Economic Implications of Inattention and Altruism

Jamie Hentall MacCuish

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Department of Economics

University College London

October 18, 2023

Declaration of Authorship

I, Jamie Hentall MacCuish, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

[redacted]

Statement of Conjoint Work

Note on the joint work in Jamie Hentall MacCuish's thesis "Economic Implications of Inattention and Altruism":

The chapter " Parental Altruism and Transfers of Time and Money - A Lifecycle Perspective" is co-authored work between Helena Uta Bolt, Eric French, Jamie Hentall MacCuish and Cormac O'Dea. Each author contributed equally. The chapters, "Costly Attention and Retirement" and "Disentangling Risk and Intertemporal Preferences with Costly Information Acquisition" are single-authored by Jamie Hentall MacCuish.

Signed:

[redacted]

Professor Fabien Postel-Vinay, Primary Supervisor

Abstract

This thesis explores the economic implications of increasing the psychological realism used in economic models. It does this along two dimensions. The first two chapters investigate the economic implications of limited attention, and the third those of altruism between parent and child. The first chapter shows that incorporating limited attention into a model of retirement explains both observed mistaken pension beliefs and the large drop in employment at pension eligibility age in spite of weak economic incentives to stop working precisely at that age. The second chapter documents a previously undocumented attribute of inattention: it can separate risk and time preferences. Standard time-additive expected-utility theory famously implies that the coefficient of relative risk aversion and the elasticity of intertemporal substitution are inverse reciprocals. This complete lack of independence between two key risk and time parameters is at the heart of some of the most famous puzzles in finance. This chapter shows costly attention can separate them because it introduces a new reason for agents to dislike risk unrelated to the curvature of utility that determines the intertemporal elasticity of substitution: the utility cost of reducing uncertainty. The third chapter jointly studies different channels through which altruistic parents invest in their children - by spending time with them to foster cognitive skills, by paying for their education, and by making monetary transfers.

Impact Statement

This thesis impacts policy and academic discourse. The academic impact is theoretical, methodological, and quantitative.

Chapter 3 provides a novel theoretical insight: costly attention can separate risk and time preferences in a time-additive expect-utility framework. Time-additive expect-utility remains the de-facto economic model of human behaviour and is widely believed to remove independence between two key risk and time preferences parameters. This chapter shows that allowing for costly attention this is not the case. As it provides a novel insight into the dominant paradigm in economics the potential academic impact of this chapter is far-reaching. Since this lack of independence is at the heart of the equity premium and risk-free rate puzzles, the implications are not restricted to the academic but extend to real financial markets.

Chapter 2 makes a methodological contribution by developing a solution method for dynamic rational inattention models with endogenous heterogeneous beliefs. Previously rational inattention could handle either dynamics or belief heterogeneity, my method handles both. Dynamic models of rational inattention, where people choose how much to learn, naturally give rise to differing beliefs, previous work has needed to make highly specific assumptions to avoid belief heterogeneity. Hence this methodological contribution has a potentially far-reaching academic impact.

Chapters 2 and 4 contribute to the academic literature by impacting our quantitative understanding. Chapter 2 shows mistaken pension beliefs arising from costly attention contribute significantly to explaining the puzzling drop in employment at the pension age. Chapter 4 uses a unique dataset following a single cohort from birth to estimate the extent of parental altruism and impacts our understanding of this deep parameter.

Chapters 2 and 4 are the most directly policy-relevant. Chapter 2 studies the implication of increasing the state pension age and Chapter 4 answers the question of why rich parents tend to have rich children.

Acknowledgements

Firstly, I would like to thank my family because a long endeavour like a PhD thesis seems the perfect opportunity to acknowledge all the personal support I have received. Thank you to my wife, Mary, for supporting and putting up with me throughout these long years. I know you made many sacrifices too. I see them, I appreciate them, and I love you. Then I would like to thank my parents for all the encouragement they gave me throughout my life. Few things could be more tautologically true, but I would not be where I am without you. Thank you to my mother for instilling in me a sense of right and wrong, and a particular thank you to my father, my own personal giant whose shoulders I stood on. Thanks for all the copy editing, Dad. Finally, I would like to thank Eleanor, you didn't make this any easier, but you made it more meaningful.

I have received tremendous academic support to make it here as well, more than can possibly be listed here, and so I will restrict myself to three people who provided the most support. Firstly, my primary supervisor Fabien Postal-Vinay, who always listened, encouraged and engaged with what I asked him, thank you. Secondly, my advisor Eric French for opening opportunities for me and encouraging me to "keep at it" throughout. Finally, my co-author Cormac O'Dea "the nicest person in economics", for going above and beyond in giving his time and support. Thank you for pulling me out of a few dark spots.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis enriches the psychology of the agents used in economic model along two dimensions allowing for: one, the scarce and costly nature of attention and, two, the altruistic ties between parent and child. Over the following chapters, it does so by comparing the predictions of enriched models of human behaviour to observational data and then derives the policy implications of these enrichments. Chapters 2 and 3 consider the implications of costly attention, and Chapter 4 those of altruism.

One important consequence of inattention is that households cannot take advantage of features of benefits they are unaware of. This impact of inattention has been very well documented in terms of straightforward take-up of benefits but if we are unaware of features of benefits this will impact our behaviours in more ways, be that an inability to leverage the insurance value of benefits or apparently surprising behaviour when we update our information. Chapter 2 explores these consequences of inattention in an area where misbeliefs offer strong evidence of their importance: the pension benefit system. More specifically it asks whether costly attention to an uncertain and potentially changeable pension policy can explain both observed mistaken beliefs and a known puzzle in the retirement literature: the large drop in employment at pension eligibility ages despite weak economic incentive to stop working precisely then.

To investigate these questions in Chapter 2 I use a recent reform to the female State Pension Age (SPA) in the UK to provide necessary variation to estimate both the employment drop at SPA and the extent of objective uncertainty about the SPA. The data used to do this comes from the English Longitudinal Study of Aging (ELSA) a rich panel survey dataset linked to administrative records that contains good information about people's pension beliefs. This allows me to estimate a structural model of rational inattention in which households are free to learn about a changeable SPA any way they want in exchange for paying a utility cost to receive more precise information about the SPA. I find that despite only a small cost of attention being required to explain observed mistaken beliefs, large employment consequences follow because those close to retirement are close to their participation margin. As a result, the model can explain a significant proportion of the excessively large drop in employment at SPA.

Chapter 3 steps back from data to investigate some theoretical implications of costly attention whilst still considering its ability to explain observed phenomena. Standard time-additive expected-utility theory famously implies that the coefficient of relative risk aversion and the elasticity of intertemporal substitution are inverse reciprocals. This complete lack of implied independence between two key parameters determining risk and time preferences is at the heart of two of the most famous puzzles in the finance literature, the equity premium puzzle and the risk-free rate puzzle, because observed risk aversion and intertemporal substitution do not fit this tight relationship. Hence, although theoretical in nature, this inverse reciprocity has important implications for understanding observed behaviour. Chapter 3 shows, at least for two-period two-state-of-the-world models costly attention can break this complete lack of independence. The intuition behind these results is that costly attention introduces a new reason for agents to dislike risk unrelated to the curvature of utility over consumption because risk, or uncertainty, directly and negatively enters their utility function reflecting the cognitive cost of learning. This chapter shows that this new reason for agents to dislike risk does not affect their preference for intertemporal substitution and so breaks the problematic inverse reciprocal relationship. I then show that, at least for stylised models, this property allows you to simultaneously solve the equity premium and risk-free rate puzzles.

Chapter 4 then moves onto parental altruism towards their offspring. Parents have multiple ways of investing in their children. During childhood, they can spend time with their children, thus fostering their cognitive skills (Cunha et al. (2006), Heckman and Mosso (2014)). They can also pay for their child's education (Belley and Lochner (2007), Abbott et al. (2019)). Lastly, they can give cash transfers (Castaneda et al. (2003), De Nardi (2004)). Using data from the National Child Development Study (NCDS) we estimate a dynamistic lifecycle model in which one generation cares about the next altruistically and can impact the welfare of the second generation by giving them direct cash transfers, sending them to further education, and investing time in them as children.

We find only modest dynamic complementarity between early-time investments in children and later-time investments. However, like Delaney (2019) and Daruich (2018), we find substantial complementarities between final childhood ability at age 16 and education in wages. Among men with a college education, a one standard deviation increase in cognitive ability leads to an additional 19 % in wages. Among those with the lowest level of education, this premium, at 9%, is much smaller. As a result, high-ability individuals are more likely to select into education than their low-ability counterparts. This dynamic complementarity, in combination with self-selection into education, is a key mechanism that perpetuates income inequality across generations. High-income households, who have more resources to send their child to college, have higher returns to investing in their child's ability than their low-income counterparts; thus they invest more in their children. Second, we find that more than a quarter (28%) of the variance of lifetime wages can already explained by the characteristics of the family before an individual is even born. By the time individuals are 23, their characteristics can explain up to 62% of the variance in lifetime wages. Thus, more than half of the lifetime variability in wages is realized by age 23.

Chapter 2

Costly Attention and Retirement

2.1 Introduction

Most people are confused about pensions. One example is they frequently mistake the age from which they can receive pension benefits by multiple years, as seen in Figure 2.3 . Mistaken pension beliefs are so common they may seem unsurprising. They are, however, incompatible with standard complete information models. Widespread mistaken beliefs about financially important policies suggest incomplete information resulting from information frictions or cognitive limitations.

Ignoring information frictions limits us in understanding policy uncertainty's impact on people's decisions. People are not only unsure how policy may change, as complete information suggests, often they do not know current rules. These mistaken beliefs are easier to rationalise if we acknowledge that government policy is objectively uncertain. Governments change policies making people's mistakes about them unsurprising. How this interplay between objective policy uncertainty and subjective mistaken beliefs impacts retirement is the focus of this paper.

Specifically, I embed rationally inattentive households in a lifecycle model which generates mistaken beliefs and helps explain the excess employment sensitivity puzzles. This puzzle, documented in multiple countries [1], is that people exit employment at pension-eligibility ages whilst benefit systems offer only weak incentives to stop working precisely then. Mistaken beliefs increase the wealth

[1]For example in the US by Behaghel and Blau (2012), in Germany by Seibold (2021), and in Switzerland by Lalive et al. (2017)

and uncertainty shocks received from the resolution of pension uncertainty upon reaching these eligibility ages. These increased shocks help explain the excessive employment reaction. Costly attention stands out from alternative explanations of the puzzle by also explaining stated beliefs.

I use recent reforms that increased the UK female State Pension Age (SPA) to identify the effects of the SPA on employment. UK institutional features rule out most explanations for exits from employment at pension eligibility ages: forcing an employee to retire due to age is illegal, and state pension receipt is not conditional on employment status. Liquidity constraints provide a motive to retire at the SPA since the inability to borrow against pension income prevents intertemporal substitution. High-wealth women, however, also exit employment at the SPA, rendering this explanation, at best, incomplete. Costly attention penalises acquiring the information to optimally substitute over time, creating a new barrier to intertemporal substitution.

This paper first documents the pertinent facts concerning mistaken beliefs and excess employment sensitivity. Next, it builds a model with information frictions, in the form of costly attention, that accounts for these facts. The model incorporates costly attention, modelled using rational inattention (e.g. Sims, 2003), to the stochastic SPA, capturing objective policy uncertainty, into a dynamic life-cycle model of retirement (e.g. French, 2005). This generates mistaken beliefs that help explain retirement choices.

Endogeneity of beliefs drives the relationship between retirement and mistaken beliefs but complicates the model by introducing a state (beliefs) and a choice (learning strategy) that are both high-dimensional. To the best of my knowledge, solving a structural rational inattention model with endogenous heterogeneous beliefs is a first. I develop a general-purpose solution method for dynamic rational inattention models with endogenous heterogeneous beliefs that overcomes these complications. This method extends the algorithmic recommendations of Armenter et al. (2019) to dynamic models using theoretical results from Steiner et al. (2017) and overcomes the computational complications using the sparsity proven to be a general property of rational inattention models by Caplin et al. (2019).

The English Longitudinal Study of Ageing (ELSA), a panel survey, provides data to study mistaken beliefs and their impact on employment since it contains self-reported and true SPAs. Reform-affected women are substantially mistaken about their SPA less than four years from it, most being out by over a year. These mistakes predict employment responses: women who are more mistaken in their late 50s about their SPA have a smaller response upon reaching it in their early 60s. This pattern suggests endogenous learning: women who do not care about the SPA neither learn nor respond to it.

I estimate the model using two-stage simulated method of moments, targeting asset and employment profiles. Policy uncertainty and costly attention increase the employment response to the SPA compared to a complete information baseline, explaining 30%-74% of the shortfall. By exploiting the SPA belief data, I separately identify beliefs and preferences: a solution to the belief-preference identification problem (e.g. Manski, 2004) that avoids the common need to assume people are well-informed. The mean household is willing to pay £15.37 to learn today's SPA, so estimated attention costs are small, which is in line with other evidence (e.g. Chetty, 2012). Despite small attention costs, information letters about current pension entitlement pass cost-benefit analysis as their marginal cost is close to £1. Large employment changes result from small attention costs because people near retirement are close to their participation margin.

Pension eligibility ages are seen as a key to increasing old-age labour force participation, a common policy objective (see Landais et al., 2021). Relative to complete information, costly attention increases the employment response *at* the SPA, so one may naively conclude it makes the SPA a better tool to achieve this goal. The opposite is often true. Policy experiments, comparing increases in employment resulting from increases in SPA in versions of the model with and without information frictions, show costly attention increases the employment response *at* the SPA by intertemporally shifting part of the informed agent's employment response forward but can decrease the overall response. Informed agents increase labour supply immediately; those subject to costs of learning, being less informed, do not respond until nearer their SPA. Ignoring costly attention overstates the SPA's effectiveness at increasing old age employment by up to 27%. This illustrates another reason to send policy information letters: informed individuals' behaviour is more predictable.

An extension throws light on another puzzle: 87% of people observed claim the state pension as soon as eligible despite an actuarially advantageous benefit increase for deferring. It introduces a claiming decision, policy uncertainty over the adjustment for deferring, and a cost of learning about this adjustment. Together these create a new incentive to claim that addresses this puzzle: claiming removes the need to pay attention to policy, thus increasing the proportion of early claimers helping to explain the deferral puzzle.

The paper is structured as follows. Section 2.2 reviews literature. Section 2.3 outlines institutional context and data and presents descriptive and reduced-form analysis. Section 2.4 presents the model, starting with a standard model of complete information and then building in objective pension policy uncertainty and a cost of attention to this uncertain policy. Section 2.5 discusses the solution method. Section 2.6 discusses estimation and Section 2.7 model fit and implications. Section 2.8 presents the extension addressing the deferral puzzle. Section 2.9 concludes.

2.2 Related Literature

The main contribution of this paper is embedding costly attention into a lifecycle model of retirement to explain the excess employment sensitivity puzzle whilst accommodating observed beliefs. To do this, it builds on two literatures: dynamic lifecycle models of retirement and rational inattention, but it is also deeply connected to works documenting excess employment sensitivity and pension beliefs. The most relevant papers from each strand, and from the wider literature, are reviewed below and the contributions to each explained.

Dynamic lifecycle models of retirement Dynamic lifecycle models of retirement have a history stretching back to Gustman and Steinmeier (1986) and Burtless (1986), and this paper includes the features this literature identifies as key that are relevant in the UK. Computational limitations led early works to ignore uncertainty and borrowing constraints, but more recent work finds them crucial. Rust and Phelan (1997) introduced uncertainty into a dynamic lifecycle model along with a formulation of incomplete markets that ruled out all savings. French (2005) reintroduced saving whilst maintaining incomplete markets through a borrowing constraint, alongside other innovations such as a fixed cost of work to help explain the retirement phenomena. Gustman and Steinmeier (2005) allow for time preference heterogeneity; van der Klaauw and Wolpin (2008) model Medicare; and French and Jones (2011) add uncertain medical expenses onto these innovations. Much of this literature is US-focused, and some of its concerns are not relevant in the UK context I study (e.g. medical insurance). The key features I include from this literature are uncertainty, borrowing constraints, and individual heterogeneity, and the most similar paper is O'Dea (2018) who estimates a model of males in the UK.

Rational inattention This paper relies on recent theoretical advances from the rational inattention literature to model costly attention and contributes back to this literature a novel application and quantitative techniques. Rational inattention traces its heritage back to Sims (2003). Initially used to add costly attention to macroeconomic models (e.g. Luo, 2008; Maćkowiak and Wiederholt, 2009, 2015)), recently, its domain of application has expanded. In decision theory, Caplin and Dean (2015) develop a revealed preference test for rational inattention; in game theory Ravid (2020) analyses ultimatum bargaining with rational inattentive buyers; and in a field experiment, Bartoš et al. (2016) explain job market discrimination. A series of papers starting with Matějka and McKay (2015) analyse general classes of models with rationally inattentive agents. They solve static discrete choice models with rationally inattentive agents. Steiner et al. (2017) extends these results to dynamic discrete choice models, which is key to solving the dynamic rational inattention model with the endogenous heterogeneous beliefs that result from embedding costly attention into a lifecycle model. Turning the theoretical solutions of Steiner et al. (2017) into a practical solution methodology for rich quantitative models is a contribution of this paper making it the first, to the best of my knowledge, to solve a model with endogenous heterogeneous beliefs. Key to bridging this gap between elegant theory and practical solution methodology are two papers. Caplin et al. (2019) show rational inattention generically implies consideration sets, implying solutions are sparse and provide conditions for sparsity which help to reduce computation. When sparsity does not provide a shortcut, I follow Armenter et al. (2019) in using sequential quadratic programming to solve the within-period rational inattention problem. By applying rational inattention to rich micro data, this paper joins a frontier in the literature (e.g. Macaulay, 2021; Porcher, 2020) and extends it by allowing for endogenous heterogeneous beliefs, which those papers avoid by assuming complete information sharing.

Excess employment sensitivity Employment being more sensitive to statutory pension ages than standard models predict is a puzzle observed in multiple countries; this paper provides evidence for it in the UK. Lumsdaine et al. (1996) document the excess employment sensitivity puzzle in the US, and much of the lifecycle models of retirement literature was dedicated to explaining it. The consensus was that liquidity constraints explained the retirement spike at the 62 early retirement age, and Medicare eligibility explained the spike at the 65 full retirement age (Rust and Phelan, 1997; French, 2005; Gustman and Steinmeier, 2005; French and Jones, 2011). The ability to test these explanations was limited as the US early and full retirement ages remained unchanged until 2004, when the full retirement age increased, providing the variation to estimate its impact on employment. Larger effects were detected than predicted by standard models (Mastrobuoni, 2009) and part of the age 65 spike followed the full retirement age despite Medicare eligibility remaining at 65 (Behaghel and Blau, 2012), undermining Medicare eligibility as its sole cause. [2] Ageing populations forced other governments to increase statutory pension ages with similar results: increases in pension age induce larger labour supply response than standard models predict. This is documented in Austria by Manoli and Weber (2016), in Germany by Seibold (2021), and in Switzerland by Lalive et al. (2017). I document an excess employment sensitivity puzzle in the UK by using the female state pension age reform building on the work of Cribb et al. (2016), principally by

[2] These insights were not found incorrect, rather post-reform data did not support them completely explaining employment sensitivity.

using richer data to rule out potential standard complete information explanations for the employment response.

Belief data The use of belief data is growing (Koşar and O'Dea, 2022), and pension beliefs are an interesting case as mistakes are easy to detect; by using mistaken pension beliefs to identify attention costs, this paper contributes to this growth. The earliest papers to investigate pension knowledge (e.g. Bernheim, 1988; Gustman and Steinmeier, 2001) look at individual forecast errors about the level of pension benefit. Forecast errors conflate misprediction of future rule changes with mistaken beliefs about current policy, and disentangling them requires information on people's knowledge of current pension rules. Manski (2004) documents precisely one such study, that finds much individual uncertainty about their benefits is explained by a lack of understanding of current Social Security formula. Rohwedder and Kleinjans (2006) study the dynamics of forecast errors and find they shrink as individuals approach retirement, providing evidence of learning. Crawford and Tetlow (2010) look at self-reported SPAs and find large errors common; Amin-Smith and Crawford (2018) document these mistakes are predictive of the employment response to the SPA. This paper finds similar patterns to Crawford and Tetlow (2010) and Amin-Smith and Crawford (2018), prevalent mistaken beliefs predictive of labour supply, together with a similar pattern of learning to that found by Rohwedder and Kleinjans (2006). I use these patterns to identify attention which represents a novel use of belief data, since most papers use belief data to identify parameters which individuals hold private information about whilst maintaining the assumption of complete information.

Wider Literature Policy uncertainty plays an important role in this paper, and so it relates to others investigating policy uncertainty, such as Baker et al. (2016). Of particular relevance, Luttmer and Samwick (2018) measure the welfare cost of individuals' perceived uncertainty about their social security benefits. This paper also belongs within the tradition of behavioural public economics (Bernheim and Taubinsky, 2018; Chetty, 2015), some of the most related works from this literature are those that also consider the implications of limited information such as Fuster et al. (2022), Lockwood (1991), or Taubinsky and Rees-Jones (2018).

2.3 Institutional Context, Data, and Analysis

Explaining the puzzlingly large employment response to the UK state pension age (SPA) is the goal of this paper. It identifies this response to the SPA using a reform to the female SPA which Section 2.3.1 outlines while explaining what makes it illuminating of excess employment sensitivity. Section 2.3.2 discusses the data. Sections 2.3.3 - 2.3.4 provide descriptive and reduced form analysis, Section 2.3.3 documenting the excess employment sensitivity puzzle, and Section 2.3.4 documenting erroneous beliefs about the SPA as well as their relationship to employment sensitivity to the SPA.

