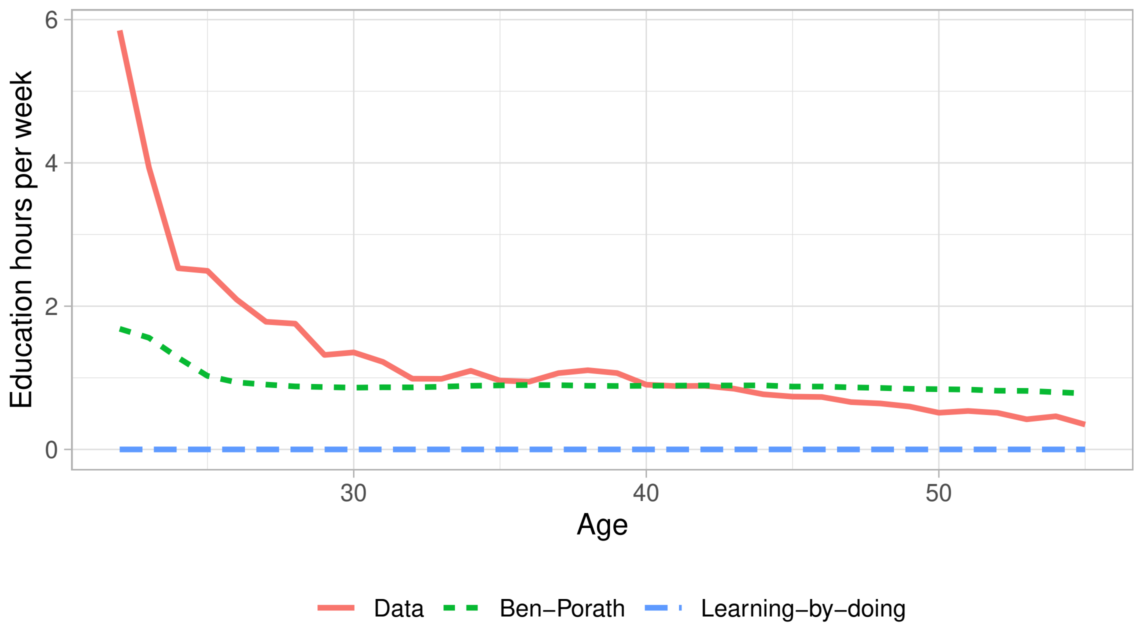



Source: UKHLS and model simulations. Line depicts unweighted average of log-wages conditional on each age. Average log-wage is calculated for sub-sample with positive work hours at each age. Wages are corrected for inflation using CPI, such that the measured wage is in 2021 prices.

Figure 3.7 presents the fit of both models to average life-cycle profiles of educational investments. There is no return to educational investments in the *learning-by-doing* model, so investments are trivially zero throughout. The *Ben-Porath* model, on the other hand, could potentially generate realistic profiles, but fails to do so. In contrast to the high levels of investment that we observe early in working life, the model generates almost completely flat average investments; there is a small decline in the first five years of working life, following which average investments are almost exactly 1 hour per week. This feature of the *Ben-Porath* simulation is likely to be a direct result of the extreme concavity of the human capital accumulation function with respect to education hours. The low value of $\phi_2$ ensures that increasing education hours has a very limited impact on the accumulation of human capital, and therefore the average level of investment responds very little to changes in wages or the return to human capital, resulting in a flat life-cycle profile.

In combination with the parameter estimates above, these results suggest that neither model adequately captures some of the key dynamics that we observe in the data. The major

Figure 3.7: Education hours fit for Ben-Porath and learning-by-doing models




Source: UKHLS and model simulations. Line depicts unweighted average of education hours conditional on each age.The definition of education hours is discussed in Section 3.3.1. All averages are taken over the full sample, including individuals who report no education hours.

qualitative failings of the model are:

- **Ben-Porath:** Average educational investments are very flat over working life, and never attain the high levels in the early 20s that we observe in the data. Wages grow too slowly throughout working life.
- **Learning-by-doing:** Labour hours in the early years of working life are much higher than those we observe in the data and they do not increase as wages grow. Average wages increase from the late 30s onwards, at the same time as average labour hours are recovering following child-birth. Wages grow too slowly throughout working life.

In Section 3.8, I show that a model containing both mechanisms of human capital accumulation can address these shortcomings, and therefore provide a much closer fit to observed data.

## 3.8 Mixed model estimation results

Table 3.6 reports the estimated parameters for the mixed model, with the parameters from the two pure models also provided for comparison. The mixed model has two additional degrees of freedom relative to the pure models. First, I estimate parameter (1), $\alpha$, alongside the other parameters. This allows the human capital accumulation function to depend on both educational investments and labour hours simultaneously. Second, I include both parameter (2) and parameter (3) in the estimation simultaneously. These parameters control the curvature of human capital next period with respect to educational investments and labour hours this period.

Table 3.6: Estimated model parameters

| # | Parameter | Description | *BP* | *LBD* | *Mixed* |
|---|---|---|---|---|---|
| (1) | $\alpha$ | Weight in HC on BP vs. LBD | 1.000 | 0.000 | 0.711 |
| (2) | $\phi_1$ | Curvature of $\tilde{h}_{t+1}$ wrt. $n_t$ | - | 0.769 | 1.078 |
| (3) | $\phi_2$ | Curvature of $\tilde{h}_{t+1}$ wrt. $e_t$ | 0.056 | - | 0.917 |
| (4) | $\gamma_2$ | Curvature of $u$ wrt. $n_t + e_t$ | 0.571 | 0.453 | 0.550 |
| (5) | $\chi(1)$ | Util. of $n_t + e_t$: single, no child | 68.109 | 17.422 | 67.039 |
| (6) | $\chi(2)$ | Util. of $n_t + e_t$: single, child under 5 | 113.674 | 75.331 | 225.548 |
| (7) | $\chi(3)$ | Util. of $n_t + e_t$: single, child 6 and 12 | 70.031 | 44.194 | 122.692 |
| (8) | $\chi(4)$ | Util. of $n_t + e_t$: couple, no child | 36.696 | 18.348 | 71.646 |
| (9) | $\chi(5)$ | Util. of $n_t + e_t$: couple, child under 5 | 79.507 | 33.430 | 120.626 |
| (10) | $\chi(6)$ | Util. of $n_t + e_t$: couple, child 6 and 12 | 54.827 | 25.174 | 92.416 |
| (11) | $\mu_{h_0}$ | Mean of $\ln(h_0)$ | 2.352 | 2.351 | 2.303 |
| (12) | $\sigma_{h_0}$ | Standard deviation of $\ln(h_0)$ | 0.272 | 0.252 | 0.275 |
| (13) | $\mu_a$ | Mean of $\ln(a)$ | 0.418 | -1.970 | $-1.782$ |
| (14) | $\sigma_a$ | Standard deviation of $\ln(a)$ | 0.314 | 0.148 | 0.093 |
| (15) | $\rho_{h_0,a}$ | Correlation, $\ln(h_0)$ & $\ln(a)$ | -0.754 | -0.561 | $-0.781$ |

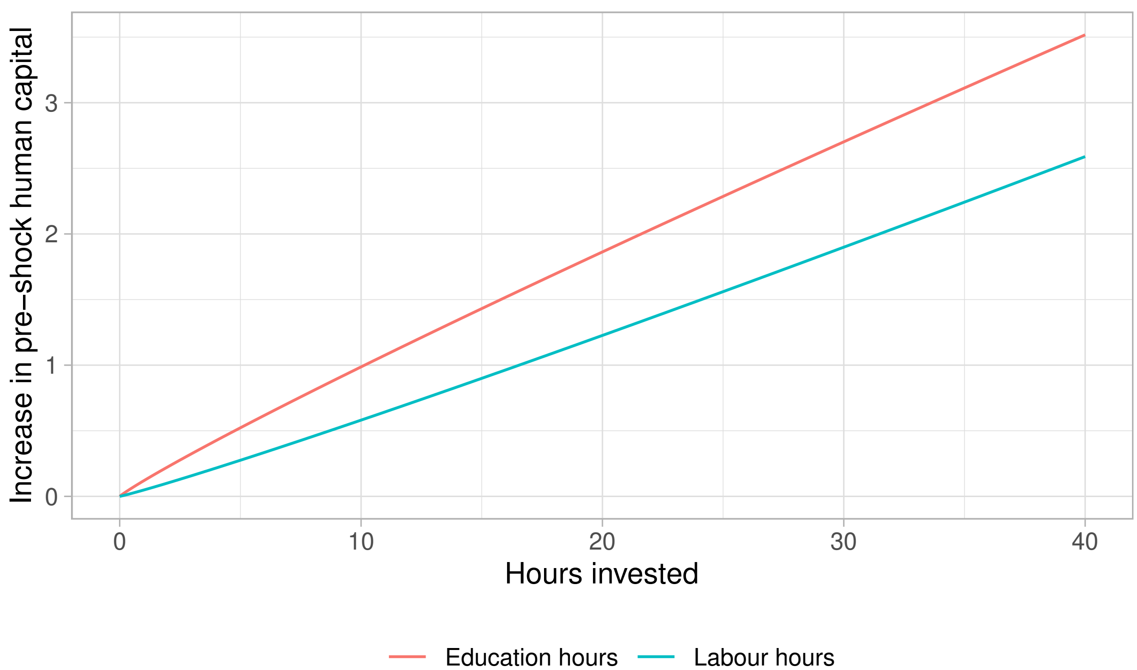
Notes: BP = Ben-Porath, LBD = Learning-by-doing. Parameter (1) is not estimated in either the BP or LBD model, but is estimated in the mixed model. Parameter (2) is not estimated in the Ben-Porath model, since $n_t$ is excluded from the human capital accumulation. Similarly, parameter (3) is not estimated in the learning-by-doing model.

Parameter (1) determines the weight on educational investments relative to labour hours in the human capital accumulation function. The parameter value indicates that one hour spent in education will generate around 2.5 times as much human capital than one hour spent in work. However, because because the human capital accumulation function is non-linear in both education hours and labour hours, this ratio will differ as the time invested is increased or decreased.

Parameter (2) and (3) are the exponents of $n_t$ and $s_t$ in the human capital accumulation function. To illustrate the impact of these parameters in combination with parameter (1), Figure 3.8 shows the pre-shock increase in human capital that results from different labour hours or educational investments for an individual of median ability.

In contrast to the pure *learning-by-doing* model, the human capital accumulation function is estimated to be slightly convex with respect to labour hours. Working 40 hours per week generates approximately 2.1 times more human capital than working 20 hours per week. Human capital accumulation remains concave in education hours, but the concavity is much less pronounced than in the pure *Ben-Porath* model. Investing 8 hours per week in education yields 88% more human capital than investing 4 hours per week. For a realistic range of hours, educational investments always yield more human capital than labour hours, but the ratio of the returns decreases as the hours invested grows.

Figure 3.8: Human capital from labour hours and educational investments




Source: Model simulations. Line depicts the pre-shock increase in human capital for investing the number of hours indicated on the x-axis in either education or labour. All values are calculated for an individual of median ability. Human capital is measured in £ per hour.

The remaining parameters are more comparable to the values estimated in the pure models. Parameter (4), which determines the curvature of utility with respect to non-leisure time, has a similar value in the mixed model to the value estimated in the *Ben-Porath* model. The same patterns in the dis-utility of non-leisure are visible, with young children greatly increasing the dis-utility of non-leisure for both single mothers and mothers in couples. Again, as in both pure models, single mothers have significantly higher dis-utility from non-leisure than mothers in couples. The values of parameters (5) to (10) are generally much larger than in the *Ben-Porath* model, despite a similar value of $\gamma_2$. This is because, in comparison to that model, labour hours have a human capital benefit as well as a earnings benefit. For a given wage, the dis-utility of work must be higher to generate a similar level of labour hours.

The values of the parameters (11) through (15), which determine the initial distribution of human capital and ability, also have similar values to those found when estimating the pure models. The unconditional distribution of initial human capital is almost identical across all three models. Average ability, which controls the rate at which wages grow, is slightly higher than levels found in the *learning-by-doing* model. Finally, the correlation between ability and initial human capital is negative, as in both pure models.

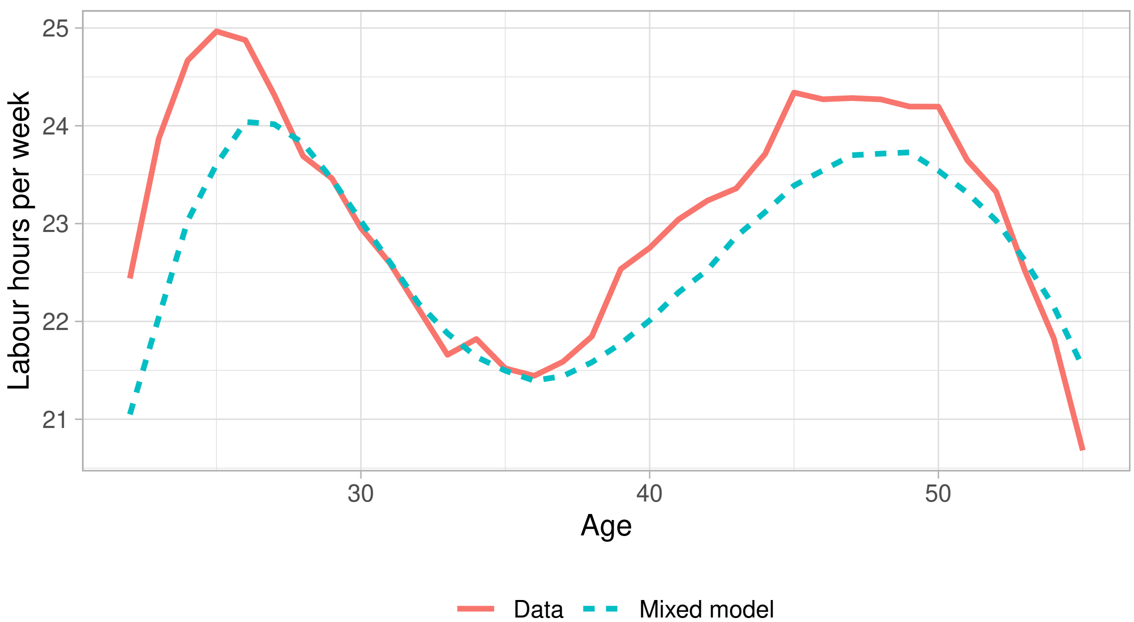
Figure 3.9 presents the fit of the mixed model to average life-cycle profiles of labour hours. As with both pure models, I am able to replicate both the decline in labour hours that begins in late 20s, as well as the recovery in labour hours from the late 30s. However, the mixed model also better replicates the qualitative features of labour hours profiles towards the beginning and end of working life. At the start of working life, labour hours grow rapidly, from around 21 hours per week to 25 hours per week. These levels remain slightly below the levels observed in the data, but the increase in hours is comparable and occurs over the same ages. Labour hours also decline at a similar rate in the 50s to the rate observed in the data.

Figure 3.10 presents the fit of the mixed model to average life-cycle profiles of wages. Again, the mixed model provides a much better fit towards the beginning and end of working life than either pure model. Wage levels and growth rates closely match the data throughout the 20s. Simulated wages also decline at an appropriate rate from approximately age 50. Wages still plateau a few years earlier than observed in the data, and there is some growth in wages in the 40s that is not visible in the data.

Figure 3.11 presents the fit of the mixed model to average life-cycle profiles of educational investments. Unlike the *Ben-Porath* model, the mixed model can provide a very accurate fit to average educational investments. The simulations replicate the observed high levels of investment early in working life, as well as the steady decline from age 40.

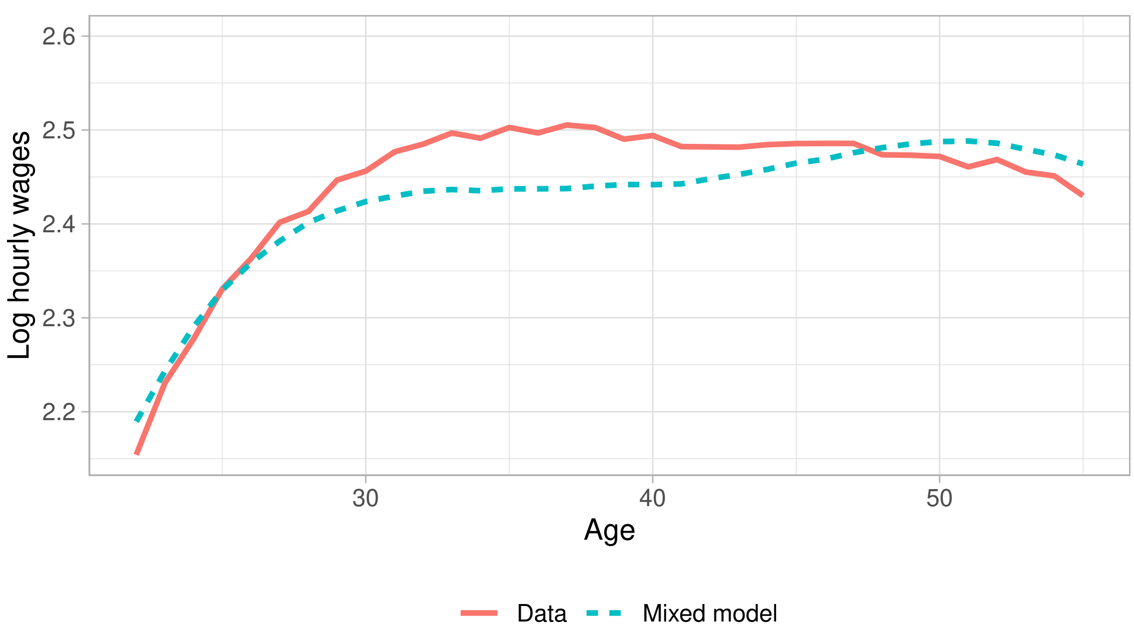
Why does the mixed model fit the data significantly better? First, consider the educational

Figure 3.9: Labour hours fit for mixed model



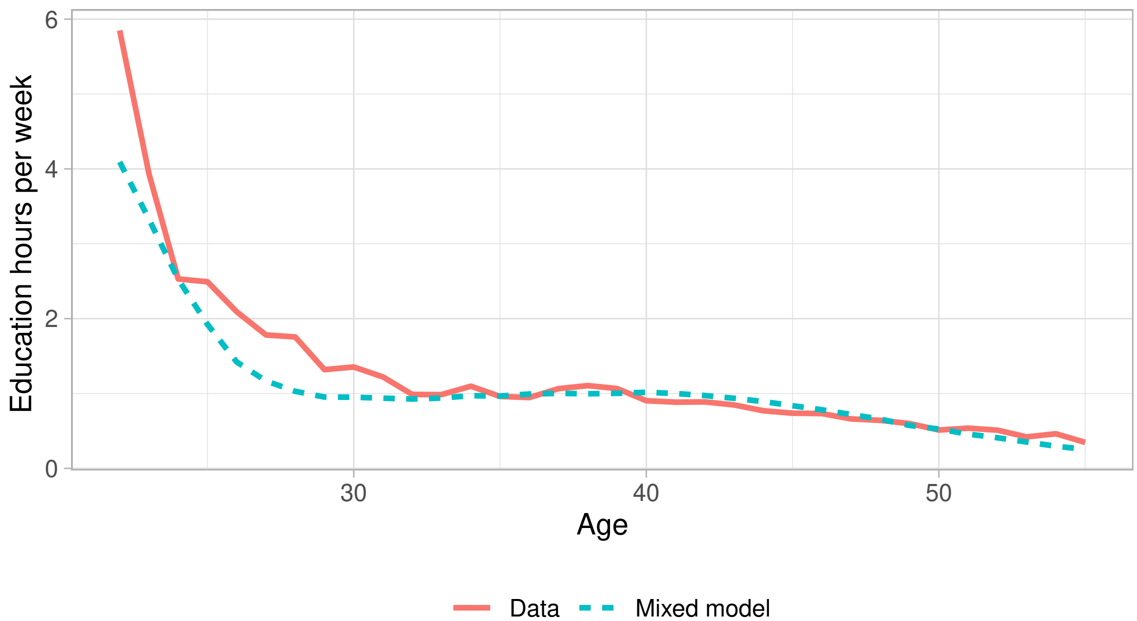

Source: UKHLS and model simulations. Line depicts unweighted average of labour hours conditional on each age.The definition of labour hours is discussed in Section 3.3.1. All averages are taken over the full sample, including individuals who report no labour hours.

Figure 3.10: Log wage fit for mixed model




Source: UKHLS and model simulations. Line depicts unweighted average of log-wages conditional on each age. Average log-wage is calculated for sub-sample with positive work hours at each age. Wages are corrected for inflation using CPI, such that the measured wage is in 2021 prices.

Figure 3.11: Education hours fit for mixed model




Source: UKHLS and model simulations. Line depicts unweighted average of education hours conditional on each age.The definition of education hours is discussed in Section 3.3.1. All averages are taken over the full sample, including individuals who report no education hours.

investment profiles. When estimating the *Ben-Porath* model, these investments are assumed to be the only source of human capital growth. Despite relatively modest investments in middle-age, they must remain sufficiently productive to compensate for depreciation. The observed investments are unresponsive to life-cycle changes in demographics, despite these changes causing large shifts in the utility of leisure time and therefore labour hours. To fit these facts, the human capital accumulation function needs to be highly concave with respect to education hours. This generates extremely inflexible investment levels.

However, a more natural explanation is possible when labour hours are included as an additional source of human capital. The opportunity cost of educational investments is proportional to the wage level, as discussed in Section 3.5.1. At the beginning of working life, when wages are low and there are many periods of working life remaining, educational investments are inexpensive and widely used to accelerate wage growth. As wages get closer to their peak, the cost of educational investments increases and women switch to labour hours as the primary source of human capital. Educational investments decline rapidly. Throughout the rest of working life, labour hours are the primary source of human capital, compensating for losses from depreciation. Educational investments are mainly used following a negative human capital shock, which temporarily decreases the opportunity cost of education. The addition of human capital accumulation through work hours therefore allows the model to match the rapid initial decline in education investments, without requiring that the low investments levels at older ages are themselves sufficient to compensate for depreciation and generate gradual wage growth.

Figure 3.12 and 3.13 illustrates this dynamic. Figure 3.12 plots labour hours and education hours relative to log wages, separately for women aged 22 to 30 and women aged 31 to 40. The graph demonstrates that educational investments are primarily used by individuals with low wages, across both age groups. Individuals in the top half of the wage distribution rely almost exclusively on labour hours to accumulate human capital. Educational investments are much smaller for women in their 30s than for the younger group, even when conditioning on wages. This could be driven by a number of factors: fewer remaining periods of working life, higher utility of leisure, or lower labour supply and therefore less human capital utilisation.

Figure 3.13 plots average labour hours and educational hours life-cycles versus the life-cycles of individuals who receive a large negative human capital shock (below 5th percentile of the shock distribution) at age 30. Following a large negative shock, women substantially reduce labour hours and increase education hours.

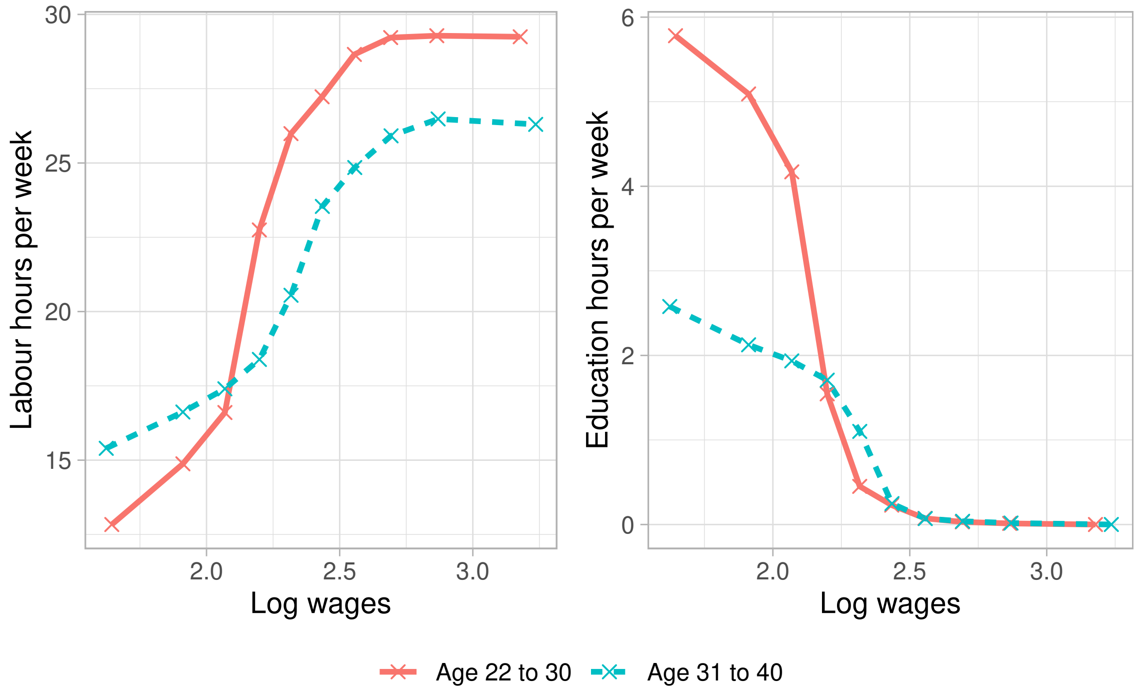
Including educational investments and labour hours in the human capital accumulation function also explains the rapid increase in labour hours early in working life, as women switch their educational investments to labour hours as wages increase. Finally, having both sources of human capital allows the model to better replicate the rapid increase in wages in the first few years of working life, as this is the only period when most women are utilising both sources of human capital.

## 3.9 Policy analysis

### 3.9.1 Labour supply elasticities

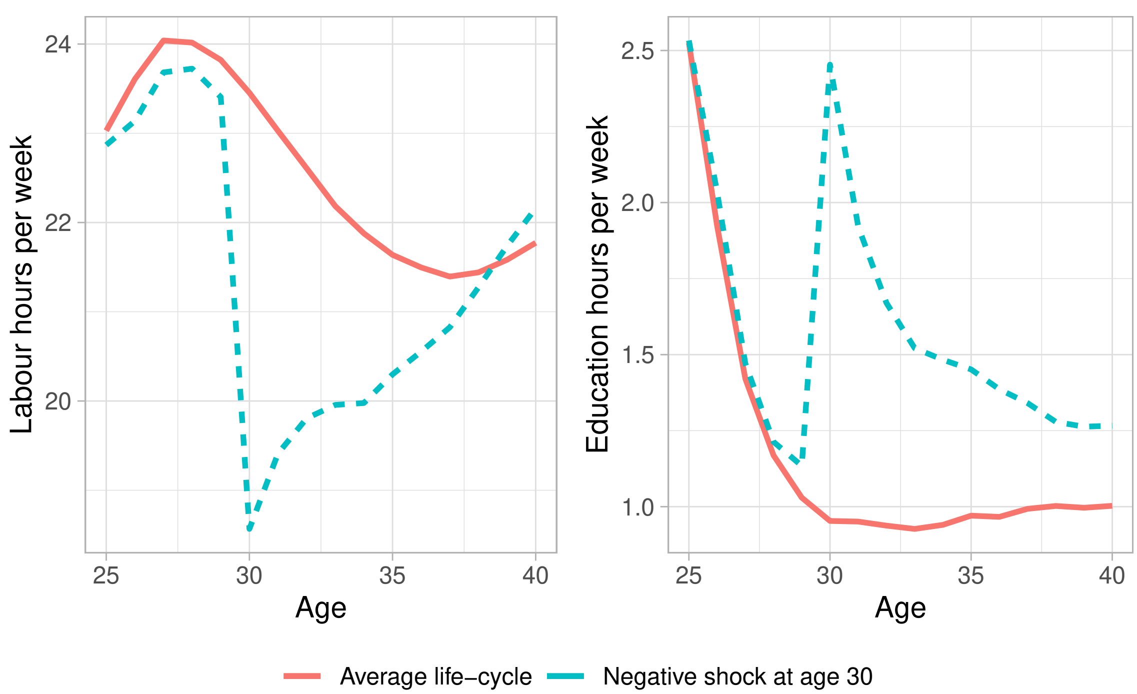
In this section, I will consider the interaction of each model with the benefit system. As discussed in Section 3.5.2, there are two main reasons to expect that the impact of the benefit system will vary depending on the assumed human capital accumulation mechanism. First,

Figure 3.12: Labour hours and educational investments by log-wage decile




Source: Model simulations. Log-wage deciles are calculated for the simulated sample of women aged between 22 and 40. Average wages and labour/education hours are then calculated conditional on decile and age group. Each line joins these points for the appropriate age group.

Figure 3.13: Response to human capital shock

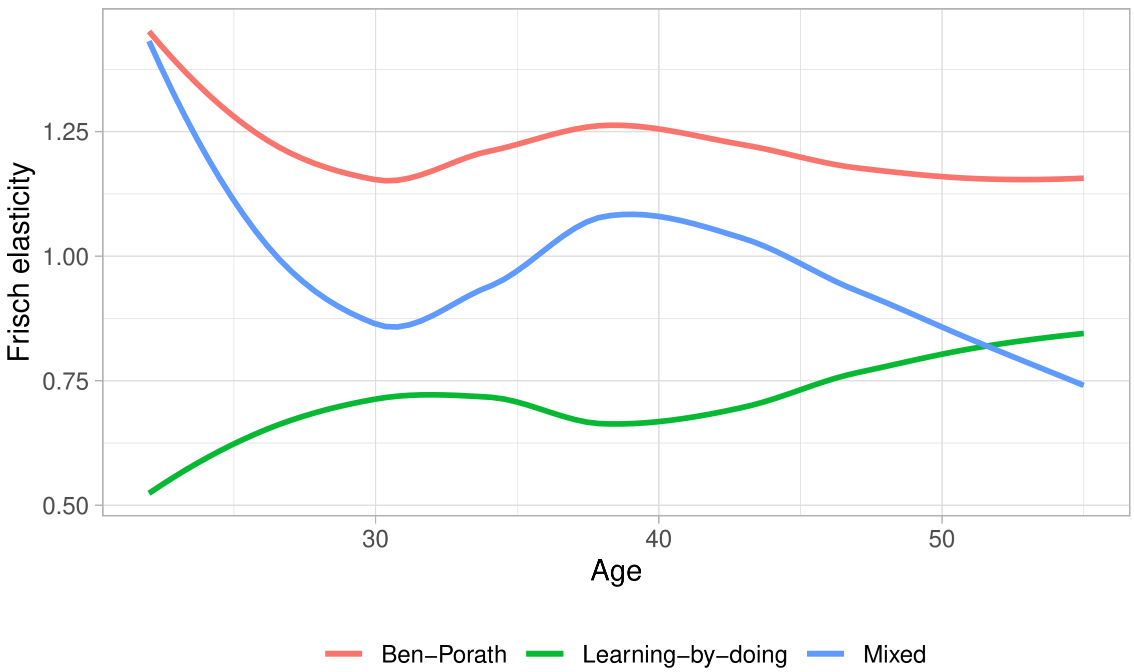



Source: Model simulations. Average life-cycle line depicts unweighted average of education hours conditional on each age. Negative shock line is calculated identically, but conditions on the subsample who receive a below 5th percentile shock to human capital at age 30.

the alternative models will generate different labour supply elasticities. These elasticities differ both in levels and in their evolution over the life-cycle. Second, the models differ in the impact of high marginal tax rates on the incentives to accumulate human capital.