2.3.1 Institutional Context

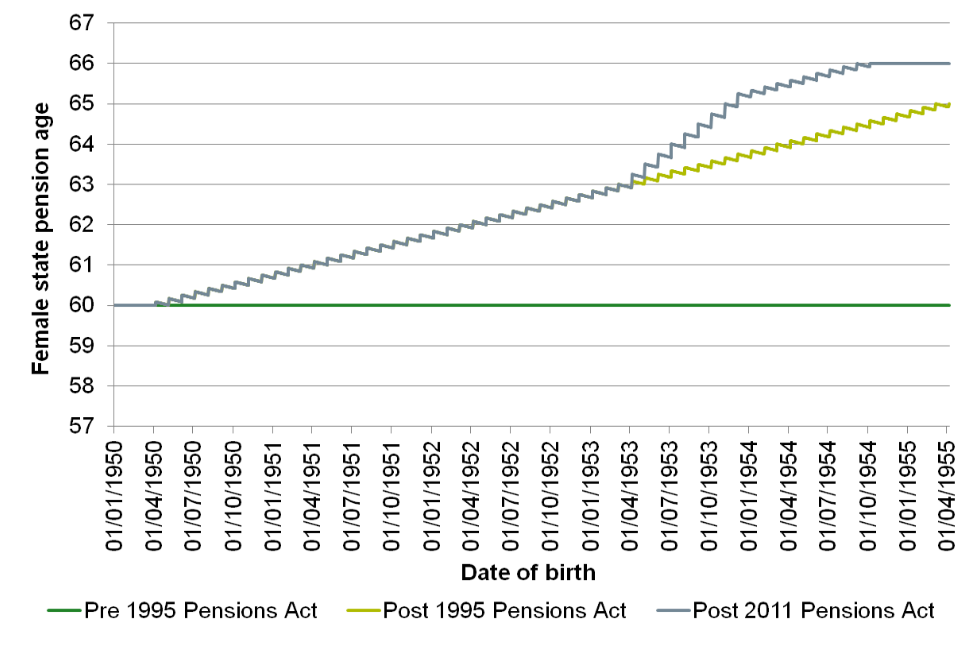
The UK State Pension Age (SPA) is the earliest age at which retirement benefits, known as the state pension, can be claimed. In other words, it is the Early Retirement Age of the UK pension system. The SPA is the sole focal age of the state pension system. Deferral of receipt does increase the generosity of the benefit; however, during the period considered, this was without a cap on the deferral duration and so did not imply an effective Full Retirement Age. [3]

The UK State Pension came into force in 1948, with a SPA of 65 for men and 60 for women. This remained unchanged until the Pensions Act 1995 legislated for the female SPA to gradually rise from 60 to 65, one month every two months, over the ten years from April 2010. The Pensions Act 2011 accelerated the rate of change of the female SPA from April 2016 so that it equalises with men's by November 2018. It additionally legislated an increase to both the male and female SPA to 66 years, phased in between December 2018 to October 2020. Figure 2.1 summarises how these changes affect women in different birth cohorts.

This UK SPA reform is a convenient context to study the excess employment sensitivity puzzle, as many possible explanations for labour market exits at the early

[3] Despite a generous actuarial adjustment, deferral was rare, implying another puzzle, discussion of which is deferred to Section 2.8.

Figure 2.1: SPA by Date of Birth under Different Legislation



Note: State Pension Age for women under different legislation. Source: Pensions Act 1995, schedule 4 (http://www.legislation.gov.uk/ukpga/1995/26/schedule/4/enacted); Pensions Act 2007, schedule 3 (http://www.legislation.gov.uk/ukpga/2007/22/schedule/3); Pensions Act 2011, schedule 1 (http:/www.legislation.gov.uk/ukpga/2011/19/ schedule/1/enacted).

retirement age are ruled out. Firstly, firms cannot force employees to retire solely based on age: this would be classed as illegal age discrimination under UK law[4]. So, firm-mandated retirement cannot explain the sensitivity of employment to the SPA. Secondly, the state pension is not conditional on employment status. Individuals may claim the state pension and continue working, and many do. Thirdly, the UK pension system does not provide major tax incentives to exit the labour market at the SPA. Unlike the US system, there is no earnings test[5], and although the state pension is taxable income, a component of income tax, called National Insurance contributions, is removed upon reaching the SPA[6].

[4]The Equality Act (2006) banned mandatory retirement below age 65, which is greater than the highest SPA considered in this paper. The Equality Act (2010) extended this ban to all ages with some exceptions discussed in appendix 2.A

[5]An earnings test is a feature of some social security systems that penalise working whilst claiming retirement benefits. Those unfamiliar with it need not worry as it is not a feature of the UK system; it is only mentioned to reassure those familiar with systems which include an earnings test.

[6]Cribb et al. (2013) estimate changes to an individual participation tax rate at SPA and find they do not predict the employment response at SPA.

These three facts imply the state pension is essentially an anticipatable increase in non-labour income with the SPA its eligibility age. As the reform was announced in 1995 and began in 2010, the income change was anticipatable with a horizon of at least 15 years. Hence, the puzzle is not that employment responds to the SPA reform but that the response concentrates at the SPA when so much forward notice was given. In a standard life-cycle model, with complete information and forward-looking agents, labour supply responses do not concentrate at anticipatable income changes unless liquidity constraints prevent agents from smoothing intertemporally. So, these three features remove incentives to exit the labour market at the SPA for all but the liquidity constrained because the inability to borrow against future pension income forces these people to wait for this anticipatable additional income to decrease labour supply [7]. Accordingly, I treat the ability of liquidity constraints to explain the sensitivity of employment to the SPA as synonymous with the ability of standard models of complete information to do so, and Section 2.3.3 focuses on ruling out this explanation.

2.3.2 Data

To study the labour supply response to the State Pension Age (SPA), a dataset that samples a large number of older individuals is required. To investigate the reasons for the response, rich microdata are also needed. The English Longitudinal Study of Ageing (ELSA) is the UK[8] dataset that strikes the best balance between these two desiderata, and so it forms the principal data source for this paper.

ELSA is a panel dataset at a biennial frequency containing a representative sample of the English population aged 50 and over. It is modelled on the US Health and Retirement Study (HRS) and contains rich microdata about multiple aspects of respondents' lives. Particularly relevant here, ELSA contains detailed data on labour market circumstances, earnings, and the amount and composition of asset holdings. From wave 3 onwards, ELSA collects information on people's knowledge of their SPA. Having such information is, of course, crucial to investigating

[7] A market accepting future pension benefits as collateral does not exist. Such loans are not illegal; they are just not observed.

[8] Technically, ELSA (Banks et al., 2021) only covers England and Wales.

the role played by erroneous beliefs in the excess sensitivity puzzle. ELSA requests National Insurance numbers (equivalent to a US Social Security number) and permission to link to administrative records from respondents, 80% of whom consent. Additionally, survey data on health, education, and family are instructive of motivations for retirement.

ELSA waves 1 (2002/03) through to 7 (2014/15) capture those affected by the 1995 pension age reform reaching SPA; hence I take the sample used for analysis and estimation from these waves. As this paper is concerned with the reform to the female SPA, males are dropped from the sample, except when estimating a spousal income process when females are dropped. The only selection criteria for the female sample are that I drop women aged over 75 and under 55; this contains 24,114 observations of 7,201 women. The implementation of the female SPA reform began in 2010, and so the first wave of ELSA after the implementation of the female SPA reform is wave 5. Earlier waves are important to control for pre-trends and to increase power when estimating model inputs. The oldest women affected by the reform were born on 6 April 1950. Having older cohorts is important as a control group and also informative when estimating model inputs.

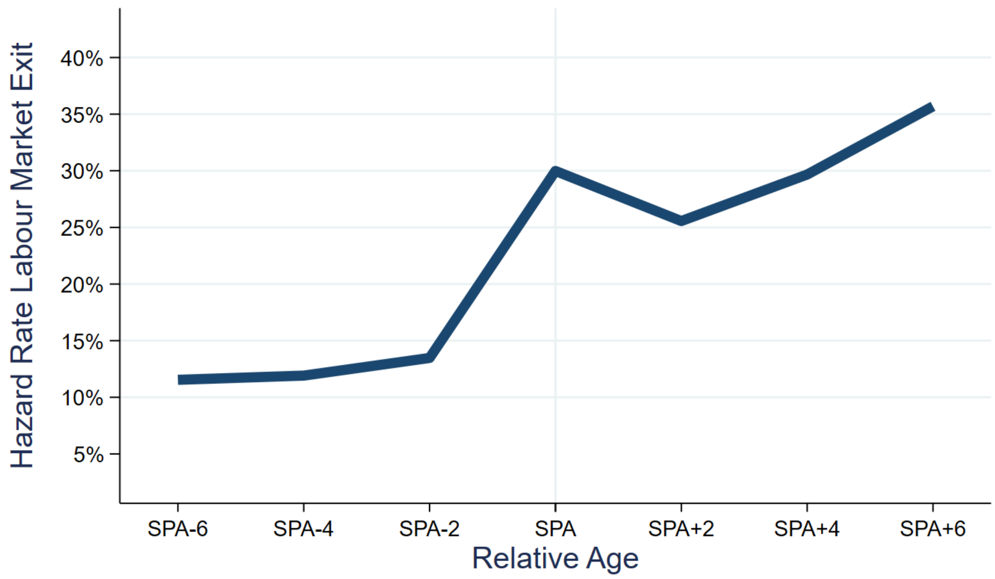
2.3.3 Excess Employment Sensitivity

Employment being more sensitive to official retirement ages than implied by incentives is a puzzle documented in multiple countries (see Section 2.2). This section presents evidence of this puzzle in relation to the UK SPA. Liquidity constraints being essentially the only standard complete information mechanism that could generate this sensitivity to the SPA (see Section 2.3.1), I focus on demonstrating that liquidity constraints alone do not explain the puzzle.

Figure 2.2 captures the fundamentals of the excess employment sensitivity puzzle. It plots the average hazard rate of exiting employment at an age from the SPA. A large jump in exits at the SPA is observed. By adjusting the SPA at the monthly cohort level, the female SPA reform allows more careful identification of the employment response to the SPA.

To do this, I build on Cribb et al. (2016), who use this reform to identify the

Figure 2.2: Fraction exiting labour employment



Note: Pooled average faction exiting employment market at ages relative to the SPA. Data was plotted at two yearly intervals due to the biennial frequency of ELSA waves.

labour supply response to the SPA and find it significant. They argue against constraints driving their results because, whilst homeowners are less likely to be constrained than renters, the effects of the SPA on their employment are indistinguishable. The focus of Cribb et al. (2016) was documenting the response to the SPA rather than explaining it, and homeownership is a coarse proxy for being liquidity constrained, as equity in one's own home is illiquid. So, I use the richer data in ELSA to investigate motives for the employment response to the SPA, in particular, ruling out liquidity constraints. This results in the most detailed evidence to date of an excess employment response to the UK SPA.

The main estimating equation used in this section is presented in equation 2.1. It is a regression of the probability of employment (y_{it}) on: an indicator of being below the SPA; a set of quarterly age, cohort, and date dummies; and a vector of controls[9] leading to the following specification:

[9]The full list of controls used is: a full set of marriage status, years of education, education qualifications, and self-reported health dummies; partner's age; partner's age squared; dummies for partner eligible for SPA; and assets of the household.

Table 2.1: Effect of SPA on Employment: Heterogeneity by Wealth

	(1)	(2)	(3)
Below SPA	**0.080**	**0.061**	**0.114**
s.e	(0.0183)	(0.0215)	(0.0283)
p=	.000	.006	.000
Below SPA×(NHNBW.>Med.)			**-0.053**
s.e			(0.0354)
p=			.137
Obs.	23,638	6,930	23,638
Cohorts	132	90	222

Notes: Column (1) shows the results of running the two-way fixed effect specification in 2.1 as a random-effects model with controls used: a full set of marriage status, years of education, education qualifications, and self-reported health dummies; partners age; partners age squared; the aggregate unemployment rate during the quarter of interview; dummies for partner eligible for SPA, and for being one and two years above and below SPA; and assets of the household. Column (2) repeats this regression on the subsample with above median Non-Housing Non-Business Wealth (NHNBW) in the last interview before their SPA. Column(3) tests whether the different treatment effects observed in columns (1) and (2) are different by introducing an interaction between being below the SPA and having above-median NHNBW.

$$Pr(y_{it}) = \alpha \mathbb{1}[age_{it} \leq SPA_{it}] + \sum_{c \in C} \gamma_c \mathbb{1}[cohort_i = c] + \sum_{a \in A} \delta_a \mathbb{1}[age_{it} = a] + \sum_{d \in D} \kappa_d \mathbb{1}[date_{it} = d] + X_{it}\beta \quad (2.1)$$

This form assumes cohort-and-date-constant age effects and age-and-date-constant cohort effects, and cohort-and-age-constant age effects. Given these assumptions, which are just a rephrasing of the parallel trends assumption, the parameter α is a difference-in-difference estimator of the treatment of being below the SPA. I test this parallel trends assumption by interacting the fixed effects and the Wald test fails to reject the null that these interactions are zero (p = 0.9451). This treatment is administered to all, but variation in the duration of treatment is induced by the reform.

Column 1 of Table 2.1 presents the results of estimating equation 2.1. I find a 0.080 increase in the probability of being in work from being below the SPA significant at the 0.1% level.

To address the question of whether liquidity constraints can explain this treatment effect, I restrict to the subsample of women from households with above median assets and repeat the analysis. Specifically, I restrict to those with above median non-housing non-business wealth (NHNBW) in the wave before they reached their SPA, as this is when the resources to smooth labour supply affect their reaction to the SPA. [10] This generates a cut-off of £29,000. The objective of this cut-off is to restrict to a group whose retirement choices are unlikely to be affected by the liquidity constraint. Given the SPA was reformed in monthly increments, and equation 2.1 controls for quarterly age and cohort fixed effects, an individual's control is someone born in the same year and quarter but a few months older past the SPA. This narrow time window makes arguing against liquidity constraints easier: women with over £29,000 in NHNBW seem unlikely to need to wait 1-3 months for the state pension to stop working. The results are in column 2 of Table 2.1. For this subpopulation, we find a treatment effect of 0.061, similar in size to results for the whole population and significant at the 1% level.

Column 3 of Table 2.1 encapsulates columns 1 and 2 in a single regression by fully interacting specification (1) with an indicator of being below the SPA and being in the subpopulation of specification (2). The interaction term is not significant at any reasonable level, indicating that the treatment effect is not significantly different between those with above and those with below-median assets. I summarise the excess employment sensitivity puzzle by the results in columns (1) and (2) and use these as auxiliary models the structural model aims to replicate.

Appendix 2.A contains robustness including restricting to more liquid assets categories and different functional forms such as dropping controls (to address bad control concerns) and having the labour supply response to the SPA vary continuously with assets. All of these specifications lead to the conclusion that, although assets matter for the labour supply response to the SPA, the effect is not strong enough for liquidity constraints to explain away the response. Appendix 2.A also considers whether factors, neglected for brevity in this section like health, private

[10] NHNBW is all wealth excluding their primary residence and personally owned business. This is an asset categorisation from Carroll and Samwick (1996). In appendix 2.A I repeat the analysis using the most liquid category from that paper VLA.

Table 2.2: Placebo Tests

	One Year Below SPA	Two Years Below SPA
Placebo Test Coefficient	**-0.013**	**-0.018**
s.e	(0.0251)	(0.0185)
$p=$	.614	.310
Obs.	7,440	7,440
Cohorts	72	72

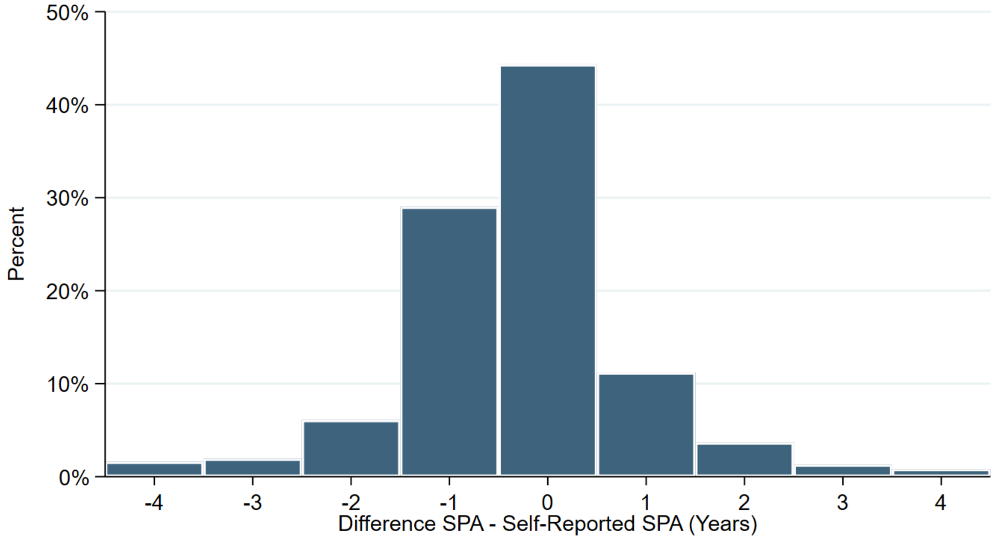
Notes: A placebo tests for a violated parallel trends assumption. I drop observations over SPA and replace the treatment with an indicator for one or two years below SPA are shown.

pension, and joint retirement, can explain the excess employment sensitivity puzzles and finds they cannot. The basic reason is that although they are important for labour supply, the SPA does not correlate with a significant change in any of them.

The traditional difference-in-difference approach used in this section makes strong assumptions about treatment effect heterogeneity. In appendix 2.A I relax these assumptions using the modern imputation approach to difference-in-difference estimation of Borusyak et al. (2021). Allowing for arbitrary heterogeneity produces estimates supportive of a static treatment effect at the SPA assumed in this section and also of average treatment effects in line with those estimated in this section. I conclude it is reasonable to give a causal interpretation to the treatment effects estimated in this section.

What follows, however, does not rest on the causal nature of these estimates. I use these regression results as an untargeted auxiliary model to a structural model, so what is important is the model's ability to replicate them not whether they are causal. What follows does depend on the reader finding these results puzzling, at least as far as standard complete information models are concerned. The placebo test results in Table 2.2, support the idea something is puzzling about the SPA. It contains the results of dropping observations over SPA and replacing the treatment in equation 2.1 with indicators of being one or two years under SPA; unlike the treatment, these coefficients are negative and insignificant. So the results in this section are detecting something specific to the SPA, which is puzzling for those with significant liquid wealth.

Figure 2.3: Mistaken SPA Beliefs of Women Subject to the Reform at Age 58



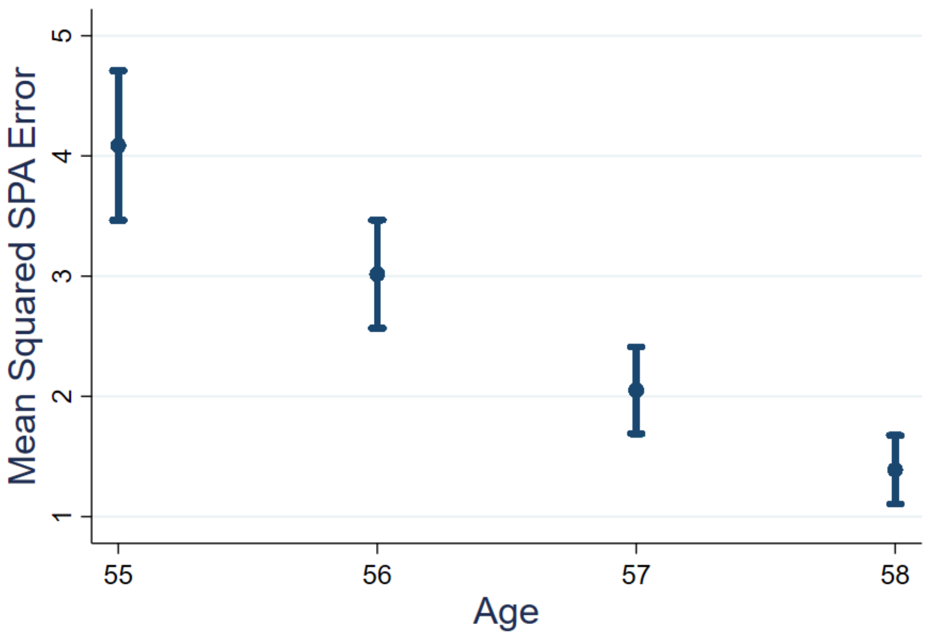
Notes: Plot of error in self-reported State Pension Age (SPA). The graph shows the frequency by which respondents gave mistaken answers about their SPA, with errors binned at the yearly level.

2.3.4 Mistaken Beliefs and Employment Sensitivity

Mistaken beliefs about one's pension are so common that few find their existence surprising. Yet, they are difficult to reconcile with frictionless information, for surely this is a topic the individual is incentivised to know about. This section documents these mistaken beliefs, specifically mistakes about the SPA, and how they relate to the excess employment sensitivity documented in Section 2.3.3.

The SPA being such a simple facet of the benefit system, confusion about it is both puzzling and simple to demonstrate. It is an exact function of date of birth, recorded in ELSA, and from wave 3, women under 60 are asked what their state pension age is. Any discrepancy demonstrates imperfect knowledge of one's SPA. Figure 2.3 shows this difference between the true and reported SPA of 58-year-old women subject to the reform. The largest group are those who know their SPA to within a year, although this contains many mistaken by a margin of months. It also leaves over 50% who are out by a year or more, striking evidence of the prevalence of mistaken pension beliefs in the UK. Appendix 2.A shows that self-reports cluster around the true SPA of each cohort; just the sort of pattern that emerges from a

Figure 2.4: Mean Squared Error in Self-reported SPA



Notes: Mean Squared Error in Self-reported SPA plotted against respondents' age.

model of costly attention. [11].

Another prediction of costly information is learning because acquired knowledge is retained, and the marginal value of knowing your SPA increases as you approach it. This prediction is supported by the data as seen in Figure 2.4 that plots, against age, the mean squared error in self-reported SPAs. A declining age profile can be seen, indicating errors shrink as these women age towards their SPA. The model uses this declining mean squared error as a moment to identify the cost of attention which is a novel contribution to the rational inattention literature that adds empirical validity.

Next we need to ask if mistaken beliefs impact employment's sensitivity to the SPA. Table 2.3 documents the heterogeneity of the labour supply response to the SPA by the degree of mistaken belief. This is found by introducing into specification

[11] Appendix 2.A also documents the distribution errors in self-reports at their natural monthly frequency

2.1 the size of the error in self-reported SPA in the last wave before the question is no longer asked at 60 and an interaction between this error and the indicator of being below the SPA. The interaction is significant and negative indicating that, on average, for each additional year the individual is out by in their SPA self-report, the labour supply response decreases by 6.2 percentage points. [12]

The existence of a relationship between mistaken beliefs and labour supply indicates they need to be studied together; the nature of the relationships indicates the endogeneity of mistaken beliefs is important. Table 2.3 show those who are least informed of the SPA before they are 60, have the smallest labour supply response upon reaching the SPA after 60. This is consistent with a model of endogenous costly information acquisition: those who care least about the SPA select the least information about it and also have the smallest labour supply response upon reaching it. In a model of exogenous information acquisition, this mechanism of selection into being informed would not exist and those who were worst informed would be so purely due to bad luck. An individual mistaken due to bad luck, unlike one mistaken due to choice, generally has a larger labour supply response upon reaching the SPA as they receive a larger shock upon the resolution of policy uncertainty that comes upon reaching the SPA. So, the negative relationship suggests an important role for the endogenous learning incorporated into the model in Section 2.4.

The excess employment sensitivity puzzle is only puzzling for standard models of complete information, deviating from standard assumptions can account for it. Two recent examples that account for this puzzle by deviating from standard assumptions are Seibold (2021), who suggests reference-dependent preferences, and Lalive et al. (2017), who suggests passive decision making. However, as models of complete information, these explanations do not account for mistaken beliefs or the correlation between these and the labour supply response to the SPA documented in Table 2.3.

[12] Appendix 2.A considers as robustness whether the direction of error in self-reported SPA and the change in self-report error size between first and last observation are important to the labour supply response. The results are consistent with the interpretation given here of beliefs being causal.

Table 2.3: Heterogeneity by SPA Knowledge

Below SPA	**0.132**
s.e	(0.0165)
p=	.000
Below SPA×(abs. Error in SPA report)	**-0.066**
s.e	(0.0142)
p=	.000
Error in SPA report	**0.040**
s.e	(0.0118)
p=	.001
Obs.	10,488
Cohorts	63

Notes: Results of running specification 2.1 with an additional interaction between absolute error in SPA self-report and an indicator of being below the SPA to pick up heterogeneity of this labour supply response along the beliefs dimension. A smaller sample size here than in Table 2.1 results from the question about SPA knowledge only being introduced in wave 3 and only being asked to individuals under 60.

2.4 Model

This section presents the model: Section 2.4.1 a baseline standard complete information model, capturing the relevant features of the UK retirement context, and Section 2.4.2 introduces two additions: objective uncertainty about government policy and costly information acquisition about this uncertain policy. This allows the model to capture the interplay between individuals' confusion about government policy and their reaction to it.

2.4.1 Complete Information Baseline

Before diving into details, a summary of key features may help orient the reader. As the model aims to explain the labour supply response to the female SPA reform, it concentrates on women. The model's decision-making unit is a household containing a couple or a single woman, but when a husband is present, they are passive as their labour supply is inelastic. The household maximises intertemporal utility from consumption, leisure, and bequests by choosing consumption, labour supply, and savings. Households face risk over i) whether they get an employment offer, ii) the wage associated with any offer, and iii) mortality. The households receive non-labour income from state and private pensions after the relevant eligibility age for each.

In more detail, households are divided into four types indexed by k, based on the high or low education status of the female and the presence or absence of a partner. Households choose how much to consume c_t, how much to invest in a risk-free asset a_t with return r, and, if not involuntarily unemployed, how much of the women's time endowment (normalised to 1) to devote to wage labour $1-l_t$ (full-time, part-time or none at all) at a wage offer w_t that evolves stochastically. Unemployment ue_t, where $ue_t = 0$ indicates employment (presence of a wage offer) and $ue_t = 1$ unemployment (the absence), also evolves stochastically. The partner's labour supply is inelastic, and so his behaviour is treated as deterministic. The wife receives the state pension once she reaches the SPA, a parameter varied to mimic the UK reform, and a private pension once she reaches the type-specific eligibility age $PPA^{(k)}$. Both pension, $S^{(k)}(.)$ the state pension and $P^{(k)}(.)$ the private pension, are treated as type-specific functions of average lifetime earning $AIME_t$ ($AIME_{t+1} = \frac{(1-l_{t+1})w_{t+1}+AIME_t t}{t+1}$) [13] . From age 60, the women face a probability s_t^k of surviving the period. Finally, households value bequests through a warm glow bequest function (De Nardi, 2004; French, 2005). Only one birth cohort is modelled at a time, and periods are indexed by age of the women t. Therefore, the full vector of model state is $X_t = (a_t, w_t, AIME_t, ue_t, t)$.

Utility The warm glow bequest motive creates a terminal condition $T(a_t)$ that occurs in a period with probability $1-s_{t-1}^{(k)}$:

$$T(a_t) = \theta \frac{(a_t+K)^{\nu(1-\gamma)}}{1-\gamma}$$

where θ determines the intensity of the bequest motive, and K determines the curvature of the bequest function and hence the extent to which bequests are luxury goods. The functional form surrounding $a_t + K$ is the utility from consumption of a household (see below), so approximately captures the utility a descendant would gain from these assets, and hence altruism as a motive for the warm-glow as well as

[13] This is average yearly earnings, to keep notation in line with the literature I use the abbreviation Average Indexed Monthly Earnings, which is the variable US Social Security depends on.

keeping parameters to a minimum.

Whilst alive, a household of type k has the following homothetic flow utility:

$$\text{where } u^{(k)}(c_t, l_t) = n^{(k)} \frac{((c_t/n^{(k)})^{\nu} l_t^{1-\nu})^{1-\gamma}}{1-\gamma}$$

where $n^{(k)}$ is a consumption equivalence scale taking value 2 if the household represents a couple and 1 otherwise. In other words, utility takes an isoelastic from, with curvature γ, over a Cobb-Douglas aggregator of consumption and leisure, with consumption weight, ν.

Initial and terminal conditions The model starts with women aged 55 because, firstly, ELSA starts interviewing people at 50 and, secondly, as the focus is retirement modelling early life-cycle behaviour would be computationally wasteful. It starts at 55 rather than 50 because this is the youngest age with significant numbers of SPA self-reports and variation in the true SPA, thus allowing me to initialise the state variables from the data for different SPA-cohorts. When age 100 is reached in the model, the woman dies with certainty.

Labour market The female log wage, w_t, is the sum of a type-specific deterministic component, quadratic in age, and a stochastic component:

$$\log(w_t) = \delta_{k0} + \delta_{k1}t + \delta_{k2}t^2 + \varepsilon_t \tag{2.2}$$

where ε_t follows an AR1 process with persistence ρ_w and normal innovation term with standard error σ_ε, and has an initial distribution $\varepsilon_1 \sim N(0, \sigma^2_{\varepsilon,55})$. The quadratic form of the deterministic component of wages captures the observed hump-shaped profile and is common in the literature.

The unemployment status of the woman ue_t evolves according to a type-specific conditional Markov process. From age 80, the woman can no longer choose to work; this is to model some of the limitations imposed by declining health.

As spousal income results from the confluence of wages, mortality and pension income, it follows a flexible polynomial in age:

$$\log(y^{(k)}(t)) = \mu_{k0} + \mu_{k1}t + \mu_{k2}t^2 + \mu_{k3}t^3 + \mu_{k4}t^4 \tag{2.3}$$

This specification averages out and abstracts away from both idiosyncratic spousal income and mortality risk. In effect, the household dies when the woman dies, and the husband's mortality risk only turns up in so far as it affects average income, as if husbands were a pooled resource amongst married women. This allows me to ignore transitions between married and single which, while important to wider labour supply behaviours of older individuals (e.g. Casanova, 2010), are of secondary importance, at best, to labour supply responses to the SPA. The function $y^{(k)}(t)$ amalgamates spousal labour and non-labour income including pensions. Both female wage and spousal income are post-tax.

Social insurance Unemployment status is considered verifiable, so only unemployed women, $ue_t = 1$, can claim the unemployment benefit b.

The wife receives the state pension as soon as she reaches the *SPA* which abstracts away from the benefit claiming decision. This is done for two reasons, both touched upon earlier. Firstly, over 85% of people claim the state pension at the SPA, so, in terms of accuracy, little is lost by this simplification. Secondly, this small fraction deferring receipt of the state pension occurs despite deferral having been actuarially advantageous during the period considered. This presents another puzzle to standard models of complete information as they generally imply acceptance of actuarially advantageous offers. This benefit claiming puzzle is taken up in Section 2.8, but deferring it until then gives this baseline model a chance of addressing the excess sensitivity puzzle.

Lifetime average earning ($AIME_t$) evolves until the woman reaches the age she starts to receive her $PPA^{(k)}$, at which point it is frozen.[14] Both the state and private pensions are quadratic in $AIME_t$, until attaining their maximum, at which point they

[14]It is frozen at this age to avoid creating the counterfactual incentive to get a new job to increase your current private pension income.

are capped. Until being capped, the pensions functions have the following forms

$$S^{(k)}(AIME_t) = sp_{k0} + sp_{k1}AIME_t - sp_{k2}AIME_t^2 \tag{2.4}$$

$$P^{(k)}(AIME_t) = pp_{k0} + pp_{k1}AIME_t - pp_{k2}AIME_t^2 \tag{2.5}$$

These pension functions abstract away from the details of state and private pension systems but capture some of the key incentives in a tractable form. The state pension is a complex path-dependent function dependent on past and current regulations, and cannot be exactly captured without detailed administrative data (see Bozio et al., 2010). This functional form captures the dependence of the state pension on working history without getting into these difficulties. Being type-specific allows $S^{(k)}(.)$ to capture indirect influences of education and marital status on the state pension; for example, being a stay-at-home mum counted towards state pension entitlement but only after a reform was enacted. Every private pension scheme is different, but the dependence of $P^{(k)}(.)$ on $AIME_t$ reflects the dependence of most defined benefit schemes on lifetime earnings. This functional form less accurately reflects the structure of defined contribution systems, which are essentially saving accounts, but saving for retirement is captured in the model with the risk-free asset and the models starts after the statutory defined contribution eligibility age beyond which they can be accessed without penalty.

Total deterministic income Combining spousal income, benefits, and private and state pension benefits into a single deterministic income function yields:

$$\begin{aligned} Y^{(k)}(t, ue_t, AIME_t) =& y^{(k)}(t) + b\mathbb{1}[ue_t = 1] + \mathbb{1}[t \geq SPA]S^{(k)}(AIME_t) \\ &+ \mathbb{1}[t \geq PPA^{(k)}]P^{(k)}(AIME_t) \end{aligned} \tag{2.6}$$

Household maximisation problem and value functions The Bellman equation for a household of type k is:

$$V_t^{(k)}(X_t) = \max_{c_t, l_t, a_{t+1}} \{u^{(k)}(c_t, l_t) + \beta(s_t^{(k)}(E[V_{t+1}^{(k)}(X_{t+1})|X_t] + (1 - s_t^{(k)})T(a_{t+1}))\} \tag{2.7}$$

Subject to the following budget constraint, borrowing constraint, and labour supply constraint:

$$c_t + (1+r)^{-1} a_{t+1} = a_t + w_t(1-l_t) + Y^{(k)}(t, ue_t, AIME_t) \tag{2.8}$$

$$a_{t+1} \geq 0 \tag{2.9}$$

$$ue_t(1-l_t) = 0 \tag{2.10}$$

2.4.2 Two Additions: Policy Uncertainty and Costly Attention

This section introduces two additions to the complete information model. Firstly, Section 2.4.2.1 introduces objective policy uncertainty in the form of a stochastic SPA, capturing the observed variation of SPAs over the life-cycle resulting from pension reform. Secondly, Section 2.4.2.2 introduces costly attention to this stochastic SPA, in the form of a disutility for more precise information, allowing the model to capture mistaken beliefs. As these additions represent innovation, Section 2.4.2.3 rounds off with a discussion.

2.4.2.1 Policy Uncertainty: the Stochastic SPA

To capture the objective policy uncertainty resulting from the fact that governments can and do change pension policy, I make the SPA stochastic. The motivation for this addition is that the SPA changes. For the women in my sample, their SPA increased by up to 6 years during their working life, a change that was not foreseeable when they began working life.

Although the SPA does change, introducing an important dimension of uncertainty, changes are not sufficiently frequent to estimate a flexible stochastic SPA process. For this reason, I impose a parsimonious functional form on the stochastic SPA:

$$SPA_{t+1} = \min(SPA_t + e_t, 67) \tag{2.11}$$

where $e_t \in \{0,1\}$ and $e_t \sim Bern(\rho)$. So each period, the SPA may stay the same or increase by one year, as the shock is Bernoulli, up to an upper limit of 67. This captures a key aspect of pension uncertainty, that in recent years governments have reformed pension ages upward but generally not downward, whilst maintaining a simple tractable form. I do not consider SPAs below the pre-reform age of 60. Hence, as the law-of-motion only allows for increases, SPA_t is bounded below by 60 and above by 67.

In the model, the variable SPA_t represents the current best available information about the age the woman will reach her SPA, and as such, the data analogue is the SPA the government is currently announcing for the woman's cohort. Only one SPA cohort is modelled at a time. So there is no conflict in having a single variable SPA_t whilst, in reality, at a given point in time, different birth cohorts have different government-announced SPAs.

2.4.2.2 Costly Attention (Rational Inattention)

The second addition is the cost of information acquisition about the stochastic SPA. This allows the model to capture the fact that people are mistaken about their SPA and that these mistaken beliefs are the results of an endogenous learning process. As such, it creates a potential for the model to replicate the patterns of learning documented in Section 2.3.4.

Directly observed vs learnable states: To make the exposition of rational inattention to the SPA as clear as possible, I introduce two notational simplifications. I group decisions into a single variable $d_t = (c_t, l_t, a_{t+1})$ and all states other than the SPA into a single state variable $X_t = (a_t, w_t, AIME_t, ue_t, t)$.[15] The stochastic SPA SPA_t is separated because, unlike other state variables, it is not directly observed by the household. Instead, the household must pay a utility cost to receive more precise information about the SPA, as outlined below. The other stochastic state variables, w_t and ue_t being directly observed can be interpreted as these variables being more salient. Rather than any of the other myriad burdens on people's attention I focus on costly attention to the state pension policy because this is the uncertainty that is resolved upon reaching the SPA and hence may help explain why people respond as they do to the SPA.

[15]This is the same collection of variables in X_t as when it was defined in the baseline model. I highlight this as a notational change as I want to be explicit that X_t has not absorbed the new state SPA_t

Within period timing of learning: As the household no longer directly observes SPA_t, it is a hidden state. It is still a state as it is payoff relevant, but since the household does not observe it, it cannot enter the decision rule. This introduces a new state variable $\underline{\pi_t}$ the belief distribution the household holds about SPA_t. Since the household chooses how much information about the SPA to acquire, its choice can be thought of as a two-step process: first choosing a signal distribution and then conditional on the signal draw choosing actions. Although subject to a utility cost of information, the choice of signal is unconstrained; the household is free to learn about SPA_t however they want. More precisely, a household with non-hidden states X_t and $\underline{\pi_t}$ is free to choose any conditional distribution function $\underline{f_t}[X_t, \underline{\pi_t}](z|SPA_t)$ for its signal $z_t \sim Z_t$ given the value of the hidden state SPA_t.

The household is rational, and so $\underline{\pi_t}$ is formed through Bayesian updating on their initial belief distribution $\underline{\pi_{55}}$ given the full history of observed signals draws z^t. Specifically, the posterior is formed as:

$$Pr(spa|z_t) = \frac{f_t(z_t|spa)\pi_t(spa)}{Pr(z_t)} \tag{2.12}$$

Then the prior at the start of next period $\underline{\pi_{t+1}}$ is formed by applying the law of motion of SPA_t, equation 2.11, to this posterior.

Entropy and mutual information: The cost of attention is directly proportional to the mutual information, defined below, between signal and SPA. Mutual information is the expected reduction in uncertainty, as measured by the entropy, about one variable resulting from learning the value of another. Entropy, in this information theoretic sense, is a measure of uncertainty that captures the least space[16] needed to transmit or store the information contained in a random variable.

Definition 2.4.1 (Entropy/conditional entropy)**.** *The entropy $H(.)$ of $X \sim P_X(x)$ is minus the expectation of the logarithm of $P_X(x)$, $H(X) = E_X[-\log(P_X(x))]$. Conditional entropy is $H(X|Y) = E_Y[H(X|Y = y)]$.*

[16] If the logarithm is taken with respect to base 2 then entropy measure this space in bits, but the base of the logarithm is unimportant as changing base only changes the unit of measure. One application, that may help intuition, is by using these concepts; a computer is able to compress a file.

Definition 2.4.2 (Mutual Information). *The mutual information between $X \sim P_X(x)$ and $Y \sim P_Y(y)$ is the expected reduction in uncertainty, as measured by entropy, about X from learning Y (equally about Y from learning X) : $I(X,Y) = H(X) - H(X|Y)$.*

Utility: Incorporating information costs, utility takes the form:

$$u^{(k)}(d_t, \underline{f_t}, \underline{\pi_t}) = n^{(k)} \frac{((c_t/n^{(k)})^{\nu} l_t^{1-\nu})^{1-\gamma}}{1-\gamma} - \lambda I(\underline{f_t}; \underline{\pi_t})$$

where the constant of proportionality λ is the cost of attention parameter, and given the above definitions we can expand $I(\underline{f_t}; \underline{\pi_t})$:

$$I(\underline{f_t}; \underline{\pi_t}) = \sum_z \sum_{spa} \pi_t(spa) f_t(z|spa) \log\Big(\pi_t(spa) f_t(z|spa)\Big) - \sum_{spa} \pi_t(spa) \log(\pi_t(spa))$$

Revelation of uncertainty: Upon reaching SPA_t, the woman learns her true SPA_t and starts receiving the state pension. Therefore the household knows that if they are not in receipt of the woman's state pension benefits, she is below her SPA. This avoids issues with the budget constraint when households do not know the limits on what they can spend. That arriving at SPA_t in the model provides a positive informational shock reflects the reality of the UK pension system; the only communication received by all cohorts in the sample was a letter sometime in the six months before their SPA. That uncertainty is resolved upon reaching SPA_t is a key model mechanism explaining why women have a labour supply response upon reaching the SPA.

Dynamic programming problem: The full set of states for the model is $(X_t, SPA_t, \underline{\pi_t}) = (a_t, w_t, AIME_t, ue_t, t, SPA_t, \underline{\pi_t})$ and its Bellman equation:

$$\begin{aligned} &V_t^{(k)}(X_t, SPA_t, \underline{\pi_t}) = \\ &\max_{d_t, \underline{f_t}} E\left[u^{(k)}(d_t, \underline{f_t}, \underline{\pi_t}) + \beta(s_t^{(k)} V_{t+1}^{(k)}(X_{t+1}, SPA_{t+1}, \underline{\pi_{t+1}}) + (1 - s_t^{(k)}) T(a_{t+1}))\right] \end{aligned} \tag{2.13}$$

subject to the same constraints 2.8 - 2.10 as the baseline model and where now the utility function includes a cost of information that is directly proportional to the mutual information between the signal and the household's current state of knowledge about the SPA $\underline{\pi_t}$, as explained above.

One problem hidden in this Bellman equation is the formation of next-period beliefs, which, due to Bayesian updating, depends upon the full distribution of signals. This means that the continuation value is not known until the solution is known; this problem will be taken up in Section 2.5.

2.4.2.3 Discussion of Costly Attention to the Stochastic SPA

This self-contained section discusses reasons for modelling the cost of attention as I have and interpretations of two new features: the cost of attention and the choice of signal function.

Expected Entropy Reduction Attention Cost: A cost of information acquisition is included to accommodate mistaken beliefs which predict employment responses to the SPA. As utility costs of information are uncommon in the life-cycle literature, the reasons for the functional form may be unfamiliar and unclear. I offer three reasons for the choice.