Figure 3.14 plots the labour supply response in each model to a small, temporary and anticipated increase in the wage rate at a particular age. Since the wage perturbation is small, there is a very limited impact on the marginal utility of wealth. We therefore interpret the response as a "wealth constant" or Frisch elasticity.

Figure 3.14: Frisch elasticity over the life-cycle




Source: Model simulations. Frisch elasticity at a specific age is calculated by re-solving the model with a 0.5% increase in rental rate of human capital at the relevant age. Labour supply in the permeated model is then compared to the baseline to calculate an elasticity. This exercise is repeated for each age.

Focusing first on the two pure models, I find that the *Ben-Porath* model exhibits higher elasticities at all ages than the *learning-by-doing* model. This is partly due to the higher value of $\gamma_2$ that I estimate for this model. However, the additional educational hours margin also increases the responsiveness of labour hours to wages, since women can substitute from educational investment while maintaining the same amount of leisure time.

For both models, elasticities evolve over the life-cycle due to changes in average demographics and in the return to additional human capital investment. As the remaining periods of working life reduce, the value of accumulating additional human capital declines. In the *learning-by-doing* model, where human capital is accumulated through labour hours, this increases the elasticity of labour supply as earnings increasingly constitute the primary benefit of work. In the *Ben-Porath* model, where human capital accumulation competes with labour hours, this reduces the elasticity of labour supply.

The labour supply elasticities in the mixed model lie in-between the two pure models for most of the life-cycle. At the beginning of working life, elasticities in the mixed model are comparable to the *Ben-Porath model*. As women age, they reduce educational investments and elasticities tend towards those found in the *learning-by-doing* model.

### 3.9.2 Varying benefit parameters

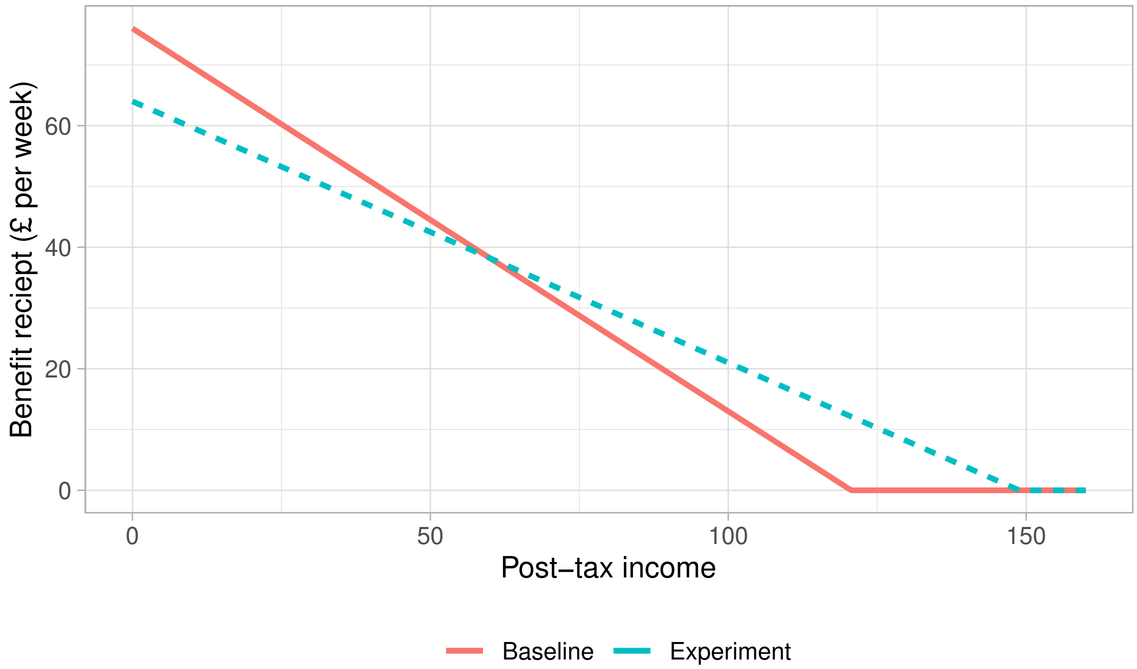
To better understand the interaction between human capital accumulation and benefits, I consider the impact of substituting the baseline benefit schedule for a flatter schedule. As discussed in Section 3.6, the baseline benefit schedule is based on the UK universal credit in the 2020/21 tax year. This schedule included a basic allowance of £75 per week for single individuals over the age of 25 and a taper rate of 63%. In this section I implement a 20 percentage point reduction in the taper rate, from 63% to 43%. To keep government expenditure approximately constant, I reduce the basic allowance to £64 per week. This value was calculated to ensure equivalent expenditure on benefits in the mixed model. The increase in allowance for couples and children are maintained at their previous levels. This reform slightly increases expenditure on benefits in the *learning-by-doing* model and slightly decreases expenditure in the *Ben-Porath* model, but in both cases expenditure on benefits is within 3% of the baseline level. Figure 3.15 compares the baseline benefit schedule and the reduced taper experiment.

The reform considered is similar to the changes in the benefit schedule introduced in the UK in 2021, when a temporary increase in the basic allowance introduced during the COVID-19 pandemic was retracted and the benefit withdrawal rate was reduced by 8%,

#### Learning-by-doing

Figure 3.16 shows the impact of the reduced taper rate in the *learning-by-doing* model on labour hours and wages. Reducing the taper rate decreases the effective tax rate for

Figure 3.15: Benefit schedule comparison




Notes: Comparison of benefit receipt between the model baseline and the policy experiment for a household composed of a single adult.

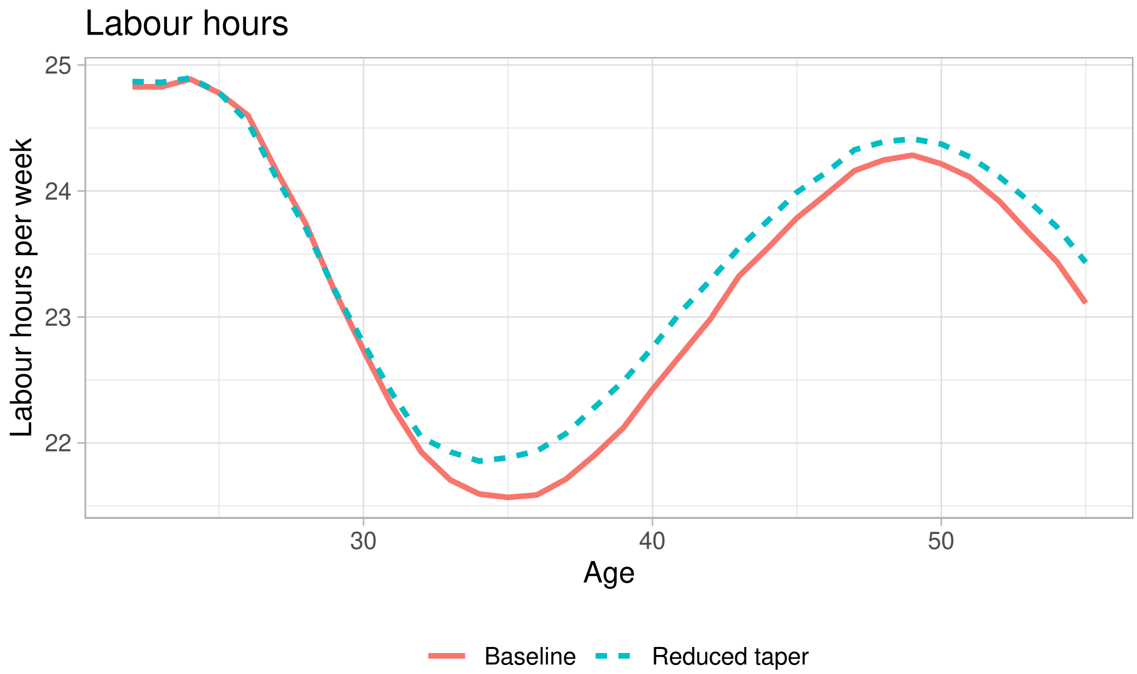
individuals on benefits, increasing labour hours at all ages. As noted above, labour supply elasticities at the beginning of working life are relatively low and the impact on labour hours is consequently small. For the first few years of working life, the primary motivation for working is the accumulation of financial and human capital, and labour hours are therefore relatively unresponsive to changes in the effective tax rate. As women age, the impact on the labour hours increases. Wages also increase across working life, as women accumulate more human capital due to increased work experience.

**Ben-Porath**

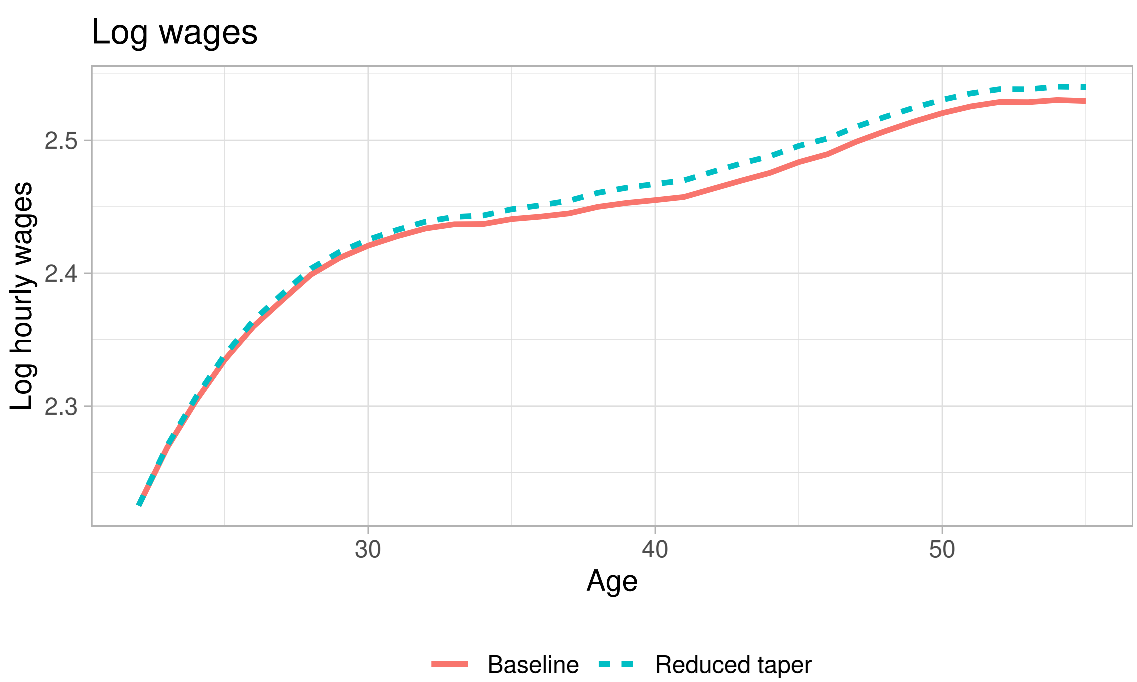
Figure 3.17 shows the impact of the reduced taper rate in the *Ben-Porath* model on labour hours, education hours and wages. As in the *learning-by-doing* model, reducing the effective tax rate increases labour hours at all ages. The increase in labour hours is larger in the *Ben-Porath* model, reflecting the higher labour supply elasticities discussed above. However, the higher labour hours do not translate into increased wages.

Instead, the reduction in the effective tax rate increases the opportunity cost of investment

Figure 3.16: Impact of reduced taper in learning-by-doing model




Source: Model simulations. Line depicts unweighted average of labour and education hours conditional on each age. All averages are taken over the full sample, including individuals who report no labour or education hours.




Source: Model simulations. Line depicts unweighted average of log-wages conditional on each age. Average log-wage is calculated for sub-sample with positive work hours at each age.

in human capital for individuals on benefits. Consequently, education hours fall across the life-cycle. As discussed in Section 3.7, the estimated human capital accumulation function in the pure *Ben-Porath* model is highly concave in education hours. As a result, education hours are inelastic with respect to changes in the effective tax rate. The reform therefore has only a small effect on education hours, and the change in education hours has, in turn, a very limited impact on wages.

**Mixed model**

Figure 3.18 shows the impact of the reduced taper rate in the mixed model on labour hours, education hours and wages. The impact on labour and education hours are qualitatively similar to those in the *Ben-Porath* model. Reducing the effective tax rate increases the earnings return to work for individuals on benefits, and therefore also increases opportunity cost of education. Women increase their labour hours and decrease their education hours at all ages.

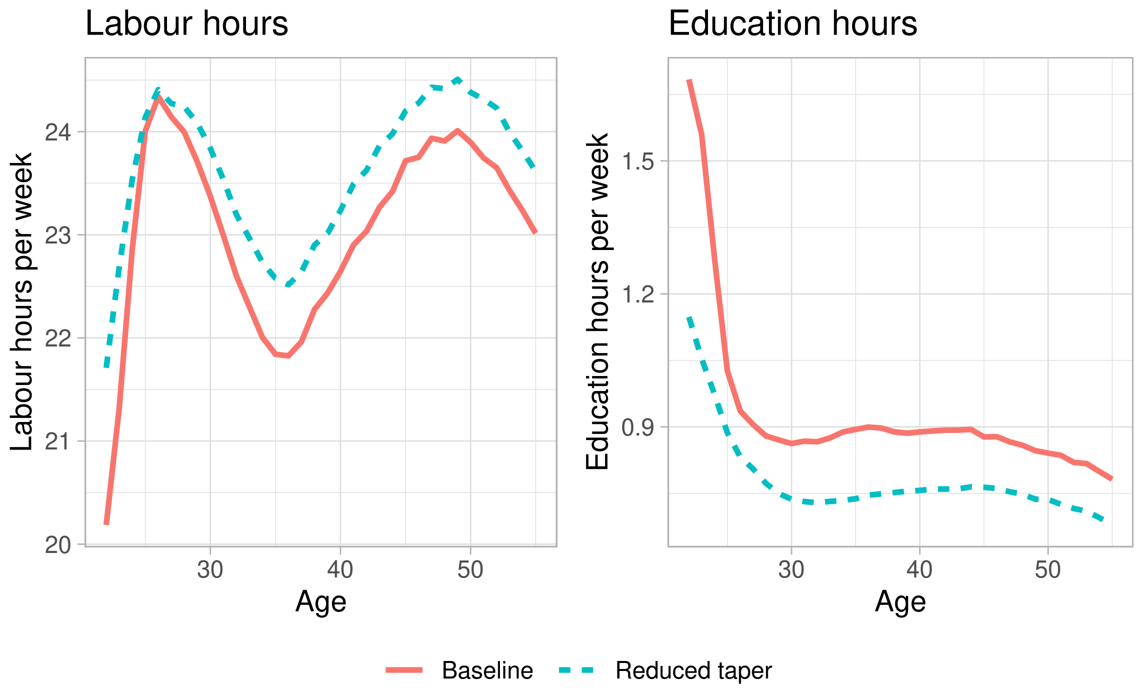
While the impact on labour hours is comparable to that found in the *Ben-Porath* model, the decrease in education hours is much more significant. Educational investments early in working life are reduced from 4 hours per week to 1 hour per week. The increase in labour is initially insufficient to compensate for the resulting loss in human capital, and wages are reduced in early working life. However, from age 40, increased work experience fully compensates for the reduction in education and wages are higher than in the baseline.

As with the labour supply elasticities, behaviour in the mixed model therefore approximates a *Ben-Porath* model at early ages, when wages are low and education investments are an important component of human capital accumulation. Later in the life-cycle, when wages are higher and education investments are low, the predicted impact of benefit reforms are more similar to those found in the *learning-by-doing* model.

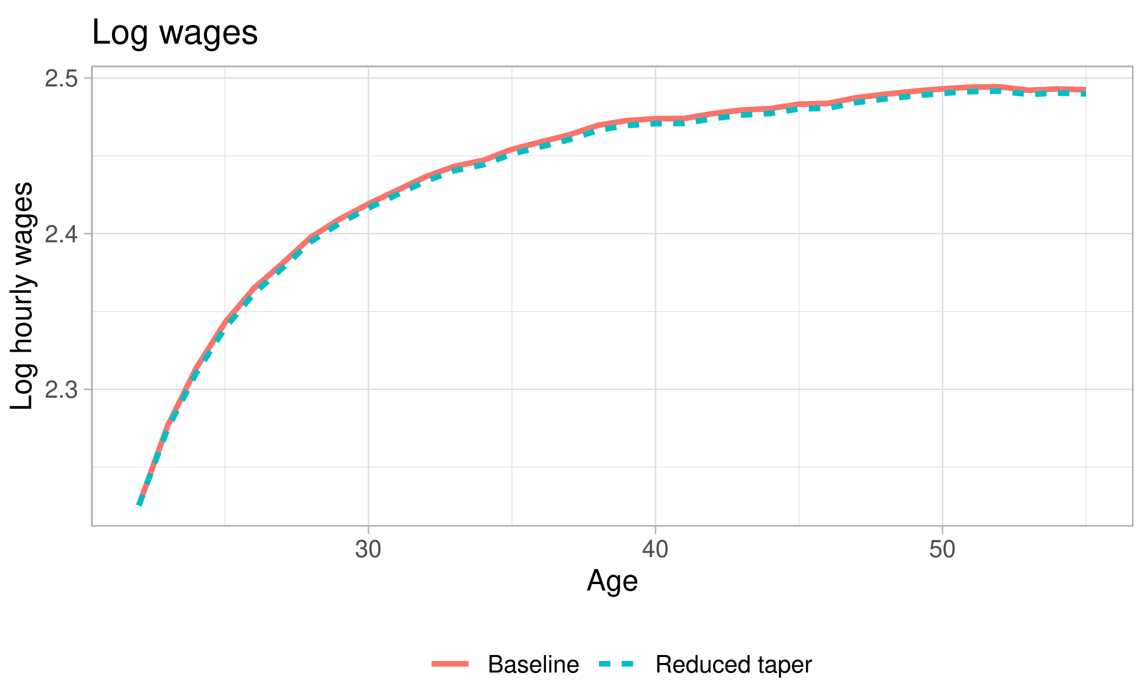
## 3.10 Conclusion

There are three primary findings of this paper. Firstly, neither *learning-by-doing* nor *Ben-Porath* models are able to replicate women's life-cycle profiles of wages, labour hours and education hours. Both pure models understate initial wage growth. The *learning-by-doing*

Figure 3.17: Impact of reduced taper in the Ben-Porath model

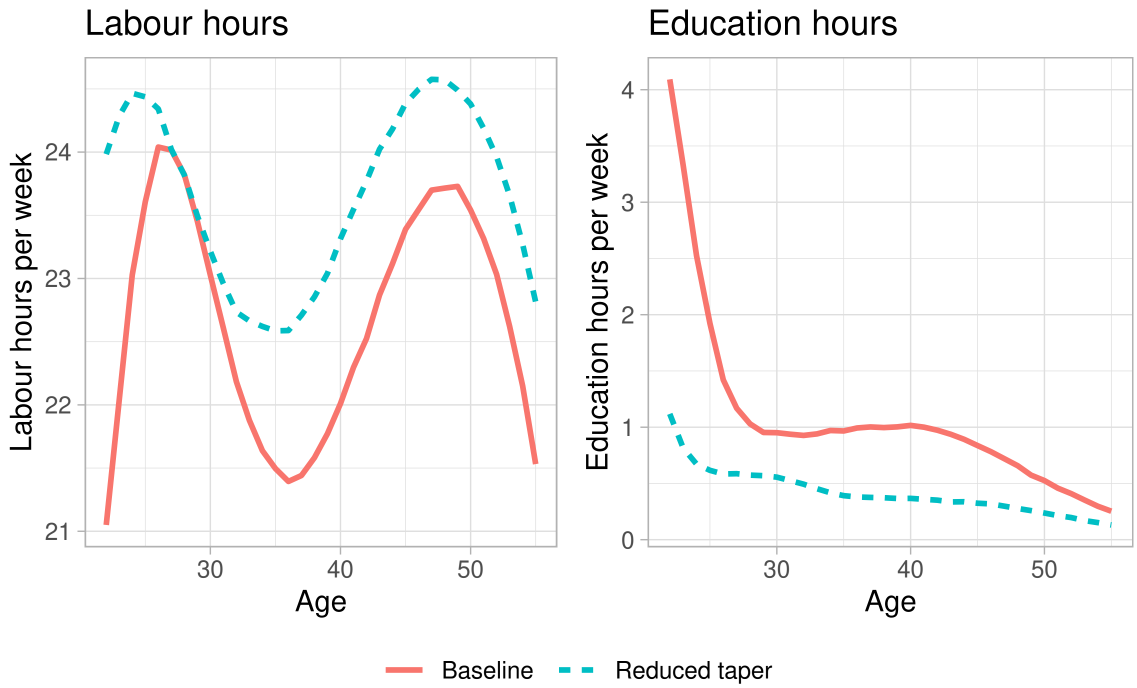



Source: Model simulations. Line depicts unweighted average of labour and education hours conditional on each age. All averages are taken over the full sample, including individuals who report no labour or education hours.

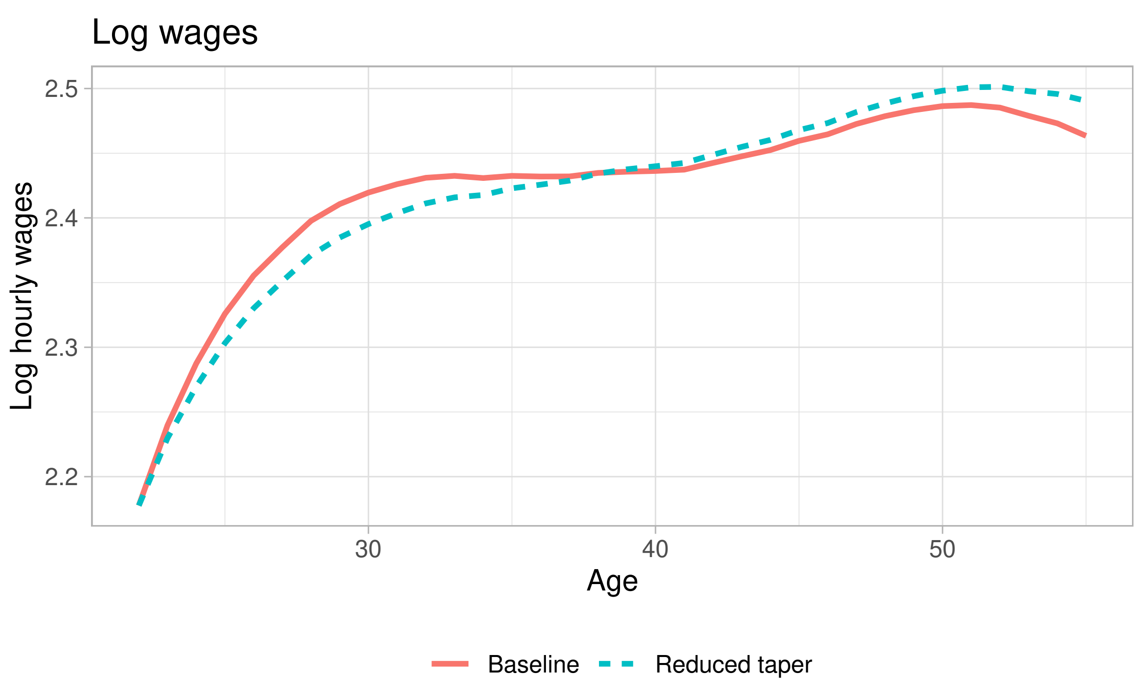



Source: Model simulations. Line depicts unweighted average of log-wages conditional on each age. Average log-wage is calculated for sub-sample with positive work hours at each age.

Figure 3.18: Impact of reduced taper in the mixed model




Source: Model simulations. Line depicts unweighted average of labour and education hours conditional on each age. All averages are taken over the full sample, including individuals who report no labour or education hours.




Source: Model simulations. Line depicts unweighted average of log-wages conditional on each age. Average log-wage is calculated for sub-sample with positive work hours at each age.

model also cannot capture the increase in labour hours observed early in the life-cycle, whereas the *Ben-Porath* model generates education profiles that are too flat over working life. Incorporating both human capital accumulation mechanisms within the same model substantially improves model fit, and also results in more plausible parameter estimates.

Secondly, the impact of benefit reform varies based on the assumed human capital accumulation mechanism. This is because (a) matching the same data with alternative models generates large differences in implied labour supply elasticities and (b) benefit withdrawal generates high effective tax rates, which will impact future wages differently depending on whether human capital is a by-product of labour or a competing use of time. *Ben-Porath* models generate relatively high labour supply elasticities that fall as workers approach retirement, but effective tax rates decrease the cost of human capital accumulation. *Learning-by-doing* models generate relatively low labour supply elasticities that increase as workers approach retirement, but high effective tax rates increase the cost of human capital accumulation.

To illustrate this point, I simulate a reform to the benefit system that reduces withdrawal rates while maintaining overall expenditure on benefits. This reform reduced the effective tax rate for individuals on benefits. When human capital accumulation occurs through labour supply, reducing the effective tax rate has a small, positive impact on both labour hours and wages. When human capital accumulation requires costly investment, reducing the effective tax rate has a larger positive impact on labour hours, but reduces education hours and therefore wages.

Finally, behaviour in my preferred model, which incorporates both modes of human capital accumulation, approximates a *Ben-Porath* model early in the life-cycle and a *learning-by-doing* model late in the life-cycle. This is true both in terms of labour supply elasticities and in terms of the projected impact of reductions in the effective tax rate. As wages increase and women approach retirement, costly education investments become a less important component of the human capital investment portfolio and behaviour increasingly tends towards a pure *learning-by-doing* model. Reforms to the benefit system that increase work incentives may, counter-intuitively, decrease wages early in working life when education is a important driver of wage growth.

The findings of this paper suggest that analysis of the benefit system needs to incorporate a clear understanding of the mechanics of human capital accumulation. I have presented a positive analysis of a particular reform. Normative questions, such as whether Pareto-improving reforms to low income support could be identified or how optimal low-income support varies under different social welfare functions, offer promising areas of future research.

# Chapter 4

# Wages, Experience, and Training of Women over the Life Cycle[1]

## 4.1 Introduction

Women's careers are marked by interruptions related to childbirth and the resulting loss in labor market experience. This, together with the fact that women often work part time while children are growing up, underlies an increasing wage gap relative to men as well as to women who continue an uninterrupted career as full time workers. The question we address in this paper is whether work-related training has a role to play in reducing this wage gap and whether it can be used to help reintegrate women in the labor market following a long absence.

In this paper we specify a model of female labor supply over the lifecycle including the choice to obtain work-related training. In our model, women enter the labor market after completing education. In each period they face a working hours and savings choice. Marriage, separation and children arrive exogenously with a probability estimated from the data and depending on prior children, age and marital status. The evolving family structure over the lifecycle is a key feature because it affects the incentives and preferences of women for work and training. While working their human capital grows through experience, at a rate depending

[1] This chapter is based on a paper co-authored with Richard Blundell (University College London & Institute for Fiscal Studies), Monica Costa Dias (University of Bristol & Institute for Fiscal Studies) and Costas Meghir (Yale University & Institute for Fiscal Studies)

on whether work is part time or full time. Job separations imply a loss in human capital and hence earnings. During their working life they may also participate in work-related training, which is paid for by deductions from their earnings but increases human capital and therefore wages in future periods. While we recognise that part of the cost of training and part of the return may accrue to the firm, we do not explicitly model incidence. However, we do not impose that the worker enjoys the full return to training: we allow the data to determine the returns to training episodes for the worker based on wage data.

Our focus is on the two human capital enhancing activities, working and training. Each of these activities responds to incentives in a different way, which poses interesting policy questions. For example, passive learning in work is encouraged by any factor increasing the incentives to work, such as in-work benefits (EITC in the US, WFTC in the UK). By making working more desirable, these work conditioned policies may also mechanically increase the amount of active work-related training over the life-cycle. Perhaps more interestingly, by topping up low pay benefits can indirectly subsidise the cost of training associated with foregone earnings, see Heckman, Lochner, and Cossa (2003). The design of the subsidy may also interact with the return to training in ways that may increase or reduce its return. Understanding the importance of work-related training for human capital and wages is thus central to designing policy that could help reduce the earnings costs of children on women. In turn, this discussion also reveals that policy reforms changing incentives to work, and to work more, may also affect training rates. In such case, they can be used to identify the effects of training on future wages. We will exploit such variation together with our model to quantify these effects.

Our basic data source is the UK BHPS, a long panel running since 1991 with key labor market and household information. Importantly, it includes detailed information on the incidence and intensity of training. This information is similar to one of the first systematic analyses of work related training by Altonji and Spletzer (1991). We supplement this with information on welfare and tax systems in the UK over many years, which allows us to construct the precise budget constraint that an individual is facing in each year of work. This leads us to our identification strategy: our data includes multiple cohorts, entering the labor market at different times. Each is facing a different welfare and tax system implying changes in incentives. During their lifetimes they face reforms that affect a number of cohorts but at different ages. This generates exogenous variation in the incentives that people face at different parts of the distribution. Thus individuals of different cohorts and education groups face both different work and training incentives. This is the key idea that underlies our identification strategy and provides the variation we need to estimate the model.

Our findings point to a potentially important role for training women who completed high school level education but did not go on to complete University. We show that it can have a role in reducing the wage loss that arises from part time work post children. Moreover, policies that subsidize the training of recent mothers from this group can increase their disposable income (beyond the taxation required to fund it) as well as overall welfare. We also find that a modest subsidy pays for itself by incentivising full-time work both during the eligibility period and after it. Finally, while training can play some role in reducing the labor market costs of children, this cost remains quite large even after systematic training policies. Other policies that would reduce the incidence of part time work, such as better childcare availability, may have a more important role to play.

The paper proceeds as follows: In the next section we describe our data, followed by a description of the institutional framework. We then carry out an empirical analysis to investigate how incentives related to the tax and welfare system affect training. Having shown that training is indeed sensitive to such incentives we specify our model and describe our estimation approach, which uses the simulated method of moments. This section is followed by the description of the results including our counterfactual simulations. We then offer some concluding remarks.

## 4.2 Data

Estimation uses the 18 yearly waves of the British Household Panel Survey (BHPS), a longitudinal dataset following the lives of families and their offshoots from 1991 to 2008. The survey started with a representative sample of 5,050 households living in Great Britain; it was later replenished in 1997 and 2001 with 1,000 households from the former European Community Household Panel, and in 1999 with two samples of 1,500 households each from the Welsh and Scottish extensions.[2] Except for some attrition, all household members in the original samples remain in the sample until the end of the period. Other individuals have also been added to the sample, as they formed families with original members of the panel or were born into them.

The BHPS collects detailed demographic information that we use to characterise the dy-

[2] An additional sub-sample from Northern Ireland was added in 2001 but is not used here.

namics of family formation, as well as socio-economic information mapping the education attainment, labour supply, earnings, training events, childcare expenditures and assets of all household members aged 16 and above. In 1992, 2001 and 2002, the BHPS contains an additional module on lifetime histories that we use to recover the employment history of adult respondents since they first started to work. Respondents also report retrospective information on family background, including measures of parental education, number of siblings, sibling order, whether they lived with parents when aged 16, books at home during childhood, etc. We synthesise this information into two indices of socio-economic background that will be used to qualify individual earnings capacity and choices.

Our observation unit is women who have completed education, are aged 19 to 60, and for whom we observe complete employment histories. The histories of women who return to full-time education to acquire additional qualifications are truncated. We also truncate the histories of those who become self-employed at any point during the sample period, from that moment onwards. Finally, we exclude women who are not UK citizens or who are ever observed claiming disability benefits. The records of women in the cleaned sample are then linked to information on a present partner and children as relevant.

Our final sample is an unbalanced panel of 7,359 women and 55,591 observations. We arrange them into three groups by highest level of completed education, corresponding to less than high-school, high-school qualifications and equivalent, and 3-year college degree and above.[3] Table 4.1 shows the sample composition by family type and education of the woman.

We consider both the extensive and intensive margins of labour supply and discretise the distribution of labor supply to 3 points: not working for pay, which we take to be 0 hours in paid work per week and corresponds empirically to the cases of workers doing less than 5 weekly hours of work; working part-time, which we take to be 20 hours of work per week and combines all those doing 5 to 20 hours; and full-time work, which we take to be 40 weekly hours and combines workers doing 21 or more hours per week. The underlying measure of weekly hours we use is for usual hours in main job, including paid and unpaid overtime. We also consider only employees, and delete the paths of workers becoming self-employed, from that moment onwards. More details on data selection can be found in the Online Appendix.

Wages are measured on a per-hour rate by dividing weekly earnings in main job including

[3] In the UK, these levels correspond, respectively, to GCSE qualifications (which are acquired at the end of secondary school, at age 16) and below, A-levels qualifications (obtained at the end of high-school, aged 18) and equivalent, and 3-year University degree and higher.

Table 4.1: Sample size and distribution of family types by education

| | Education | | | |
|---|---|---|---|---|
| | Less than high school | High school | University | Total |
| Family type (%) | | | | |
| Single, no kids | 15.1 | 21.0 | 24.7 | 18.2 |
| Couple, no kids | 34.6 | 33.6 | 35.6 | 34.4 |
| Single, with kids | 11.1 | 7.9 | 4.6 | 9.2 |
| Couple, with kids | 39.2 | 37.5 | 35.1 | 38.1 |
| Employment (%) | | | | |
| Full-time (>20 hours) | 53.2 | 68.9 | 77.3 | 61.2 |
| Part-time (5-20 hours) | 21.2 | 15.6 | 11.6 | 18.2 |
| Nr of individuals | 3,921 | 2,377 | 1,061 | 7,359 |
| Nr of observations | 30,802 | 17,419 | 7,370 | 55,591 |

Notes: BHPS data for the years 1991 to 2008.

paid overtime by weekly hours also in main job (including any overtime, as detailed above). Since our model does not deal with macroeconomic fluctuations, we net out aggregate wage growth from the wage rates and from all monetary values of the tax and benefit system, described below in Section 4.3. We also trim the wage rate distribution, on the 2nd and 98th percentiles, to limit the importance of measurement error in earnings and working hours.

**Training Data.** One distinctive feature of the BHPS is that it includes a detailed description of all work-related training taking place during the year prior to the interview among those currently employed. This measure of training is an umbrella to a wide variety of education activities meant to increase or improve skills in work and that can be pursued while working full or part time hours. It includes part-time college or university courses, evening classes, employer-provided courses either on or off the job, government training schemes, open university courses, correspondence courses and work experience schemes, but excludes full-time education. Work-related training amounts to over 80% of all recorded training episodes, of which 96% happen among those in paid work at the time of the interview. The data documents the purpose of the training (whether induction training in a new job, to gain skills for current job or to prepare for some new job in the future), its total duration, who paid for any direct costs, where it took place and whether it lead to any qualification.

Our measure of training is an indicator for whether the respondent has had strictly more than 40 hours of training over the previous year. In calculating the total time in training over the year, we have excluded instances of training for induction in a new job or where the participants report as it being unrelated to work. Specifically, we only consider training spells meant to increase the skills workers need in their current job (for example by learning a new technology), or to prepare for a new job; we exclude training meant to help workers getting started in their current job (induction training) or to develop skills generally (not work-related). We also exclude the 4% of cases where trainees are not working. For the remaining instances of training we first convert total duration – which can be reported in months, weeks, days or hours – into hours assuming 8 or 4 hours in a day for those in full-time or part-time hours, respectively. We then exclude all training episodes that result in 40 hours or less of training per week since they seem likely to capture minor work-based certification programmes, such as first-aid training.[4] Conditional on our selection, 76% of the training we account for leads to formal qualifications. This we take as suggestive evidence that the training considered here is human capital enhancing and transferable across jobs and firms.

Table 4.2 briefly describes training spells among women, by education. We show figures for our measure of training, labelled 'selected', and for a similar measure constructed on all work-related training, labelled 'any'. Panel (a) of the Table shows that training is a common event, with between 17 and 37 percent of employed women receiving some form of training in each year. It is also much more common among those in the middle and top education groups. Our more demanding measure of training accounts for just over 40% of all training spells. These are non-negligible investments, with a median length of between 80 and 96 hours per year, or between 2 and 3 full-time weeks (panel (b)). In a working year of 48 weeks, the median training duration amounts to an average of about 2 hours of job-related training per week.

Panels (c) and (d) focus in Table 4.2 narrow the sample to include only trainees under our preferred definition. Women who have not completed high-school education are more likely to receive training at work (50%) than either High School educated women (36%) or University educated women (28%). University educated women are often trained at work, at private training centres. Around one-quarter of training occurs at a university or further

[4] In robustness checks, we have included induction related training and used a continuous training hours measure. The life-cycle patterns and our regression analysis shows (discussed below) are not qualitatively affected.

Table 4.2: Training descriptives for women by education (BHPS)

| | **Education** | | | |
|---|---|---|---|---|
| | **Less than High School** | **High School** | **University** | **Total** |
| ***(a)* Training rates for employed (%)** | | | | |
| Any training | 17.1 | 33.4 | 37.0 | 27.4 |
| Selected training | 5.4 | 14.3 | 16.2 | 11.1 |
| ***(b)* Median hours of training for trainees (hrs per year)** | | | | |
| Any training | 24 | 40 | 40 | 32 |
| Selected training | 80 | 96 | 88 | 88 |
| ***(c)* Where did training take place (selected training, %)** | | | | |
| At work | 50.3 | 36.4 | 28.6 | 36.3 |
| College/university | 22.8 | 27.6 | 26.2 | 26.5 |
| Other | 26.9 | 35.9 | 45.2 | 37.2 |
| ***(d)* Who paid explicit fees, if charged (selected training, %)** | | | | |
| Fees paid by employer | 69.3 | 71.0 | 71.5 | 70.9 |
| No fees paid by employer | 30.7 | 29.0 | 28.5 | 29.1 |

Notes: BHPS data for the years 1991 to 2008. All figures exclude instances of education or training spells that are not work-related. Training measured only for those in work at the time of the interview. 'Selected training' further excludes induction training and instances of training that add up to 40 or fewer hours of training in the course of one year.

education college, across all three education groups. When explicit fees are charged for training, these fees are paid by the employer in between 69 and 72 percent of instances. However, this measure does not account for additional costs of training, such as the loss of income that could result from fewer working hours.

## 4.3 Institutional background

The personal tax and welfare benefit systems operating in the UK during the 90s and 00s all consist of a small set of individual-based taxes and a larger set of benefits that are mostly means-tested on family income. Within the same structure, the period saw numerous reforms to the specific parameters determining entitlement to benefits and tax liabilities. The most significant was the sequence of reforms to the benefits of families with children that ocurred between Autumn 1999 and April 2002, which introduced the Working Families Tax Credit

(WFTC) and changed the Income Support (IS) benefits for low-income families. We exploit these reforms in addition to other smaller changes in taxes and benefits to identify the returns to work experience and training and to study how welfare policy may affect training. We do so by modelling women and their families living through two tax and benefit systems that are representative of the main institutional features over the period of the data: that operating in April 1995, describing the policy environment of the 90s, and that finally implemented in April 2002, after the WFTC-IS reform was completed. Here we describe the main features of these systems; a more comprehensive discussion of the taxes and transfers in the UK can be found in Adam, Browne, and Heady (2010) and Blundell, Costa Dias, Meghir, et al. (2016).

In terms of tax liabilities, the main instruments targeting families are the Income Tax and the National Insurance contributions. The basic structure of these taxes remained unaltered over the period. Income Tax is progressive, a step function over four income brackets. The 1995 system comprised of a personal income disregard that was not taxed, and rates 20% (starting), 25% (basic) and 40% (higher) that were gradually applied to additional fractions of personal income. The period saw a mild tax reduction, with a modest increase in the personal income disregard and some reduction of the rates to 10%, 22% and 40%. This was partly compensated by adjustments in the basic income threshold defining the brackets at which the starting and basic rates apply, and by a small increase in the main rate of National Insurance contributions, from 10% to 11%.

The UK benefit system is more complex. We model a range of benefits, including: Job-Seekers Allowance (JSA), which is the UK unemployment benefit; Income Support (IS), a minimum income floor that carries no work or job-search requirement; Tax Credits, a benefit for working families; Child Benefit, a universal benefit for families with children; Housing Benefit (HB), which subsidises housing costs for families who live in rented accommodation; and Council Tax Benefit, which subsidises the local property tax. These benefits interact in complex ways, so it is important to consider them together.

For mothers, the key components of the public transfer system are the IS and the Tax Credits. These were also the focus of the WFTC-IS reform of 1999-2002, an intervention aimed at improving the financial circumstances of low income families with children and keep mothers in work to protect their skills and labour market attachment. The reform implemented a significant increase in the generosity and coverage of IS and Tax Credits. For lone mothers, the IS award increased by over 10% relatively to wage levels over the period and remained taxed at 100% marginal rate. Since this subsidy is not work-contingent, this aspect of the reform reduced the incentives to work of mothers. The reform of the Tax Credit benefits, however, counteracted the increase in out-of-work benefits with a generous increase in subsidies for working mothers and an expansion of the target population to higher levels of family income. This was implemented by a 25% rise (in constant wage levels) in the maximum award for lone-mothers of one child, and a drop in the withdrawal rate from 70% to 55%. Over this period, Tax Credits kept the minimum working hours eligibility rule of 16 hours per week as well as the additional award for families working at or above the 30 hours threshold.

Figure 4.1 summarises the effects of these reforms on the take-home pay of single mothers. It shows, in 2008 prices and for a lone-mother on the minimum wage of April 2004, her entitlement (on the left) and disposable income (on the right) by working hours per week. The strong incentive to work part-time hours is clearly visible both before an after the reform. It is also apparent that the reform increased the incentive to work both part-time and more hours, by increasing the award at 16 hours by more than it increased out-of-work benefits and by reducing the rate at which in-work benefits are tapered away.

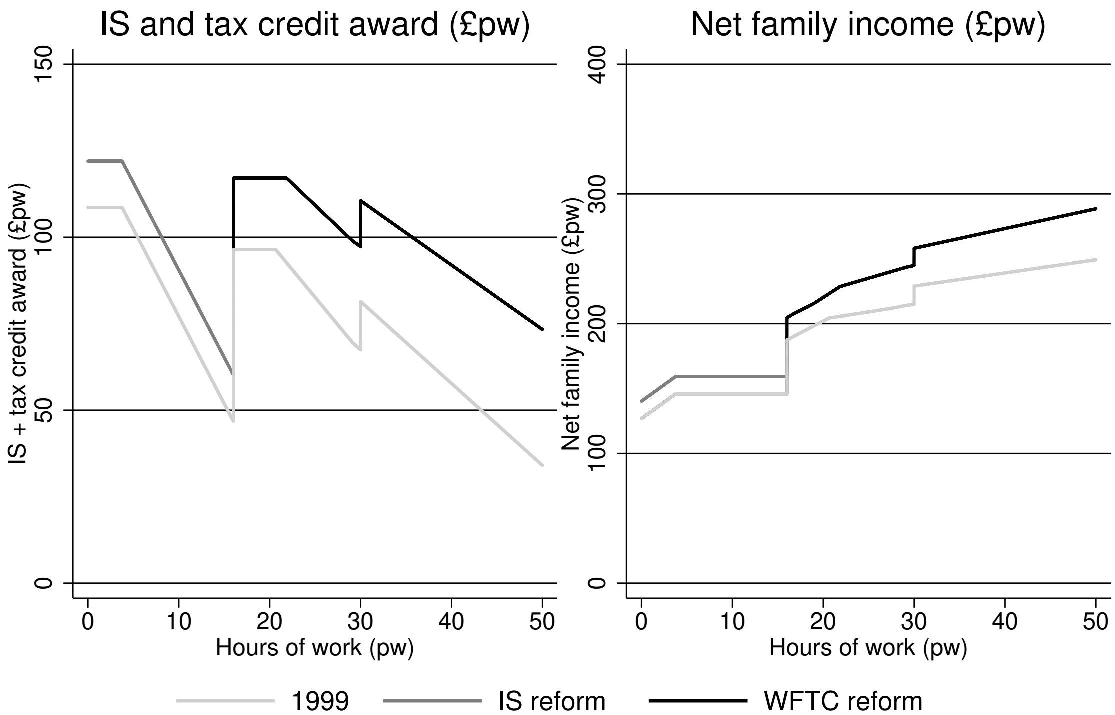
Figure 4.2 pictures the equivalent quantities for low-paid couples with one child aged 4 with one spouse working 40 hours per week at the 2004 minimum wage, by working hours of the second earner. Clearly, the reform had a much more modest effect on the disposable income of couples and, if anything, it reduced the incentives to work of the second earner in the family by taxing additional earned income more heavily.

## 4.4 Life-cycle profiles of employment and training

The life-cycle patterns of wages, labour supply and training are suggestive of how these variables are linked for women, and of the motivations behind investments in training. Figure 4.3 shows the life-cycle profile of average log hourly wages of women and men, by education. The dashed lines for women exhibit the typical strong gradient by education and a steep upward profile early in the working life, particularly for high-school and university graduates. However, women's wages quickly flatten out during their late 20s or early 30s, coinciding with the main fertility period. The flattening is permanent after that.

The solid lines for men show wages increasing with education and growing rapidly in the early years of working life. However, the wages of men continue to grow far later into working life than the wages of similarly educated women, independently of education. The

Figure 4.1: Income Support and Tax Credit for minimum wage lone parent with 1 child

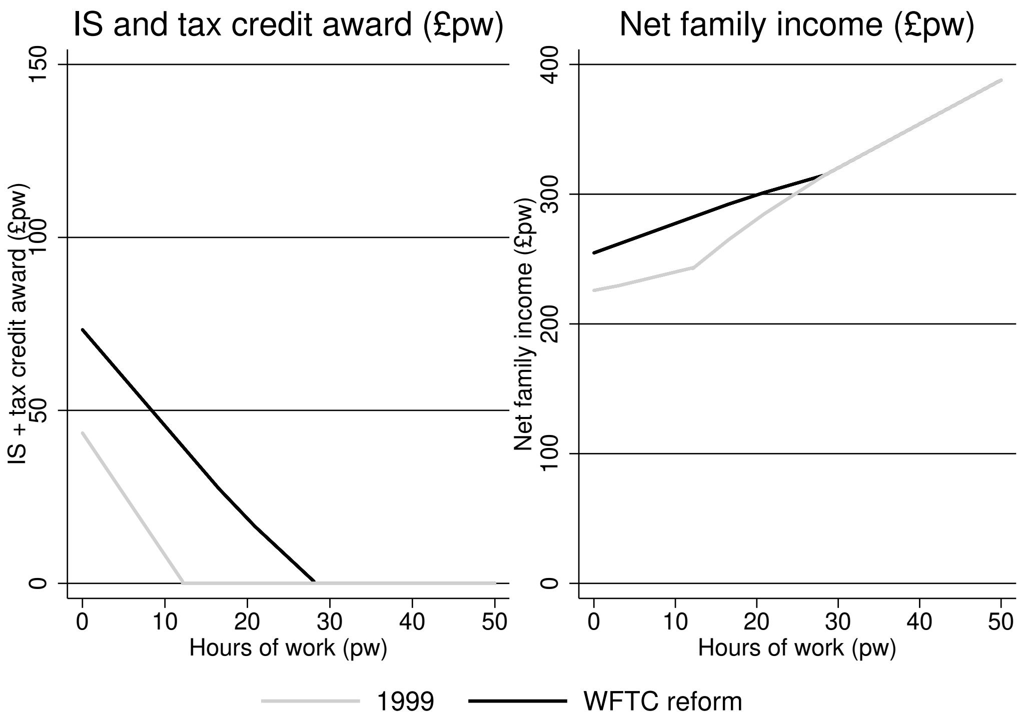



Notes: From Blundell, Costa Dias, Meghir, et al. (2016). Simulations from FORTAX for lone-mother of one child aged 4, earning the 2004 minimum wage, not paying housing rents or childcare. Graph on the right pictures the IS plus TC award, graph of the left pictures the disposable income of the family; both in 2008 prices by working hours of the mother.

continued growth of men's wages compared to a flattening of women's wage profiles opens up a gender wage gap. For low educated women, this gap is already apparent by their early 20s. For higher educated women, the gap opens in their late 20s. These patterns coincide with differences across women by education in the timing of childbirth. For instance, 51% of women with less than high school qualifications in our sample have at least one child by age 23. This compares to 4% of University educated women. University educated women only reach comparable levels at age 32, where 50% of our sample have at least one child.

This wage profile is accompanied by strong changes in labour supply. Figure 4.4 shows, on the left, that the employment rates of women dip in the middle of their working lives. The dip happens earlier and is more pronounced for the lower educated. The right panel shows the proportion working part-time among women in work. The same period witnesses a strong growth in part-time hours that persists into late working life, particularly for those with high-school qualifications and less. Overall, employment and full-time working hours

Figure 4.2: Income Support and Tax Credit for low-paid couple with 1 child



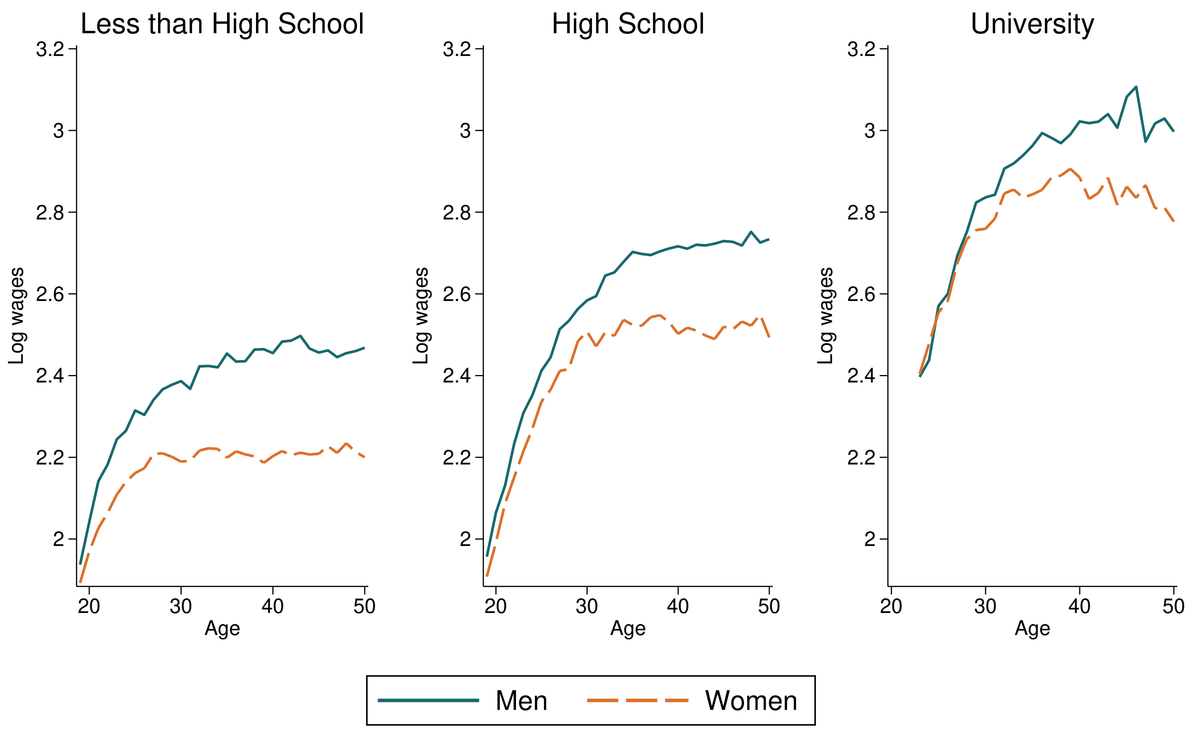

Notes: From Blundell, Costa Dias, Meghir, et al. (2016). Simulations from FORTAX for couple of one child aged 4, not paying housing rents or childcare, both spouses earning the 2004 minimum wage, one spouse working 40 hours per week. Graph on the right pictures the IS plus TC award, graph of the left pictures the disposable income of the family; both in 2008 prices by working hours of the second earner.

seem strongly complementary with education.

Blundell, Costa Dias, Meghir, et al. (2016) documented these working patterns, related them to fertility episodes and quantified their consequences for the wage progression of women with different levels of completed education. What that paper did not consider, however, is how work-related training interacts with education, labour supply, work experience and wages. Here we see training as one element of human capital, together with education and work experience. Whether these three factors are complements or substitutes in the formation of wages will have consequences for the intensity and timing of training across different groups. For instance, if training can be used to offset human capital depreciation from non-working periods then it may be more prevalent among women returning to the labour market after a long fertility-related interruption than among men of similar age.

We start investigating this by contrasting the training patterns of women and men over

Figure 4.3: Average log wages of employed women and men over the life-cycle, by education

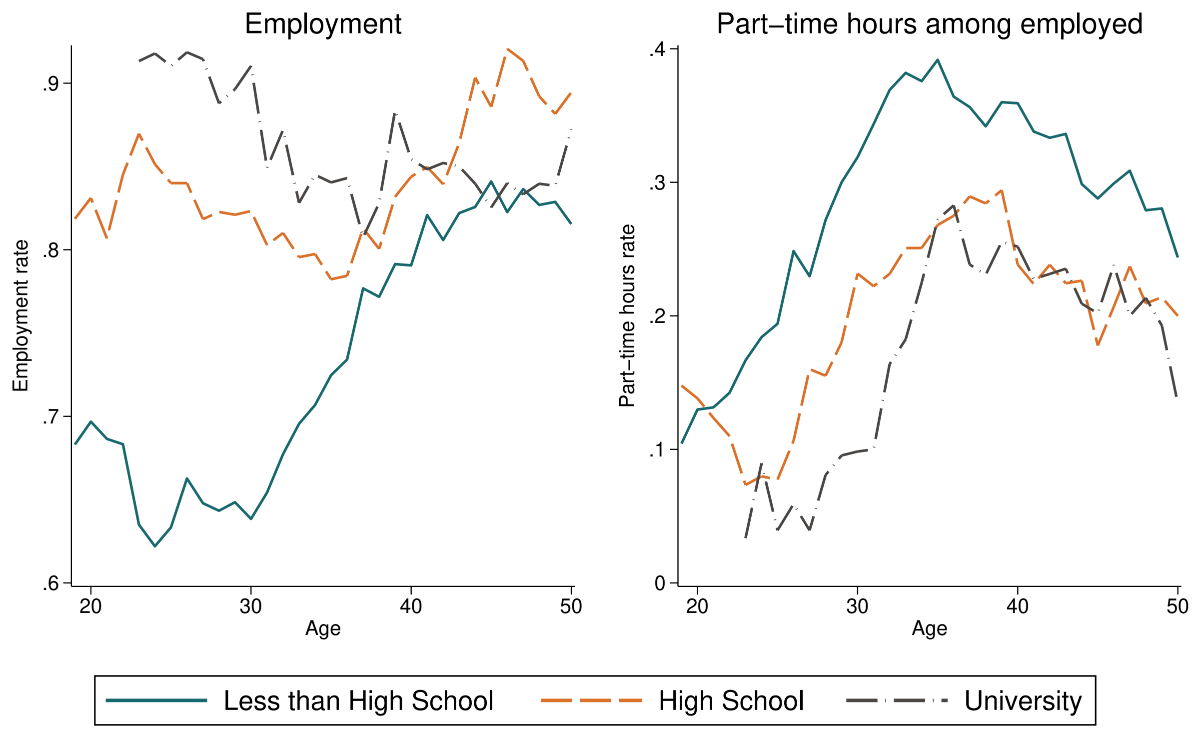



Notes: BHPS data for years 1991-2008. Real wages measured on a per-hour rate in logs.

the course of life in Figure 4.5. Panel A of this figure shows training rates by gender and education for all individuals, independently of work status (with training for those out of work always set to zero). Several features are noteworthy. First, on-the-job training is very common among High-School and University graduates. There is a clear education gradient in training, with workers with less than high school qualifications being much less likely to invest. This suggests that, like work experience, the type of training that we measure is complementary with education instead of being used to compensate for the lack of academic skills.[5] Second, despite women being much more likely to interrupt their careers during the main child-rearing period, the training rates of women and men are surprisingly similar. This holds even at the start of working life, at which point women may foresee a long career interruption linked to fertility in the near future. Third, the overall pattern of training is downward slopping, as predicted by the classical Mincer/Ben-Porath human capital framework. Noticeably, however, the slope is not monotonic for women, particularly

[5]One alternative explanation is that our measure favours training that is closer to the type that high educated people receive, and that other types of training (needed, for instance, for manual jobs) are not captured by our data.

Figure 4.4: Employment and working hours over the life-cycle, by education



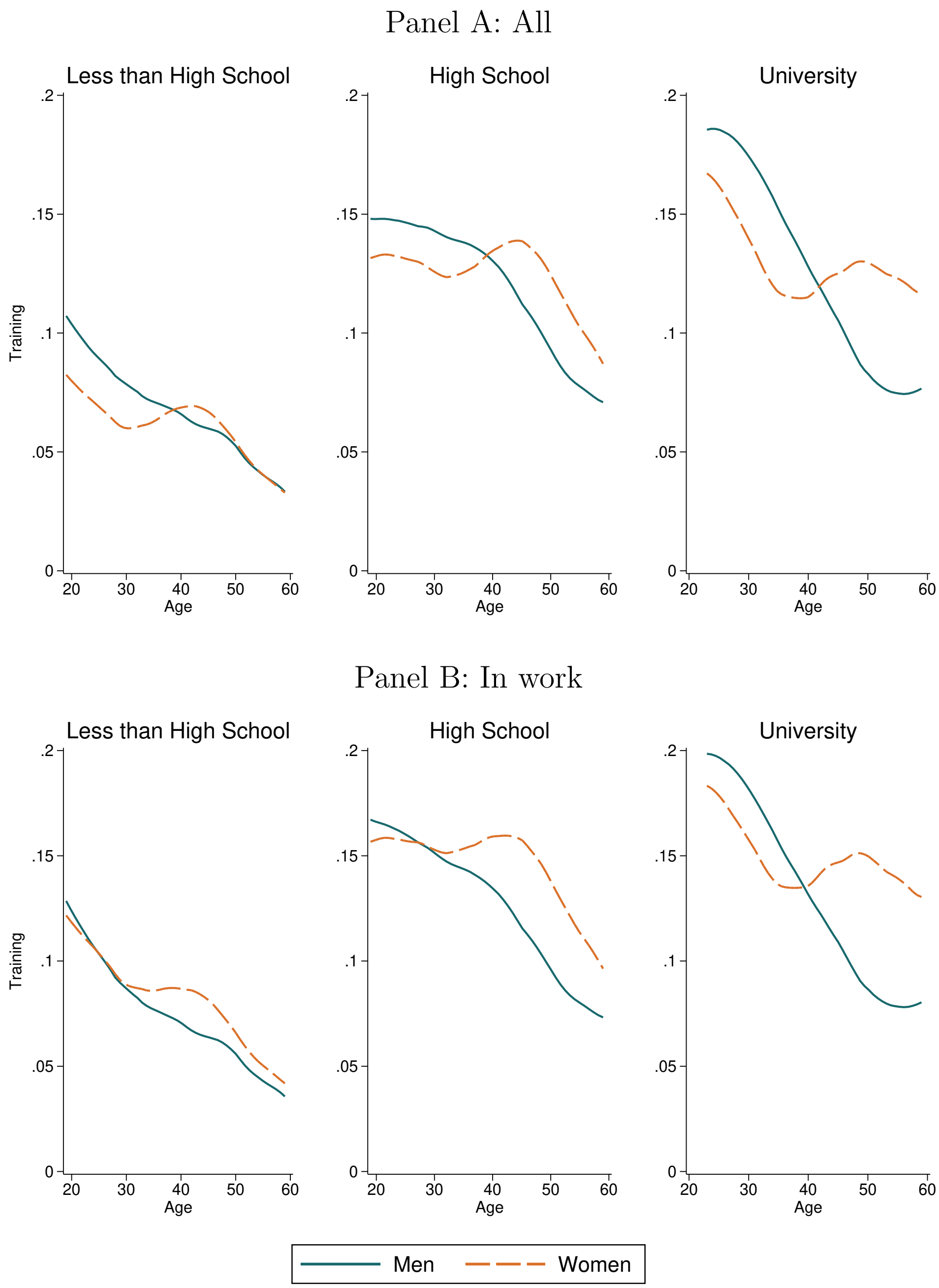

Notes: Notes: BHPS data for years 1991-2008. The graph on the left shows employment rates by age and education. The graph on the right shows the proportion of working women in part-time hours conditional on being in work, also by age and education.

so for the more educated. Instead, training rates peak for a second time when women in these education groups are in their 40s or early 50s, a period that coincides with many of them returning to full-time work.