Firstly, although this functional form is not widely used in life-cycle models, this is because most life-cycle models ignore costly information acquisition, not because any other functional form is widely used. In fact, a cost of information acquisition that is directly proportional to the mutual information is among the most common in the costly information literature leading to two important advantages.[17] It is tractable because many useful results are available for this functional form, and it follows a convention. Tractability is important in models of costly information which can be too complex to solve, and following a convention has merit because it restricts the degrees of freedom available to fit the data.

Secondly, it endogenously generates certain rules-of-thumb or heuristics observed sufficiently often to be christened as a behavioural bias. We could treat these simplifying rules-of-thumb, or heuristics, as pre-ordained behavioural rules people blindly follow. This has major disadvantages: one, this does not explain why the

[17] Caplin et al. (2017) and Fosgerau et al. (2020) are examples of papers from the costly attention literature that use other functional forms. Both can be seen as introducing more flexibility into the cost of attention function rather than completely abandoning the entropy approach.

particular rule-of-thumb and, two, it ignores the fact that people change rule-of-thumb as circumstances change. Hard-coded behavioural biases suppress a central insight of economics: people respond to incentives. Endogenising observed heuristics with a cost of attention avoids these pitfalls because, one, it explains why given heuristics are used and, two, it allows an agent to change heuristics in response to incentives. Two examples come from Kõszegi and Matějka (2020) who show this cost of attention generates both mental budgeting (quantity allocated to a category being fixed and composition changing) and naive diversification (composition being fixed and quantity allocated changing) depending on the circumstance. A third example comes from Caplin et al. (2019) who show it leads to consideration sets: ignoring many options to focus on a subset.

Thirdly, strong a priori reasons to think that a cost of cognition should depend on entropy exist. The information-theoretic concept of entropy was developed to explain how computers process information and gives a lower bound on the efficient transmission and storage of information. The computational theory of mind Mcculloch and Pitts (1943) holds the human mind is a computer. This is controversial and well outside the scope of this paper, but even its most stringent opponent would agree the brain performs some tasks like a computer, with information processing a primary candidate. So, if the brain processes information efficiently, mutual information should enter into the ideal cost of attention function. This is not to say an ideal cost of attention function would be linear in mutual information, but if it enters into the ideal then a first-order approximation in this dimension is reasonable when information processing is our focus. [18]

Interpreting the cost of attention: Costly information is modelled abstractly and so open to various interpretations but to guide the reader's intuition, I suggest two: the first broad and the second more literal.

In the broader interpretation learning about the SPA can be taken as illustrative of learning about the state pension system in general. The pension system is

[18] If the argument above is correct, one expects that entropy would have found a use in neuroscience and psychology, and indeed this is the case (for example Frank (2013) or Carhart-Harris et al. (2014).

multifaceted, and people are confused about most of these facets. The model concentrates all costs of information acquisition onto tracking one aspect of the pension benefit system, the SPA. So the model may also capture learning about these other facets and the resolution of uncertainty about them. Hence, it is possible to think of this cost of learning about the SPA as a cost of learning about pension policy more generally, and I believe the reader taking this perspective can equally draw interesting lessons from this model. In Section 2.8 I look at an extension in which the household also learns about an uncertain actuarial adjustment to deferred claiming.

The more literal interpretation of the cost of attention is as the cost of learning exclusively about your SPA. This is it captures all costs of learning your SPA: hassle costs, as well as information processing, storage, and recall. As an illustration, the author has paid the hassle cost of looking up his SPA but has not paid the cognitive cost of remembering this information. Hence, I would show up in survey data as someone with a mistaken belief and could also not use my SPA in decision-making. Therefore, including the cognitive cost of remembering and assimilating information as well as any hassle cost is the minimum data and model consistent conceptualisation.

Interpreting the choice of signal: The choice of a signal function to learn about the SPA may be difficult to conceptualise. The SPA is a number we can just look up which seems simpler than choosing a signal function. However, looking it up is a learning strategy encompassed by the choice of a signal function conception, corresponding to choosing a perfectly informative signal function.[19] In reality, carefully reading relevant regulations is not the main way people learn about government policy in general or the state pension in particular: people learn from other people or news outlets. In both examples, there is a random component, what stories newspapers run and what other people talk about, and a choice component, whether you keep reading or ask follow-up questions. This is analogous to the choice of a signal function in that it is partly a choice and partly stochastic, and so it captures much about the messy real-world learning process.

[19]Being more careful about cognitive cost, a perfectly informative signal includes looking up, remembering, and assimilating into choices.

2.5 Model Solution

By introducing a high dimensional state $\underline{\pi_t}$ (beliefs) and a high dimensional choice $\underline{f_t}$ (signal), rational inattention has complicated the model to the extent that solving it is a contribution. To achieve this I weave together recent theoretical results into a consistent solution method for dynamic rational inattention models with endogenous heterogeneous beliefs, like the one presented above. Section 2.5.1 explains how this is done, both to communicate the methodological innovations and to provide intuition of the model solution. First I explain details specific to solving the model of this paper.

2.5.1 Details Specific to this Model

All versions of the model are solved by dynamic programming, specifically backward induction, but the $\underline{\pi_t}$ and $\underline{f_t}$ alter the nature of the within period problem in the model with rationally inattentive households, in some periods. Only in some periods because $\underline{\pi_t}$ and $\underline{f_t}$ only matter before the SPA: after the SPA the true value is known and so beliefs ($\underline{\pi_t}$) and learning ($\underline{f_t}$) about the SPA are irrelevant. Periods after the SPA can be solved, like the baseline and the model with only policy uncertainty, by simple search techniques to find the optimal choice amongst the discrete options.

We proceed by backward induction from terminal age $t = 100$ using standard techniques for the within-period problem in the model with rationally inattentive households until age $t = 66$. We can proceed back to age $t = 67$ because, as SPA_t is bounded above by 67, the woman receives her state pension with certainty from this age. At $t = 66$ the household is perfectly informed meaning $\underline{\pi_t}$ is irrelevant, but SPA_t is a state variable because receipt of the state pension affects utility. If she is not in receipt of her state pension ($SPA_t > t$), she infers $SPA_t = 67$ with certainty because she knows the data generating process, just not the value of SPA_t. Otherwise she is past SPA_t ($SPA_t \leq t$) and its precise value is irrelevant. The same is true for all ages, distinctions between past SPAs do not matter, so we can solve for a single representative $SPA_t \leq t$ using standard techniques. Hence, each year we proceed backwards, the list of future SPAs we need to solve separately grows by one. At age

$t = 65$, if $SPA_t > t$ she can no longer infer its true value and so beliefs ($\underline{\pi_t}$) become a state and the choice of signal function relevant. Beliefs are a state and the signal a relevant choice for all $t \leq 65$ whenever $SPA_t > t$. These are the periods where rational inattention matters. As $\underline{\pi_t}$ is a distribution over all future SPAs ($SPA_t > t$), its points of support also grow by one with each step in the backward induction. This growth of the state space along two dimensions, relevant true SPAs and beliefs over future SPAs, continues until we reach $t = 59$. At this point, all SPAs 60-67 are future, and rational inattention is relevant regardless of the value of the SPA_t.

The solution of within period problems, when rational inattention matters, because $t < SPA_t$, is explained immediately below in Section 2.5.2. There I ignore the details presented here because they have no appreciable implications for how to solve generic dynamic rational inattention models with endogenous heterogeneous beliefs.

2.5.2 Solving Generic Dynamic Costly Attention Models with Endogenous Beliefs

Dynamic rational inattention models with endogenous heterogeneous beliefs are complicated by the presence of a high dimensional state $\underline{\pi_t}$ (beliefs distribution) and a high dimensional choice $\underline{f_t}$ (signal distribution). This section presents my solution method. I use the model of retirement decision from this paper to explain the method, but it applies to any dynamic rational inattention models with endogenous heterogeneous beliefs.

To solve the periods in which rational inattention is relevant, I leverage results from three recent theoretical papers. Most centrally, I rely on results from Steiner et al. (2017) who extend the static logit-like results for $\underline{f_t}$ from Matějka and McKay (2015) to a dynamic setting, showing dynamic rational inattention problems reduce to a collection of static problems. As such it gives me analytic results that greatly simplify dealing with the high dimensional choice $\underline{f_t}$. With the results of Steiner et al. (2017) the model is theoretically solvable but the high dimensional state $\underline{\pi_t}$ means finding that solution is practically impossible. Results from Caplin et al. (2019) help to make finding a solution feasible. They provide sufficient conditions to complement the necessary condition in Matějka and McKay (2015). Additionally, as mentioned earlier, they show rational inattention generically implies consideration sets. That is there are many actions that the household will ignore and never take. This implies the solving conditional choice probabilities, or stochastic decision rules, will be sparse. The sufficient conditions in their paper allow me to check for sparsity ex-ante which greatly reduces the computational burden. Finally, when sparsity does not provide a short-cut solution to the within period optimisation problem, I employ sequential quadratic programming to solve the optimality conditions. Using this algorithm for static rational inattention problems is an approach suggested by Armenter et al. (2019) and as Steiner et al. (2017) reduces the dynamic problem to a sequence of static ones I am able to use the same approach to the within period problem.

The rest of this section precedes as follows. Firstly, Section 2.5.2.1 gives an outline of the proof of the main results from Steiner et al. (2017). Then Section 2.5.2.2 will take the results from Section 2.5.2.1 and present my solution method.

2.5.2.1 Analytic Foundations of Solution Method

Steiner et al. (2017) show that a wide class of similar models have a logit-like solution. [20] To provide some intuition, and because an understanding of these results is needed to understand the solution methodology, in this section, I present an outline of their proof using my model as a lens through which to explain their results. Steiner et al. (2017) extend Matějka and McKay (2015)[21] to a dynamic setting and so most of what is explained here applies equally to static problems.

Key results: If we define the effective conditional continuation values as:

$$\overline{V}_{t+1}^{(k)}(d_t, X_t, SPA_t, \underline{\pi_t}) =$$
$$E[s_t^{(k)} V_{t+1}^{(k)}(X_{t+1}, SPA_{t+1}, \underline{\pi_{t+1}}) + (1 - s_t^{(k)}) T(a_{t+1}) | d_t, X_t, SPA_t, \underline{\pi_t}],$$

[20] My framework is a slight extension Steiner et al. (2017). Observable states a_t, $AIME_t$ map to the payoff relevant lagged choices in their framework but y_t, which is exogenous and whose current value is known, means I need to extend the free signal. Appendix 2.B.1 contains the details.

[21] This is a more complicated step than it may sound and to show this they had to overcome various thorny issues, stemming from the information acquisition. Although I allude to some of these complexities I mostly ignore them to give the reader the intuition for the dynamic logit-like results.

where expectations are over X_{t+1} and SPA_{t+1} and Section 2.5.2.2 describes finding $\underline{\pi_{t+1}}$, the Bellman equation 2.13 becomes:

$$V_t^{(k)}(X_t, SPA_t, \underline{\pi_t}) = \max_{d_t, \underline{f_t}} E\left[u^{(k)}(d_t, \underline{f_t}, \underline{\pi_t}) + \beta \overline{V}_{t+1}^{(k)}(d_t, X_t, SPA_t, \underline{\pi_t})\right].$$

Steiner et al. (2017) show that the solution to this model has actions that are distributed with conditional choice probabilities $d_t|SPA_t \sim \underline{p_t}(d_t|SPA_t)$ and associated unconditional probabilities $d_t \sim \underline{q_t}(d_t)$ that satisfy:

$$p_t(d|spa) = \frac{\exp\left(n^{(k)}\frac{((c/n^{(k)})^{\nu}l^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \log(q_t(d)) + \beta\overline{V}_{t+1}^{(k)}(d, X_t, SPA_t, \underline{\pi_t})\right)}{\sum_{d'\in\mathscr{C}}\exp\left(n^{(k)}\frac{((c'/n^{(k)})^{\nu}l'^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \log(q_t(d')) + \beta\overline{V}_{t+1}^{(k)}(d', X_t, SPA_t, \underline{\pi_t})\right)}, \tag{2.14}$$

$$\max_{\underline{q_t}} \sum_{spa} \pi_t(spa) \log\Bigg(\sum_{d\in\mathscr{C}} q_t(d) \exp\Bigg(n^{(k)}\frac{((\frac{c}{n^{(k)}})^{\nu}l^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \beta\overline{V}_{t+1}^{(k)}(d, X_t, SPA_t, \underline{\pi_t})\Bigg)\Bigg). \tag{2.15}$$

Sketch proof: The household does not observe SPA_t but solves the problem for an observed value of $(X_t, \underline{\pi_t})$ and all possible values of SPA_t simultaneously. They do this by selecting a signals function $\underline{f_t}(z|SPA_t)$ which gives a noisy signal of the unobserved SPA_t, and then make a decision contingent on the realisation of the signal $d(z)$.

The first step in solving this problem is to note that, since the signal encapsulates an internal cognitive process it is inherently unobservable. Hence, nothing is lost in combining the choice of a stochastic signal function $\underline{f_t}$ and a deterministic decision conditional on the signal $d(z)$ into a single choice of a stochastic decision $d_t \sim \underline{p_t}(d_t|SPA_t)$. The stochastic decision conditions on SPA_t, which the household does not directly observe because they observe the signal that is conditional on SPA_t; this is the source of the stochasticity as conditional on the signal the decision $d(z)$ is deterministic.

The next step is a revelation principle type argument. As the household is rational and pays a utility cost for information they will not select any extraneous information. All information has a cost $\lambda I(\underline{f_t};\underline{\pi_t})$, but only information that leads to a better choice has a return, therefore the household will choose a signal function that perfectly reveals their action i.e. signal and action are in a one-to-one correspondence. Therefore the $\underline{p_t}(d_t|SPA_t)$ is simply a relabelling of $\underline{f_t}(z_t|SPA_t)$. The function $\underline{f_t}$ tells you the signal seen, re-labelling with the choice taken on seeing that signal gives $\underline{p_t}$. From this it follows that $I(\underline{f_t};\underline{\pi_t}) = I(\underline{p_t};\underline{\pi_t})$, as mutual information is a function of the probabilities in a distribution, not the values of the associated random variable. Therefore we can re-write the agent's decision problem as:

$$V_t^{(k)}(X_t, SPA_t, \underline{\pi_t}) =$$
$$\max_{\underline{p_t}} E\left[n^{(k)}\frac{((c/n^{(k)})^{\nu} l^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} - I(\underline{p_t};\underline{\pi_t}) + \beta \overline{V}_{t+1}^{(k)}(d, X_t, SPA_t, \underline{\pi_t})\right].$$

As the problem is treated as discrete choice there exists a finite budget set available to the agent $\mathscr{C} \subset \mathbb{R}^2$, $\mathscr{C} = \{d_1 = (c_1, l_1), ..., d_N = (c_N, l_N)\}$. Then the problem becomes:

$$\max_{\underline{p_t}} \sum_{spa} \pi_t(spa) \sum_{i=1}^{N} p_t(d_i|spa) \Big(n^{(k)} \frac{((c_i/n^{(k)})^{\nu} l_i^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} - I(\underline{p_t};\underline{\pi_t}) + \beta \overline{V}_{t+1}^{(k)}(d_i, X_t, SPA_t, \underline{\pi_t}) \Big) \tag{2.16}$$

and from the symmetry of mutual information: [22]

$$I(\underline{p_t};\underline{\pi_t}) = \sum_{spa} \pi_t(spa) \left(\sum_{d} p_t(d|spa) \log(p_t(d|spa)) \right) - \sum_{d} q_t(d) \log(q_t(d)) \tag{2.17}$$

and $\underline{q_t}$ is the resulting marginal distribution of d:

$$q_t(d) = \sum_{spa} \pi_t(spa) p_t(d|spa).$$

[22] We have been thinking of mutual information as the expected reduction in entropy about the SPA from learning the signal, or equivalently, what action to take. That is equivalent to the expected reduction in entropy about the action from learning the SPA, which is what is expressed above.

Substituting 2.17 into 2.16, rearranging, and collapsing the repeated sums gives:

$$\max_{\underline{p_t}} \sum_{spa} \pi_t(spa) \sum_{i=1}^{N} \Big(n^{(k)} \frac{((c_i/n^{(k)})^{\nu} l_i^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \log(q_t(d_i)) - \log(p_t(d_i|spa_i)) + \beta \overline{V}_{t+1}^{(k)}(d_i, X_t, SPA_t, \underline{\pi_t}) \Big). \tag{2.18}$$

Taking $\underline{q_t}$ as given, optimality with respect to any $p_t(d|spa)$ requires the following FOC, derived from differentiating 2.18, be satisfied [23]

$$\mu(spa) = n^{(k)} \frac{((c/n^{(k)})^{\nu} l^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \log(q_t(d)) - (\log(p_t(d|spa)) + 1) + \beta \overline{V}_{t+1}^{(k)}(d, X_t, SPA_t \underline{\pi_t}),$$

where $\mu(spa)$ are the Lagrange multipliers associated with the constraint that $p_t(.|spa)$ be a valid probability distribution, $\sum_{d \in \mathscr{C}} p_t(d|spa) = 1$. Rearranging gives:

$$p_t(d|spa) = \exp\left(n^{(k)} \frac{((\frac{c}{n^{(k)}})^{\nu} l^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \log(q_t(d)) + \beta \overline{V}_{t+1}^{(k)}(d, X_t, SPA_t, \underline{\pi_t}) - \mu(spa) + 1 \right).$$

Then as $\sum_{d \in \mathscr{C}} p_t(d|spa) = 1$ we can divide the right-hand side by this sum without changing the value to eliminate the nuisance terms which gives the solution for $\underline{p_t}$:

$$p_t(d|spa) = \frac{\exp\left(n^{(k)} \frac{((c/n^{(k)})^{\nu} l^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \log(q_t(d)) + \beta \overline{V}_{t+1}^{(k)}(d, X_t, SPA_t, \underline{\pi_t})) \right)}{\sum_{d' \in \mathscr{C}} \exp\left(n^{(k)} \frac{((c'/n^{(k)})^{\nu} l'^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \log(q_t(d')) + \beta \overline{V}_{t+1}^{(k)}(d', X_t, SPA_t, \underline{\pi_t})) \right)}.$$

This derivation assumed $\underline{q_t}$ was given, but as $\underline{q_t}$ is the marginal to conditional $\underline{p_t}$ it is also chosen. The form of $\underline{q_t}$ can be found from substituting 2.14 into 2.18 and noting that the logarithm of the numerator in 2.14 cancels all other terms in 2.18

[23] Eagle-eyed readers may have noted this treats the continuation value as fixed. Showing ”one can ignore the dependence of continuation values on beliefs and treat them simply as functions of histories” was an achievement of Steiner et al. (2017) which I abstract from to give the intuition.

leaving only the summation from the denominator. So $\underline{q_t}$ can be found by solving:

$$\max_{\underline{q_t}} \sum_{spa} \pi_t(spa) \log\Big(\sum_{d' \in \mathscr{C}} q_t(d) \exp\Big(n^{(k)} \frac{((\frac{c}{n^{(k)}})^{\nu} l^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \beta \overline{V}^{(k)}_{t+1}(d, X_t, spa, \underline{\pi_t})\Big)\Big).$$

2.5.2.2 Solution Method

Being the first to solve a dynamic rational inattention model with endogenous heterogeneous beliefs, requires a new solution method. At its core the solution method is to solve 2.15 for $\underline{q_t}$ and substitute the solution into 2.14 to get $\underline{p_t}$. This basic description conceals two major hurdles which this section explains culminating in a description of the algorithm.

The first major difficulty is that next period's beliefs given actions are not known until the full probability distribution of actions is known. This is because we do not know how strong a signal of a given SPA an action is unless we know how likely they were to take that action given other possible SPAs. It follows that next period's effective conditional value function $\overline{V}_{t+1}$ is not known, even when the next period's value function V_{t+1} is known, because we do not know the beliefs tomorrow that will result from an action today. Substituting the results of 2.14 and 2.15 into the Bayesian updating formula 2.12 gives:

$$\begin{aligned} Pr(spa|d_t) &= \frac{p_t(d_t|spa)\pi_t(spa)}{q_t(d_t)} \\ &= \frac{\pi_t(spa)\exp\Big(n^{(k)} \frac{((c/n^{(k)})^{\nu} l^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \beta \overline{V}^{(k)}_{t+1}(d, X_t, spa, \underline{\pi_t}))\Big)}{\sum_{d' \in \mathscr{C}} q_t(d')\exp\Big(n^{(k)} \frac{((c'/n^{(k)})^{\nu} l'^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \beta \overline{V}^{(k)}_{t+1}(d', X_t, spa, \underline{\pi_t}))\Big)}. \end{aligned}$$

Then the prior at the start of next period $\underline{\pi_{t+1}}$ is formed by applying the law of motion of SPA_t, equation 2.11, to this posterior. Since the posterior depends not only on the exponentiated payoff but also on the $\underline{q_t}$, we need a solution ($\underline{q_t}$) to know next period's beliefs given choices and hence to know the effective conditional continuation values:

$$\overline{V}_{t+1}^{(k)}(d_t, X_t, SPA_t, \underline{\pi_t}) = E[s_t^{(k)} V_{t+1}^{(k)}(X_{t+1}, SPA_{t+1}, \underline{\pi_{t+1}}) + (1 - s_t^{(k)}) T(a_{t+1}) | d_t, X_t, SPA_t, \underline{\pi_t}] \tag{2.19}$$

Steiner et al. (2017) evade this difficulty by removing the beliefs from the state space and replacing them with the full history of actions. They can do this because, given initial beliefs, the full history of signals, or equivalently actions, perfectly predicts the beliefs in period t. This is an inspired move for a theory paper and is a key step in extending Matějka and McKay (2015) to the dynamic case.[24]For applied work, it is basically a non-starter. It involves introducing redundant information into the state space because if two action histories lead to the same beliefs they do not truly represent different states. Redundant information in the state space is problematic because the curse of dimensionality means this is often the binding constraints to producing rich models. What moves this here from problematic to a non-starter is that this redundant information grows exponentially with the number of periods.