Conceivably, these patterns can be mechanically driven by the life-cycle of employment among women. Specifically, since female employment rates drop markedly during the main childrearing periods and recover once children are older, lower training rates at that stage and their subsequent pick up may just reflect that movement out and back into work. Panel B refutes that hypothesis by showing similar life-cycle variation in training rates among those in work.

Figure 4.6 provides further insight on the timing of training by plotting its frequency around the birth of the first child. It shows that the training rates are flat around the time of first birth for women with less than high school qualifications, seemingly unaffected by childbirth.

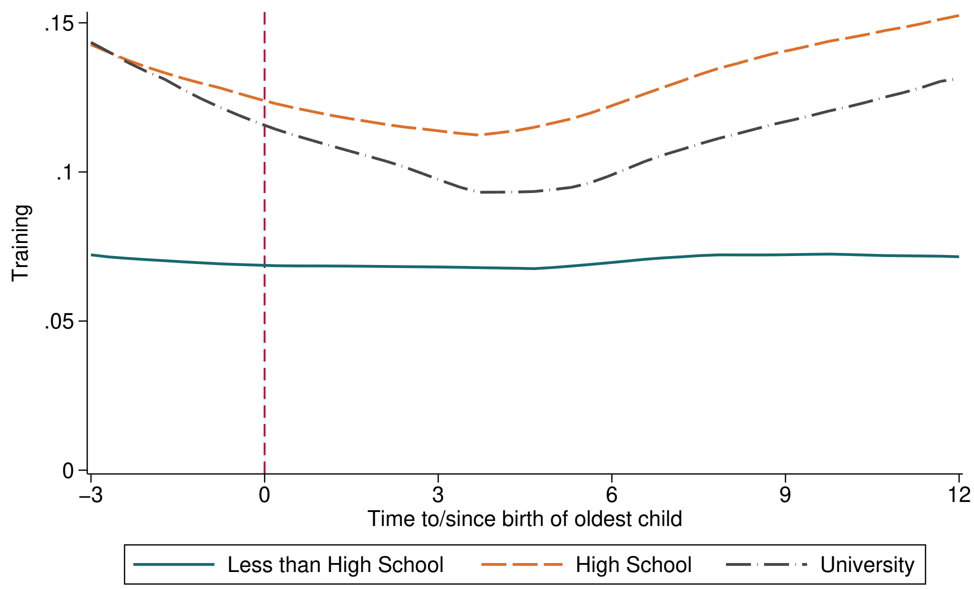
Figure 4.5: Training rates over the life cycle, by gender and education




Notes: BHPS data for years 1991-2008. The training variable is an indicator for having had 40 or more hours of work-related training over the last 12 months. Panel A shows training rates for the entire population, by age, gender and education. Panel B additionally conditions on working at least 5 hours per week on an usual week, which is the measure of employment used in this paper. Lines are smoothed using a Epanechnikov kernel.

In contrast, the training rates of women with high school or university qualifications vary significantly around chilbirth, first declining to reach a minimum while the child is very young and later partly recovering as the child moves to primary and secondary schools.

Figure 4.6: Training rates among mothers and mothers-to-be in paid work, by time to/since birth of first child and education




Notes: BHPS data for years 1991-2008. The training variable is an indicator for having had more than 40 hours of work-related training over the last 12 months.

These patterns suggests a role for training in offsetting some of the losses in human capital and earnings capacity due to career interruptions, at least among mothers with High School qualifications or more. It is unlikely though that training alone will be enough to close the kind of gender differences in pay shown in figure 4.3. Even if the returns to training are similar to those from additional years of formal education, training spells are generally much shorter and so we would expect an effect that is proportionally adjusted. But training may, nevertheless, speed up gains in skills that women lose during working interruptions and make work more valuable for them.

The life-cycle patterns of training also suggest a role for public policies subsidising working mothers that has received little attention so far (one notable exception being Heckman, Lochner, and Cossa (2003). Specifically, working incentives targeting mothers – such as the UK Tax Credits that we described before or the US Earned Income Tax Credit – may have unforeseen effects on the take up of training through various channels. First, by making working more desirable they may mechanically increase the amount of training over the entire life-cycle. Second, by increasing the number of periods that women are in work, wage subsidies will also increase the number of periods over which women will reap the return from training, hence overall increasing the total return to the investment. Third, by topping up low pay, the benefits may indirectly subsidise the cost of training associated with foregone earnings. And finally, the design of the subsidy may interact with the return to training among subsidised women in ways that may increase or reduce its return.

## 4.5 Training responses to work incentives

One observation from the discussion in the previous sectin is that reforms in incentives to work may provide useful exogenous variation to identify the impact of training on the earnings of women. Existing studies have mostly focused on the impact of tax reforms on employment and hours. For instance, it has been shown that the WFTC reform affected the labour supply of lone-mothers (e.g. Blundell, Costa Dias, Meghir, et al. 2016; Brewer et al. 2006). Here we show that the various reforms to the tax and benefit system that happened in the UK over the 90s and 00s, of which the WFTC reform is a prominent example, also affected the probability that women take-up training.[6] This implies that tax and benefit variation can be used to help identify the returns to training in the context of a life-cycle model.

Our empirical specification is very simple. We estimate the following regression model of training $T$ on a set of three simulated income variables that describe how working incentives change over time for different families, in response to policy changes:

$$T_{it} \quad = \quad \mathbf{1}\left[\gamma_0 + \gamma_1 \hat{Y}_{it}^O + \gamma_2 \hat{Y}_{it}^P + \gamma_3 \hat{Y}_{it}^F + \gamma_4' X_{it}' + \epsilon_{it}^T \geq 0\right] \tag{4.1}$$

In the above, the dependent variable $T_{it}$ is an indicator for having had more than 40 hours of training over the last 12 months for woman $i$ at time $t$, and $\left(\hat{Y}_{it}^O, \hat{Y}_{it}^P, \hat{Y}_{it}^F\right)$ are the respective simulated income variables. They measure family disposable income for three scenarios of female labour supply, respectively not working (superscript $O$), working part-time hours (respectively $P$) and working full-time hours ($F$). We use the tax system in place in period $t$ to simulate these incomes based on predicted female wages (on her age and education) and details of the demographics of the family.[7] $\hat{Y}$ single out how policy reforms differentially affect

[6] We supplemented the variation in the monetary incentives to work with local variation in the availability of training captured by a Bartik instrument. We found that geographical variation to be too weak to drive training rates and dropped it.

[7] We use the IFS micro-simulation program Fortax, which provides a detailed description of the taxes

the resources of families of different types depending on the labour supply of women. We also control for a set of other covariates $X$, which includes time dummies, a quadratic polynomial in age, indicators for family composition, two indices that summarise parsimoniously a set of observed variables characterising the socio-economic background of the woman.[8] These variables are meant to control for variation in the disposable income variables not induced by policy reforms.

Table 4.3: Regression of training on simulated income

| | (1) Less than High School | (2) High School | (3) Degree |
|---|---|---|---|
| Simulated income: no work ($\hat{Y}^O$) | -0.000254** | -0.000280 | -0.000178 |
| | (0.0000881) | (0.000145) | (0.000209) |
| Simulated income: part-time ($\hat{Y}^P$) | 0.000606*** | 0.000524* | 0.000751* |
| | (0.000146) | (0.000238) | (0.000376) |
| Simulated income: full-time ($\hat{Y}^F$) | -0.000705*** | -0.000878*** | -0.000960*** |
| | (0.000105) | (0.000150) | (0.000223) |
| Observations | 30383 | 17260 | 7328 |
| Demographic Controls | Yes | Yes | Yes |
| Family Background Controls | Yes | Yes | Yes |
| Year Dummies | Yes | Yes | Yes |
| Age polynomial (2nd order) | Yes | Yes | Yes |
| F-Test on Instruments | 20.93 | 15.53 | 8.312 |
| F-Stat p-val | 0.00 | 0.00 | 0.00 |

Notes: BHPS data for years 1991-2008. Outcome variable is an indicator for whether the woman has taken more than 40 hours of work-related training during the year that precedes the interview. Estimates show effects of simulated family disposable income for different levels of female labour supply on the probability of taking up training. The simulations are constructed using a detailed microsimulation model for the UK. We use the tax system in place in period $t$ to simulate these incomes based on predicted female wages (on her age and education) and details of the demographics of the family. The regressions also control for year dummies, demographic characteristics (including a quadratic in age and dummies indicating family composition) and family background (including the first two principal components drawn from a collection of variables that describe the childhood household of each individual and an indicator for whether this information is missing). The F-statistics at the bottom of the table test the joint significance of the three simulated income variables. Standard errors, shown in parentheses under estimates, are clustered at the individual level.

Table 4.3 displays the results, focussing on the income variables. It shows that changes in incentives to work strongly affect the probability that women enrol in significant amounts

and benefits operating at each time period.

[8] The indices are the first and second principle components of a set of observed retrospective variables on parental background, from when the woman was 16 years of age. They summarise information on the education of both parents (five levels each), number of siblings and sibling order (dummies for no siblings, three or more siblings, and whether respondent is the first child), books in childhood home (three levels) and whether lived with both parents when aged 16.

of training. The F-statistics at the bottom of the table show that this is especially true for the two bottom education groups. This is not unexpected since public policies target the bottom of the income distribution and are, therefore, more effective in influencing choices at that margin.

Estimates in Table 4.3 are for all women, regardless of their employment status. Since the type of training that we are considering only happens among those in work, it could be thought that our estimates are effectively capturing the effects of monetary incentives to work on employment, and through employment on training. To check this possibility, we estimated the same regression model for the restricted sample of women in paid work. Results are shown in Table A1 in the Online Appendix. They demonstrate that this is not the case, particularly for women in the middle education group. For them, the F-statistics that we estimate is still strong (at 8.8). The effect of the simulated income variation is weakest for college graduates (F-statistic of 6.4) and in between the two for the group of women with less than high school qualifications (7.3). Given the strength of the policy variation in affecting the training rates of high school graduates and the fact that training is very prevalent among women this group as well, our focus will be on this group for the remainder of the paper.

## 4.6 The model

We study training choices and their value for earnings through the lens of a life-cycle model of labour supply and human capital (HC) formation. Our model builds on the life-cycle model of female education, labour supply and experience capital of Blundell, Costa Dias, Meghir, et al. (2016) by integrating on-the-job training in the process of HC formation and by adding a layer of heterogeneity that shapes the returns to HC investments. Here, however, we focus on a homogeneous education group.

### 4.6.1 Overview of the model and its key components

We consider the adult life of women, after completing education. Following our discussion of training incidence and training incentives, we focus on the key group of women who completed high school but did not complete a degree. Our model considers labour supply, training, consumption and savings choices of women from the moment they enter the working life at the age of 19. Adult life is split in two periods, the working period and the post-retirement period. Retirement is assumed to happen deterministically at the age of 60. Once retired, women stop working and live out of the savings they accumulated during working life.

All women initiate their adult life as singles with no children. They are characterised by various dimensions of *ex-ante* permanent heterogeneity, some observed and others not. The observed heterogeneity is captured by two indices of family background, describing the socio-economic conditions of their parental home when they were aged 16. These affect their productivity in and preferences for work. The other component of observed heterogeneity is the cohort to which women belong. Different cohorts are affected by different sequences of work incentives shaped by the policy reforms, which may affect their working and training choices.

*Ex-ante* unobserved heterogeneity is two-dimensional. It includes one ability component, which directly affects wages, and one preference component, which drives the utility costs of working hours and training. We assumed that these two dimensions of heterogeneity are perfectly correlated. The structure of the unobserved heterogeneity terms is clearly specified below, when we set out preferences and wages.

During their working life, women decide in each period whether to work and for how many hours, whether to invest in training if they are working, and how much to consume today and save for the future. Labour supply is modelled in three hour-points, corresponding to not working, working part-time and full-time. Training is fixed at 2 hours per week, the median value of the distribution of training conditional on it exceeding 40 hours over the previous year, or 1 full-time working week worth of training.

Working has present and future returns, in the form of earnings and experience capital respectively. Earnings are proportional to the number of working hours net of time in training, with an hourly wage rate that depends on the stock of human capital, the woman's ability type and a persistent productivity shock. Human capital is represented by a single index, and is endogenous in our model. It accumulates over the life cycle through working experience and training episodes; it depreciates during out-of-work periods, formalising the idea that career interruptions carry long-term consequences for earnings capacity.

In a competitive labour market framework with general training, workers bear the full cost of training and capture its entire return. However, firm-specific training and labour market frictions may change this result, instead creating the grounds for firms and workers to share the costs and returns from training (Acemoglu and Pischke 1999; Lentz and Roys 2015). In our model, we do not explicitly consider the role of firms and the labour market in determining how the cost and return to the investment is shared between workers. We assume that training carries a monetary cost equal to foregone earnings due to time taken away from work, and that it bears a return through HC that is reflected in future wages. However, we also allow training to carry a utility cost that may partly capture, in a reduced-form sense, incidence in the cost of training. It also captures other drivers of training, such as actual preferences, effort or congestion in training places. In the same vein, the contribution of training to the HC index also has a reduced-form interpretation. It represents a combination of its effect on the accumulation of skills and the sharing of their productive value with the firm. Training may also contribute to employer learning about productivity as in Altonji and Spletzer (1991). They conclude that training has a mixed role, both as enhancing human capital and compensating for the depreciation of skills acquired in formal education, but also as a mechanism that supports employer learning. However, the nature of the data does not allow them to estimate the relative importance of these factors.

In our framework we give a pure human capital interpretation to the effects of training. Investments in training are driven by various mechanisms that also determine their timing and return. Crucially, if wages are concave in HC then the monetary cost of training is lower and returns are larger when HC is low. This creates stronger incentives to invest at the start of the working life – when there is also a longer period ahead to bear returns, as in a Ben-Porath model – and when returning to work after long separations, to compensate for the depreciation of skills.

Other key components of the model also create rich interactions with employment and training choices and their returns. One is the stochastic process of family formation and dissolution, which maps out the formation and dissolution of couples and fertility episodes. The model reproduces the empirical marital sorting patterns and fertility histories of women whose highest education qualification is high school (Chiappori, Costa Dias, and Meghir 2018; Chiappori, Iyigun, and Weiss 2009).

Finally, choices of consumption are restricted by liquidity constraints. The family budget is determined not only by the earnings of the woman but also by those of a present partner, tax liabilities and public transfers. In particular, the model embeds a detailed description of the personal taxes and benefits operating in the UK and how they change over the sample period. This is implemented using the micro-simulation tool FORTAX (Shaw 2011).

### 4.6.2 Female wages and human capital

We consider the problem of a woman aged $t$ and, for simplicity of notation, omit the individual index. If working, this woman draws a per-hour wage that depends on the human capital she accumulated so far ($\kappa$), indicators for whether the family background factors are above or below their median in the population ($x_1, x_2$), permanent ability type $\omega$, and an idiosyncratic persistent productivity shock $\nu$. The latter follows an AR(1) process with normal innovations $\zeta$ and initial value drawn from a normal distribution. Formally, the wage equation is

$$\begin{aligned} \ln w_t &= b_0 + b_1 x_1 + b_2 x_2 + (\gamma_0 + \gamma_1 x_1 + \gamma_2 x_2) \ln(\kappa_t + 1) + \omega + \upsilon_t \\ \text{where } \upsilon_t &= \rho \upsilon_{t-1} + \zeta_t \end{aligned} \tag{4.2}$$

We allow for classical measurement error in wages by defining observed wages $w^m$ as follows

$$\ln w_t^m = \ln w_t + \xi_t \quad \text{where} \quad \xi_t \sim \text{iid.}$$

Gross pay $y$ depends on workings hours $h$. Women can choose to work either 0 hours, 18 hours or 38 hours, representing out-of-work, part-time and full-time hours respectively. Total working time also depends on whether the woman takes time to train as follows

$$y_t = w_t (h_t - d_t \bar{h}_d) \tag{4.3}$$

where $d$ is an indicator for training and $\bar{h}_d$ is training time, which is exogenously set to 2 hours per week.

Human capital $\kappa$ is accumulated in work, at a rate that depends on working hours and training status, and depreciates at a constant rate $\delta$ per period. The human capital process is

$$\begin{aligned} \kappa_{t+1} &= \kappa_t (1-\delta) + g_1(h_t) + g_2(h_t) k_t + \tau_1 d_t + \tau_2 d_t k_t \\ \kappa_{\underline{t}} &= 0 \end{aligned} \tag{4.4}$$

$g_1$ and $g_2$ define how human capital accumulates with work. We allow for the human capital gains from work to depend on the number of working hours and to vary linearly with human capital accumulated so far. Both $g_1$ and $g_2$ are set to 0 if the woman is not working, and $g_1$ is also set to 1 if she works full-time; other values are estimated. $\tau_1$ and $\tau_2$ measure the human capital return to training, which we also allow to vary linearly with the stock of human capital. The woman starts her working life at time $\underline{t}$ with an initial stock of human capital equal to zero.

Our model of wages and human capital formation implies that training is both cheaper and draws larger returns (if, as expected, $\gamma_0+\gamma_1 x_1+\gamma_2 x_2<1$ and $\tau_2 \leq 0$) when human capital is low. This reinforces the incentive to invest young in order to bear the returns for longer. It also makes training investments more valuable after the long career interruptions common among mothers of young children, if these interruptions carry a significant loss of skills that would be implied by a large depreciation rate $\delta$.

The wage equation also exhibits complementarity between human capital and ability, implying that high ability workers have more to gain from training activities that enhance human capital. But since high ability workers also pay a higher cost in terms of foregone earnings, the overall effect of ability on training take-up is ambiguous.

### 4.6.3 The employment and earnings of the spouse

Let $m_t = 0,1$ be an indicator for the presence of a partner at time $t$. We denote his characteristics and outcomes by adding a 'tilde' to his variables. Although his labour supply choices and human capital process are not endogenously modelled, we adopt a stochastic specification that captures the main features of the richer female model.

The spouse at time $t$ is characterised by his education $\tilde{s}_t$ and his productivity level $\tilde{\upsilon}$. The distribution of his education reproduces that observed empirically among spouses of high-school graduated women. To limit the size of the state space, his age is assumed to equal that of the woman, $t$. If working, his wage rate is

$$\begin{aligned}
\ln \tilde{w}_t &= \tilde{b}_{\tilde{s}} + \tilde{\gamma}_{\tilde{s}} \ln (t-18) + \widetilde{\upsilon}_t & (4.5)\\
\text{where } \tilde{\upsilon}_t &= \tilde{\rho}_{\tilde{s}} \tilde{\upsilon}_{t-1} + \tilde{\zeta}_t. & (4.6)
\end{aligned}$$

$\widetilde{\upsilon}$ is the productivity shock, initially drawn from a $\tilde{s}_t$-specific normal distribution when the couple is formed and later modelled as a $\tilde{s}$-specific auto-regressive process with normal iid innovations $\tilde{\zeta}$. As for women, we interpret transitory wage shocks as measurement error and specify the observed wages of the spouse as

$$\ln \tilde{w}_t^m = \tilde{w}_t^m + \tilde{\xi}_t \quad \text{where } \tilde{\xi} \sim \text{iid}.$$

In line with the empirical evidence, we consider only two labour supply points for men in couples: they are either not working, in which case their working hours $\tilde{h}$ are set to zero, or working full-time hours, with $\tilde{h} = 40$. Their employment process is

$$\begin{aligned} \text{In new couples:} \quad & \text{Prob}\left[\tilde{h}_t = 40 \mid t, \tilde{s}_t, m_{t-1} = 0\right] = \psi_0\left(t, \tilde{s}_t\right) && (4.7) \\ \text{In existing couples:} \quad & \text{Prob}\left[\tilde{h}_t = 40 \mid t, \tilde{s}_t, \tilde{h}_{t-1}, m_{t-1} = 1\right] = \psi_1\left(t, \tilde{s}_t, \tilde{h}_{t-1}\right) && (4.8) \end{aligned}$$

### 4.6.4 The budget constraint

Family resources include both the earnings of the woman, those of a present partner and net public transfers. Let $a_t$ represent the stock of assets that the family brings into period $t$. Each period choices are limited by a liquidity constraint ruling-out borrowing. The budget constraint is formalised in terms of the evolution of assets:

$$\begin{aligned} a_{t+1} &= (1+r)a_t + y_t + m_t\tilde{h}_t\tilde{w}_t - T\left(w_t, h_t, X_t\right) \\ a_{t+1} &\geq 0 \quad \text{and} \quad a_{\underline{t}} = 0 \quad \text{and} \quad a_{\bar{t}+1} = 0 \end{aligned} \tag{4.9}$$

In the above expression, $r$ is the risk-free interest rate, $\underline{t}$ is the start of working life, and $\bar{t}$ is the last period of life, set at 10 years after the retirement age of 60. We assume that women enter their working life with no assets, which is consistent with empirical evidence, and that any remaining assets have no value after $\bar{t}$.

$T$ is the tax and benefit function. It depends on the wage rate of the woman, her working hours (because the UK tax credits have an hours rule) and on all other state variables characterising the demographic and financial circumstances of the family, summarised in $X$. In particular, $X$ includes presence of children and age of the youngest child, marital status, whether present partner is working and his wage rate. We use the detailed microsimulation tool, Fortax, to calculate $T$.[9]

[9] Fortax describes most of the UK personal taxes and benefits and how they changed over the period we model, including income tax, social security contributions, and the main subsidies for working-age families, namely income support, job-seekers allowance, tax credits, housing benefit, council tax benefit, child benefit.

### 4.6.5 The dynamics of family formation

We adopt a flexible Markov model to capture the dynamics of fertility, marriage and divorce. To preserve computational tractability while representing the key drivers of female labour supply, we only keep track of the age of the youngest child but allow for multiple fertility events. Let $t^k$ denote the age of the youngest child in the family. Childbirth is represented by re-setting $t^k$ to zero and happens at a rate that depends on the woman's age, whether she has other children (denoted by the indicator $n^k$) and the age of the youngest, and whether she is married $(m)$

$$\text{Prob}\left[t^k = 0 \mid t, n^k_{t-1}, t^k_{t-1}, m_{t-1}\right] \tag{4.10}$$

It is assumed that a child lives with her parents until turning 19, at which point she deterministically leaves her parents' home.

The probability that a woman marries or remains married to a man of education $\tilde{s}$ depends on her past marital circumstances, her age, whether she has children, and the education of her spouse if he is present in the previous period,

$$\text{if single at } t-1: \qquad \text{Prob}\left[m_t = 1, \tilde{s} \mid t, m_{t-1} = 0, n^k_{t-1}\right] \tag{4.11}$$

$$\text{if married to man } \tilde{s} \text{ at } t-1: \qquad \text{Prob}\left[m_t = 1, \tilde{s} \mid t, m_{t-1} = 1, \tilde{s}, n^k_{t-1}\right] \tag{4.12}$$

Otherwise she will be single at time $t$.

### 4.6.6 Utility and value functions

In each period $t$ of her working life, the woman decides about total family consumption $(c)$, savings $(a)$, her own labour supply and training investments to maximise her lifetime utility. Working life starts at $\underline{t} = 19$ for our sample of High School graduates. It ends deterministically at 60 when the woman retires, after which family savings fund an additional 10 years of consumption.

We assume intertemporal separability in preferences. The per-period utility of her choices depends on her preference type, $\theta$, and a subset of the state variables $X_t$ that characterise her circumstances at age $t$:

$$u\left(c_t, h_t, d_t;\, \theta, X_t\right) = \frac{\left(c_t/n_t\right)^{\mu}}{\mu} \exp\left\{U\left(h_t, d_t, \theta, X_t\right)\right\}. \tag{4.13}$$

In the above expression, $n$ is the equivalence scale, factoring in family size,[10] and $\mu$ is the parameter determining both the degree of risk aversion and the elasticity of intertemporal substitution.

The function $U$ reflects how the value of additional consumption varies with working hours and training status by family composition for women of different $\theta$ types. We decompose it into two additive terms, one relating only to working hours, $U_h$, and the other driving the utility cost of training, $U_T$:

$$U\left(h, d, \theta, X\right) \;=\; U_h\left(\theta, X_1\right) + d \times U_T\left(h, \theta, X_2\right) \tag{4.14}$$

with $(U_h, U_T)$ defined as follows

$$U_h\left(X_1\right) \;=\; \begin{cases} 0 & \text{for } h = 0 \\ l_h(\theta) + \alpha_h X_1 & \text{for } h = 18, 38 \end{cases} \tag{4.15}$$

$$U_T\left(h, \theta, X_2\right) \;=\; l_T(\theta) + \alpha_T X_2 + \alpha_{T,h}. \tag{4.16}$$

In the above, we denote by $X_1$ and $X_2$ the two relevant subsets of state variables (not mutually exclusive) that directly affect preferences for working hours and training, respectively, and by $(\alpha_h, \alpha_T)$ their associated parameters. $X_1$ includes a full set of interactions between marital status and whether she is a mother, indicators for age of youngest child in bands (0-2, 3-5, 6-10), and the background factors $(x_1, x_2)$. $X_2$ includes indicators for whether or not she is a mother and age of youngest child in bands. Equation (4.16) also includes an interaction term between working hours and training status $(\alpha_{T,h})$. Heterogeneity in preferences $\theta$ takes two values, for low and high preferences for work, and is assumed perfectly correlated with heterogeneity in ability $\omega$. The terms $(l_h(\theta), l_T(\theta))$ measure the importance of unobserved preferences for work and training in driving choices.

The intertemporal problem of the woman can now be formalised. Let $\beta$ be the discount

[10] $n = 1$ for singles, 1.6 for couples 1.4 for mother with child and 2 for a couple with children.

factor. Her problem in period $t$ of her working life is

$$V_t(\omega, \theta, X_t) = \max_{(a_\tau, c_\tau, h_\tau, d_\tau)_{\tau=t,\ldots,\bar{t}}} E_t \left[ \sum_{\tau=t}^{\bar{t}} \beta^{\tau-t} u(c_\tau, h_\tau, d_\tau;\, \omega, \theta, X_\tau) + \beta^{\bar{t}-t} b(\kappa_{\bar{t}}) \middle| \omega, \theta, X_t \right] \tag{4.17}$$

The term $b(\kappa_{\bar{t}})$ represents the value of human capital at retirement. It is meant to capture the fact that human capital will have some value post age 59, both because some women will remain active in work and because human capital is valuable outside work as well. This values is specified as follows,

$$b(\kappa_{\bar{t}}) \quad = \quad \phi_1 \frac{(\phi_2 + \kappa_{\bar{t}})^\mu}{\mu}$$

The maximisation problem in 4.17 is conditioned by the budget constraint (4.9), the female wage and human capital processes (4.2)-(4.4), the dynamics of employment and wages of a present partner (4.5)-(4.8) and the dynamics of family formation (4.10)-(4.12). The woman starts her working life as a single woman with no children.

## 4.7 Estimation

We estimate the subset of model parameters driving female wages, human capital formation and preferences for working hours and training using the method of simulated moments. The values for all other parameters are taken from Blundell, Costa Dias, Meghir, et al. (2016). These include the subset of parameters defining the pre-determined family dynamics, male employment and male wages. A description of their estimation procedure and the full set of estimates can be found in their Web Appendix B. Three other parameters are set at typical values in the literature: the parameter regulating the curvature of the utility function $\mu$ is set at $-0.56$, implying a risk aversion coefficient of 1.56; the risk-free interest rate $r$ is set at 0.015 and the discount factor $\beta$ at 0.98, together implying that agents are mildly impatient (Attanasio, Low, and Sánchez-Marcos 2008; Attanasio and Weber 1995; Blundell, Browning, and Meghir 1994).

Estimation relies on a set of 139 moments capturing various aspects of lifecycle behavior and wages.[11] We construct the simulated moments to reproduce their data counterparts, based

[11] The moments include full- and part-time employment and training rates by age, family demographics, socio-economic background, and interactions between calendar time and demographics; employment and

on the simulation of 5 lifetime profiles for each of the 1,443 high school educated women who are observed in BHPS with observed socio-economic background and life histories of employment. From the resulting 7,215 profiles we select a window that exactly matches the observation window of the corresponding woman in the survey data. This way, we exactly reproduce the time, age and socio-economic structure of the data.

Our estimation procedure uses the exogenous variation in the labour supply and training incentives from policy reforms. Using regression analysis, we showed in section 4.5 that such exogenous variation was important for high school graduates and may play an important role in driving the results for them (I. Andrews, Gentzkow, and Shapiro 2017).