Hence, I rely on the theoretical results of Steiner et al. (2017) that used the history of action state-space representation, but in practice, I use the more compact belief state-space representation for the actual computational work. To get around the issue that I need $\underline{q_t}$ to know $\overline{V}_{t+1}$ I use a simple guess-and-verify fixed-point strategy. First I guess a value $\underline{\tilde{q}_t}$ and solve the fixed point iteration for the effective conditional continuation value defined by substituting 22 into 23. Then given $\overline{V}_{t+1}$ I solve 2.15 for $\underline{q_t}$. If resulting $\underline{q_t}$ is sufficiently close to $\underline{\tilde{q}_t}$, I accept this solution otherwise I replace $\underline{\tilde{q}_t}$ with $\underline{q_t}$ and repeat. [25]

This solution to the first major difficulty, however, exacerbates the second, the high computational demands resulting from the high dimensional state $\underline{\pi_t}$, by increasing the computation required at each point in the state space. Here relief can be found from the results of Caplin et al. (2019), who show that generically rational inattention implies consideration sets. Hence, the solving conditional choice

[24]This allowed them to show we can ignore the dependence of continuation values on beliefs.

[25]Although, I have not proved this is a contraction mapping the fixed point iteration always converged and generally in relatively few iterations.

probabilities (CCPs) $\underline{p_t}$ are sparse. That is, various actions will never be taken. I can check for this sparsity, ex-ante, at various points in the process and remove any actions that will never be taken. This reduces the dimensionality of the optimisation in equation 2.15. Moreover, if after removing these actions we are left with a single action, then we have solved the problem without further calculation.

The simplest criterion used to cull actions is removing strictly dominated alternatives. The agent is rationally inattentive and so will never select an action that is strictly dominated in all possible realisations of the SPA. Hence, all actions that are strictly dominated across all realisations of SPA_t can be removed. This is done before making a guess for $\underline{\tilde{q}_t}$ and solving for $\overline{V}_{t+1}$, by removing any actions that are strictly dominated across all possible joint realisations of SPA_t and $\underline{\pi_{t+1}}$. Doing this before solving for $\overline{V}_{t+1}$ reduces unnecessary computational burden in the fixed point iteration needed to find that object. Having solved for $\overline{V}_{t+1}$, and hence having prediction for next period beliefs $\underline{\pi_{t+1}}$ given any action, I remove actions that are strictly dominated across all realisations of SPA_t.

Removing strictly dominated actions only uses the ordinal information encoded in the utility. Expected utility implies that utility encodes cardinal information as well, which can be exploited using the necessary and sufficient condition from Caplin et al. (2019). It is easily shown (see appendix 2.B.2) that if a there exists a decision $d^\star = (c^\star, l^\star)$ which satisfies:

$$\sum_{spa} \pi_t(spa) \frac{\exp\left(n^{(k)} \frac{((c^\star/n^{(k)})^\nu l^{\star 1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \beta \overline{V}^{(k)}_{t+1}(d^\star, X_t, spa, \underline{\pi_t})\right)}{\exp\left(n^{(k)} \frac{((c/n^{(k)})^\nu l^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \beta \overline{V}^{(k)}_{t+1}(d, X_t, spa, \underline{\pi_t})\right)} \geq 1, \tag{2.20}$$

for all other decisions $d = (c, l)$ then it is the only action taken $q(d^\star) = 1$. Unlike dropping strictly dominated alternative, which reduces the dimensionality and so makes solving equation 2.15 easier, checking equation 2.20 is only advantageous when the optimal behaviour is to take the same action in all realisations of SPA_t. As such the benefit of checking condition 2.20 depends on the problem faced and how frequently it reveals the optimal solution without needing to solve an optimisation. For the retirement model in this paper, it was found useful.

Finally, when sparsity does not provide a shortcut to a solution I employ sequential quadratic programming to solve 2.15, an approach to static rational inattention problems suggested by Armenter et al. (2019). Hence, bringing this together, a high-level summary of the solution algorithm is:

Remove d from $\mathscr{C}$ that are strictly dominated across all possible combinations of SPA_t and $\underline{\pi_{t+1}}$

if $|\mathscr{C}| = 1$ **then**

  Set $\underline{q_t}$ to degenerate distribution at $d \in |\mathscr{C}|$

else

  Set initial value of $\underline{\tilde{q}_t}$ and Error ¿ Tolerance

  while Error ¿ Tolerance **do**

    Solve for $\overline{V}_{t+1}$ given $\underline{\tilde{q}_t}$

    Remove d from $\mathscr{C}$ that are strictly dominated across all possible SPA_t given $\underline{\pi_{t+1}}$

    if $|\mathscr{C}| = 1$ **then**

      Set Error = 0 ¡ Tolerance and $\underline{q_t}$ to degenerate distribution at $d \in |\mathscr{C}|$

    else

      if there is an action d that satisfies 2.20 **then**

        Set Error = 0 ¡ Tolerance and $\underline{q_t}$ to degenerate distribution at d

      else

        Solve 2.15 using sequential quadratic programming for $\underline{q_t}$

        Set Error to distance between $\underline{q_t}$ and $\underline{\tilde{q}_t}$

        Update $\underline{\tilde{q}_t} = \underline{q_t}$

      end if

    end if

  end while

end if

Substitute $\underline{q_t}$ into 2.14 to solve for $\underline{p_t}$.

This hides many other computational complexities that arise from maximising the log sum exponential form. These can be found in appendix 2.C.

2.6 Estimation

The model is estimated by two-stage simulated method of moments. The first stage estimates, outside the model, parameters of the exogenous driving processes and the initial distribution of state variables; also, a small number of parameters are set drawing on the literature. Using the results of the first stage, the second stage estimates the remaining preference parameters $(\beta, \gamma, \nu, \kappa, \lambda)$ by the simulated method of moments.

2.6.1 First Stage

The parameters of the wage process, the state and private pension system, and the unemployment transition matrix are estimated outside the model. The curvature of the warm-glow bequest and the interest rate are taken from the literature.

Initial Conditions: To set the initial conditions of the model I need values for $a_t, w_t, AIME_t, ue_t$. Initial wages w_t are set to a draw from the estimated initial wage distribution (see below) and all agents start as employed ($ue_t = 1$). Assets a_t and initial average earning $AIME_t$ are initialised from the type-specific empirical joint distribution. For assets, the empirical counterpart used is household non-housing non-business wealth. Wave 5 of ELSA was linked to administrative data from the UK tax authority allowing me to observe the full working histories of these individuals and so construct a measure of $AIME_t$, but, as this happened for wave five and only 80% consented, this is only true for a subsample of individuals. To avoid dropping data, and to enable the model to match initial period assets, I impute $AIME_t$ with a quintic in wealth and a rich set of observed characteristics. To minimise the risk, inherent in this process, of overstating the correlation between these two key state variables I add noise onto the imputed values of $AIME_t$ that replicates the observed heteroscedasticity of $AIME_t$ with respect to assets (see appendix 2.D for more details).

Wage Equation: I assume wage data is contaminated with serially uncorrelated measurement error ($\mu_{j,t}$) leading to the following variant of equation 2.2 as data generation process:

$$\log(w_{j,t}) = \delta_{k0} + \delta_{k1}t + \delta_{k2}t^2 + \varepsilon_{j,t} + \mu_{j,t} \tag{2.21}$$

for individual j, of type k, in period t, where period t is indexed by female age and type k indicates whether high or low education and single or married. The parameters of the age-dependent deterministic component of the wage process $(\delta_{k0}, \delta_{k1}, \delta_{k2})$ are estimated by type-specific regression. The parameters of the stochastic component of the wage equation $(\rho_w, \sigma_\varepsilon, \sigma_{\varepsilon,55}, \sigma_\mu,)$ are estimated using a standard approach (e.g. Guvenen, 2009; Low et al., 2010) that chooses values that minimise the distance between the empirical covariance matrix of estimated residuals and the theoretical variance covariance matrix of $\varepsilon_t + \mu_{j,t}$.

Pension Systems: Both pensions are type-specific functions of average lifetime earnings. These are estimated on the $AIME_t$ measures constructed from administrative data, described above. However, as the state pension is relatively insensitive to education and the private pension relatively insensitive to marital status, to increase power I simplify the state pension to be marital-status-specific and the private pension education-specific. I estimate the private pension claiming age as the type-specific mean earliest age women are observed with private pension income.

Unemployment Transition Matrix I classify a woman as unemployed if she claims an unemployment benefit and estimate type-specific transition probabilities in and out of this unemployment state.

Stochastic State Pension Age: I estimate the probability of an increase in the SPA, ρ, on the cumulative changes to the original female SPA of 60 experienced by reform-affected cohorts. That is I select the ρ to minimise the mean error in SPAs given the data generating process is equation 2.11, getting an estimate of $\rho = 0.102$

Parameters Set Outside the Model The curvature of the warm-glow bequest is taken from De Nardi et al. (2010) and the interest rate from O'Dea (2018). Prices are deflated to 2013 values using the RPI. Survival probabilities are taken from the UK Office for National Statistic life tables and combined with ELSA data to estimate type-specific survival probabilities following French (2005), details in appendix 2.D.

2.6.2 Second Stage

In the second step, moments are matched to estimate the preference parameters: the isoelastic curvature (γ), the consumption weight (ν), the discount factor (β), and the bequest weight (θ), as well as the cost of attention (λ) in the version with costly attention.

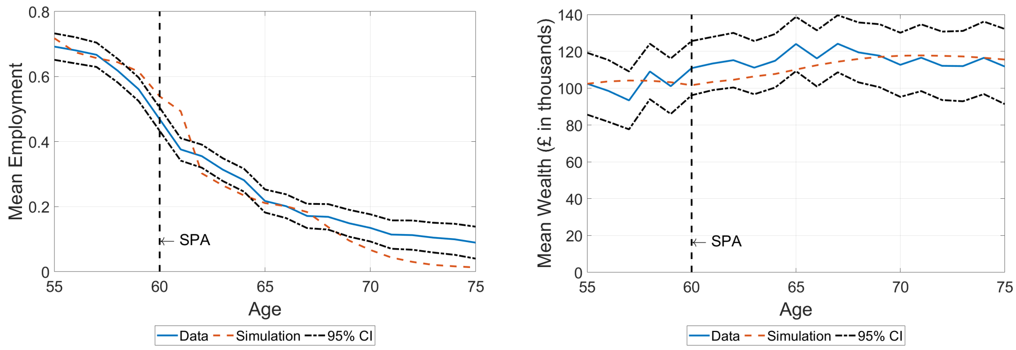
The moments used are the 42 pre-reform moments of mean labour market participation and asset holdings between 55 and 75. To avoid contamination by cohort effects or macroeconomic circumstances a fixed effect age regression was estimated which included: year of birth fixed effects, SPA-cohort specific age effects, the aggregate unemployment rate rounded to half a percentage point and an indicator of being below the SPA. The profiles used were then predicted from these regressions using average values for the pre-reform cohorts, details in appendix 2.D.

Due to the novel nature of the cost of attention parameter in the lifecycle literature, I investigated a range of values for λ alongside attempts to identify it from the reduction in self-reported SPA mean squared error between 55 and 58. Estimation of λ is done separately from targeting the other moments and holding the values of the other parameters constant. This has three principal advantages: one, it reduces computation; two, it uses the variation most directly affected by costly attention to identify λ; and, three, it does not use variation in labour supply to identify λ alleviating concerns the excess employment puzzle is directly targeted. This comes at the cost of not using all information to identify λ.

2.7 Results

Section 2.7.1 presents the goodness-of-fit to targeted moments and the model's ability to replicate the key empirical facts regarding excess employment sensitivity, mistaken beliefs and the relationship between them. Finally, section 2.7.2 presents implications of the model about patterns of information acquisition and the welfare cost of costly attention. Section 2.7.3 concludes with model policy predictions.

Figure 2.5: Fit to Targeted Profiles



(a) Employment Profile **(b)** Asset Profile

Notes: Model fit to targeted profiles. The empirical profile is for the pre-reform SPA cohort with a SPA of 60. The model was simulated with an unchanging SPA of 60, mimicking the conditions faced by this cohort.

2.7.1 Model Evaluation

This section presents the model fit and, given parameter estimates, investigates how well the model replicates the employment response to the SPA. Results of first stage estimation are in appendix 2.E.1.

The model with policy uncertainty attains a good fit to pre-reform employment and asset profile with $SPA = 60$ as shown in Figures 2.5a and 2.5b, and Table 2.4 contains the estimated parameter. The graphs for the baseline and the version with policy uncertainty together with rational inattention are very similar and are in appendix 2.E.2. However, in the response generated to the dynamic SPA, these model versions clearly distinguish themselves.

To investigate this response to the SPA, I simulate with $SPA = 60$, $SPA = 61$, and $SPA = 62$, as these are the SPAs reached in ELSA waves 1-7. Then, I repeated the regression analysis from Section 2.3.3 on the simulated data using an adaptation of equation 2.1 to the limitations of the model. That is, I estimate the treatment effect of being below the SPA on the probability of being in work using a two-way fixed effects difference-in-difference methodology that regresses on the treatment indicator, a full set of age and cohort fixed effects (not date as it is not distinct from age in the model) and the controls from the empirical specification having counter-parts in the model. As in Section 2.3.3, I repeat this on the subsample with

Table 2.4: Parameter Estimates

ν: Consumption Weight	0.439
	(-)
β: Discount Factor	0.985
	(-)
γ: Relative Risk Aversion	3.291
	(-)
θ: Warm Glow bequest Weight	100
	(-)

Notes: Estimated parameters from method of simulated moments when targeting the pre-reform labour supply and assets profiles.

above median empirical assets (£29,000) in the period before their SPA. The top panel of Table 2.5 contains the results. Column 5 repeats the empirical estimates of these treatment effects found in columns 1 and 2 of Table 2.1, hence the difference between column 5 and column 1 shows the baseline model struggles to match both the aggregate response to SPA and the correlation of this response with wealth.

This failure of the baseline reflects the excess employment sensitivity puzzle that led to investigations of policy uncertainty and costly attention. To examine their impacts separately, I introduce them sequentially. Column 2 shows policy uncertainty alone makes little to no difference. This is because the level of objective policy uncertainty is low; we observe changes to the SPA arrive infrequently. Both this version and the baseline fall short of matching the treatment effect in both the whole population and the population with above median assets at SPA but are closer to the lower treatment effect in the richer subpopulation. [26]

Column 3 shows the result with costly attention to the stochastic SPA, with a relatively arbitrary cost of attention of $\lambda = 3 \times 10^{-6}$ that fits these treatments' effect tolerably well.[27] The treatment effect in both the whole population and those with above median assets move significantly toward the data, falling in the confidence intervals.[28]

[26]Section 2.3.3 shows the response by the rich is puzzling ex-ante, Appendix 2.C shows, if we directly target treatment effects, the baseline matches that of the whole population but not of the rich. With these parameters estimated, however, the baseline struggles most with the aggregate.

[27]Larger attention cost match them slightly better, but higher values were deemed unrealistic, and improvements were marginal past a point.

[28]Conicendently, the treatment effects are the same to three decimal places, there is no type in the

Table 2.5: Model Predictions for Different Costs of Attention

	Baseline	Policy Uncert.	Costly Attention $\lambda = 3 \times 10^{-6}$	Costly Attention $\lambda = 1.3 \times 10^{-7}$	Data
	Treatment Effect being below SPA on employment				
Whole Population *[95% C.I]*	0.019	0.023	0.052	0.037	0.080 [0.044,.0116]
Assets >£29,000 *[95% C.I]*	0.027	0.032	0.052	0.040	0.061 [0.018,.0103]
	Reduction in MSE of SPA Self-Reports				
MSE drop 55-58 *[95% C.I]*	-	-	-0.85	1.53	1.69 [0.31,3.36]
Coefficient	Treatment Effect Heterogeneity by SPA Error				
Interaction *[95% C.I]*	-	-	-0.001	-0.022	-0.066 [-0.094,-034]

Notes: The columns show results from two costs of attention. The top panel shows labour supply response across the wealth distribution as per Table 2.5. The second panel shows the reduction in self-reported SPA MSE between 55 and 58. The bottom panel shows, in the interaction term, the heterogeneity in labour supply response to the SPA by self-reported SPA error at age 58.

SPA self-reports in ELSA offer an opportunity to improve on this arbitrary value of λ, as they offer clear and direct identifying variation. Exploiting this, I identify λ from the reduction in mean squared error in self-reported SPAs between ages 55 and 58 for the cohort with a SPA of 60, the cohort simulated during estimation. The middle panel of Table 2.5 shows these numbers. Column 4 shows that for a smaller cost of attention $\lambda = 1.3 \times 10^{-7}$, the degree of learning, captured by the reduction in mean squared error, is well matched. For this lower value of λ, the fit for the employment response to the SPA is worse, although it still improves on the fit of the baseline and the model with only policy uncertainty. For the larger value of λ in column 3, we see that knowledge of the SPA actually gets worse between age 55 and 58, indicating that if any learning is happening, it is getting drowned out by drift from agents updating with the known laws of motion.

That people better informed of their SPA in their late 50s have a smaller labour supply response at their SPA in their 60s (see Table 2.3) was an impetus to investigating the role of endogenous information in this excess sensitivity puzzle. A natural question is whether the model replicates this relationship. Two countervailing model forces exist linking SPA knowledge to the labour supply response at the SPA. On the one hand, SPA knowledge is endogenous, implying those whose actions de-

table.

pend least on the SPA acquire the least information about it. On the other, comparing two ex-ante equivalent households where, by luck, one ended up worse informed than the other, the worse informed household receives a larger shock upon discovering their SPA and so has a larger reaction. Which dominates determines whether the model generates a positive or negative relationship between SPA knowledge and the labour supply response at the SPA. The bottom panel of Table 2.5 shows a negative relationship for the smaller λ and no relationship for the larger.

The fact that the model replicates the key facts from the data indicates it has the mechanism required to explain the data. A limitation is that it requires different values of λ to replicate different facts. Two explanations spring to mind: the levels of some incentives may be misaligned, or other mechanisms may contribute to the excessive employment response. Unidimensional policy uncertainty is a massive simplification, many aspects of pension policy are uncertain, which may explain misalignment of incentives. Section 2.8 enriches pension policy uncertainty, but data limitations make the extension necessarily more speculative. Section 2.7.3.2 introduces a norm to retire at SPA alongside costly attention.

2.7.2 Model Implications

Size of informational frictions λ is not easily interpretable having as natural units of utils per bit. Gabaix (2019) discusses this difficulty and suggests converting attention cost to implied misperceptions of prices, but this approach is not applicable when the object subject to attentional costs is not traded as with the SPA. It is widely appreciated that utils are not interpretable; it is less widely appreciated that expressing attention cost per bit exaggerates the cost because learnable bits of information are scarcer in models than in reality. To account for both these issues, I express λ as the compensating assets that increase a household's utility as much as learning their SPA today.

Table 2.6 shows summary statistics of the distribution of compensating assets for $\lambda = 1.3 \times 10^{-7}$ and $\lambda = 3 \times 10^{-6}$. For $\lambda = 1.3 \times 10^{-7}$, the costs range from £0.32 at the 5th percentile to £37.74 at the 95th with a mean of £15.37, but there is substantial heterogeneity. The correlations with assets are negative, indicating

Table 2.6: Summary Statistics of Attention Cost Converted to Compensating Assets (£)

λ	Mean	SD	Median	5% Percentile	95% Percentile	Cor. with Assets
1.3×10^{-7}	£15.37	£11.56	£12.43	£0.32	£37.74	-0.18
3×10^{-6}	£59.93	£97.24	£23.75	£0.18	££262.34	-0.23

Notes: Distribution of costs of attention as measured by compensating assets producing equivalent utility to learning your SPA today. Shown for two different costs of attention.

that information frictions impose the highest cost on the poorest members of society. For $\lambda = 3 \times 10^{-6}$, qualitatively a similar picture arises, but the costs are higher. These costs give an upper bound on welfare gains from reducing uncertainty by, for example, sending information letters. The gains are modest, but so are the marginal costs. Once the fixed administrative and technological costs have been paid, marginal costs can include little more than postage which costs £0.68.

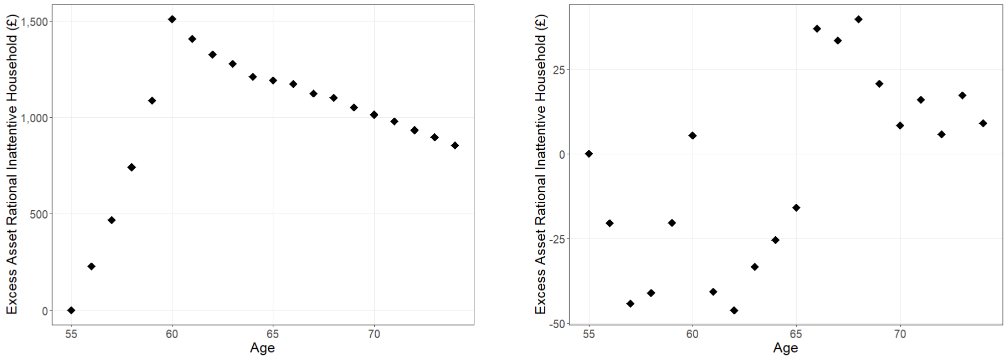
Who learns and when The model allows us to investigate who learns and when. To investigate this, I run a pooled regression of bits of information acquired on the states of the model and indicators of age for women 55-59 with a SPA of 60. The results are unsurprising and, for this reason, are in the appendix, Table 2.E.4. On average: a positive amount of information is gathered, those with more alternative resources (assets and labour income) gather less information, and those with higher future pensions (AIME) gather more. Most information is acquired at age 55, declining until age 58 and then jumping back up at age 59, immediately before the SPA.