Within the model we use the policy variation by considering four tax and benefit systems, namely the ones operating in April 1995, 1999, 2002 and 2004. The reforms are unannounced.

Our moments include pre- and post-2002 measures of employment, working hours and training that explicitly capture the variation induced by the reform. Responses to the reform are likely to vary by cohort, as they are differently exposed to the reform, and individual permanent characteristics. We exploit these interactions to identify the value of working and training for future wages, by explicitly modeling the differential exposure to the reforms of different cohorts and by allowing responses to depend on socio-economic background.

The estimates of the model parameters are the set of parameter values $\Theta$ that minimise the following expression

$$\sum_{\kappa=1,\ldots,K} \frac{\left(M^d_{\kappa,N} - M^s_{\kappa,S}\left(\Theta\right)\right)^2}{\operatorname{Var}\left(M^d_{\kappa,N}\right)} \tag{4.18}$$

where $K$ is the total number of moments used in estimation, $M^d_{\kappa,N}$ is the estimate of moment $\kappa$ from $N$ observations of observed data and $M^s_{\kappa,S}$ is the corresponding moment calculated on $S$ model simulations for parameter values $\Theta$.[12] We calculate asymptotic standard errors following Gourieroux, Monfort, and Renault (1993).

hours transition rates by family demographics and past wages; the mean, variance and percentiles of the wage distribution over the course of life and at entrance into working life; the correlation between wages and socio-economic background, years of work, working hours, training and past wages; the growth rate of wages by past working hours, training and socio-economic background.

[12] It is implicit in the maximisation criterion that we are not using the optimal asymptotic weighting matrix, following the suggestion of Altonji and Segal (1996). Instead, we use the diagonal matrix of inverse variances of the moments, which are bootstrapped using 1,000 replications.

## 4.8 Parameter estimates and implications for behaviour

### 4.8.1 Wages, human capital and the return to training

Table 4.4 shows estimates of the female wage process. Estimates in panel A of the Table are for the wage rates at the start of working life ($b_0$) and the return to human capital ($\gamma_0$). Socio-economic background has a relatively small (but statistically significant at conventional levels) effect on starting wages; in turn, the return to human capital does not vary significantly with socio-economic background. Our estimate of the return to human capital in wages ($\gamma$) is, as expected, smaller than 1. Combined with the log linear specification of the wage equation, this implies that the return to one additional unit of human capital decrease with the stock already accumulated. Finally, unobserved heterogeneity in the wage rates ($\omega$) is important (see estimates in Panel B of the Table). Our estimates indicate that being high ability raises the wage rate by 24 log points compared to the average.

Table 4.4: Wage parameters

| | Parameter Value | St. Error |
|---|---|---|
| *Panel A: Wage coefficients* | | |
| Intercept, $\exp(b_0)$ | 6.86 | (0.065) |
| increment: high factor 1, $\exp(b_1)$ | 0.64 | (0.093) |
| increment: high factor 2, $\exp(b_2)$ | -0.31 | (0.028) |
| Return to human capital, $\gamma_0$ | 0.27 | (0.004) |
| increment: high factor 1, $\gamma_1$ | -0.04 | (0.005) |
| increment: high factor 2, $\gamma_2$ | 0.03 | (0.004) |
| *Panel B: Unobserved heterogeneity in ability, $\omega$* | | |
| $\omega$ type I: wage effect | 0.24 | (0.012) |
| $\omega$ type I: probability | 0.79 | (0.002) |
| *Panel C: Distribution of persistent productivity shock $\nu$* | | |
| Persistence of productivity, $\rho$ | 0.95 | (0.002) |
| St. dev. of productivity innovation, $\zeta_t$ | 0.12 | (0.003) |
| St. dev. of initial productivity, $\upsilon_0$ | 0.27 | (0.007) |

Uncertainty in wages is characterised by the persistent unobserved productivity process $\nu$. Our estimates in Panel C suggest that though this process is highly persistent, with autocorrelation coefficients of around 0.95, there is a high level of wage uncertainty. There is also substantial heterogeneity in initial wages.

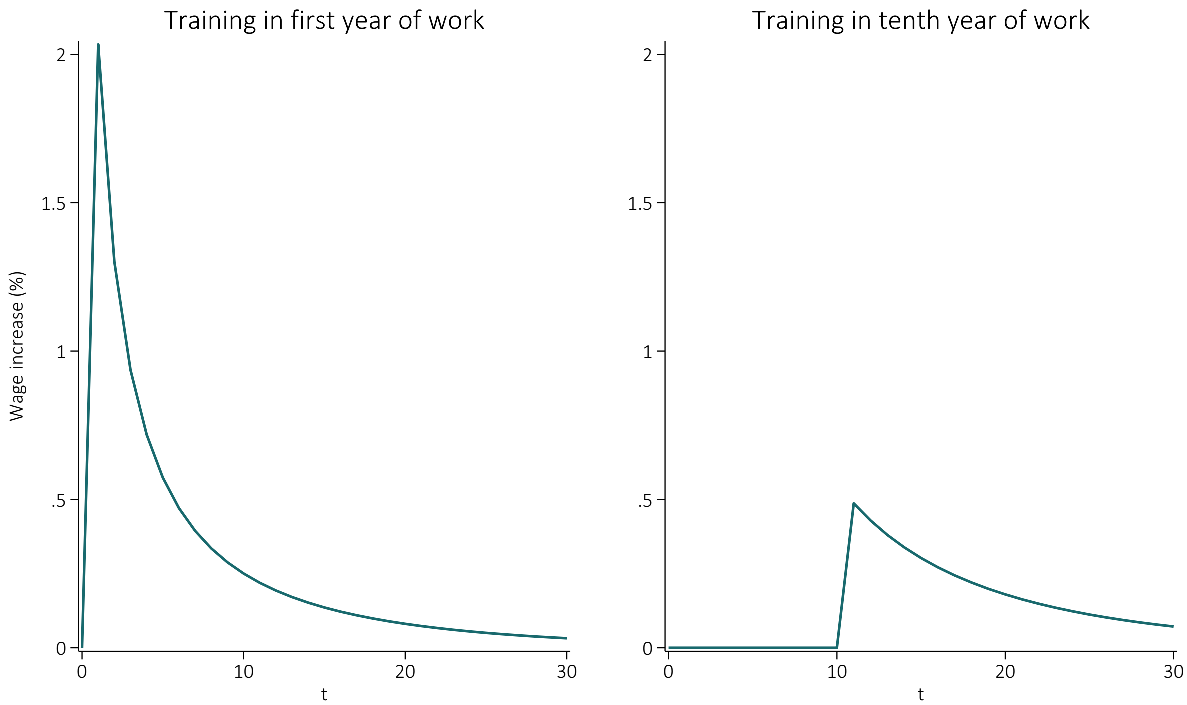
Training affects wages through its impact on human capital. Our estimates show the incremental effect of training over work experience for the duration of training; i.e. they show how much more human capital workers gain if they choose to take time away from working and use it to train instead. The top row of Table 4.5 shows the estimate of this effect for women at the start of working life, when they have not yet accumulated human capital from work ($\tau_1$). Our estimate suggests that, at that stage of the working life, training increases human capital by 16 percent of the return to one year of full-time work (which is normalised to 1). We allow for more flexibility in how training affects human capital, and hence wages, by adding an interaction term with the stock of human capital ($\tau_2$ in the second row of the table). Our estimates, however, suggest that this term is not needed.

Table 4.5: Parameters in the human capital accumulation process

| | Parameter Value | St. Error |
|---|---|---|
| training, $\tau_1$ | 0.16 | (0.008) |
| training $\times$ human capital, $\tau_2$ | 0.00 | (0.004) |
| part-time, $g_1(18)$ | 0.13 | (0.009) |
| part-time $\times$ human capital, $g_2(18)$ | 0.00 | (0.003) |
| full-time $\times$ human capital, $g_2(38)$ | -0.02 | (0.005) |
| depreciation rate ($\delta$) | 0.08 | (0.002) |

The magnitude of the effect of training is slightly larger than the human capital return from working part-time hours, which are estimated to be 13 percent of the full-time return at the start of working life ($g_1(18)$ in second row of the Table). We also allow for an interaction term with the stock of human capital ($g_2(18)$), and again find no evidence of the need to allow for more flexibility in how part-time hours affect human capital and wages. The only interaction of the stock of human capital that is statistically significant at conventional levels is that with full-time hours ($g_2(38)$), but even there the effect is small. Our estimate shows that one additional unit of human capital reduces the human capital return to full time hours by 2%. Since human capital never increases beyond 12 in simulations, at the maximum this parameter is responsible for a 24% drop in the human capital return to one year of full-time work.

Figure 4.7: Wage return to one episode of training while working full-time, by education



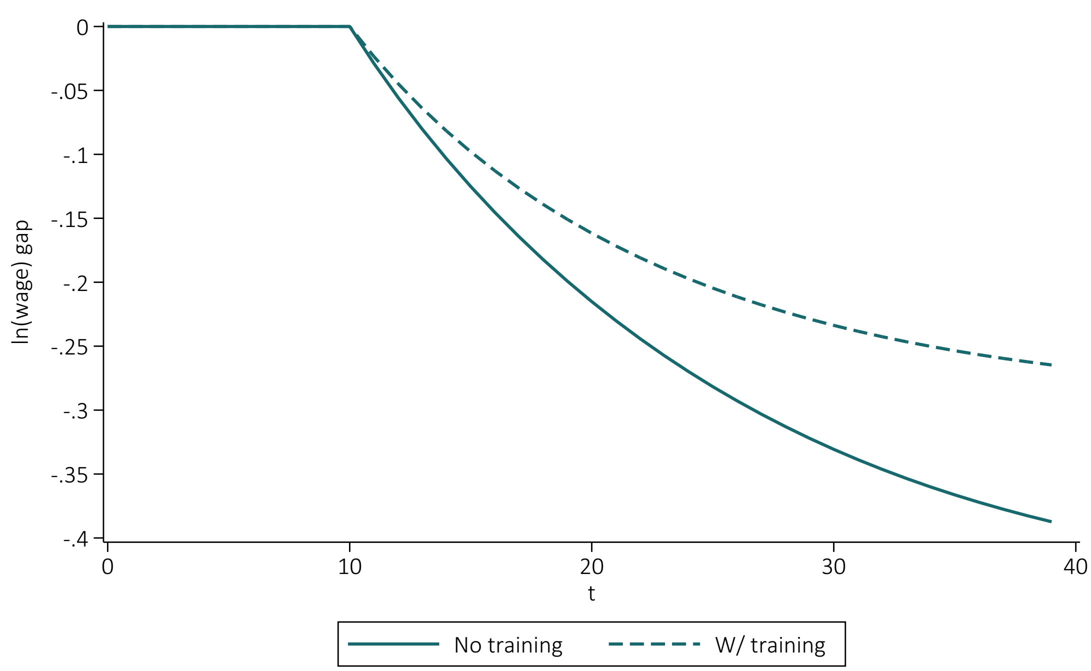

Notes: Percentage change in wage rates due to single episode of training in years 1 (LHS panel) and 10 (RHS panel) of full-time work. Agent is assumed to have no human capital at t = 0 except for that acquired through formal education and is working full-time over the entire period.

The size of the impact of training on wages depends on the interactions between its impact on human capital (determined by $\tau_1$) and its wage return (determined by a combination of $(\gamma_0, \gamma_1, \gamma_2)$ for different groups), the depreciation rate ($\delta$), and the stock of human capital at the time of training. Figure 4.7 illustrates the overall short- and long-term wage effects of one episode of training taking place at different stages of the working life. The plot on the left shows the impulse response to one training episode in year 1 of working life for women in full-time hours; the plot of the right shows the equivalent figure if training happens after 10 years of full-time work.

There is a modest but not insignificant initial effect on wage rates that, however, declines quickly as the additional human capital depreciates over time. The initial effect is much more pronounced if training is taken earlier in the working life, prior to the building up of human capital with working experience and consistent with decreasing marginal returns to investments in human capital. For instance, training increases the wage rate by 1.5% if taken in the first period of work, but only by 0.4% if taken after 10 years of working full-time. The falling returns to training with accumulated human capital is an important determinant of the timing of training in our model.

Our estimates of the wage impact of training can be compared with estimates of the impact of one additional year of education found in the broader literature once adjusted for the relatively small number of hours spent in training. Assuming that school requires thirty hours of study per week and takes place over forty weeks, a year of schooling requires 1,200 hours of time investment. This is approximately 12 times longer than the 100 hours corresponding to a training episode within our model. Card (1999) surveys the vast literature on returns to education and finds estimates implying increases in wages of between 5% and 15% associated with an additional year of high school, or approximately 0.4% to 1.3% per 100 hours invested. Blundell, Dearden, and Sianesi (2005) estimate a wage return of 24% for the two years of education differentiating High School graduates from those who leaving school at 16 (with less than high school qualifications) in the UK context, or approximately 1% per 100 hours invested. Our estimates of the initial return from training at the start of working life fall on very similar values.

Figure 4.8: Training and the wage penalty from working part-time hours, by education




Notes: Solid lines represent the wage penalty, in log points, from moving to continuous part-time work after 10 years of continuous full-time work. The dotted lines factor in continuous training starting in year 10, together with part-time working hours.

In Figure 4.8 we document the extent by which training can offset the part time penalty in wages. The diagram compares the loss in wages that results from a shift from full-time work to (a) part-time work (solid line) or (b) part-time work plus training (dashed line). It represents how the impact of training compares with that of part-time hours. The solid lines in the figure show that part-time work is associated with a large wage penalty. The dashed lines show that taking training together with part-time hours offsets almost one third of the part-time penalty.

### 4.8.2 Utility parameters and the cost of training

Tables 4.6 and 4.7 show estimates of the parameters driving the utility cost of work and training as defined by the index functions $U_h$ and $U_T$ in equations 4.15 and 4.16. In both Tables, a positive parameter reflects higher costs of working or training.

Table 4.6: Parameters determining utility cost of working

| | Parameter Value (1) | St. Error (2) | Parameter Value (3) | St. Error (4) |
|---|---|---|---|---|
| | *Utility Parameters in $U_h$* | | | |
| | Full-Time Employment ($\alpha_{38}$) | | Part-Time Employment (increment: $\alpha_{18} - \alpha_{38}$) | |
| Singles, no children | 0.56 | (0.006) | -0.37 | (0.004) |
| Single mothers | 0.47 | (0.011) | -0.22 | (0.009) |
| Married, no children | 0.33 | (0.014) | -0.23 | (0.015) |
| Married mothers | 0.34 | (0.013) | -0.24 | (0.012) |
| Child aged 0-2 | 0.16 | (0.009) | -0.07 | (0.008) |
| Child aged 3-5 | 0.11 | (0.010) | -0.05 | (0.009) |
| Child aged 6-10 | 0.06 | (0.010) | -0.04 | (0.006) |
| Spouse working | -0.07 | (0.013) | 0.08 | (0.012) |
| High background factor 1 | 0.02 | (0.008) | 0.00 | (0.005) |
| High background factor 2 | 0.03 | (0.008) | -0.02 | (0.005) |
| $l_h(\theta)$ type I | -0.38 | (0.178) | 0.00 | (0.005) |

In order to rationalise the observed employment rates at the given monetary incentives to work, the model requires working to carry a utility cost for all groups (see estimates in columns 1 and 2 of Table 4.6). The costs are lower for married women than for single women, partly offsetting differences in incentives to work between the two groups due to spouse's income and benefit entitlement. Moreover, a working spouse brings down the utility cost

Table 4.7: Parameters determining utility cost and benefits of training and the terminal value of human capital

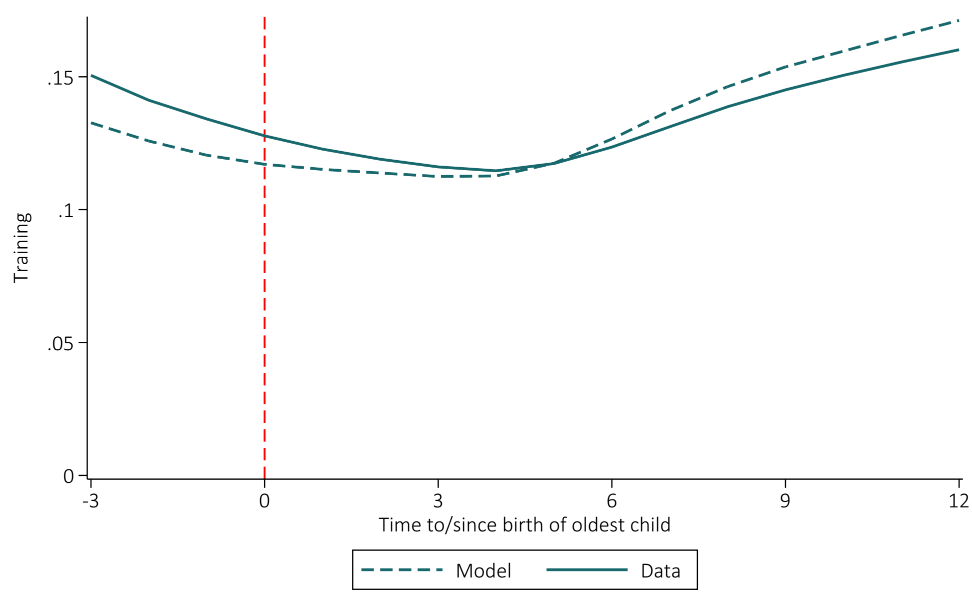
| | | Parameter Value | St. Error |
|---|---|---|---|
| | *Utility Parameters in $U_T$, $(\alpha_T, \alpha_{Th})$* | | |
| (1) | Single, no children | 0.002 | (0.010) |
| (2) | Single mothers | 0.007 | (0.008) |
| (3) | Married, no children | -0.002 | (0.015) |
| (4) | Married mothers | -0.002 | (0.016) |
| (5) | Child aged 0 to 2 | 0.010 | (0.025) |
| (6) | Child aged 3 to 5 | 0.004 | (0.014) |
| (7) | Child aged 6 to 10 | 0.003 | (0.008) |
| (8) | Spouse working | 0.009 | (0.014) |
| (8) | High background factor 1 | 0.004 | (0.007) |
| (8) | High background factor 2 | 0.002 | (0.005) |
| (9) | Part-time interaction | 0.016 | (0.006) |
| (10) | $l_T(\theta)$ type I | -0.028 | (0.003) |
| | *Terminal Value of Human Capital* | | |
| (11) | Scale parameter, $\phi_1$ | 0.05 | (0.009) |
| (12) | Curvature parameter, $\phi_2$ | 0.21 | (0.137) |

of working, a result in line with past research showing complementarity in spouses' leisure (Blundell, Pistaferri, and Saporta-Eksten 2016). Mothers of young children, particularly of pre-school age, also face higher costs of working. Columns 3 and 4 of the Table report estimates for the incremental effects of working part-time hours, showing that part-time is less onerous in utility terms than full-time hours.

Estimates for the parameters governing the utility cost of training are shown in Table 4.7. We have fixed the monetary cost of training to equal the foregone wage for 2 hours of training per week, or 104 hours per calendar year, which corresponds in the data to the median level of training among trainees undergoing more than one week of training over the year. The utility cost of training is identified from the discrepancy between the predicted take up of training (if costs were zero) and the actual take up.

Most parameters in the utility of training are small and mostly not statistically significant at conventional levels: the utility cost of training does not seem to depend on the demographic structure of the household or even on the family background factors. Perhaps this is not surprising since most of the cost associated to the household structure relates to the decision to work or not and once that has been paid it is no longer relevant for the training decision itself. However, the interaction with part-time hours (row 9) shows that training is more costly when women are doing short working hours; and the unobserved heterogeneity term (row 10) shows that the group with higher preferences for work also has a positive preferences for training (which mirrors a higher training cost for those with lower preferences for work). Thus, given our estimated returns to training our model rationalizes observed training cost as a preference for training among higher ability women (which constitute 80% of the population according to estimates in Panel B of Table 4.4). A model that admits search frictions or other imperfections could provide a structural interpretation of this since in that case the firm and the worker share the costs of training.

Figure 4.9: Model versus data – Training incidence among working mothers, by time since/to birth of oldest child and maternal education




Our model implicitly points to two additional mechanisms explaining the life-cycle patterns of training. First, families with children have higher needs and may be more likely to face liquidity constraints. In those circumstances, the foregone earnings associated with training may be an especially high cost to pay that could drive training rates down during that period of life. And second, the expected return to training may be negatively affected by motherhood as higher career intermittency limits women's ability to reap its full return before depreciation eventually washes out the human capital gains from training.

Figure 4.10: Model versus data – Training incidence over the life-cycle among working women, by maternal age and education

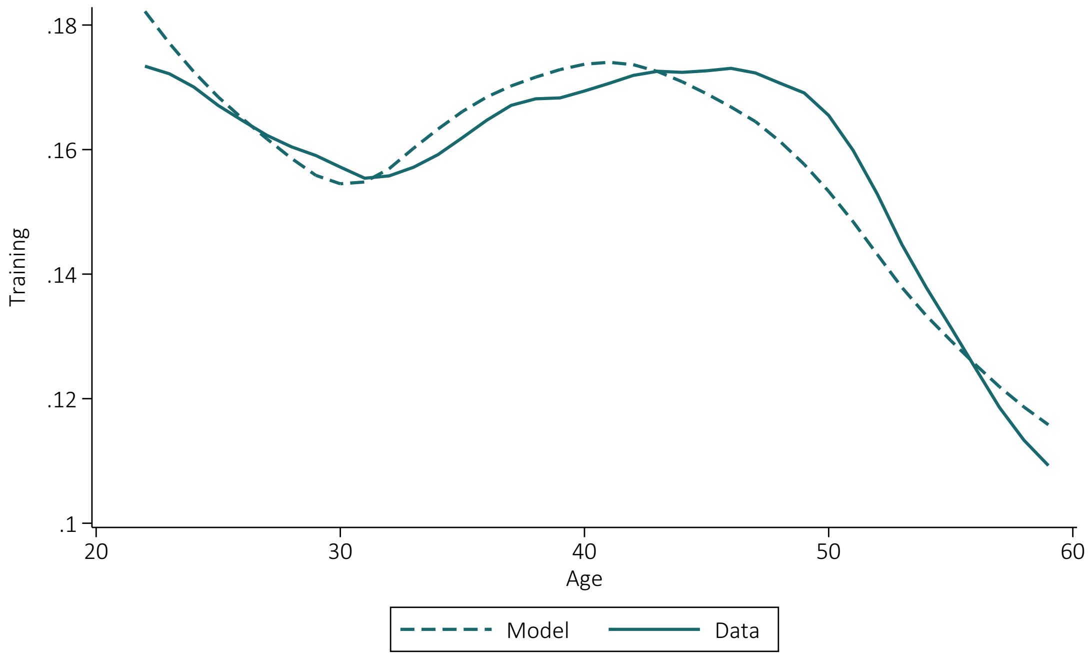
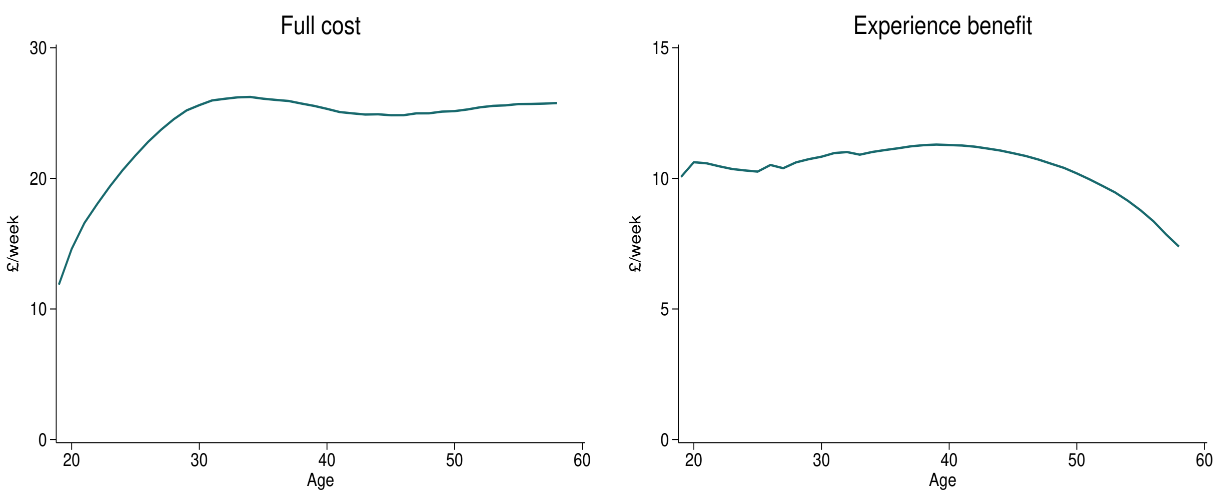
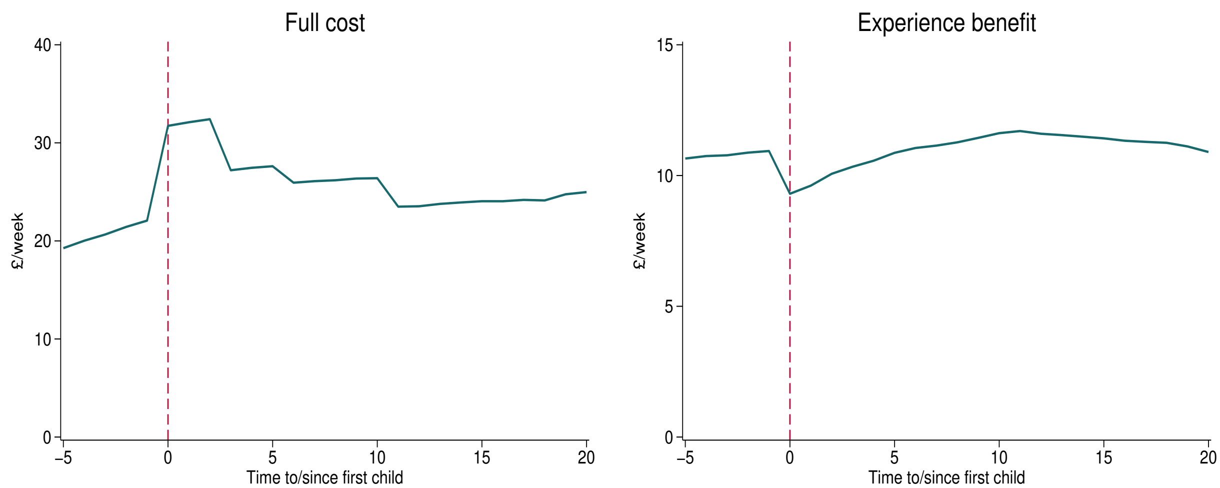



Figure 4.11: Monetised total cost of and experience return to training across whole population, by age and education




(a) Monetized cost of training (b) Consumption value of extra human capital

Figure 4.12: Monetised cost of and experience return to training across whole population, by time to/since first birth and education




(a) Monetised cost of training (b) Consumption value of extra human capital

Notes: Left-hand panels show average monetary compensation required to equalise period utility between (1) working full-time and not training and (2) working full-time and training. Right-hand panel shows the average monetary deduction required for an individual to be indifferent to receiving additional human capital equivalent to one unit of training.

Figure 4.11 plots age profiles for the average total cost of training on the left, including both the monetary cost associated with lost labour time and the monetized direct utility cost. We compare this to the consumption value of the additional human capital acquired through one episode of training on the right. In line with the observed training rates, average cost exceeds average return by a factor of 2 for most age groups. Figure 4.12 plots similar figures but by time to/from the birth of the first child. The life-cycle variation is strongly associated with the dynamics of family demographics through employment behaviour rather than through the utility cost of training. The returns to training also change around childbirth but by a much more modest amount, and then slowly recover as the child grows up.

Finally, the last two rows in table 4.7 show the parameters associated with the terminal value of human capital at the time of retirement. The scale parameter $\phi_1$ is positive, which implies that human capital is valuable in retirement.

Table 4.8: Model simulations – Employment and training responses to changes in wages and the monetary cost of training

| | level (%) | 5% permanent decrease in net earnings | 5% permanent decrease in training cost |
|---|---|---|---|
| | | *(a) Employment* | |
| All women | 85.7 | -2.0 | 0.0 |
| By family demographics | | | |
| Singles, no kids | 93.3 | -1.4 | 0.0 |
| Single mothers | 69.7 | -4.2 | 0.0 |
| Couples, no kids | 94.3 | -0.8 | -0.1 |
| Mothers in couples | 79.0 | -2.6 | 0.0 |
| | | *(b) Training conditional on employment* | |
| All women | 16.7 | 0.3 | 2.1 |
| By family demographics | | | |
| Singles, no kids | 16.3 | 0.0 | 3.0 |
| Single mothers | 10.3 | 0.7 | 2.9 |
| Couples, no kids | 18.4 | 0.3 | 1.6 |
| Mothers in couples | 16.6 | 0.2 | 1.9 |

### 4.8.3 Responses of employment and training to changes in prices

We use the model to quantify responses to changes in the monetary incentives to work and train. Table 4.8 shows responses in employment rates (Panel A) and training rates among employed women (Panel B) to changes in the wage rates (column 2) and in the earnings foregone while training (column 3). Column 1 provides a sense of scale by displaying the simulated levels of employment and training by family demographics. All simulations are run under the 2002 tax system.

Column 2 reports average immediate response to an unanticipated and permanent 5% decline in the post tax wage rate starting at each age in the 23 to 50 interval. Overall, this change leads to a 2% decline in employment on a base of 85.7% displaying the dominance of the wealth effect. The response is larger for mothers, particularly single mothers, than it is for women without children, reflecting their larger labour supply elasticities. While training responses to changes in the wage rates are smaller than those of employment, they are nevertheless important given current training rates. A permanent drop in wages reduces future returns to training, but that is offset by the negative impact it has on the current cost of training. We find that the latter dominates leading to a small overall increase in training rates particularly for single mothers. Finally, column 3 in the Table shows that the training responses to a drop in the cost of training are large, particularly for single women. In turn, employment does not respond to changes in training incentives.