2.7.3 Model Predictions

2.7.3.1 Optimal Savings

Retirement preparedness is a concern, particularly in countries where private savings accounts form a growing fraction of retirement income, but the academic literature fails to reach a consensus on whether households under or over save for retirement (e.g. Scholz et al., 2006; Crawford and O'Dea, 2020). As this paper takes age 55 assets as given, its ability to answer this question is limited, but comparing the savings of the rationally inattentive household to the frictionless benchmark

Figure 2.6: Excess Saving



(a) $\lambda = 3 \times 10^{-6}$ **(b)** $\lambda = 1.3 \times 10^{-7}$

Notes: Excess saving relative to model with only policy uncertainty with the true SPA set at 60 throughout.

(i.e. with only policy uncertainty) illuminates how informational frictions impact retirement preparation.

Figure 2.6a shows that when attention cost are highs ($\lambda = 3 \times 10^{-6}$), households oversave for retirement. Rather than learning at this high cost, the SPA households insure against policy uncertainty. For these simulations, the SPA was kept constant at 60 throughout, and as can be seen, once the households reach the SPA and the policy uncertainty is resolved, the households begin to run down their assets. Figure 2.6b show excess saving when the cost of attention is low ($\lambda = 1.3 \times 10^{-7}$). The mistakes are, unsurprisingly, much smaller, and there is not such a clear pattern. Averaging across households, they slightly undersave for retirement and then increase their assets later. This additional saving later in life results from lower accumulated lifetime earnings meaning lower pension income.

2.7.3.2 Increasing Old Age Participation with the Pension Ages

Rising old-age dependency ratios have made increasing labour force participation of older individuals a policy priority of governments around the world (e.g. OECD, 2000; Barr and Diamond, 2009; Landais et al., 2021), and statutory retirement ages are seen as a key tool to achieve this. Costly attention increases the responsiveness of employment at the SPA, so it seems natural it makes the SPA a more effective tool to achieve this goal. This is not necessarily the case.

Table 2.7: Additional Mean Employment from Increasing SPA from 60

Post reform SPA	Without Passive Household				Fraction of Passive Household			
	Policy Uncert.		Costly Atten.		Policy Uncert.		Costly Atten.	
	Prime (55-65)	All (55-79)	Prime (55-65)	All (55-79)	Prime (55-65)	All (55-79)	Prime (55-65)	All (55-79)
61	0.07	0.09	0.06	0.08	0.11	0.12	0.11	0.12
62	0.15	0.18	0.15	0.18	0.23	0.26	0.23	0.26
63	0.18	0.22	0.19	0.24	0.32	0.35	0.30	0.34
64	0.41	0.50	0.33	0.43	0.63	0.71	0.50	0.59
65	0.40	0.49	0.33	0.42	0.63	0.70	0.50	0.58

Notes: Mean additional years of employment over different horizons (55-65 and 55-79) resulting from increasing the SPA from 60 to the age in the first column for the model with and without costly attention. The right panel additionally includes a fraction of naive-passive agents who retiree at their SPA regardless.

The left panel of Table 2.7 shows the change in mean employment resulting from increasing the SPA from 60 to a SPA in the range 61-65 for the model with $\lambda = 1.3 \times 10^{-7}$ and with policy uncertainty alone. Averages are shown over prime working years (55-65) and all working life (55-79). Focusing on the SPA increase to 65, with costly attention mean employment increases by 0.33 years over 55-65 and 0.42 years over 55-79; this compares to 0.40 and 0.49 additional years in the model without informational frictions. Neglecting costly attention overestimates the employment increase by 23% over 55-65 and 15% over 55-79. The reason for a smaller response with costly attention is the rationally inattentive household is less aware of SPA changes, so increases labour supply less in the lead-up to their new SPA. Post their new SPA, the rationally inattentive agent works more to compensate. This reduces but does not overturn the difference over 55-79 because of imperfect intertemporal substitutability and lower employment at older ages.

Costly attention generates a smaller overall increase in employment but a larger response at the SPA because much of the bunching at SPA reflects intertemporal shifting of employment. Fully aware households immediately internalise changes to the SPA, increasing labour supply when the woman is in her 50s. The rationally inattentive household only partially responds until they realise, much closer to the SPA, the need to make up for lost time. Figure 2.7 illustrates these dynamics for an increase of the SPA to 62.

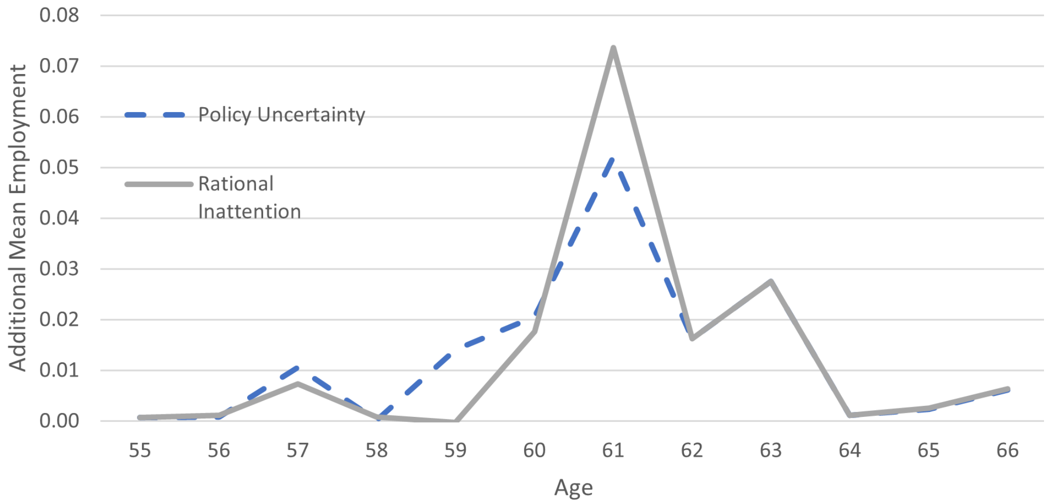
Costly attention only partially explains excess employment sensitivity (see Table 2.5), leaving room to doubt these labour supply predictions. Also, evidence framing effects, or norms, affect the employment response to pension ages (e.g. Seibold, 2021; Gruber et al., 2022) should be considered alongside evidence presented in this paper for the importance of mistaken beliefs. A simple way to include a norm to stop working at the SPA, alongside costly attention, is to have a fraction of passive decision makers in the style of Chetty et al. (2014) or Lalive et al. (2017). I introduce a fraction of naive-passive agents, not anticipating their passive retirement, and compare predictions of models with and without costly attention after the inclusion of these naive-passive agents. The model without costly attention requires 11% of women are passive to replicate the employment response to the SPA, whilst with $\lambda = 1.3 \times 10^{-7}$ requires only 8%.

The left panel of Table 2.7 shows these results. Focusing again on the SPA increase to 65, costly attention again predicts smaller employment increases, but now the difference is larger. When the SPA is 60, passive retirees decrease employment post 60, so a larger passive share amplifies the employment increase over 60-64 from changing the SPA from 60 to 65. Post 65, a larger passive share mutes the increase, but due to generally declining employment, it is the earlier effect that dominates. The model with costly attention predicts an additional 0.58 years of mean employment between 55 and 79, in contrast to 0.70 without costly attention, and an additional 0.50 years of mean employment between 55 to 65, compared to 0.63 without. Neglecting costly attention overestimates the employment increase by 27% over 55-65 and 20% over 55-79. So, ignoring costly attention leads to a 27% over prediction of the employment response to the SPA increase during prime working years and a 20% overall over prediction.

2.8 Extension: Deferral Puzzle

Section 2.7 shows that, even though the model can explain each feature of the data in isolation, it requires different costs of attention to replicate different features

Figure 2.7: Increase in Employment from Increasing SPA 60 to 62



Notes: For each model, the difference in employment increase between simulations of households with a female SPA of 60 and those with a female SPA of 62.

pointing toward misalignment in incentives. As the stochastic State Pension Age (SPA) understates policy uncertainty about the State Pension, this is a natural place to look for this misalignment.

For this reason, I introduce learning and uncertainty about another aspect of the state pension system: the actuarial adjustments to benefits from deferring. Combined with a claiming decision, this not only helps to align incentives by making the model more realistic but also helps explain the deferral puzzle (detailed below). Since the adjustments rate becomes irrelevant upon claiming, rational inattention to this aspect of the pension system speaks directly to this puzzle because calculations implying deferral is actuarially favourable ignore the attention benefits of claiming: claiming removes the need to pay attention to this adjustment rate. The model of Section 2.4.2 does not incorporate such a mechanism for two reasons. Firstly, it does not include a benefit-claiming decision. Secondly, the only source of uncertainty subject to an attention cost is the SPA, and once it is reached, the uncertainty resolves, irrespective of claiming. The simplest extension that contains this new incentive to claim is presented in the rest of this section, along with some results.

2.8.1 Deferral Puzzle

By deferral puzzle, I mean the fact deferral of state pension benefits was uncommon despite an extremely generous adjustment between April 2005 and April 2016. During this period, state pension benefits increased by 1% for every 5 weeks deferred implying an annual adjustment of 10.4%. This is an extremely generous actuarial adjustment, and yet 86.7% of women observed over their SPA in ELSA during the period had claimed by their first post-SPA interview.

What exactly constitutes actuarially fair depends on life expectancy and the interest rate, but at all plausible levels, this adjustment was generous. For the women who reached their SPA during this window, life expectancy at SPA was somewhere in the range 23 to 25 years. Taking the conservative estimates for mean life expectancy of 23 years, a benefit adjustment of 10.4% p.a. deferred is advantageous at any interest rate up to 9%. During this period, the Bank of England base rate never exceeded 5.75% and from March 2009 until the end sat at the historic low of 0.5%. Hence, at any plausible commercial interest rate, an adjustment of 10.4% was actuarially advantageous.

Even for the small group of women observed deferring, the duration of deferral was short. Sticking to the conservative estimates of 23 years of life expectancy at SPA and the upper bound of 5.75% for the interest rate implies an optimal deferral of 9 years. The median observed deferral is 2 years, and 99.54% of deferrers claimed within 8 years of the SPA.

Of course, these calculations are all done for mean life expectancy, which masks the heterogeneity in life expectancy. However, heterogeneity alone is not a plausible explanation as it would mean 86.7% of women had significantly below mean life expectancy, implying implausible skewness in the distribution of life expectancy at SPA.

2.8.2 Model and Estimation

Benefit claiming is a binary decision and having claimed is an absorbing state: once an individual claims the state pension, they cannot unclaim. Benefit claim is only an option once past the SPA, and, to keep the problem tractable, an upper limit of

Table 2.8: Parameter Estimates - Extension

ν: Consumption Weight	0.5310
	(-)
β: Discount Factor	0.9852
	(-)
γ: Relative Risk Aversion	2.0094
	(-)
θ: Warm Glow bequest Weight	20,213
	(-)

Notes: Estimated parameters from method of simulated moments for the model extension with a stochastic deferral rate and a benefit claiming decision.

70 is placed on deferral.

Stochastic deferral adjustment is modelled as iid with two points of support. Having only two points of support limits the growth of the state space resulting from solving the model with different values of the adjustment rate to a factor of two. Having the uncertainty be iid means that beliefs do not enter as a state variable. Instead, the true probabilities form beliefs in each period: yesterday's learning is not relevant to today's state of the world. This also avoids a fundamental identification problem as there is no data on beliefs about adjustment rates. As benefit claiming is an absorbing state, an indicator of having claimed or not also expands the state space.

The two points of support are chosen as 10.4% and 5.8%, the actuarial adjustment from 2006 to 2016 and post-2017 respectively. The probability of being offered the higher actuarial adjustment of 10.4% is chosen to match the average actuarial adjustment since 1955, resulting in a probability of 0.415. Deferral rules are taken from Bozio et al. (2010) and since earlier deferral rules were previously stated in absolute rather than percentage terms, the ONS time series of state pension spending going back to 1955 (https://www.gov.uk/government/publications/benefit-expenditure-and-caseload-tables-2021) is used to work out implied average percentage deferral adjustments.

The model with policy uncertainty, to the stochastic SPA and adjustment rate, is then re-estimated to match the same pre-reform employment and assets profiles

Table 2.9: Model Predictions - Extension with benefit claiming and uncertain deferral

	Costly Attention	Data
Population	Treatment Effect for being below SPA on employment	
Whole Population	0.0416	0.080
Assets >Median(£29,000)	0.0903	0.061
Age	Variance of SPA Answers	
55	2.985	2.852
58	1.795	1.180
Coefficient	Treatment Effect Heterogeneity by SPA Error	
Treatment Effect	0.0532	0.157
Interaction	-0.0111	-0.023

Notes: Costly attention refers to the model with, additionally, a cost of information acquisition about the stochastic policy. The top panel shows labour supply response across the wealth distribution as per Table 2.5. The second panel shows the reduction in self-reported SPA between 55 and 58. The bottom panel shows, in the interaction term, the heterogeneity in labour supply response to the SPA by self-reported SPA error at age 58.

with a constant realisation of 10.5% for the deferral adjustment, which was the deferral rate these cohorts faced. Parameter estimates are in Table 2.8 and, for these values, only 6.2% of individuals claim the state pension before the mandatory claiming age of 70, much lower than the 99% plus claiming seen in the data.

Next, I introduce costly attention with a cost of attention corresponding to approximately £10 of consumption to the median consuming household to be fully informed. This increased the number voluntarily claiming to 22.2%, approximately a fourfold increase on the model without informational frictions, but still short of the rate observed in the data. As can be seen in Table 2.9, this cost of attention produced a relatively good fit along all dimensions of interest.

2.9 Conclusion

This paper shows that accommodating one empirical regularity, mistaken beliefs, into a model of retirement helps explain another puzzling one, the sensitivity of employment to the State Pension Age (SPA). These mistaken beliefs result from learning about an objectively uncertain and changeable pension policy whilst subject to information frictions. This interplay of objective policy uncertainty and subjective beliefs generates a larger employment response at the SPA: reaching SPA resolves the policy uncertainty, and the size of the resulting shock is larger because of mistaken beliefs.

In doing so, I am the first, to the best of my knowledge, to solve a dynamic rational inattention model with endogenous heterogeneous beliefs, and weaving together recent theoretical results into a solution method is a contribution of this paper. This allow me to show endogenous beliefs are crucial to explaining the observation that people more mistaken about their SPA have smaller employment responses upon reaching it: they are mistaken because they choose not to learn about it as it is irrelevant to their actions.

I use data on these mistaken beliefs to identify the cost of attention, hence estimate a model of learning on belief data. This approach to the belief-preference identification problem avoids loading all explanations onto preferences. The mean household's estimated willingness to pay to learn today's SPA is £15. Since the marginal costs of communicating pension policy via information letters are closer to £1, the model indicates this policy is welfare improving. The small cost of attention generates relatively large changes in employment because households near retirement are close to the participation margin.

Policy experiments comparing changes in employment resulting from SPA increases in versions of the model with and without information frictions show costly attention increases the employment response *at* the SPA by intertemporally shifting part of the informed agent's employment response forward but can decrease the overall response. Informed agents increase labour supply immediately; those subject to costs of learning, being less well-informed, respond nearer their SPA. Ignoring costly attention overstates the SPA's effectiveness at increasing old age employment by up to 27%, which illustrates another reason to send policy information letters. Informed individuals' behaviour is more predictable.

Finally, I present an extension of the main model with a mechanism to explain another puzzle: that people do not take up more than actuarially advantageous deferral options. The insight offered by this extension is that the assertion that deferral is actuarially advantageous ignores the cost of paying attention to pension policy which claiming avoids. Hence this assertion omits an incentive to claim.

Appendix

2.A Additional Empirical Details

2.A.1 Additional Institutional Details

2.A.2 Equity Acts

The equality Act (2006) banned mandatory retirement below age 65. Since all women observed to age past their SPAs in ELSA waves 1-7 had a SPA of 60-63, it would have been illegal for their SPAs to have coincided with a compulsory retirement age. The equality Act (2010) went further and banned all compulsory retirement ages with a handful of specific exceptions known as employer justified retirement ages (EJRA).

Since these EJRA need to be over 65 and all SPAs considered in the empirical section are below this age, they are not strictly relevant to the empirics. However, some background and anecdotes about them may illustrate how strict UK age discrimination law is as regards forcing people to retire. Seldon v Clarkson, Wright and Jakes (2012) clarified exactly when EJRA are justified. It laid out three criteria an EJRA must meet: one, the reason justifying the EJRA must be an objective of public interest (e.g. intergenerational fairness), not just of the firm; two, this objective must be consistent with the social policy aims of the state; and, three, an EJRA must be a proportionate means to achieve this objective.

The plaintiff in Seldon v Clarkson, Wright and Jakes (2012) was a partner in a law firm, and it was judged that this EJRA was justified. Documented cases of EJRAs are relative few; apart from partners in law firms, two of the most discussed EJRA are at the UK top Universities: Oxford and Cambridge. Other UK univer-

Table 2.A.1: Effect of SPA on Employment: Heterogeneity by Wealth

Below SPA	**0.332**
s.e	(0.0096)
p=	.000
Below SPA× NHNBW	**-3.97×10^{-7}**
s.e	(7.42e-08)
p=	.000
Obs.	7,947
Indv.	3,846

Notes: Table shows the results of running the two-way fixed effect specification in 2.1 interacting a continuous measure of NHNBW with the treatment and all fixed effects and controls.

sities appear to have removed compulsory retirement requirements ages, and interestingly Oxford recent lost an employment tribunal that judged that their EJRA was not justified. Ewart v University of Oxford (2019) found that although the objective of Oxford's EJRA (intergenerational fairness) was valid, an EJRA was not a proportionate way to achieve this due to limited demonstrated effectiveness weighed against its clearly detrimental impacts. Hopefully, this goes someway to illustrate that UK law treats forced retirement very seriously as age discrimination and that the few expectations made are precisely that: exceptional.

2.A.3 Excess Employment Sensitivity

2.A.3.1 Continuous interaction

Only considering two asset groups, above and below median assets, is an arbitrary dichotomisation and leads to a loss of information. For this reason, Table 2.A.1 shows results for a specification containing an interaction between being below the SPA with the continuous NHNBW variable. As can be seen, this interaction term is highly significant but tiny, an additional £$(\frac{0.01}{3.97\times10^{-7}})$ or £25,118 of NHNBW is required to decrease the treatment effect by 1 percentage point. This indicates, unsurprisingly, that wealth does impact how important the SPA is to someone's retirement decision but that liquidity constraints cannot completely explain the sensitivity of labour market exits to the SPA. For example, these results imply a woman from a household at the 95% percentile of the distribution, with £409,000 in NHNBW, would experience a significant treatment effect of a 0.162 increase in her probability of being in work from being below the SPA. NHNBW of £409,000 seems ample to smooth labour supply over the horizon of one to three months. So, although wealth matters for the impact of the SPA on employment, it seems liquidity constraints cannot explain away the effect.

2.A.3.2 Restricted Asset Categorisation

As the goal of investigating treatment effect heterogeneity by asset holdings is to understand the role played by liquidity constraints, the main text is restricted to NHNBW. However, parts of NHNBW can be illiquid, and so in Table 2.A.2 repeat the analysis but for a more restricted asset category, very liquid asset, which is only assets that can reasonably be liquidated in a matter of weeks. As can be seen, the results are qualitatively very similar to those using NHNBW and do not support liquidity constraints alone explaining away the treatment effect. The treatment effect for those with above median assets is still positive, and although the difference between the two subgroups is now significant, column 4, containing the continuous interaction terms, shows that, again, this heterogeneity is too weak for the treatment to be completely explained by liquidity constraints.

2.A.3.3 Bad Control Concerns

Bad controls concerns are particularly important in the case of DID. Some take the view that only time-invariant controls should be included because controls imply that we are imposing parallel trends conditional on that variable.

To address these concerns here, I take a broad brush solution and run a version of the mode without any controls, showing that qualitatively the conclusions drawn are not impacted by the presence or otherwise of controls.

Table 2.A.3 shows the results of this exercise of dropping controls. As can be seen, the results are very little changed from those with controls.