The parameters in Table 4.8 are key to inform policy as they reflect the potential responses in employment and training to reforms changing the monetary incentives to do so. They are consistent with the observed effects of the WFTC reform on employment and training. These are described in the set of moments we used to identify the model, and displayed at the bottom 8 rows of tables A2, A3 and A4. We can see that the model closely fits the employment and training rates before and after the reform for all family types.

## 4.9 Counterfactual simulations and discussion

### 4.9.1 Subsidized training for mothers

We now investigate the long-term impacts of subsidizing training for mothers of young children, who may have especially loose links to the labor market. The policy could impact the labor market outcomes of these mothers in two ways. First, by increasing training rates among eligible mothers, it may help recover some of the losses in productive human capital associated with career interruptions once mothers return to work. Second, the subsidy may also reduce the duration of career breaks by indirectly promoting employment during the early stages of motherhood. The results from the previous sections suggest that mothers are especially sensitive to the cost of training and that training has modest but positive effects on wages, so the question is whether subsidizing training could help close the cost of child-rearing for mothers.

We compare outcomes under the 2002 tax and benefit system with three modified regimes that introduce training subsidies. In all three cases, mothers of children aged 7 or younger are entitled to subsidies of different levels of generosity if they decide to take up training.

Our simulations quantify the long-term effects of these policies for women living through the new regimes over their entire lives. All effects are calculated under revenue neutrality, with any costs being recovered through adjustments in the basic tax rate from the tax liabilities net of benefit entitlements of this group of women and their partners. The way one achieves revenue neutrality is relevant since, for example, changing the tax rate to fund subsidies has its own incentive effects.

Table 4.9 shows model predictions of the effects of subsidized training on training rates, employment, hours, wages, savings, income and welfare. The first column displays the effects of a £500 lump-sum subsidy for mothers of children aged 0 to 7 in training. The second column increases this to £1500, and in the final column the subsidy provides full compensation for the monetary cost of training, which includes foregone earnings. The subsidy policy is made revenue neutral with a change in the basic tax rate.

Under our assumption of standard training units of 2 hours per week, the £500 annual subsidy amounts to approximately £5 per hour. This is not a trivial subsidy, making up about 40% of the average hourly wage rate of eligible mothers. However, it is more modest than other work related subsidies such as Tax Credits because it only supports a limited amount of training.

The results in the first column show that training rates respond strongly to the subsidy during the eligibility period (panel (a)), as suggested by the responses to a drop in the monetary cost of training described in table 4.8. Moreover, the effect quickly fades to 0 in later periods when mothers lose eligibility (panels (c) and (e)).

The subsidy is timed to coincide with the fall in training we observe around the birth of first child. Figure 4.13 shows, as an example, the impact of the subsidy on the prevalence and timing of training. The fall in training at the time of childbirth, which is observed in the data and replicated by our baseline model, is completely offset by the subsidy. As a result, training rates decline gradually over the lifecycle, resembling the male training profiles discussed above (see Figure 4.5).

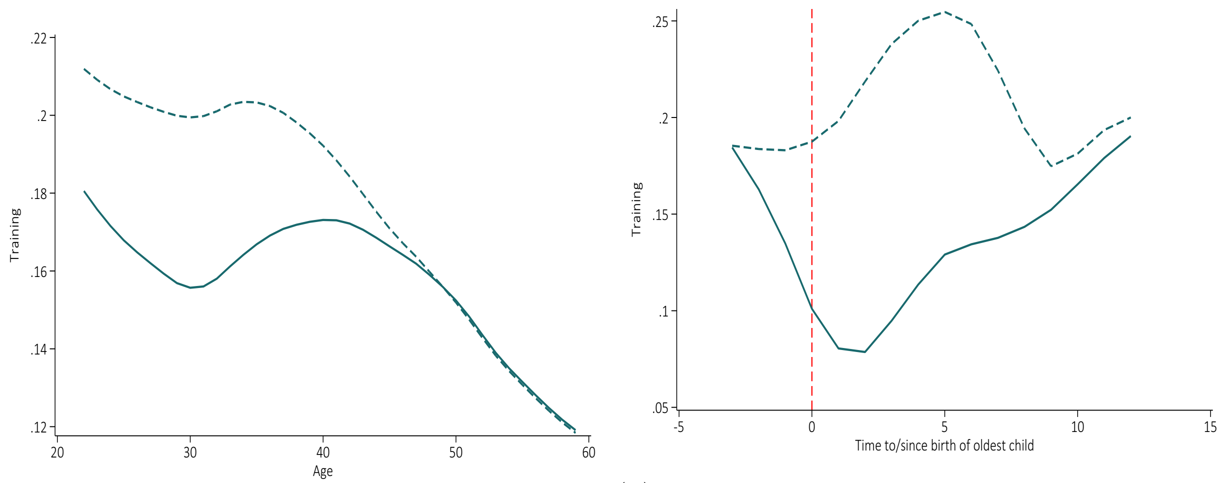
The least generous subsidy has a small impact on full time employment, increasing it by 0.64 percentage points during the first 7 years of the child, which corresponds to the period of entitlement (panel (a)). All of this extra time in paid work comes from those who were previously doing part-time work, resulting in a net effect on employment close to zero. The small net response in employment is aligned with predictions of how employment responds to changes in the cost of training, detailed in Table 4.8.

Table 4.9: Impact of training subsidies

| | Annual training subsidy | | |
|---|---|---|---|
| | £500 | £1500 | Full compensation |
| | (*a*) Child aged 0-7 | | |
| Training | 9.33 | 40.95 | 19.84 |
| Employment | 0.02 | 3.14 | 0.01 |
| Full-time | 0.64 | 1.37 | 1.35 |
| Part-time | -0.62 | 1.77 | -1.34 |
| | (*b*) Child aged 8 | | |
| Assets (%) | -0.12 | 0.51 | 0.69 |
| Wages (%) | 0.31 | 2.60 | 1.36 |
| | (*c*) Child aged 8-18 | | |
| Training | -0.19 | -0.38 | -0.31 |
| Employment | 0.10 | 0.52 | 0.15 |
| Full-time | 0.41 | -0.30 | 0.20 |
| Part-time | -0.31 | 0.82 | -0.05 |
| | (*d*) Child aged 19 | | |
| Assets (%) | -0.05 | 0.24 | 0.29 |
| Wages (%) | 0.14 | 0.72 | 0.44 |
| | (*e*) Child aged 19+ | | |
| Training | 0.00 | -0.03 | 0.03 |
| Employment | 0.21 | 0.26 | 0.23 |
| Full-time | 0.72 | 0.26 | 0.62 |
| Part-time | -0.51 | 0.00 | -0.39 |
| | (*f*) Lifetime outcomes | | |
| Disposable income (%) | 0.24 | 0.35 | 0.23 |
| Consumption equivalent (%) | 0.83 | 0.74 | 0.77 |
| | (g) Revenue neutrality adjustment | | |
| Basic income tax change | -0.02 | 0.5 | 0.15 |

Notes: Calculations based on model simulations. Column 1 shows the effects of a £500 yearly subsidy, while columns 2 shows similar calculations for an yearly subsidy of £1,500. Column 3 shows simulated figures for a subsidy that exactly covers foregone earnings of trainee mothers. In all cases, only mothers of children aged 0 to 7 are entitled to the subsidy if taking training. Age of the child in panels (a) to (e) refers to the youngest child in the family. The change in disposable income (panel (f)) is net of the tax adjustment. The consumption equivalent in the same panel is calculated at the start of working life to keep expected lifetime utility constant. All changes are in percentage points unless otherwise stated.

Figure 4.13: Training over lifecycle for High School educated under £500 subsidy




(a) Training rate by age

(b) Training rate by time since birth of oldest child

Panel (b) shows that the cumulative effect of the additional training and full time work on the wage rates of women at the end of the eligibility period is positive, with mothers benefiting from a 0.31% increase in wages. This demonstrates that the policy has a small but not negligible impact on the human capital of mothers at the end of the eligibility period. The subsidy also reduces savings by a modest 0.12% when the child reaches 8 years of age. This suggests that the additional human capital the woman accumulated over this period will make future work more likely and she will need to rely less on savings.

Indeed we find that the small impacts on human capital and assets at the end of the eligibility period drive similarly small dynamic effects. Panels (c) and (e) confirm that that the policy slightly increases employment after the eligibility period, and that all increase is on the full-time margin. These responses drive an increase in the lifetime disposable income of the families of these women by 0.24% (panel (d)) and a larger increase in equivalised consumption of 0.83%.[13] Since the counterfactual simulation is revenue neutral, all these responses are net of the tax adjustment. In the case of this less generous policy, we find that it pays for

[13]The value of the consumption compensation ($\iota$) is the solution to:

$$EV_0 \quad = \quad E\sum_t \beta^{t-\underline{t}} \frac{((1-\iota)\, c_{1t}/n_{1t})^{\mu}}{\mu} \exp\left\{U(h_{1t}, d_{1t}, \theta, \omega, X_{1a})\right\}$$

where the index 0/1 stands for the pre/post-reform solutions and the value function is evaluated at different stages in life for different rows. The equation can be solved for $\iota$, yielding: $\iota = 1 - \left(\frac{EV_0}{EV_1}\right)^{\frac{1}{\mu}}$.

itself. By bringing more women into full-time work for an extended period, the government raises in extra taxes the funds required to implement the subsidy (panel (g)).

Column 2 of the Table shows similar results for a more generous lump-sum subsidy of £1,500 per year, or about 120% of the pay of eligible mothers during training episodes. The additional generosity comes with a high price, requiring an increase of 0.5pp in the basic tax rate to balance the public budget. For comparison, Blundell, Costa Dias, Meghir, et al. (2016) calculations suggest that funding for the 2002 Tax Credit scheme in the UK adds 0.9pp to the basic tax rate. Despite its cost, which is fully borne by this population of women and their partners, our simulations show that this policy is welfare increasing and drives up disposable income by more than the less generous policy. These effects result from the strong impact that the policy has on the training rates of eligible mothers, which increase by 41 percentage points, and their employment rates, which also increase by 3 percentage points. The combined effect of these responses result in higher wages at the end of the eligibility period, by 2.6% (panel (b)) that drive later employment gains and persistent increases in wages (by 0.72% when the child reaches 19 years of age).

The lump-sum subsidies provide a stronger incentive for those in low pay, who may also benefit less from training if they are from the low ability group or on a flatter wage trajectory induced by low (persistent) productivity shocks. We therefore re-designed the subsidy to exactly cover the foregone earnings of trainee mothers of children aged 0 to 7. Results for this policy are displayed in column 3 of Table 4.9. Because this type of design incetivises training among higher paid mothers, it also ends up being more expensive for each trainee than the generous £1,500 lump sum subsidy, costing £1,600 per trainee. However, it draws fewer women into training than the lump-sum benefit because it is less generous for lower paid women. So in the end the cost of such policy is smaller than that of the more generous lump-sum transfer, requiring an increase of 0.15pp in the basic tax rate to balance the public accounts. Its effects lie between the figures in the first two columns of the table, for the two lump-sum subsidies.

## 4.10 Conclusions

We have estimated a lifecycle model of female labor supply, and human capital accumulation through work experience and training. Our main aim has been to understand the role that job training can have in offsetting the loss of experience resulting from having children, which leads to an increasing wage gap for women with children.

Training can be important for wages and we show that it can partly offset the wage gap attributable to the prevalence of part time work and non-employment following a return to the labor market after having children.

Finally, we evaluate a policy of subsidizing training for mothers with children younger than 8. All policies are revenue neutral and funded by increasing taxes. A fixed modest subsidy of £500 increases the take up of training substantially and leads to small but persistent gains in wages, lifetime disposable income and welfare. It also pays for itself. We also consider other less effective and more expensive approaches.

This paper has ignored the all important question of incidence for the costs of training as well as for the returns. In a classical competitive labor market workers pay for general training and wages fully reflect returns to investment (Becker 1964). But in the presence of frictions this may not occur; firms and workers may share both the returns and the costs of training. While here we measure correctly the returns to the individual and attribute some of the costs to them we have not considered the returns to the firm of individuals being trained or how the firms and the workers may share the costs. This is a central question, all the more so if we are to understand why college graduates have such high levels of job training but little or no observed return. In a follow up paper we are investigating this issue based on a model inspired by Acemoglu and Pischke (1999).

# Chapter 5

# Conclusion

Chapters 2 to 4 each contain their own conclusions. Here, I briefly set out some areas of future research.

Chapter 2 develops a two-sided empirical matching model and uses the model to explore the impact of various higher education policies in on intergenerational mobility. While gender was not the focus of the paper, the associated empirical work identified gender disparities in prior attainment, subject selection and earnings. In particular, women study Arts, Humanities and Social Science subjects at a much higher rate than men, even when conditioning on their prior attainment. These subjects offer relatively weak labour market returns. Encouraging increased uptake of Science, Technology, Engineering and Mathematics among women is an area of policy interest, and a similar methodology could be used to investigate the determinants of subject selection and the extent to which differences in subject selection accounts for the gender wage gap among graduates.

In addition, while I did discuss the impact of abolishing tuition fees, I did not explore the parameters of the student loan system in detail. Over the last five years, several changes have been made to the repayment conditions for individuals with student loans, including to the interest rate and the repayment threshold. The model presented in Chapter 2 is an ideal laboratory for exploring the fiscal and behavioural consequences of these changes, as well as identifying welfare-improving reforms to the loan system.

Chapter 3 establishes that changes to the benefit system will feed-through to wages via human capital accumulation. Several reforms to the benefit system were introduced in the

United Kingdom over 2021. First, the temporary "uplift" to the basic allowance, introduced during the COVID-19 pandemic, was removed. The work allowance was expanded. Finally, in November 2021, the benefit withdrawal rate was reduced by 8 percentage points. These reforms are similar to the policy considered in the Chapter 3, which combined a reduction of the withdrawal rate with a decrease in the basic allowance. However, accurately capturing their effects requires a richer depiction of the benefit system than I developed in that paper, expanded to incorporate housing benefit and the work allowance. I am currently working on these extensions for a paper that will explicitly assess the impact of these recent reforms.

The framework in the paper could also be applied to consider broader, normative questions. Rather than projecting the impact of a specific reform, the model could be used to identify Pareto-improving reforms to low income support or to understand how optimal low-income support varies under different social welfare functions. This would relate the paper to the existing literature on optimal taxation with endogenous human capital, such as Blandin and Peterman (2019) and Stantcheva (2017), and offers a promising avenue for future research.

Finally, both Chapter 3 and Chapter 4 assume that the financial costs of skill investment are fully incident on the trainee. In the presence of labour market frictions, this may not be the case. Firms and workers may share both the returns and the costs of training. This could help explain some anomalies in my results, such as the high training rate and low return to training among university educated women described in Chapter 5. A more complete accounting would additionally consider the returns to the firm of individuals being trained and how the firms and the workers may share training costs. Incorporating training into a search model could therefore be a valuable next step in my research agenda.

# Bibliography


Acemoglu, Daron and Jorn-Steffen Pischke (Feb. 1999). "Beyond Becker: Training in Imperfect Labour Markets". In: *The Economic Journal* 109.453, pp. 112–142.

Adam, Stuart, James Browne, and Christopher Heady (2010). "Taxation in the UK". In: *Dimensions of Tax Design: The Mirrlees Review*. Ed. by James Mirrlees et al. Oxford University Press.

Agarwal, Nikhil (2015). "An empirical model of the medical match". In: *American Economic Review* 105.7, pp. 1939–78.

Altonji, Joseph G. and Lewis M. Segal (July 1996). "Small-Sample Bias in GMM Estimation of Covariance Structures". In: *Journal of Business & Economic Statistics* 14.3, p. 353.

Altonji, Joseph G. and James R. Spletzer (Oct. 1991). "Worker Characteristics, Job Characteristics, and the Receipt of On-the-Job Training". In: *Industrial and Labor Relations Review* (1), pp. 58–79.

Andrews, Isaiah, Matthew Gentzkow, and Jesse M. Shapiro (June 2017). "Measuring the Sensitivity of Parameter Estimates to Estimation Moments". In: *The Quarterly Journal of Economics* 132.4, pp. 1553–1592.

Andrews, Rodney J, Scott A Imberman, and Michael F Lovenheim (2017). *Risky Business? The Effect of Majoring in Business on Earnings and Educational Attainment*. Tech. rep. National Bureau of Economic Research.

Anelli, Massimo (2018). "Returns to elite university education: a quasi-experimental analysis". In: *Journal of the European Economic Association*.

Arcidiacono, Peter (2004). "Ability sorting and the returns to college major". In: *Journal of Econometrics* 121.1, pp. 343–375.

— (2005). "Affirmative action in higher education: How do admission and financial aid rules affect future earnings?" In: *Econometrica* 73.5, pp. 1477–1524.

Attanasio, Orazio, Hamish Low, and Virginia Sánchez-Marcos (Aug. 2008). "Explaining Changes in Female Labor Supply in a Life-Cycle Model". In: *American Economic Review* 98.4, pp. 1517–1552.

Attanasio, Orazio and Guglielmo Weber (Dec. 1995). "Is Consumption Growth Consistent with Intertemporal Optimization? Evidence from the Consumer Expenditure Survey". In: *Journal of Political Economy* 103.6, pp. 1121–1157.

Azmat, Ghazala and Ştefania Simion (2020). "Charging for Higher Education: Estimating the Impact on Inequality and Student Outcomest". In: *The BE Journal of Economic Analysis & Policy* 1.ahead-of-print.

Belfield, Chris et al. (2017). "Higher education funding in England: Past, present and options for the future". In: *IFS Briefing Note BN211*.

Black, Dan A and Jeffrey A Smith (2006). "Estimating the returns to college quality with multiple proxies for quality". In: *Journal of labor Economics* 24.3, pp. 701–728.

Black, Sandra E, Kalena E Cortes, and Jane Arnold Lincove (2015). "Academic undermatching of high-achieving minority students: Evidence from race-neutral and holistic admissions policies". In: *American Economic Review* 105.5, pp. 604–10.

Blandin, Adam and William B. Peterman (Oct. 2019). "Taxing capital? The importance of how human capital is accumulated". In: *European Economic Review* 119, pp. 482–508.

Blundell, Richard, Martin Browning, and Costas Meghir (Jan. 1994). "Consumer Demand and the Life-Cycle Allocation of Household Expenditures". In: *The Review of Economic Studies* 61.1, pp. 57–80.

Blundell, Richard, Monica Costa Dias, David Goll, et al. (Jan. 2021). "Wages, Experience, and Training of Women over the Life Cycle". In: *Journal of Labor Economics* 39.S1, S275–S315.

Blundell, Richard, Monica Costa Dias, Costas Meghir, et al. (Sept. 2016). "Female Labor Supply, Human Capital, and Welfare Reform". In: *Econometrica* 84.5, pp. 1705–1753.

Blundell, Richard, Lorraine Dearden, and Barbara Sianesi (July 2005). "Evaluating the effect of education on earnings: models, methods and results from the National Child Development Survey". In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 168.3, pp. 473–512.

Blundell, Richard, Luigi Pistaferri, and Itay Saporta-Eksten (Feb. 2016). "Consumption Inequality and Family Labor Supply". In: *American Economic Review* 106.2, pp. 387–435.

Bohacek, Radim and Marek Kapicka (Jan. 2008). "Optimal human capital policies". In: *Journal of Monetary Economics* 55.1, pp. 1–16.

Bovenberg, A. Lans and Bas Jacobs (Dec. 2005). "Redistribution and education subsidies are Siamese twins". In: *Journal of Public Economics* 89.11-12, pp. 2005–2035.

Brewer, Mike et al. (Dec. 2006). "Did working families tax credit work? The impact of in-work support on labour supply in Great Britain". In: *Labour Economics* 13.6, pp. 699–720.

Britton, Jack et al. (2021). "How much does degree choice matter?" In: *IFS Working Paper w21/24*.

Broecke, Stijn (2012). "University selectivity and earnings: Evidence from UK data on applications and admissions to university". In: *Economics of Education Review* 31.3, pp. 96–107.

Campbell, Stuart et al. (2021). "Matching in the dark? Inequalities in student to degree match". In: *Journal of Labor Economics*.

Card, David (1999). "The Causal Effect of Education on Earnings". In: *Handbook of Labor Economics*. Elsevier, pp. 1801–1863.

Chetty, Raj et al. (2020). "Income Segregation and Intergenerational Mobility Across Colleges in the United States". In: *Quarterly Journal of Economics*.

Chevalier, Arnaud (2011). "Subject choice and earnings of UK graduates". In: *Economics of Education Review* 30.6, pp. 1187–1201.

Chiappori, Pierre-André, Monica Costa Dias, and Costas Meghir (Oct. 2018). "The Marriage Market, Labor Supply, and Education Choice". In: *Journal of Political Economy* 126.S1, S26–S72.

Chiappori, Pierre-André, Murat Iyigun, and Yoram Weiss (Dec. 2009). "Investment in Schooling and the Marriage Market". In: *American Economic Review* 99.5, pp. 1689–1713.

Chowdry, Haroon et al. (2012). "Fees and student support under the new higher education funding regime: what are different universities doing?" In: *Institute for Fiscal Studies Briefing Note*.

Conesa, Juan Carlos, Sagiri Kitao, and Dirk Krueger (Feb. 2009). "Taxing Capital? Not a Bad Idea After All!" In: *American Economic Review* 99.1, pp. 25–48.

Corak, Miles (2013). "Income inequality, equality of opportunity, and intergenerational mobility". In: *Journal of Economic Perspectives* 27.3, pp. 79–102.

Costa, Carlos E. da and Marcelo R. Santos (Feb. 2018). "Age Dependent Taxes with Endogenous Human Capital Formation". In: *International Economic Review* 59.2, pp. 785–823.

Dale, Stacy B and Alan B Krueger (2002). "Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables". In: *The Quarterly Journal of Economics* 117.4, pp. 1491–1527.

Dale, Stacy B and Alan B Krueger (2014). "Estimating the effects of college characteristics over the career using administrative earnings data". In: *Journal of Human Resources* 49.2, pp. 323–358.

Dearden, Lorraine, Emla Fitzsimons, and Gill Wyness (2014). "Money for nothing: Estimating the impact of student aid on participation in higher education". In: *Economics of Education Review* 43, pp. 66–78.

Delavande, Adeline and Basit Zafar (2019). "University choice: The role of expected earnings, nonpecuniary outcomes, and financial constraints". In: *Journal of Political Economy* 127.5, pp. 2343–2393.

Denning, Jeffrey T, Benjamin M Marx, and Lesley J Turner (2019). "ProPelled: The effects of grants on graduation, earnings, and welfare". In: *American Economic Journal: Applied Economics* 11.3, pp. 193–224.

Diamond, William and Nikhil Agarwal (2017). "Latent indices in assortative matching models". In: *Quantitative Economics* 8.3, pp. 685–728.

Dillon, Eleanor W and Jeffrey A Smith (2017). "Determinants of the match between student ability and college quality". In: *Journal of Labor Economics* 35.1, pp. 45–66.

— (2020). "The consequences of academic match between students and colleges". In: *Journal of Human Resources* 55.3, pp. 767–808.

Epple, Dennis, Richard Romano, and Holger Sieg (2006). "Admission, tuition, and financial aid policies in the market for higher education". In: *Econometrica* 74.4, pp. 885–928.

French, Eric and John Bailey Jones (2004). "On the distribution and dynamics of health care costs". In: *Journal of Applied Econometrics* 19.6, pp. 705–721.

Fu, Chao (2014). "Equilibrium tuition, applications, admissions, and enrollment in the college market". In: *Journal of Political Economy* 122.2, pp. 225–281.

Gourieroux, Christian, Alain Monfort, and Eric Renault (1993). "Indirect Inference". In: *Journal of Applied Econometrics* 8, S85–S118.

Griffith, Amanda L and Donna S Rothstein (2009). "Can't get there from here: The decision to apply to a selective college". In: *Economics of Education Review* 28.5, pp. 620–628.

Hastings, Justine S, Christopher A Neilson, and Seth D Zimmerman (2013). *Are some degrees worth more than others? Evidence from college admission cutoffs in Chile*. Tech. rep. National Bureau of Economic Research.

— (2018). *The effects of earnings disclosure on college enrollment decisions*. Tech. rep. National Bureau of Economic Research.

Heckman, James J., Lance Lochner, and Ricardo Cossa (2003). "Learning-by-doing versus on-the-job training: using variation induced by the EITC to distinguish between models of skill formation". In: *Designing Inclusion: Tools to Raise Low-end Pay and Employment* *in Private Enterprise*. Ed. by Edmund S. Phelps. Cambridge University Press, pp. 74–130.

Hoxby, Caroline M and Christopher Avery (2013). "Low-income high-achieving students miss out on attending selective colleges". In: *Brookings Papers on Economic Activity, Spring*.

Hübner, Malte (2012). "Do tuition fees affect enrollment behavior? Evidence from a 'natural experiment'in Germany". In: *Economics of Education Review* 31.6, pp. 949–960.

Kapička, Marek (Apr. 2015). "Optimal Mirrleesean Taxation in a Ben-Porath Economy". In: *American Economic Journal: Macroeconomics* 7.2, pp. 219–248.

Kapička, Marek and Julian Neira (Oct. 2019). "Optimal Taxation with Risky Human Capital". In: *American Economic Journal: Macroeconomics* 11.4, pp. 271–309.

Kapor, Adam (2020). "Distributional effects of race-blind affirmative action". In: *Job Market Paper, Yale University*.

Karabarbounis, Marios (Apr. 2016). "A Road Map for Efficiently Taxing Heterogeneous Agents". In: *American Economic Journal: Macroeconomics* 8.2, pp. 182–214.

Keane, Michael P and Kenneth I Wolpin (1997). "The career decisions of young men". In: *Journal of political Economy* 105.3, pp. 473–522.

— (2001). "The effect of parental transfers and borrowing constraints on educational attainment". In: *International Economic Review* 42.4, pp. 1051–1103.

Kirkeboen, Lars J, Edwin Leuven, and Magne Mogstad (2016). "Field of study, earnings, and self-selection". In: *The Quarterly Journal of Economics* 131.3, pp. 1057–1111.

Lentz, Rasmus and Nicolas Roys (Nov. 2015). *Training and Search On the Job*. Tech. rep.

Marx, Benjamin M and Lesley J Turner (2018). "Borrowing trouble? human capital investment with opt-in costs and implications for the effectiveness of grant aid". In: *American Economic Journal: Applied Economics* 10.2, pp. 163–201.

Mountjoy, Jack and Brent Hickman (2020). "The Returns to College (s): Estimating Value-Added and Match Effects in Higher Education". In: *University of Chicago, Becker Friedman Institute for Economics Working Paper*.

Neill, Christine (2009). "Tuition fees and the demand for university places". In: *Economics of Education Review* 28.5, pp. 561–570.

Peterman, William B. (July 2016). "The effect of endogenous human capital accumulation on optimal taxation". In: *Review of Economic Dynamics* 21, pp. 46–71.

Sá, Filipa (2019). "The effect of university fees on applications, attendance and course choice: evidence from a natural experiment in the uk". In: *Economica* 86.343, pp. 607–634.

Scott-Clayton, Judith and Basit Zafar (2019). "Financial aid, debt management, and socioeconomic outcomes: Post-college effects of merit-based aid". In: *Journal of Public Economics* 170, pp. 68–82.

Shaw, Jonathan (May 2011). *FORTAX: UK tax and benefit system documentation*. Tech. rep.

Skrondal, Anders and Sophia Rabe-Hesketh (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. Chapman and Hall/CRC.

Stantcheva, Stefanie (July 2015). "Learning and (or) Doing: Human Capital Investments and Optimal Taxation". In: *NBER Working Paper* 21381.

— (Dec. 2017). "Optimal Taxation and Human Capital Policies over the Life Cycle". In: *Journal of Political Economy* 125.6, pp. 1931–1990.

Wiswall, Matthew and Basit Zafar (2014). "Determinants of college major choice: Identification using an information experiment". In: *The Review of Economic Studies* 82.2, pp. 791–824.

Zimmerman, Seth D (2019). "Elite colleges and upward mobility to top jobs and top incomes". In: *American Economic Review* 109.1, pp. 1–47.

# Appendix A

# Can Higher Education Policy Boost Social Mobility? Evidence From an Empirical Matching Model

## A.1 Individual skill measurement

We estimate two skill variables, summarising each individuals quantitative and communication skills. These skill measures aren't directly observed in our data but we assume that exam results, which we do observe, are noisy indicators of these underlying skills. We first estimate a set of structural equations capturing the relationship between the latent skills and observable exam results. Then, given each individuals exam results, we predict their underlying skills. For more details on the approach applied, see Skrondal and Rabe-Hesketh (2004).

The skill measures are based on two sets of exam results. The first set of exams are the Key Stage 2 *Standard Assessment Tests* (SATs). These exams are taken by all English students at age 11. In our data, we can observe the numerical score in three exams:

- Mathematics
- Reading
- Writing

We supplement these measures with *General Certificate of Secondary Education* (GCSE) grades. GCSEs are sat at age 16. Most students sit between 8 and 11 GCSEs across a range of subjects. The only compulsory subjects are Mathematics, English Language and Science, but most schools require that students also sit English Literature, at least one GCSE in a humanities subject (Geography or History), one in modern foreign languages subject (such as French or Spanish) and one in an arts subject (such as Music or Drama). We observe the letter grade (A* to G) for each GCSE. We use the following exams when estimating skills:

- English Language
- English Literature
- Mathematics
- Science
- Humanities

where humanities is measured as the maximum of grades in Geography and History. We choose this relatively small subset of GCSE subjects to minimise missing data, which would complicate both estimation and prediction.

We model each grade as a function of (latent) quantitative and communication skill:

$$M_i^k = \zeta_0^k + \zeta_m^k S_i^q + \zeta_c^k S_i^c + \epsilon_i^k$$

where $k$ indexes the measure, $S_i^q$ is quantitative skill and $S_i^c$ is communication skill. $S_i^q$ and $S_i^c$ are assumed to be jointly normally distributed. $\epsilon_i^k$ is unobserved and assumed to be uncorrelated across measures. We also assume that the variance and covariance of the underlying latent variables are the same for all cohorts, but allow the parameters of each measurement equation ($\zeta_0^k$, $\zeta_m^k$, $\zeta_c^k$ and $\sigma_{\epsilon^k}$) to differ between cohorts. This allows us to capture grade inflation.