2.A.3.4 Imputation Approach to DID

Using a two-way fixed effects regression to estimate difference-in-difference models assumes treatment effect heterogeneity across time and across units. When the

Table 2.A.2: Effect of SPA on Employment: Heterogeneity by VLA

	(1)	(2)	(3)	(4)
Below SPA	**0.080**	**0.047**	**0.139**	**0.331**
s.e	(0.0223)	(0.0391)	(0.0339)	(0.0096)
p=	.038	.022	.000	.000
Below SPA×(VLA.>Med.)			**-0.092**	
s.e			(0.0380)	
p=			.016	
Below SPA× VLA.				**-5.27×10^{-7}**
s.e				(9.25e-08)
p=				.000
Obs.	23,641	6,707	23,641	23,641
Cohort	132	90	132	132

Notes: Column (1) shows the results of running the two-way fixed effect specification in 2.1 as a random-effects model with controls used: a full set of marriage status, years of education, education qualifications, and self-reported health dummies; partners age; partners age squared; the aggregate unemployment rate during the quarter of interview; dummies for partner eligible for SPA, and for being one and two years above and below SPA; and assets of the household. Column (2) repeats this regression on the subsample with above median Very Liquid Assets (VLA) in the last interview before their SPA. Column(3) tests whether the different treatment effects observed in columns (1) and (2) are different by introducing an interaction between being below the SPA and having above median VLA. Column(4) includes an interaction between being below SPA and a continuous measure of VLA.

timing of treatment induces the variation in treatment, as is the case in this paper, violations of these assumptions can lead to estimated treatment effects being nonsensical combinations of the individual level treatment effect. This issue, and related issues, have been flagged by a recent wave of literature, but thankfully this literature also proposes a solution that relaxes these assumptions.

Here I implement the imputation approach of Borusyak et al. (2021). This approach allows for never treated but does not allow for always-treated units. To be consistent with this, I redefine treatment as being over the SPA, and in a first step, verify that this only changes the sign of the results in the main text, as we would expect.

Figure 2.A.1 shows the dynamic treatment effects before and after the SPA. There is no indication of violated parallel trends of anticipation effects as none of the pre-SPA treatment effects are significantly different from zero. Indeed jointly

Table 2.A.3: Effect of SPA on Employment: Heterogeneity by NHNBW no controls

	(1)	(2)	(3)	(4)
Below SPA	**0.081**	**0.068**	**0.107**	**0.334**
s.e	(0.0190)	(0.0223)	(0.0311)	(0.0095)
p=	.000	.003	.001	.000
Below SPA×(NHNBW.>Med.)			**-0.039**	
s.e			(0.311)	
p=			.016	
Below SPA× NHNBW.				**-3.76×10^{-7}**
s.e				(8.16e-08)
p=				.000
Obs.	23,613	7,273	23,613	23,613
Cohort	132	100	132	132

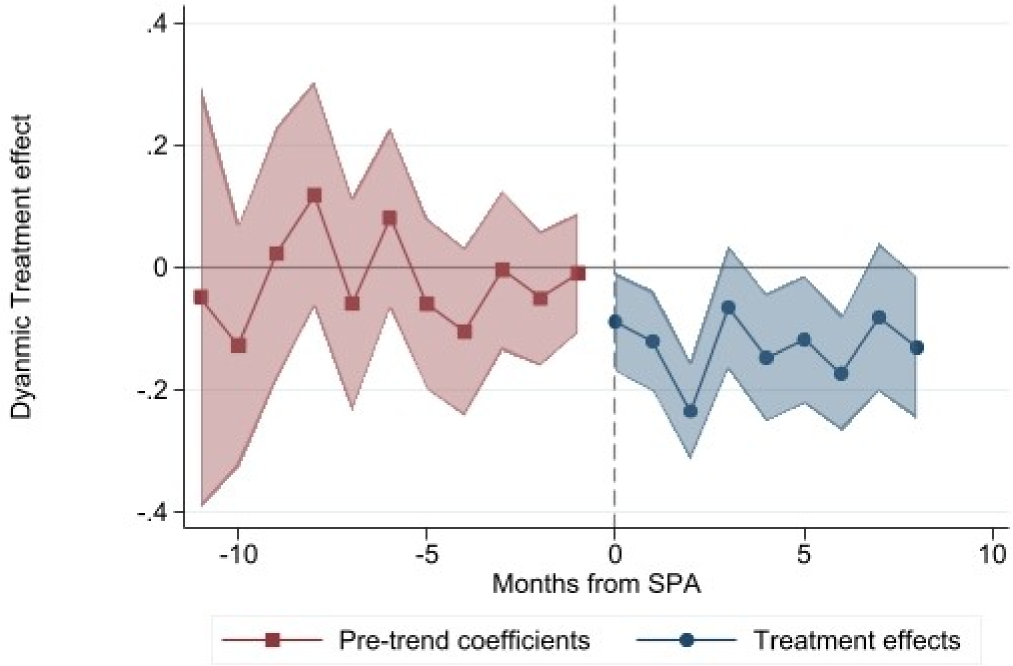
Notes: Column (1) shows the results of running the two-way fixed effect specification in 2.1 as a random-effects model with controls used: a full set of marriage status, years of education, education qualifications, and self-reported health dummies; partners age; partners age squared; the aggregate unemployment rate during the quarter of interview; dummies for partner eligible for SPA, and for being one and two years above and below SPA; and assets of the household. Column (2) repeats this regression on the subsample with above median Very Liquid Assets (VLA) in the last interview before their SPA. Column(3) tests whether the different treatment effects observed in columns (1) and (2) are different by introducing an interaction between being below the SPA and having above median VLA. Column(4) includes an interaction between being below SPA and a continuous measure of VLA.

testing for violations of parallel trends fails to reject the null of parallel trends ($p =$ .799). Conversely, 7 of 9 post-SPA treatment effects are individually significant, and we can easily reject the null ($p = .000$) of them being jointly zero. The graph also doesn't provide much indication that the post-SPA treatment effects differ from each other, although we can reject that hypothesis ($p = .198$).

Figure 2.A.2 looks at whether these individual treatment effects vary between waves. The treatment effects look quite uniform across waves, although, again, we can reject this hypothesis ($p = .137$). However, neither violation of homogeneity seems to serve, and generally, the graphs look supportive of the interpretation of a homogenous treatment effect that turns on at the SPA (as assumed in the baseline), although the statistical test show this is only an approximation.

If you are more concerned about the violations of homogeneous treatment effects, then these results show that even allowing for arbitrary heterogeneity, there is something special happening at the SPA which is difficult to explain in standard

Figure 2.A.1: Dynamic Treatment Effects by Time from SPA



Notes: The average at a given time from SPA of the dynamic individual level treatment effects estimated using the imputation approach.

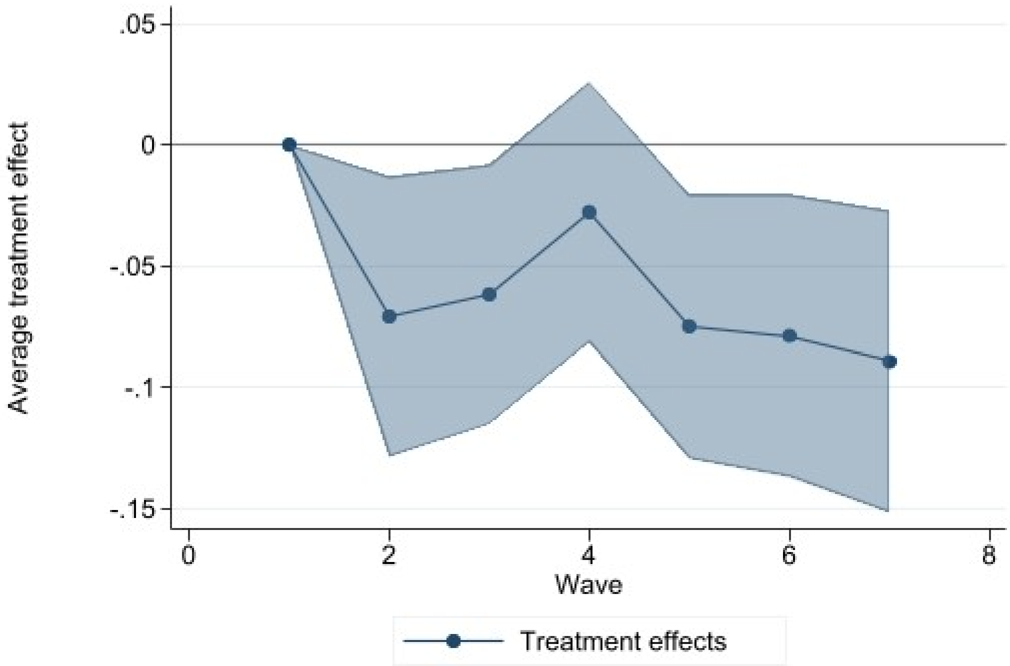
complete information models.

2.A.3.5 Health, Wealth, and Private Pensions

The rest of the section is concerned with addressing other potential explanations for the sensitivity of employment to the SPA in a standard complete information framework. Specifically, I consider if wealth, health, or private pensions can explain the labour supply response to the SPA.

Wealth effects play an important role in determining labour supply, and women who have a later SPA are lifetime poorer. The puzzle is not that they have a higher labour supply; the puzzle is that their labour supply response should be concentrated at the SPA, the change in SPA having been announced over 15 years prior to any affected individual reaching their SPA. In standard complete information life-cycle models, the affected individuals should have a higher labour supply due to the wealth effect, but the response should be spread over their life, not concentrated at the SPA itself. In equation 2.1 differences in lifetime wealth, including those in-

Figure 2.A.2: Average Treatment Effect by Wave



Notes: The within wave average of the individual level treatment effects estimated using the imputation approach.

duced by SPA differences, between year-of-birth cohorts are absorbed by the cohort effects. Hence, the only wealth differences the treatment effect will detect are between individuals with the same year of birth. To generate the observed treatment effect only with wealth difference induced by the SPA within the same year-of-birth cohort, the wealth effect would have to be massive. To see this, note the control for an individual is someone with the same age to within a quarter; the treatment effect only picks up a very short-run response whilst the wealth effect generates a response that is spread out over the life-cycle. Under the assumption this labour supply response is generated purely by a wealth effect, we can calculate an implied marginal propensity to earn out of unearned income (MPE). The implied MPE is about -0.3. This is on the high end of estimates in the modern literature (e.g. Cesarini et al., 2017), but becomes impossibly high when you factor in that this should only be catching the final two-to-three month tail end of a labour supply response that is spread out over 15-20 years. Wealth effects explaining away the treatment also seems inconsistent with the limited impact of wealth on the treatment effect;

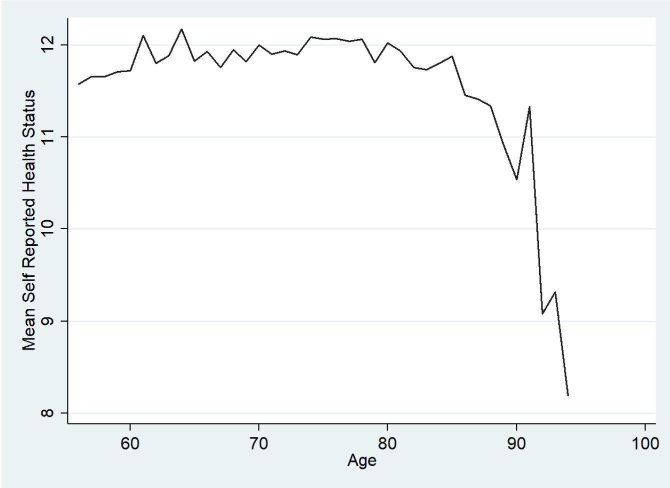


Figure 2.A.3: Self Reported Health Profile

Table 2.A.4: Heterogeneity by Health

	Coeff	s.e.	p=
Below SPA	0.093	0.0429	0.030
Below SPA×(V.good Health)	-0.027	0.0371	0.461
Below SPA×(Good Health)	0.016	0.0390	0.689
Below SPA×(Fair Health)	-0.056	0.0422	0.186
Below SPA×(Poor Health)	-0.145	0.0495	0.003

as wealth increases, the change induced by the SPA represents a smaller fraction of their total assets. Hence, we would expect the treatment effect to decrease more sharply with wealth.

Health is a major determinant of retirement behaviour (e.g. De Nardi et al., 2010). However, there is no reason to suspect it interacts with the SPA, so no reason for it to explain employment's sensitivity to the SPA. Furthermore, during the period studied, the SPA was in the range 60-63, and, at the mean, health status does not start to deteriorate until later in life. This can be seen in Figure 2.A.3, which shows the age profile of health status. All the same, as it is such an important factor in retirement Table 2.A.4 looks at heterogeneity in labour supply response to the SPA by health status. As can be seen, the labour supply response is only significantly different for those with the poorest health group. This group only make up ¡7 % of the sample, and if dropped, do not qualitatively change the results.

Finally, the timing of private pension eligibility is important for retirement

Table 2.A.5: Effect of SPA on Employment:
Less than £2,000 in DB scheme

Below SPA		**0.117**
	s.e	(0.0369)
	p=	.002
Below SPA×(NHNBW.>Med.)		**-0.049**
	s.e	(0.0592)
	p=	.413
Obs.		3,735
Indv.		2,197

choices. However, occupational pension schemes are very unlikely to have adjusted their pension ages in line with the female SPA because private pensions do not generally offer different eligibility ages to men and women[29], and this reform only changed the female SPA. Still, checking for a correlation between the SPA and normal pension ages (NPA) of private pension schemes would be desirable. Checking this directly in ELSA is complicated by the fact that only self-reported NPAs are available. For the SPA, where alongside self-reports, we know an individual's true SPA, these self-reported ages are unreliable, as is documented in Section 2.3.4. However, only defined benefit pension systems have NPAs, as defined contribution schemes can be accessed from age 55. Hence, dropping everyone with over £2,000 in a defined benefit scheme from the sample rules out an unlikely correlation between the female SPA and pension schemes NPAs from explaining the results. This is done in Table 2.A.5, and as can be seen, despite the loss of power, the treatment remains present and significant.

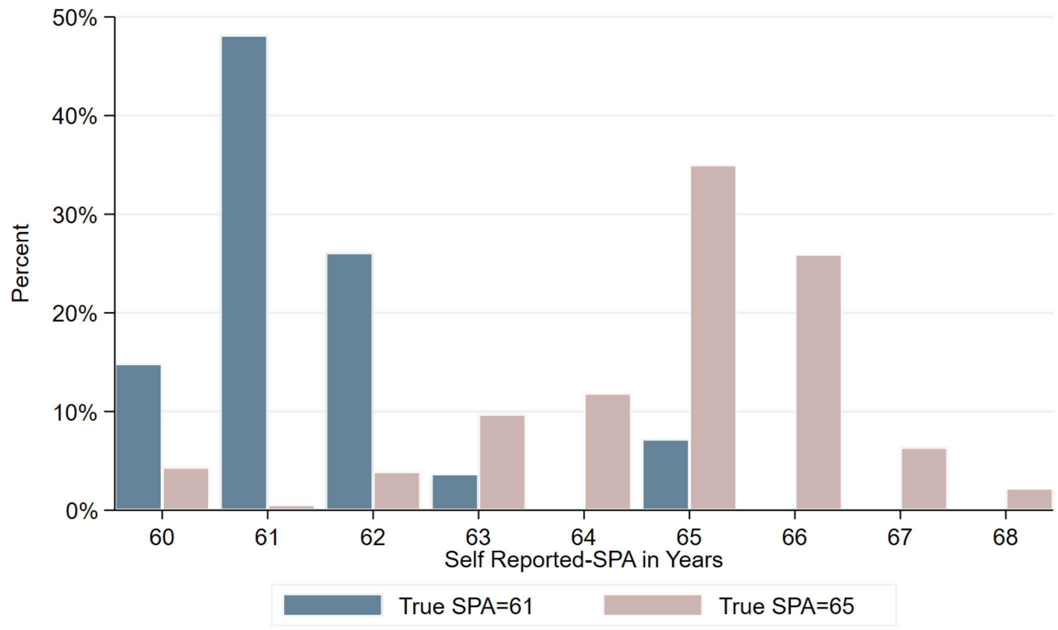
2.A.4 Mistaken Beliefs and Excess Employment Sensitivity

2.A.5 Descriptinves Beliefs

Mistaken beliefs could take on many forms. People could simply not update from the pre-reform SPA of 60 or might cling to other salient numbers like the male SPA

[29]Indeed it is likely to be illegal to do so on the grounds of that it would be discriminatory. For example, the 2012 European Court of Justice ruling known as Test-Achats explicitly outlawed charging men and women differently for the same insurance.

Figure 2.A.4: SPA Beliefs by SPA-cohort

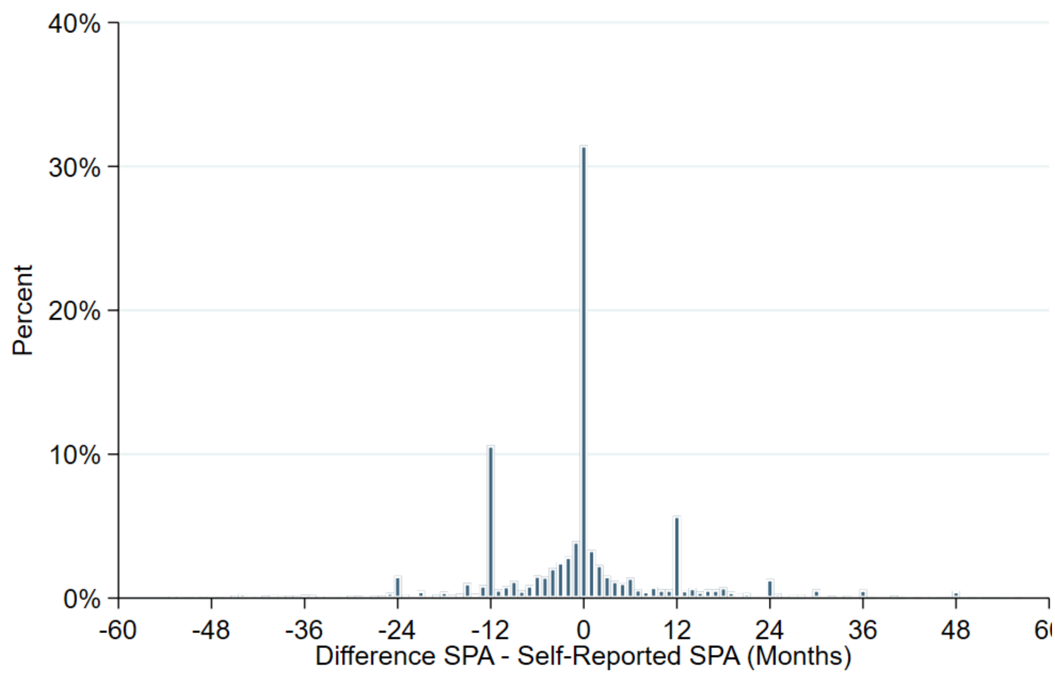


Notes: Self Perceived SPA for two SPA-cohorts. One with a rounded SPA of 61 and one with a rounded SPA of 65.

of 65. To get at these distinctions, Figure 2.A.4 plots reported SPAs for two SPA cohorts, one with a true SPA of 61 and one with a true SPA of 65. Although there is a slight increase around other salient ages, the dominant pattern is that the self-reports cluster around the true SPA for each cohort, looking very much like a noisy signal of the true SPA. Just the sort of pattern we would expect to emerge from a model of costly information acquisition.

Figure 2.A.5 shows that error in self-reported SPA at age 58 that was documented in the main text, but here at the true monthly frequency. Little that is relevant to the model is added by looking at the lower level of variation. We see that 31% know their own SPA to precisely the right month. The main thing we can glean from this graph that we cannot when the date is binned at a yearly frequency is that the spike every twelve months here show that people display an unsurprising round number bias.

Figure 2.A.5: Mistaken SPA Beliefs of Women Subject to the Reform at Age 58 (monthly)



Notes: Plot of error in self-reported SPA. The graph shows the frequency by which respondents gave mistaken answers about their SPA with errors at the true monthly level of SPA variation.

2.A.6 Treatment Effect Heterogeneity by Beliefs

I have interpreted the fact those who are more mistaken about their SPA in their late 50s have a smaller labour supply reaction upon reaching their SPA in their early 50s as evidence of the importance of beliefs. However, this fact is consistent with beliefs proxying for some unobserved heterogeneity; for example, if those who are more mistaken have lower cognitive skills leading them to work more pre-SPA, in turn forcing them to work more in old age.

In this section, I present two ways that treatment effects vary with beliefs that are consistent with selection into SPA knowledge driving the results and that are harder to explain with an appeal to unobserved heterogeneity.

Firstly, those whose beliefs imply they will receive a positive shock (i.e. they overestimate the SPA) are the ones to have the largest labour supply response to it. This can be seen in Table 2.A.6, which shows how the treatment effect varies

Table 2.A.6: Treatment Effect Heterogeneity by Direction of SPA Self Report Error

Below SPA		**0.785**
	s.e	(0.0276)
	$p=$	.006
Below SPA×(Under Estimate SPA)		**-0.013**
	s.e	(0.0466)
	$p=$	.786
Below SPA×(Knows SPA)		**0.058**
	s.e	(0.0331)
	$p=$	.082
Obs.		10,488
Cohorts		63

between three groups: those who think their SPA is sooner than it is, those who correctly state their SPA, and those who think it is further away than it is. The last group is the excluded category, and we can see that for this group, the treatment effect is positive and significant at 0.785. Then for those who underestimate their SPA, we see that their predicted treatment effect is smaller and loses significance when the sum of coefficients is tested jointly ($p = .0999$). So those who receive a negative shock do display a labour supply response to the SPA.

Secondly, those whose knowledge of the SPA got better had the largest labour supply response to the SPA. This can be seen in Table 2.A.7 which shows how the treatment effect varies between three groups: those whose SPA self-reports get worse between the first and last time they are asked, those whose self-reports stay the same, and those whose self-reports get better. It can be seen that those whose knowledge improves have the largest labour supply response to the SPA.