To identify the system of measurement equations we need to make some additional exclusion restrictions. We exclude quantitative skill from the measurement equation for English Language or English Literature GCSEs, and the communication skill from the measurement equation for Mathematics GCSE. Parameters are estimated using full information maximum likelihood. Once we have estimated the parameters, we predict the skill measures for each individual using the empirical Bayesian modal approach detailed in Skrondal and Rabe-Hesketh 2004.

## A.2 Course quality measurement

Course quality is a summary measure derived using the following variables:

- **Academic services spend:** Expenditure on all academic services (such as library and computing facilities) divided by the number of full-time equivalent students.
- **Facilities spend:** Expenditure on all student facilities (such as sports, careers services, counselling) divided by the number of full-time equivalent students.
- **Student-staff ratio:** The total number of undergraduate and postgraduate students divided by the number of academic staff.
- **Student satisfaction:** Average across all satisfaction scores as measured in in the National Student Survey.
- **Research quality:** Score of the relevant department in the Research Excellence Framework.

Data is collected from the *Complete University Guide*. Academic services spend, facilities spend and student-staff ratio are measured at the university level. Research quality and student satisfaction are both measured at the course level, varying across fields within each university. Fields within the *Complete University Guide* data are more disaggregated than the three broad fields that we consider in our analysis (STEM, LEM and AHSS). To create measures defined for our broad fields, we take a weighted average over the disaggregated fields using number of undergraduates enrolled in the relevant field at the relevant university in 2010. Finally, we take the first principal component of these five variables as our measure of course quality. The weights assigned to each of the variables (after standardisation) are reported below:

| Variable | Weight in 1st principal component |
|---|---|
| Academic services spend | 0.483 |
| Facilities spend | 0.397 |
| Student-staff ratio | -0.545 |
| Student satisfaction | 0.195 |
| Research quality | 0.523 |

## A.3 Student number caps

Higher Education Funding Council for England (HEFCE) provides information on the student number controls in place for each English university from 2010 to 2014. These controls applied to new full-time UK and EU undergraduate students and post-graduate students undertaking teacher training. Universities that exceeded the cap had their teaching grants reduced in an effort to prevent over-recruitment and control government expenditure on the HE sector.

A number of exemptions from the student number controls were in place at the beginning of the period studied. For instance, students who recently completed a full-time foundation (two-year) degree were exempt from the caps, as were any students who were studying for a qualification of an equivalent or lower level to one that they already possessed. Students who transferred courses within an institution were generally not counted against the student number controls, unless they were previously studying part-time. Some degrees which were necessary training for particular public sector professions, such as medicine or nursing, were covered in a seperate cap system.

Additional exemptions were introduced in 2012 and 2013. Students with grades AAB at A-level, or other entry qualifications which are equivalent to or higher than such A-level grades, were removed from the cap in 2012. This exemption was extended to students with ABB grades in 2013. In both cases, all institutions had their cap reduced by the number of AAB or ABB students that they recruited in the previous year, to maintain the same overall level of tightness in the HE market. Finally, in 2013, HEFCE introduced a “flexibility range”, allowing universities to exceed their cap by 3% without receiving any penalties. This effectively increased the caps by 3% for all universities.

While it is relatively straightforward to identify first-year UK and EU domiciled HE students for each university using HESA data, it is less straightforward to replicate the system of exemptions. We cannot, for instance, distinguish a student who has newly arrived in an institution from one who has transferred from another course, nor can we identify students who have previously completed a foundation course or an equivalent qualification. As such, we can only measure the number of students counted against the cap each year with some error.

Universities did not, in general, perfectly match their cap. The majority of offers are made over 6 months before students are due to arrive of campus and are conditional on attaining specific grades at A-level. Universities need to estimate the proportion of offers that are accepted, the proportion of accepted offers whose conditions are met, and the proportion of students who, after accepting the offer and meeting the conditions, actually turn up to enrol. The clearing system, which provides a secondary market after grades are realized, can help universities fill spare places if they find that fewer offers than expected have been accepted, but cannot be used to compensate for students failing to enrol.

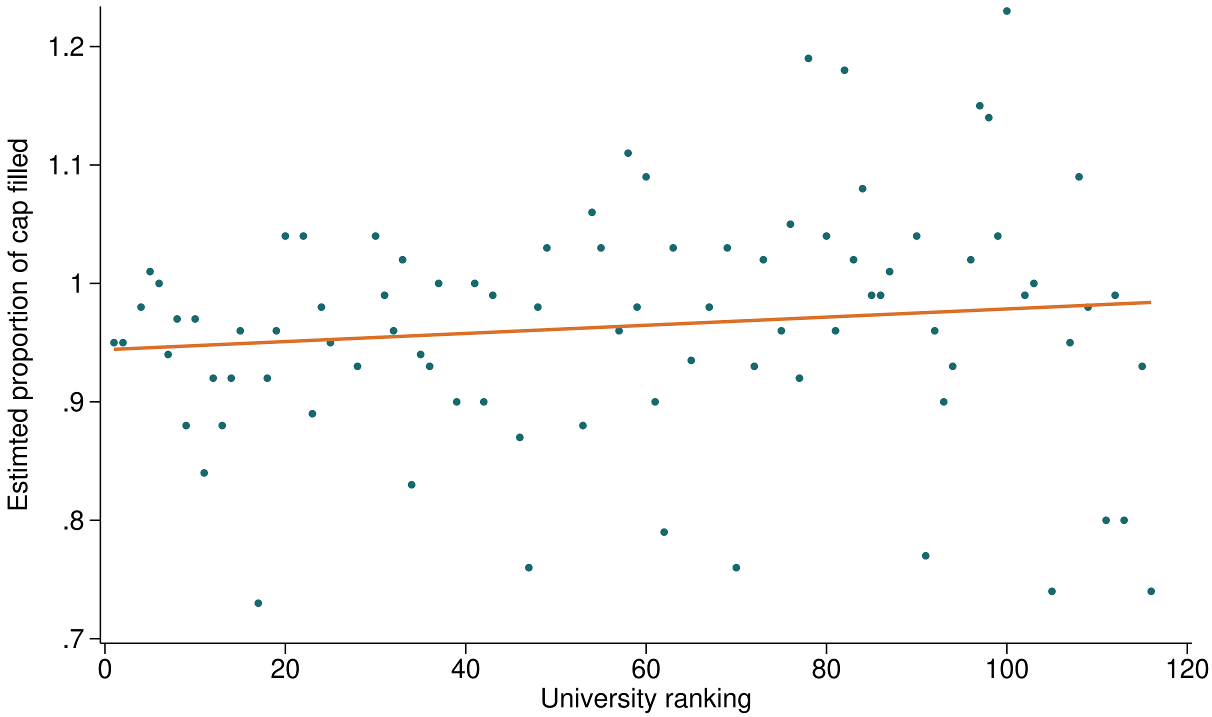
Table A.1 shows both our best estimate for the tightness of the overall market from 2010 to 2014. This is calculated as the total number of new undergraduate enrollees divided by the total number of places available through the student control system. We have adjusted the places available to account for the exemption of high-attaining students and the introduction of the flexibility range. While there is some variation year on year, the overall student number control closely matches the total number of new enrollees each year.

Table A.1: Student number caps

| Year | Total new undergraduates | Total student number control | Average tightness |
|---|---:|---:|---:|
| 2010 | 285,225 | 297,972 | 0.957 |
| 2011 | 300,795 | 295,432 | 1.018 |
| 2012 | 258,705 | 269,504 | 0.959 |
| 2013 | 284,845 | 278,685 | 1.022 |
| 2014 | 293,390 | 278,890 | 1.052 |

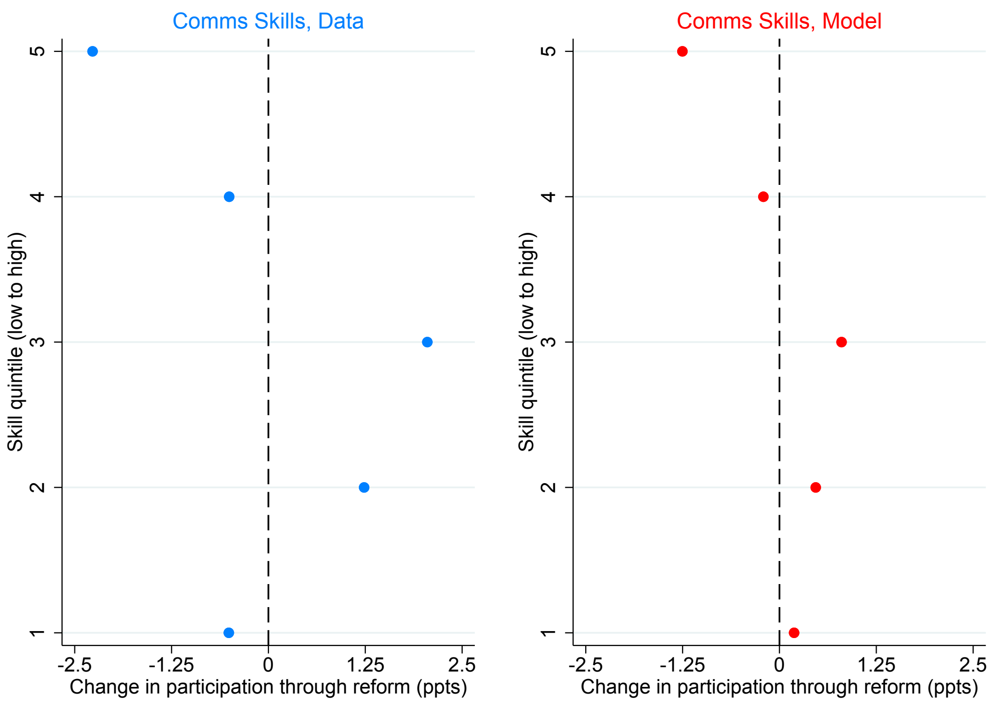
Figure A.1 plots the estimated proportion of cap filled for individual universities in 2010 relative to their rank in the Guardian University Guide in the same year. As noted above, the tightness of individual institutions is measured with significant error, and we are therefore likely to be overstating the extent to which individual institutions miss or exceed their cap. However, the figure demonstrates that, on average, universities are filling all their places and there is no visible relationship between the measured tightness of the university and its ranking.

Figure A.1: University ranking versus estimated proportion of cap filled




## A.4 Additional figures

Figure A.2: Reform effects by ability: model vs. data (communications skills)




Note: See note to Figure 2.14.

# Appendix B

# Human Capital Accumulation and Benefit Design

## B.1 Model solution

The Bellman equation during retirement is:

$$V_t(k_t, f_t) = \max_{c_t} \; u_{\text{ret}}(c_t, f_t) + \beta E\left[V_{t+1}(k_{t+1}, f_{t+1})\right]$$

subject to:

$$\begin{aligned} u_{\text{ret}}(c_t, f_t) &= \frac{\left(c/\psi(f_t)\right)^{1-\gamma_1}}{1-\gamma_1} \\ c_t + k_{t+1} &= PP(f_t) + Rk_t \\ k_{t+1} &\geq 0 \end{aligned}$$

where the expectation is taken over family type. This model can be be solved backwards to find the expected value of retirement conditional on financial capital and family type, which I denote as $V_{\text{ret}}(k_t, f_t)$. I solve this function for all family types and for a discrete grid over financial capital.

The Bellman equation during the final period of working life is then given by:

$$V_t(k_t, h_t, a, f_t, n_{m,t}, \epsilon_t) = \max_{c_t, n_t, e_t} u(c_t, n_t, e_t, f_t) + \beta E\left[V_{\text{ret}}(k_{t+1}, f_{t+1})\right]$$

with the expectation taken over family types. In all prior periods, the Bellman equation is:

$$V_t(k_t, h_t, a, f_t, n_{m,t}, \epsilon_t) = \max_{c_t, n_t, e_t} u(c_t, n_t, e_t, f_t) + \beta E\left[V_{t+1}(k_{t+1}, h_{t+1}, a, f_{t+1}, n_{m,t+1}, \epsilon_{t+1})\right]$$

where the expectation is taken over family type, human capital shock, partner productivity and partner employment. In both cases, the maximisation is subject to the following constraints:

$$\begin{aligned}
u(c_t, n_t, e_t, f_t) &= \frac{\left(c/\psi(f_t)\right)^{1-\gamma_1}}{1-\gamma_1} - \chi(f_t)\frac{(n_t + e_t)^{1-\gamma_2}}{1-\gamma_2} \\
\tilde{h}_{t+1} &= h_t + a\left(\alpha e_t^{\phi_1} + (1-\alpha)n_t^{\phi_2}\right) \\
h_{t+1} &= \exp(z_t) \times \tilde{h}_{t+1} \\
\ln y_{f,t} &= \ln(h_t) + \ln(n_t) \\
\ln w_{m,t} &= \beta_0 + \beta_1 t^{\frac{1}{2}} + \beta_2 t + \epsilon_t \\
\epsilon_t &= \rho\epsilon_{t-1} + \xi_t \\
y_{m,t} &= 1(n_{m,t} > 0) \times 40 w_{m,t} \\
c_t + k_{t+1} &= T_l(y_{f,t}) + T_l(y_{m,t}) + B\left(T_l(y_{f,t}) + T_l(y_{m,t}), f_t\right) + Rk_t \\
k_{t+1} &\geq 0
\end{aligned}$$

As with the retirement period, I solve this model backwards over a discrete grid in financial capital, human capital, ability and partner productivity. Partner employment and family type are discrete variables, so no further approximation is necessary.

Due to the design of the benefit system, agents in the model generally face non-monotonic effective marginal tax rates. When working hours are near zero, an additional pound of gross labour earnings increases net income by 37 pence (at the current UK benefit withdrawal rate). As labour hours increase, individuals will eventually stop receiving benefits and their effective marginal tax rates falls, before climbing again due to the progressivity of the income tax schedule. This structure complicates the solution of the model, creating discontinuties in policy functions and kinks in the value function.

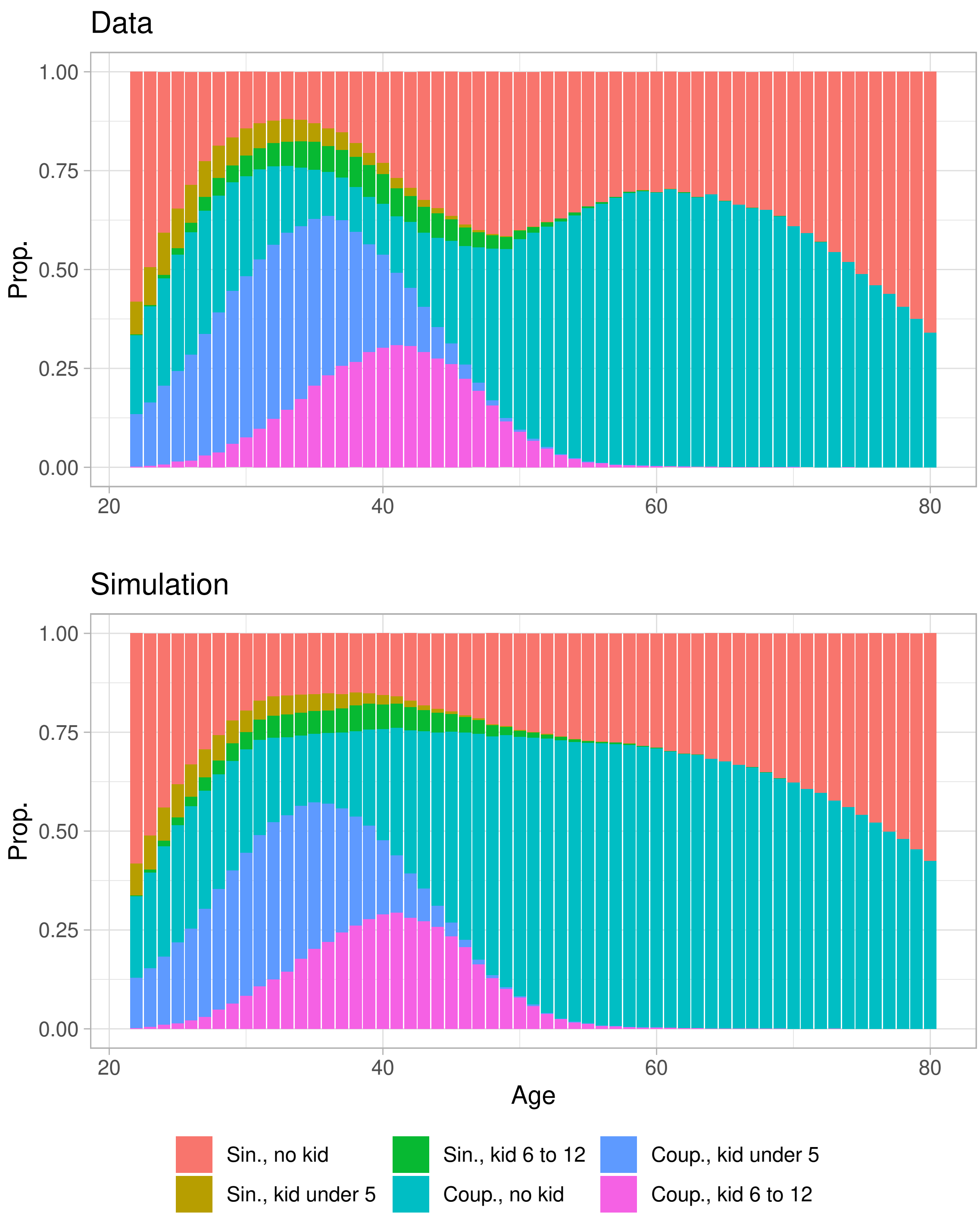
To keep the solution of the model feasible, I discretize both the labour hours and education hours choice. While the choice set is discrete, I keep the size of the set large to ensure that it still captures the impact of incentives generated by the benefit system on the intensive margin. Labour hours can be selected in four hour increments from 0 to 48 hours per week. Education hours can be selected from the set $\{0, 0.25, 0.5, 1, 2, 4, 8, 16, 32\}$. This set has been chosen to have more density at low hours, reflecting the observed distribution of education hours choices. Given the discretization of the hours choices, the model can be solved by value function iteration. At each state space, I first solve for optimal consumption and expected value conditional on each potential choice of labour and education hours. I then select the labour and education choice that maximises expected value within this set.

## B.2 Exogenous processes

To model demographic transitions, I estimate separate family-type transition matrices for each age in my sample. Once the agent enters retirement, I assume that they have no children in the household but I continue to allow the number of adults in the household to change due to, for instance, separation or death of a partner. I use these transition matrices when solving the model and when generating simulated moments. Figure B.1 compares the family type distribution observed in the data (top panel) with the simulated family type distribution (bottom panel). The fit is generally good, although I underestimate the proportion of single women without children aged between 40 and 50 and overestimate the proportion of women in couples without children over the same age range. This is likely because the estimation approach described above implicitly assumes that demographic transitions are stationary over time; instead, due to declining divorce rates, women observed in my data between the ages of 40 and 50 experienced higher rates of separation in their 20s and 30s (before I observe them) than the women who I observe early in their life-cycle.

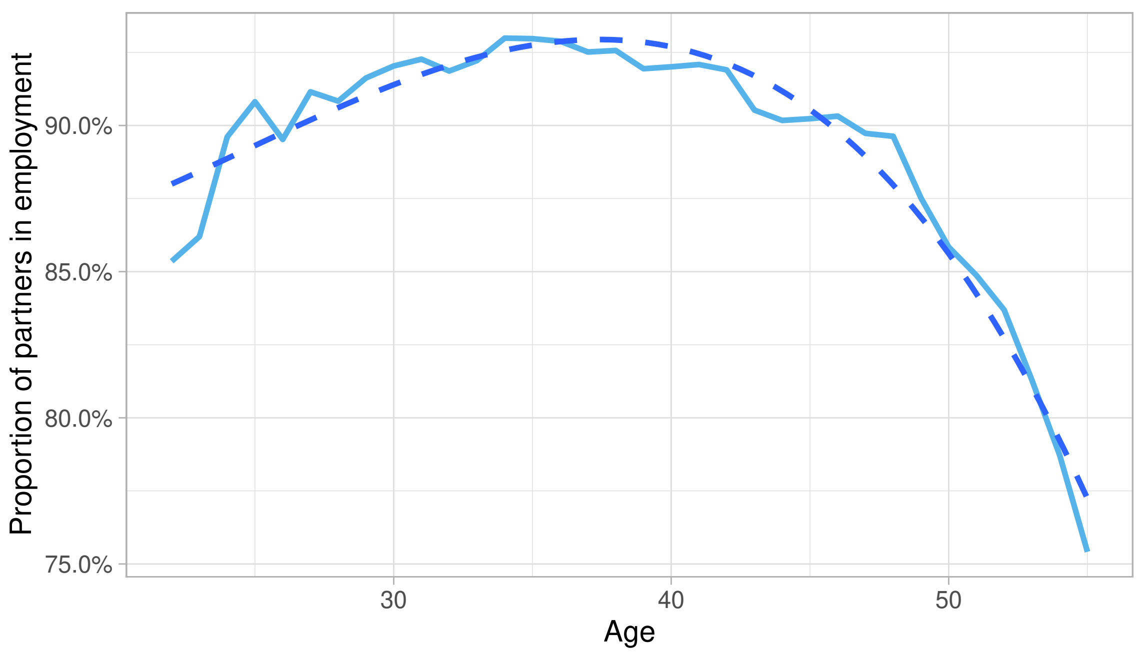
Partner employment probability is modelled as a third-degree polynomial function of women's

Figure B.1: Family type data versus simulations



age. Figure B.2 shows both average partner employment probability at each age and the estimated polynomial. Conditional on working, the vast majority of partners in our data work full-time hours; average hours are 39.25 per week and 90% of employed partners work 30 hours or more. We assume in the model that working partners provide 40 hours of labour per week.

Figure B.2: Partner employment over life-cycle




The empirical model for partner wages is stated below. This is identical to the structural model discussed in Section 3.4, but with the inclusion of an additional measurement error term $\zeta_t$. To estimate the relevant parameters, I first regress partner's wages on women's age and the square root of women's age. I then remove the time-trend to isolate $\epsilon_t + \zeta_t$. Finally, I estimate the remaining parameters, $\rho$ and $\sigma^2_\xi$, to match the auto-correlation matrix of the constructed residual.

$$\begin{aligned} \ln \tilde{w}_{m,t} &= \beta_0 + \beta_1 t^{\frac{1}{2}} + \beta_2 t + \epsilon_t + \zeta_t \\ \epsilon_t &= \rho \epsilon_{t-1} + \xi_t \end{aligned}$$

# Appendix C

# Wages, Experience, and Training of Women over the Life Cycle

Estimation is based on all 18 yearly waves of the British Household Panel Survey (BHPS), covering the period from 1991 to 2008. Apart from those who are lost through attrition, all families in the original 1991 sample and subsequent booster samples remain in the panel from then onwards. Other individuals have been added to the sample in subsequent periods, sometimes temporarily, as they formed families with original interviewees or were born to them. All members of the household aged 16 and above are interviewed. We select the sample of women in all types of family arrangement observed while aged 19 to 59.

Some definitional and data preparation procedures should be mentioned for clarity. Employment is determined by present labor-market status and excludes self-employment. The paths of women who report being self-employed are deleted from that moment onwards. This partly eliminates the trajectories of 889 women of the original sample of 7,755 women, dropping 6,569 individual-year observations. Similarly, we truncate the paths of women who report returning to full-time education after they have entered the labor market. 764 individuals are observed returning to full-time education, for whom we drop their pathways from that moment onwards, amounting to 4,737 individual-year observations. We start with 67,399 and end with 56,093 individual-year observations after the cleaning trajectories after they cross self-employment or full-time education. We also drop observations for women for whom the age of children are missing or look wrong. This leaves 55,591 individual-year observations in our final dataset.

Only women working 5 or more hours per week are classified as employed. We consider employment choices from the age of 19 for women with secondary and high school education, and from the age of 22 for women with university education. Working hours refer to the usual hours in main job including overtime. We discretized labor supply using a three-point distribution: not working (0 to 4 hours per week, modeled as 0 hours), working part-time (5 to 20 hours per week, modeled as 18 hours), and working full-time (21 hours or more per week modeled as 38 hours). The employment status and working hours observed at one point in the year are assumed to remain unaltered over the entire year. Earnings are the usual gross weekly earnings in the main job. (Hourly) wage rates are the ratio of weekly earnings to weekly hours capped at 70. The wage distribution is trimmed at percentiles 2 and 99 from below and above, respectively, and only for women working at or above 5 hours per week to reduce the severity of measurement error in wage rates.

Wage rates are detrended using the aggregate wage index and all other monetary parameters in the model, including all monetary values in the annual sequence of tax and benefit systems, were deflated using the same index. To construct this index, we run three regressions, one for each education level, of trimmed wages on time dummies and dummies of Scotland and Wales. We create three education-specific wage indices from the coefficients in time. Then we aggregate these indices using the distribution of education for the entire population of workers aged 25-59 in the sample to form the wage index. Any real monetary values (using the CPI) are then rescaled using this index.

Family type includes four groups: single women and couples without children, lone mothers, and couples with children. Women are assumed to have children only after finishing education, once entering the labor market. Cumulated work experience is measured in years. Individual assets at the beginning of adult life are the total of savings and investments net of debts. They are truncated at zero, never allowed to be negative.

Our full data set remaining after the sample selection procedure described above, used for the descriptive graphs and tables, is an unbalanced panel of 7,359 women observed for some varying period during the years 1991 to 2008. A great deal of information is collected for them, including family demographics, employment, working hours and earnings as well as those of a present partner, women's demographics such as age and education, demand for childcare and its cost.

Within the model, we focus on "high-school" educated women. These women have completed A-levels or equivalent qualifications, which are acquired at the end of high school at the age of

18. They do not possess a first degree level or post-graduate level qualification. High-school educated women are 32% of the individuals in our full sample and 31% of observations. Moreover, for inclusion in our model sample we require observation of historical data on the characteristics of their parental home when they were aged 16, including whether lived with parents, parent's education, employment status, number of siblings and sibling order, books at home. Of the 2,377 individuals with high-school education, we observe the family background for 1,443 (60.7%). These individuals form the basis for our moment estimates and the initial conditions of the model.

Figure C.1 and C.2 replicate Figure 4.5 from the main text, using a slightly different method. Rather than smoothing training rates using a local polynomial, we have binned individuals into five year age ranges and presented the average training rate for each bin alongside the 95% confidence interval. Figure C.1 presents the training rates of men, while Figure C.2 presents the training rates of women. In each case, Panel A presents the unconditional training rates and Panel B presents training rates conditional on working. The training rates of men appear to decline steadily over the lifecycle, whereas training rates of women decline at first but increase somewhat during their 40s.

For completeness, we have included below an alternative version of Table 4.3 in the main text. Whereas the table presented in the main text includes all individuals in our sample, including those who are not employed, Table C.1 conducts the same regression conditioning on working more than 5 hours per week. The sample is significantly smaller and the instruments lose some power, particularly for low education individuals. However, simulated full-time income retains strong explanatory power for training among the employed.

We also present some additional graphs showing the fit of the model in terms of employment and part-time hours over the lifecycle (Figure C.3) and over age of the oldest child (Figure C.4).

Tables C.2 to C.18 display the full list of data moments used in estimation, together with their simulated counterparts and the normalized (by the data standard error) differences between the two. Estimation used 139 moments, which fall into the following categories:

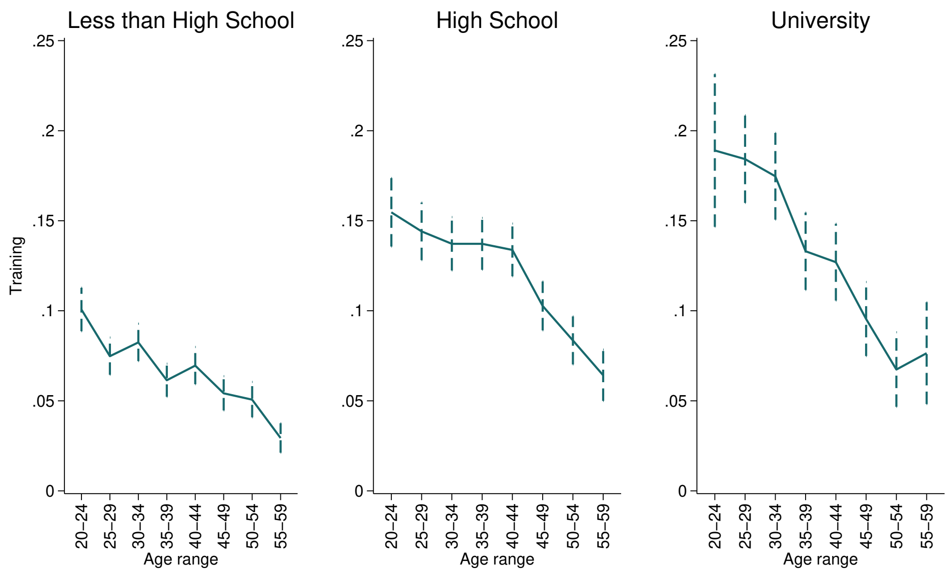
- Mean employment, part-time hours and training conditional on demographics (Table C.2, C.3 and C.4)
- Mean employment and training conditional on age band (Table C.5 and C.6)

- Transition rates from unemployment to employment conditional on demographics (Table C.7)
- Transition rates from employment to unemployment conditional on demographics and wage decile (Table C.8)
- Mean, variance and quantiles of log wage at entrance to working life (Table C.9)
- Log wage regression in first differences on training dummy and change in log experience (Table C.10)
- Log wage regression on lagged wage, family background, log years of work experience and lagged log years of work experience (Table C.11)
- Log wage regression on training, experience and working status last period (Table C.12)
- Log wage regression on age and family background (Table C.13)
- Mean yearly change in wages conditioning on working status last period (Table C.14)
- Mean wages and proportion of population with wages below pre-defined empirical wage deciles, conditional on working hours and training (Table C.15, C.16 and C.17)
- Mean log wages conditional on family background (Table C.18)

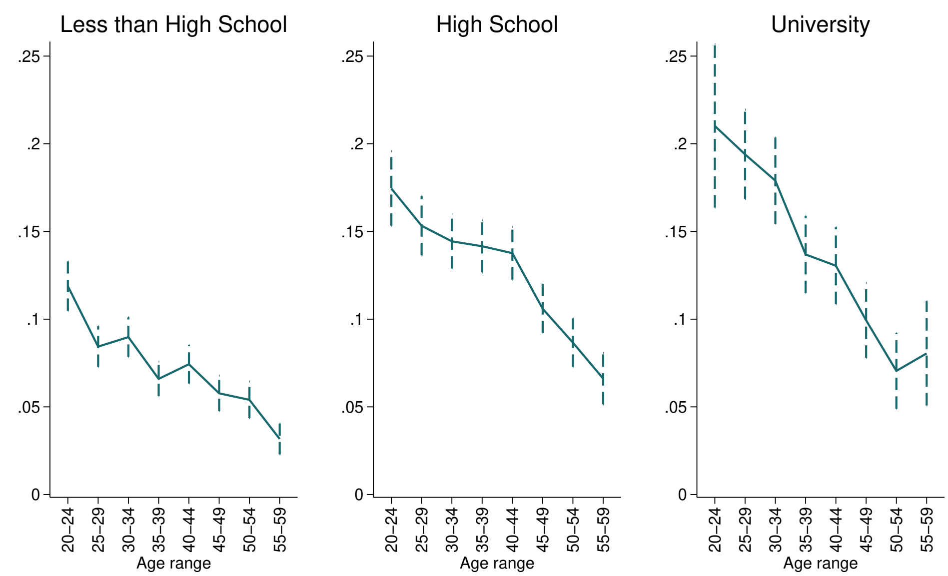
All moments are constructed from the BHPS and are education-specific. Among the 139 simulated moments, 19 fall outside the 95% confidence interval for the respective data moment, but many amongst these are very similar to their BHPS counterparts.

Figure C.1: Training rates for men over the lifecycle

Panel A: All



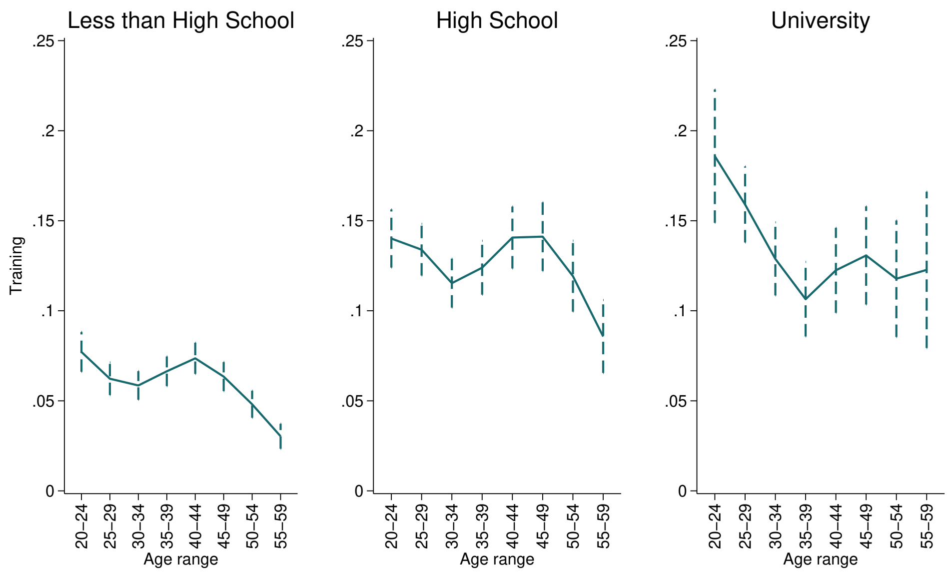

Panel B: In work



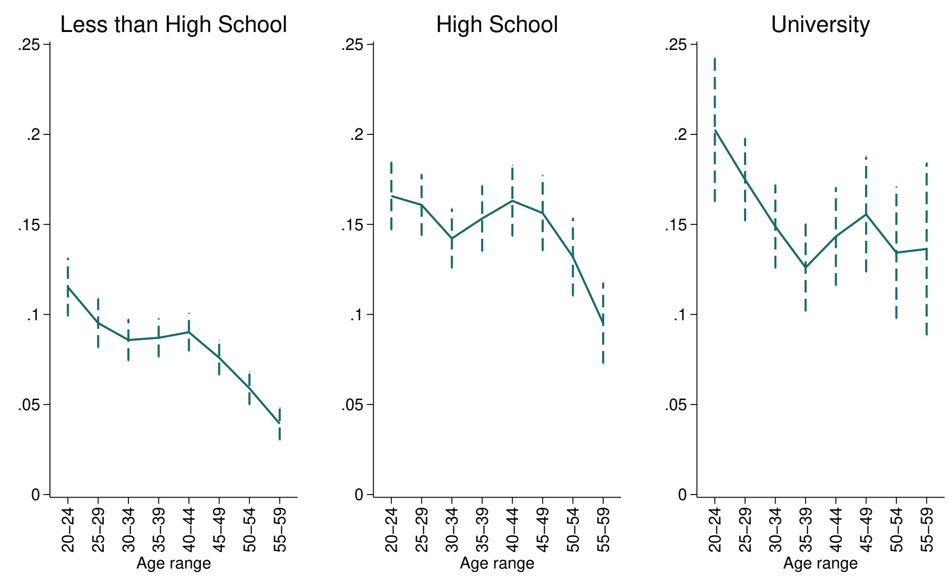

Notes: BHPS data for years 1991-2008. The training variable is an indicator for having had 40 or more hours of work-related training over the last 12 months. Panel A shows training rates for the entire population, by age and education. Panel B additionally conditions on working at least 5 hours per week on an usual week, which is the measure of employment used in this paper. Dashed line shows the 95% confidence interval.

Figure C.2: Training rates for women over the lifecycle

Panel A: All




Panel B: In work




Notes: BHPS data for years 1991-2008. The training variable is an indicator for having had 40 or more hours of work-related training over the last 12 months. Panel A shows training rates for the entire population, by age and education. Panel B additionally conditions on working at least 5 hours per week on an usual week, which is the measure of employment used in this paper. Dashed line shows the 95% confidence interval.

Table C.1: Regression of training conditional on employment

| | (1) Secondary | (2) High School | (3) Degree |
|---|---|---|---|
| Sim Income: 0 hours | -0.0000393 | -0.000134 | -0.000133 |
| | (0.000123) | (0.000171) | (0.000237) |
| Sim Income: 20 hours | 0.000436** | 0.000389 | 0.000711 |
| | (0.000198) | (0.000288) | (0.000434) |
| Sim Income: 40 hours | -0.000668*** | -0.000793*** | -0.000922*** |
| | (0.000148) | (0.000181) | (0.000255) |
| Observations | 22739 | 14658 | 6537 |
| Demographic Controls | Yes | Yes | Yes |
| Family Background Controls | Yes | Yes | Yes |
| Wave Dummies | Yes | Yes | Yes |
| F-Stat | 7.259 | 8.787 | 6.369 |
| F-Stat p-val | 0.0000751 | 0.00000862 | 0.000282 |

Standard errors in parentheses

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

Notes: BHPS data. Outcome variable indicates whether the individual is observed in more than 40 hours of work-related training. Sample is conditioned on working at least 5 hours a week. Standard errors are clustered at the individual level. Demographic controls include a quadratic in age and dummies indicating family composition. Family background controls include the first two principal components drawn from a collection of variables that describe the childhood household of each individual and an indicator for whether this information is missing.

Figure C.3: Employment over life-cycle

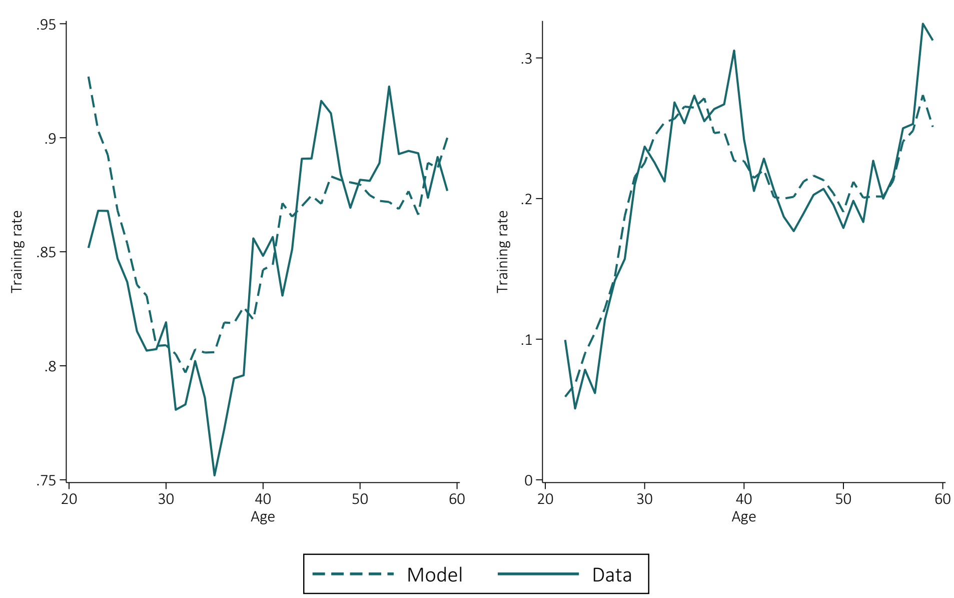


Table C.2: Mean employment during working life

| Moment | Data | Simulated | SE data | Norm. SE diff |
|---|---|---|---|---|
| Panel A: Averages by demographics | | | | |
| All | 0.837 | 0.829 | 0.010 | 0.791 |
| Single women, no child | 0.914 | 0.920 | 0.011 | 0.505 |
| Married women, no child | 0.938 | 0.935 | 0.010 | 0.300 |
| Lone mothers | 0.682 | 0.644 | 0.041 | 0.948 |
| Married mothers | 0.731 | 0.730 | 0.019 | 0.039 |
| Partner working | 0.837 | 0.847 | 0.012 | 0.883 |
| Youngest child 0-2 | 0.602 | 0.570 | 0.025 | 1.309 |
| Youngest child 3-5 | 0.713 | 0.714 | 0.024 | 0.047 |
| Youngest child 6-10 | 0.777 | 0.762 | 0.025 | 0.616 |
| Youngest child 11+ | 0.854 | 0.870 | 0.023 | 0.683 |
| Family background: factor 1 | 0.825 | 0.834 | 0.014 | 0.633 |
| Family background: factor 2 | 0.841 | 0.850 | 0.014 | 0.612 |
| Panel B: Impact of benefit reform | | | | |
| Pre-1999: single women, no child | 0.084 | 0.076 | 0.018 | 0.439 |
| Pre-1999: married women, no child | 0.110 | 0.098 | 0.013 | 0.917 |
| Pre-1999: lone mothers | -0.251 | -0.217 | 0.053 | 0.648 |
| Pre-1999: married mothers | -0.101 | -0.079 | 0.016 | 1.398 |
| Post-1999: single women, no child | 0.071 | 0.104 | 0.015 | 2.291 |
| Post-1999: married women, no child | 0.093 | 0.114 | 0.013 | 1.657 |
| Post-1999: lone mothers | -0.081 | -0.156 | 0.044 | 1.699 |
| Post-1999: married mothers | -0.111 | -0.119 | 0.015 | 0.551 |

Notes: Moments in Panel A are average employment rates (measured as working five or more hours per week) among individuals with the listed demographic and background characteristics. Moments in Panel B are the deviation in percentage points of each of the family types employment rates from average employment rates in the period indicated.

Table C.3: Mean part-time employment during working life

| Moment | Data | Simulated | SE data | Norm. SE diff |
|---|---|---|---|---|
| Panel A: Averages by demographics | | | | |
| All | 0.164 | 0.170 | 0.009 | 0.723 |
| Single women, no child | 0.062 | 0.074 | 0.011 | 1.104 |
| Married women, no child | 0.096 | 0.100 | 0.012 | 0.348 |
| Lone mothers | 0.173 | 0.205 | 0.036 | 0.901 |
| Married mothers | 0.277 | 0.272 | 0.016 | 0.317 |
| Partner working | 0.191 | 0.193 | 0.011 | 0.160 |
| Youngest child 0-2 | 0.257 | 0.280 | 0.020 | 1.163 |
| Youngest child 3-5 | 0.321 | 0.316 | 0.024 | 0.200 |
| Youngest child 6-10 | 0.281 | 0.280 | 0.024 | 0.010 |
| Youngest child 11+ | 0.196 | 0.165 | 0.027 | 1.116 |
| Family background: factor 1 | 0.158 | 0.135 | 0.012 | 1.928 |
| Family background: factor 2 | 0.179 | 0.187 | 0.013 | 0.590 |
| Panel B: Impact of benefit reform | | | | |
| Pre-1999: single women, no child | -0.115 | -0.106 | 0.018 | 0.502 |
| Pre-1999: married women, no child | -0.068 | -0.070 | 0.014 | 0.158 |
| Pre-1999: lone mothers | -0.054 | -0.010 | 0.035 | 1.264 |
| Pre-1999: married mothers | 0.128 | 0.115 | 0.015 | 0.805 |
| Post-1999: single women, no child | -0.089 | -0.086 | 0.015 | 0.201 |
| Post-1999: married women, no child | -0.068 | -0.070 | 0.012 | 0.169 |
| Post-1999: lone mothers | 0.059 | 0.075 | 0.044 | 0.350 |
| Post-1999: married mothers | 0.100 | 0.088 | 0.014 | 0.813 |

Notes: Moments in Panel A are average rates of part-time hours (measured as working between 5 and 20 hours a week) among employed individuals with the listed demographic and background characteristics. Moments in Panel B are the deviation in percentage points of each of the family types part-time hours rates from average part-time hours rates in the period indicated.

Table C.4: Mean training during working life

| Moment | Data | Simulated | SE data | Norm. SE diff |
|---|---|---|---|---|
| Panel A: Averages by demographics | | | | |
| All | 0.160 | 0.169 | 0.007 | 1.291 |
| Single women, no child | 0.173 | 0.185 | 0.014 | 0.901 |
| Married women, no child | 0.173 | 0.174 | 0.010 | 0.087 |
| Lone mothers | 0.162 | 0.177 | 0.028 | 0.545 |
| Married mothers | 0.138 | 0.141 | 0.011 | 0.334 |
| Partner working | 0.152 | 0.155 | 0.008 | 0.370 |
| Youngest child 0-2 | 0.086 | 0.080 | 0.013 | 0.436 |
| Youngest child 3-5 | 0.127 | 0.119 | 0.016 | 0.503 |
| Youngest child 6-10 | 0.159 | 0.149 | 0.017 | 0.606 |
| Youngest child 11+ | 0.192 | 0.196 | 0.018 | 0.212 |
| Family background: factor 1 | 0.158 | 0.148 | 0.009 | 1.053 |
| Family background: factor 2 | 0.165 | 0.163 | 0.010 | 0.188 |
| Part-time hours | 0.069 | 0.069 | 0.008 | 0.018 |
| Panel B: Impact of benefit reform | | | | |
| Pre-1999: single women, no child | 0.026 | 0.030 | 0.017 | 0.252 |
| Pre-1999: married women, no child | 0.009 | 0.005 | 0.010 | 0.343 |
| Pre-1999: lone mothers | 0.032 | 0.035 | 0.044 | 0.072 |
| Pre-1999: married mothers | -0.030 | -0.034 | 0.012 | 0.330 |
| Post-1999: single women, no child | 0.000 | 0.009 | 0.014 | 0.674 |
| Post-1999: married women, no child | 0.015 | 0.010 | 0.009 | 0.538 |
| Post-1999: lone mothers | -0.006 | -0.014 | 0.025 | 0.340 |
| Post-1999: married mothers | -0.017 | -0.019 | 0.010 | 0.190 |

Notes: Moments in Panel A are average training rates (measured as spending more than 40 hours in training over the last 12 months) among employed individuals with the listed demographic and background characteristics. Moments in Panel B are the deviation in percentage points of each of the family types training rates from average training rates in the period indicated.

Figure C.4: Employment of mothers

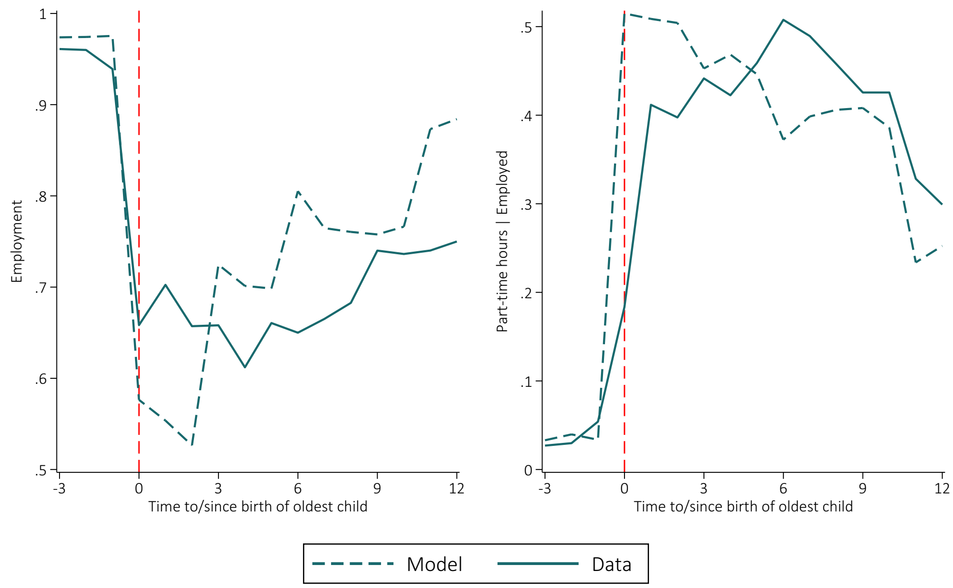



Table C.5: Employment by age

| Moment | Data | Simulated | SE data | Norm. SE diff |
|---|---|---|---|---|
| 20 - 24 years | 0.850 | 0.919 | 0.015 | 4.552 |
| 25 - 29 years | 0.822 | 0.840 | 0.016 | 1.106 |
| 30 - 34 years | 0.794 | 0.806 | 0.018 | 0.641 |
| 35 - 39 years | 0.792 | 0.813 | 0.020 | 0.997 |
| 40 - 44 years | 0.855 | 0.850 | 0.020 | 0.231 |
| 45 - 49 years | 0.895 | 0.865 | 0.020 | 1.470 |
| 50 - 54 years | 0.893 | 0.868 | 0.022 | 1.099 |
| 55 - 59 years | 0.886 | 0.884 | 0.026 | 0.099 |

Notes: Moments are average employment rates (measured as working five or more hours per week) for individuals in each of the age bands indicated.

Table C.6: Training by age

| Moment | Data | Simulated | SE data | Norm. SE diff |
|---|---|---|---|---|
| 20 - 24 years | 0.168 | 0.214 | 0.015 | 2.994 |
| 25 - 29 years | 0.165 | 0.152 | 0.013 | 1.016 |
| 30 - 34 years | 0.152 | 0.158 | 0.013 | 0.449 |
| 35 - 39 years | 0.163 | 0.165 | 0.015 | 0.136 |
| 40 - 44 years | 0.170 | 0.160 | 0.014 | 0.704 |
| 45 - 49 years | 0.176 | 0.161 | 0.018 | 0.887 |
| 50 - 54 years | 0.160 | 0.144 | 0.018 | 0.883 |
| 55 - 59 years | 0.113 | 0.111 | 0.020 | 0.121 |

Notes: Moments are average training rates (measured as spending more than 40 hours in training over the last 12 months) for employed individuals in each of the age bands indicated.

Table C.7: Transition rates from unemployment to employment

| Moment | Data | Simulated | SE data | Norm. SE diff |
|---|---|---|---|---|
| All | 0.251 | 0.275 | 0.017 | 1.372 |
| Single women, no child | 0.408 | 0.301 | 0.048 | 2.220 |
| Married women, no child | 0.187 | 0.218 | 0.035 | 0.869 |
| Lone mothers | 0.212 | 0.283 | 0.019 | 3.805 |

Notes: Moments are average transitions from unemployment (working less than 5 hours) to employment (working at least 5 hours) for individuals with the listed demographic characteristics.

Table C.8: Transition rates from employment to unemployment

| Moment | Data | Simulated | SE data | Norm. SE diff |
|---|---|---|---|---|
| All | 0.050 | 0.051 | 0.003 | 0.528 |
| Single women, no child | 0.027 | 0.025 | 0.003 | 0.852 |
| Married women, no child | 0.086 | 0.148 | 0.016 | 3.809 |
| Lone mothers | 0.082 | 0.080 | 0.007 | 0.339 |
| $w_{t-1}$ below 1st decile | 0.124 | 0.102 | 0.014 | 1.505 |
| $w_{t-1}$ below median | 0.068 | 0.072 | 0.005 | 0.726 |
| $w_{t-1}$ below 9th decile | 0.050 | 0.055 | 0.003 | 1.609 |

Notes: Moments are average transitions from employment (working at least 5 hours) to unemployment (working less than 5 hours) for individuals with the listed demographic characteristics or with wages in the previous period below the indicated quantile.

Table C.9: Log wage at entrance to working life

| Moment | Data | Simulated | SE data | Norm. SE diff |
|---|---|---|---|---|
| Mean | 2.123 | 2.134 | 0.038 | 0.295 |
| Variance | 0.149 | 0.137 | 0.015 | 0.784 |
| Mean: high background factor 1 | 2.137 | 2.151 | 0.045 | 0.316 |
| Mean: high background factor 2 | 2.055 | 2.117 | 0.059 | 1.051 |
| $w_t$ below 1st quartile | 0.250 | 0.223 | 0.043 | 0.625 |
| $w_t$ below median | 0.500 | 0.487 | 0.051 | 0.251 |
| $w_t$ below 3rd quartile | 0.750 | 0.813 | 0.045 | 1.391 |

Notes: Moments are mean and variance of wages at entrance to working life, mean of wages at the entrance to working life conditional on the indicated background characteristic, and the proportion of individuals with wages below specific quantiles of the empirical wage distribution at entrance to working life.

Table C.10: Log wage regression in first differences

$$\Delta \ln(w_t) = \beta_0 + \beta_1 \Delta \ln(\kappa_t + 1) + \beta_2 d_{t-1} + \epsilon_t$$

| Moment | Data | Simulated | SE data | Norm. SE diff |
|---|---|---|---|---|
| Diff in log years of work exp: $\Delta \ln(\kappa_t + 1)$ | 0.189 | 0.235 | 0.019 | 2.431 |
| Lagged training dummy: $d_{t-1}$ | -0.000 | 0.005 | 0.006 | 0.843 |

Notes: Moments are coefficients of the regression shown above, where $\kappa_t$ is the observed years of full-time work experience and $d_{t-1}$ is a dummy for spending more than 40 hours in training over the last year.

Table C.11: Log wage regression on accumulated experience and lagged wages
$\ln(w_t) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 \ln(w_{t-1}) + \beta_4 \ln(1+\kappa_t) + \beta_5 \ln(1+\kappa_{t-1}) + \epsilon_t$

| Moment | Data | Simulated | SE data | Norm. SE diff |
|---|---|---|---|---|
| Constant | 0.424 | 0.400 | 0.032 | 0.773 |
| High background factor 1: $x_1$ | 0.016 | 0.012 | 0.006 | 0.648 |
| High background factor 2: $x_2$ | -0.001 | 0.005 | 0.006 | 0.959 |
| Lagged log wages: $\ln(w_{t-1})$ | 0.802 | 0.810 | 0.010 | 0.828 |
| Log years of work exp: $\ln(1+\kappa_t)$ | 0.174 | 0.221 | 0.055 | 0.836 |
| Lagged log years of work exp: $\ln(1+\kappa_{t-1})$ | -0.139 | -0.192 | 0.048 | 1.103 |
| Variance of $\epsilon_t$ | 0.053 | 0.055 | 0.002 | 0.907 |
| First-order auto-corr of $\epsilon_t$ | -0.011 | -0.014 | 0.001 | 3.587 |

Notes: Moments are coefficients of the regression shown above, where $x_1$ and $x_2$ are dummy variables indicating above median family background factors and $\kappa_t$ is the observed years of full-time work experience. Sample for regression is conditional on being employed last period, since we cannot observe $w_{t-1}$ for unemployed individuals.

Table C.12: Log wage regression on lagged experience, working hours and training
$\ln(w_t) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 \ln(1+\kappa_{t-1}) + \beta_4 1(h_{t-1} = 38) + \beta_5 1(h_{t-1} = 18) + \beta_6 d_{t-1} + \epsilon_t$

| Moment | Data | Simulated | SE data | Norm. SE diff |
|---|---|---|---|---|
| Constant | 1.939 | 1.878 | 0.038 | 1.575 |
| High background factor 1: $x_1$ | 0.059 | 0.062 | 0.022 | 0.135 |
| High background factor 2: $x_2$ | 0.020 | 0.028 | 0.022 | 0.332 |
| Log years of work exp: $\ln(1+\kappa_t)$ | 0.162 | 0.157 | 0.011 | 0.446 |
| Lagged full-time dummy: $1(h_t = 38)$ | 0.241 | 0.284 | 0.025 | 1.709 |
| Lagged part-time dummy: $1(h_t = 18)$ | -0.023 | -0.015 | 0.030 | 0.251 |
| Lagged training dummy: $d_{t-1}$ | 0.089 | 0.120 | 0.013 | 2.433 |

Notes: Moments are coefficients of the regression shown above, where $x_1$ and $x_2$ are dummy variables indicating above median family background factors, $\kappa_t$ is the observed years of full-time work experience, $h_{t-1}$ indicates full-time or part-time working hours last period and $d_{t-1}$ is a dummy for spending more than 40 hours in training over the last year.

Table C.13: Log wage regression on age and family background
$\ln(w_t) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 t + \epsilon_t$

| Moment | Data | Simulated | SE data | Norm. SE diff |
|---|---|---|---|---|
| Constant | 2.109 | 2.229 | 0.041 | 2.949 |
| High background factor 1: $x_1$ | 0.053 | 0.047 | 0.023 | 0.232 |
| High background factor 2: $x_2$ | 0.015 | 0.020 | 0.021 | 0.274 |
| Age: $t$ | 0.114 | 0.078 | 0.009 | 3.995 |

Notes: Moments are coefficients of the regression shown above, where $x_1$ and $x_2$ are dummy variables indicating above median family background factors and $t$ is age in years.

Table C.14: Mean yearly change in log wages given working hours at $t-1$

| Moment | Data | Simulated | SE data | Norm. SE diff |
|---|---|---|---|---|
| Working full time at $t-1$ | 0.030 | 0.021 | 0.002 | 4.313 |
| Working part-time at $t-1$ | -0.016 | 0.009 | 0.006 | 4.227 |
| Not working at $t-1$ | -0.002 | -0.006 | 0.012 | 0.378 |

Notes: Moments are coefficients of regression of mean yearly change in wages on dummies variables indicating working hours last period. Mean yearly change in wages is measured as wages this period minus wages when last observed in employment divided by number of years since last observed in employment. It is therefore observed for any individual who has been employed in at least one previous period.

Table C.15: Other moments in log wages conditional on full-time work

| Moment | Data | Simulated | SE data | Norm. SE diff |
|---|---|---|---|---|
| Mean log wages | 2.603 | 2.597 | 0.011 | 0.568 |
| $w_t$ below 1st decile | 0.100 | 0.102 | 0.006 | 0.438 |
| $w_t$ below 1st quartile | 0.250 | 0.258 | 0.010 | 0.787 |
| $w_t$ below median | 0.500 | 0.523 | 0.013 | 1.793 |
| $w_t$ below 3rd quartile | 0.750 | 0.763 | 0.011 | 1.139 |
| $w_t$ below 9th decile | 0.900 | 0.892 | 0.007 | 1.164 |

Notes: Moments are mean log wages for individuals working full-time (measured as working more than 20 hours a week) and the proportion of individuals with wages below specific quantiles of the empirical wage distribution of full-time workers.

Table C.16: Other moments in log wages conditional on part-time work

| Moment | Data | Simulated | SE data | Norm. SE diff |
|---|---|---|---|---|
| Mean log wages | 2.382 | 2.342 | 0.020 | 2.019 |
| $w_t$ below 1st decile | 0.100 | 0.083 | 0.009 | 1.934 |
| $w_t$ below 1st quartile | 0.250 | 0.212 | 0.015 | 2.431 |
| $w_t$ below median | 0.500 | 0.493 | 0.022 | 0.302 |
| $w_t$ below 3rd quartile | 0.750 | 0.830 | 0.019 | 4.150 |
| $w_t$ below 9th decile | 0.900 | 0.972 | 0.012 | 5.763 |

Notes: Moments are mean log wages for individuals working part-time (measured as working between 5 and 20 hours a week) and the proportion of individuals with wages below specific quantiles of the empirical wage distribution of part-time workers.

Table C.17: Other moments in log wages conditional on training

| Moment | Data | Simulated | SE data | Norm. SE diff |
|---|---|---|---|---|
| Mean log wages | 2.660 | 2.675 | 0.014 | 1.103 |
| $w_t$ below 1st decile | 0.100 | 0.093 | 0.009 | 0.784 |
| $w_t$ below 1st quartile | 0.250 | 0.268 | 0.015 | 1.278 |
| $w_t$ below median | 0.500 | 0.547 | 0.019 | 2.545 |
| $w_t$ below 3rd quartile | 0.750 | 0.764 | 0.016 | 0.887 |
| $w_t$ below 9th decile | 0.900 | 0.871 | 0.010 | 3.071 |

Notes: Moments are mean log wages for individuals in training (measured as spending more than 40 hours in training over the last 12 months) and the proportion of individuals with wages below specific quantiles of the empirical wage distribution of trainees.

Table C.18: Mean log wages by family background

| Moment | Data | Simulated | SE data | Norm. SE diff |
|---|---|---|---|---|
| High background factor 1 | 2.552 | 2.554 | 0.014 | 0.082 |
| High background factor 2 | 2.578 | 2.571 | 0.015 | 0.487 |

Notes: Moments are mean log wages for individuals with the indicated background characteristics.