2.B Additional Mathematical Details

2.B.1 Extending Steiner, Stewart, and Matejka (2017)

My model does not quite fit into the framework of Steiner et al. (2017) because I have made slightly different assumptions about the information the agents receive costlessly. In this section, I first present a quick summary of their model in which I highlight the assumption that is not compatible with my model. Then I discuss

Table 2.A.7: Treatment Effect Heterogeneity by Learning

Treatment Effect SPA Knowledge Gets Worse	**0.0.40**
s.e	(0.0312)
p=	.203
Treatment Effect SPA Knowledge Stays Same	**0.076**
s.e	(0.0242)
p=	.002
Treatment Effect SPA Knowledge Gets Betters	**0.123**
s.e	(0.0225)
p=	.000
Obs.	10,488
Cohorts	63

mapping my model into their framework and where it fails. Next I present my alternative assumption that alows me to resurrect the results of Steiner et al. (2017) and I present a proof of the key lemma starting with this different timing assumption about costless information. For comparability, in this section, I adopt much of the notation of Steiner et al. (2017), and the notation is not related to the rest of the paper.

Steiner, Stewart, and Matejka (2017) model summary: There is a payoff relevant state $\theta_t \in \Theta_t$ evolving according to measure $\pi \in \Delta(\prod_t \theta_t)$ and agents must make a payoff relevant decision from a choice set D. Before making a decision d_t the agent can choose any costly signal about θ^t on signal space X. *After making a decision the agent observes a costless signal $y_t \in Y$, $y_t \sim g_t(y_t|\theta^t, y^{t-1}, d^t)$, where it is assumed that at each d^t, $y_t \perp x^t|(\theta^t, y^{t-1})$.* Agents get gross flow utilities $u(d^t, \theta^t)$ that can depend on the whole history of state and actions but suffer a utility cost for more precise information $\propto I(\theta^t, x_t|z^{t-1})$ where $z^t = (x^t, y^t)$. The sets Θ_t, D,Y, and X are finite and that $|D_t| \leq |X_t|$.

The agent chooses information strategy $f_t(x_t|\theta^t, z^{t-1})$ and action strategies $d_t = \sigma_t(z^{t-1}, x_t)$, collectively referred to as their strategy $s_t = (f_t, \sigma_t)$ to solve

$$\max_{f,\sigma} E[\sum_{t=0}^{T} \beta^t (u(\sigma_t(z^{t-1}, x_t), \theta^t) - I(\theta^t, x_t|z^{t-1})] \tag{2.22}$$

where the expectation is taken with respect to the distribution over sequences (θ_t, z_t) induced by the prior π together with the strategy $s_t = (f_t, \sigma_t)$ and the distributions g_t of costless signals. The function $u(.,.)$ is assumed continuous.

Issue mapping the retirement model to this framework: The state pension age maps comfortably to θ_t, and the decision is consumption and labour supply jointly. In the framework of Steiner et al. (2017) actions do not affect θ_{t+1} given θ_t; however as utility can depend on the complete history of actions their framework can handle the endogenous states of my retirement model a_t and $AIME_t$ as they are both exact functions of past actions. The issue is with the wage offer w_t and the unemployment state ue_t because these are exogenously evolving states whose current value is observed costlessly. Since they are exogenously evolving states they could be included in θ_t, but then the agent could use the signal to learn about them, but this is not the case for my agent as they observe these variables costlessly. The obvious solution is to include the exact values in the costless signal. However, Steiner et al. (2017) only allows the agent to have a costless signal of previous period values when making a decision because they receive the costless signal after making their choice and so can only use it the following period. To deal with this issue I need to make slightly tweak thier assumption about the costless signal.

Alternative assumption about the costless signal: I assume the agent receives their costless signal before taking an action each period and that this can be a signal of the current values. Specifically I replace the highlighted assumption above with that assumption: *Before making choosing a signal the agent observes a costless signal $y_t \in Y_t$, $y_t \sim g_t(y_t|\theta^t, y^{t-1})$, where it is assumed that at each d^t, $y_{t+1} \perp x^t|(\theta^t, y^t)$.* So I allow the costless signal to be a signal of the current values of θ_t but restrict it from being influenced by actions. This has the knock on affect that I need to redefine the decision node z^t as $z^t = (x^t, y^{t+1})$, but otherwise my setup is identical to theirs. This of course allows me to map the retirement model variables w_t and ue_t to enter θ_t alongside the state pension age but prevent households from ever choosing to learn about them by perfectly revealing the values of w_t and ue_t in the costless signal at the start of the periods. This change in timing only affects the proof of lemma 1 from Steiner et al. (2017) and I show below that this result still holds using a slightly different strategy to prove it.

For notational convenience, let $\omega^t = (\theta^t, z^{t-1})$ be the current state and the agent's current decision node, or information about the state, then:

Proposition 1. *(Lemma 1 in SSM) Any strategy s_t solving the dynamic RI problem generates a choice rule $p_t(d_t|\omega^t)$ solving*

$$\max_p E[\sum_{t=0}^{T} \beta^t (u(d^t, \theta^t) - I(\theta^t, d_t | z^{t-1})] \tag{2.23}$$

where we redefine $z^{t-1} = (d^{t-1}, y^t)$ the expectation is with respect to the distribution over sequences (θ_t, z_t) induced by p, the prior π, and the distributions g. Conversely, any choice rule p solving 2.23 induces a strategy solving the dynamic RI problem.

Proof. We proceed in steps.

Step 1:First note that for random variable $\zeta_t \in \{x_t, d_t\}$

$$E[\sum_{t=1}^{\infty} \beta^t I(\theta^t, \zeta_t | z^{t-1})] = E[\sum_{t=?}^{\infty} \beta^t (H(\theta^t | \zeta^{t-1}, y^t) - H(\theta^t | \zeta^t, y^t))] \tag{2.24}$$

But then by the entropic chain rule and that $\theta_t \perp \zeta^{t-1} | \theta^{t-1}$

$$\begin{aligned} H(\theta^t | \zeta^{t-1}, y^t) &= H(\theta^{t-1} | \zeta^{t-1}, y^t) + H(\theta_t | \theta^{t-1}, \zeta^{t-1}, y^t) \\ &= H(\theta^{t-1} | \zeta^{t-1}, y^t) + H(\theta_t | \theta^{t-1}, y^t) \end{aligned}$$

Since at each d^t, $y_{t+1} \perp x^t | (\theta^t, y^t)$ it follows that $y_{t+1} \perp (x^t, b^t) | (\theta^t, y^t) \Rightarrow H(y_{t+1} | \theta^t, x^t, y^t) = H(y_{t+1} | \theta^t, y^t) = H(y_{t+1} | \theta^t, b^t, y^t)$, so by the symmetry of mutual information

$$\begin{aligned} H(\theta^t | \zeta^t, y^t) - H(\theta^t | \zeta^t, y^{t+1}) &= I(\theta^t; y_{t+1} | \zeta^t, y^t) = I(y_{t+1}; \theta^t | \zeta^t, y^t) \\ = H(y_{t+1} | \zeta^t, y^t) - H(y_{t+1} | \theta^t, \zeta^t, y^t) &= H(y_{t+1} | \zeta^t, y^t) - H(y_{t+1} | \theta^t, y^t) \end{aligned}$$

So 2.24 becomes

$$E[\sum_{t=1}^{\infty}\beta^t(H(\theta^{t-1}|\zeta^{t-1},y^t)-H(\theta^t|\zeta^t,y^{t+1})$$
$$-H(y_t|\zeta^t,y^t)+H(y_t|\theta^t,y^{t-1})+H(\theta_t|\theta^{t-1},y^t))]$$
$$=E[\sum_{t=1}^{\infty}(\beta^{t+1}-\beta^t)H(\theta^t|\zeta^t,y^{t+1})-\beta^t H(y_t|\zeta^t,y^t)$$
$$+\beta^t(H(y_t|\theta^t,y^{t-1})+H(\theta_t|\theta^{t-1},y^t))]$$

Step 2: Given strategy s and the choice rule generated by it p by construction they generate the same gross utilities. Hence by step 1, 2.23-2.22 is:

$$E[\sum_{t=1}^{\infty}(\beta^t-\beta^{t+1})(H(\theta^t|b^t,y^{t+1})$$
$$-H(\theta^t|x^t,y^{t+1}))+\beta^t(H(y_{t+1}|b^t,y^t)-H(y_{t+1}|x^t,y^t))]$$

But then $|B|\leq|X|<\infty \Rightarrow b^t$ is measurable wrt x^t and hence $E[H(\theta^t|b^t,y^{t+1})]\geq E[H(\theta^t|x^t,y^{t+1})]$ and $E[H(y_{t+1}|b^t,y^t)]\geq E[H(y_{t+1}|x^t,y^t)]$ and therefore 2.23≥2.22
.

Step 3: As $B\subset X$ if p is a probability choice rule then $f_t(x_t|w^t)=p_t(b_t|\omega^t)$ and $x_t=\sigma_t(z^{t-1},x_t)$ is a viable solution to 2.22. For this strategy generated by this mapping, the probability choice rule makes equation2.23=equation2.22

Step 4: If s solves 2.22 the corresponding PCR p must solve 2.23, as by step 2 the value from p in 2.23≥s in 2.22, so if p doesn't solve 2.23 ∃ PCR producing greater net lifetime utiltiy than s in 2.22. But by step 3 this produces a viable solution to 2.22 with greater net lifetime utility contradicting s being a solution to 2.22.

Step 5: If p solves 2.23 then by step 3 it produces a viable solution to 2.22 but then 2.23≥2.22 so this strategy must be the optimal solution to 2.22 □

The remainder of the results I use from Steiner et al. (2017) follow with this new definition of the decision node $z^t=(x^t,y^{t+1})$ and their proof unaltered. The reason the other proofs are unaffected by this tweaking of the definition of z^t is that the costless signal has now been restricted to be insensitive to the action chosen.

2.B.2 Finding Unique Actions Using Second Order Conditions

Using the SOC of the rationally inattentive agent's problem Caplin et al. (2019) provide an alternative formulation of the solution of the model. If the CCP satisfy equation 2.14 and for all possible actions ($\forall d = (c,l) \in \mathscr{C}$)

$$\sum_{spa} \pi_t(spa) \frac{\exp\left(n^{(k)} \frac{((c/n^{(k)})^{\nu} l^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \beta \overline{V}^{(k)}_{t+1}(d, X_t, spa, \underline{\pi_t})\right)}{\sum_{d' \in \mathscr{C}} q_t(d') \exp\left(n^{(k)} \frac{((c'/n^{(k)})^{\nu} l'^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \beta \overline{V}^{(k)}_{t+1}(d', X_t, spa, \underline{\pi_t})\right)} \leq 1, \tag{2.25}$$

with equality if $q_t d > 0$. This new condition from (Caplin et al., 2019) replaces the need for the unconditional choice probabilities to solve the log-sum-exp of equation 2.15.

If an action $d^{\star} = (c^{\star}, l^{\star})$ satisfies equation 2.20 repeated here:

$$\sum_{spa} \pi_t(spa) \frac{\exp\left(n^{(k)} \frac{((c^{\star}/n^{(k)})^{\nu} l^{\star 1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \beta \overline{V}^{(k)}_{t+1}(d^{\star}, X_t, spa, \underline{\pi_t})\right)}{\exp\left(n^{(k)} \frac{((c/n^{(k)})^{\nu} l^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \beta \overline{V}^{(k)}_{t+1}(d, X_t, spa, \underline{\pi_t})\right)} \geq 1, \tag{2.26}$$

for all $d = (c,l) \in \mathscr{C}$. That is $d^{\star}$ produces such a high exponentiated-utility in all states that in expectation its ratio to the exponentiated-utility of any other payoff is greater than 1.

If such a $d^{\star}$ exists the only way to satisfy 2.25 for $d^{\star}$ is to have $q_t(d^{\star}) = 1$.

2.C Additional Computational Details

2.C.1 Solving the Models without Costly Attention

The models are solved by backward induction starting at age 101 when the household dies with certainty. The household problem is considered as a discrete choice problem. This within-period discrete choice optimisation problem is solved by grid search, selecting the value that maximises the household's utility. States are discretised with 30 grid points for assets (a_t), 4 for average earnings ($AIME_t$), 5 for wages (w_t), two for the unemployment shock (ue_t), and in the model with policy uncertainty the state pension age (SPA_t) has 8 gris points as it ranges from 60 to 67.

A finer grid of 500 points is offered to the household when making their saving choice. This keeps the size of the state space manageable whilst not unduly constraining households and is equivalent to having a finer grid for consumption than for assets. When evaluating continuation values of off-grid values, I use linear interpolation of the value function.

2.C.2 Solving the Models with Costly Attention

Belief Distribution Costly attention introduces a high dimensional state variable in the form of the belief distribution ($\underline{\pi}_t$). To discretise the distribution, I consider all possible combinations moving probability masses of a given size between the eight possible SPAs 60-67. As no amount of Bayesian updating can change the assignment of zero probability to an outcome, I want to avoid having beliefs that assigned zero probability to SPAs in my gridpoint of beliefs, and so I imposed a minimum probability to be assigned to each SPA of 0.01 and then had the probability masses that are moved about be in addition to this minimum amount. To make this more concrete, I broke the total probability into four masses that I moved between SPAs to form the grid over beliefs. In the absence of this minimum probability of any SPA, that would mean the probability masses being moved between SPAs was of a size of 0.25. In periods in which there are eight possible SPAs, because $t < 60$ and the women have not aged past any possible SPA, these probability masses are of the size $\frac{1-0.08}{4} = 0.23$. When $t < 60$, having these four probability masses to move between 8 possible SPAs leads to a total of $\binom{7+4}{4} = 330$ grid points because each combination can be thought of as an ordering of the four masses and the breaks between the eight grid points. As the women successively age past their SPAs, this shrinks as the number of SPAs to assign a probability mass to shrinks down to $\binom{1+4}{4} = 5$ when $t = 65$. Since there is no natural ordering over $\mathbb{R}^7$, I order these numbers in lexicographic ordering, which is convenient for constructing all possible combinations of the probability masses.

High Dimensional Interpolation When the prior with which a household starts the next period is off this grid, I use k-nearest neighbour inverse distance weighting to carry out the multidimensional interpolation. I use the difference in means between the distributions as an approximation to the Wasserstein, or earth mover, metric as the concept of distance used in the inverse distance weighting. High-dimensional interpolation can be a major computational burden and also a source of approximation error. For this reason, I initially start using just two nearest grid points to interpolate over; if the guess and verify loop over the unconditional choice probabilities (q_t) fails to converge after 25 iterations, I gradually increase the number of neighbours included in the interpolation until reaching a maximum at $2^8 = 256$.

Range of Attention Costs As explained in Section 2.5.2, during periods in which rational inattention matters because $t < SPA_t$ the central equation that needs to be solved to find the households optimal decision is the following:

$$\max_{\underline{q_t}} \sum_{spa} \pi_t(spa) \log\Big(\sum_{d' \in \mathscr{C}} q_t(d) \exp\Big(n^{(k)} \frac{((c/n^{(k)})^{\nu} l^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \beta \overline{V}^{(k)}_{t+1}(d, X_t, SPA_t, \underline{\pi_t})\Big)\Big)$$

Following approaches used in the RUM literature, I normalise the payoff inside this equation. First, I do this by dividing through by the highest payoff in all possible SPAs. However, the presence of λ in this equation makes this process of exponentiating utility even more problematic. Data and not computational considerations should determine what values of λ we consider; however, the fact this parameter appears as a denominator in an exponentiated expression means that as λ gets small, the difference between exponentiated payoffs gets larger. Since a lower SPA is better than having a later one, the values inside the log associated with SPA=60 are larger, and decreasing the cost of attention exaggerates these differences. However, when λ gets small, the fact that the exponentiated payoffs associated with SPA=67 are much smaller than those associated with SPA=60 does not mean the former are not important to the optimisation because for very small values log() approaches minus infinity and its rate of change approaches infinity. So how probabilities are allocated over these outcomes when the exponentiated payoff is very small has very large implications for the value of the objective function. Therefore, we cannot ignore vanishingly small exponentiated payoffs because they have outsized implications for the logarithmic objective function. This fact, combined with the very small values of the cost of attention implied by the belief data, led me to very carefully optimise the code with respect to the storage of very small utility values, rather than just dropping them as could be more happily done with a more standard objective function. To store these smaller values, I use quadruple precision float points leading to the smallest value distinguishable from zero of 10^{-4965}. However, since compilers are optimised to conduct double precision operations, moving from double to quadruple precision leads to a much greater than a factor of two slow down in runtime. For this reason, I only use quadruple precision when absolutely necessary, checking beforehand if normalising payoff leads to an underflow so that important values would be lost and treated as zero in double precision.

Solving the within period problem Culling actions that will never be taken helps makes the sequential quadratic programming problem more stable as it reduces the dimensionality of the problem. This is done by dropping strictly dominated actions. Identifying strictly dominated actions is an interesting problem with a large related literature in computer science (Kalyvas and Tzouramanis, 2017) but since the size by choice set is not large (no larger than 1,500 resulting from 3 labour supply choice and 500 asset choices) one of the simpler algorithms, Block Nested Loop, is most efficient. The range of attention costs can make the problem unstable but the routine used to carry out the sequential quadratic programming (Schittkowski, 2014) manages the range of values needed to match the data.

2.C.3 Simulating and Estimating

My initial sample of simulated individuals is large, consisting of 50,000 random draws of individuals aged 55. Given that we randomly simulate a sample of individuals that is larger than the number of individuals observed in the data, most observations will be drawn multiple times. I take random Monte Carlo draws of assets and average lifetime earnings, which are the state variables that are observed without selection bias in the data. For wages, I exploit the model implied joint distribution of these state variables. I simulate one SPA cohort at a time, and so SPA_t is initialised to a fixed value mirroring the SPA of the cohort currently being simulated. I make the assumption that the SPA answer represents draws from an individual's belief distribution and that everyone starts at age 55 with the same beliefs. This allows me to initialise the belief distribution to the distribution of point estimates seen for SPA self-report in the ELSA data.

Given these initial conditions, I simulate the choice of the individual households using the decision rule found when solving the model and the exogenous process estimate in the first stage. I then aggregate the simulated data in the same way we aggregate the observed data and construct moment conditions. I describe these moments in greater detail in appendix 2.D. The method of simulated moments procedure delivers the model parameters that minimise a GMM criterion function, which we also describe in Appendix 2.D. To find the minimum of the resulting objective function, I first sample the parameter space using Sobol sequencing and then search for a minimum using the BOBYQA (Powell, 2009) routine at promising initial conditions.

2.D Additional Econometric Details

2.D.1 Imputing AIME

Average lifetime earnings are only observed for some of the women in my sample. In order to be able to initialise the model from the joint distribution of $AIME_{55}$ and a_{55} I impute the missing observations. First, I regress $AIME_{55}$ on a quintic in NHNBW plus a very rich set of additional controls that include variables on health, education, location, labour market behaviour, housing tenure, cohort, age, wage, and measure of cognitive ability. This includes as much information as possible to impute $AIME_{55}$.

However, merely using these predictions for imputation will likely overstate the correlation between $AIME_{55}$ and a_{55}; for this reason, I add noise to the imputed variable to replicate observed heteroscedasticity. To do this, I run regressions of the non-imputed $AIME_{55}$ values on a quintic of NHNBW without the controls (because the model does not contain the other variables) and then regress the squared residuals on the same polynomial of NHNBW. Since the imputed $AIME_{55}$ are by construction homoscedastic, adding a noise term with variance given by this last regression replicates the heteroscedasticity seen in the regression of $AIME_{55}$ on the quintic of NHNBW.

2.D.2 Type-specific Mortality

Heterogeneity in life expectancy has important implications for the behaviour of older individuals (e.g. De Nardi et al., 2009), but death is often poorly recorded in survey data. For this reason, I include type-specific mortality but do not rely on the recording of death in ELSA to estimate it; instead, combining ELSA with ONS survival probabilities following French (2005). That is, I estimate type-specific death using Bayes' rule:

$$Pr(death_t|type=k) = \frac{Pr(type=k|death_t)}{Pr(type=k)} Pr(death_t)$$

Where $Pr(type=k|death_t)$ and $Pr(type=k)$ are taken from ELSA and $Pr(death_t)$ are taken from the ONS life-tables. If measurement error effects all types equally estimates of $Pr(type=k|death_t)$ from ELSA are unbiased unlike those of $Pr(death_t|type=k)$ and deals with the measurement error issue.

2.D.3 Generating Profiles

To avoid contamination by cohort effects or macroeconomic circumstances, a fixed effect age regression was estimated, which included: year of birth fixed effects, SPA-cohort specific age effects, the aggregate unemployment rate rounded to half a percentage point and an indicator of being below the SPA. More specifically, the following regression equation was estimated:

$$y_{it} = U_t + \sum_{c \in C} \gamma_c \mathbb{1}[cohort_i = c] + \sum_{s \in S} \mathbb{1}[SPA_i = s](\sum_{a \in A} \delta_{a,s} \mathbb{1}[age_{it} = a])$$

where $cohort_i$ is the year-of-birth cohort of an individual, SPA_i is her final SPA, $age_{i,t}$ her age in years, U_t aggregate unemployment to half a per cent, and the outcome variable y_{it} is either assets or employment depending on which profile is being calculated.

The profiles used were then predicted from these regressions using average values for the pre-reform cohorts. This controls for cohort effects and the effects of macroeconomic circumstances by setting their impact on the targeted profiles to their average value whilst also allowing for the key variation in behaviour between SPA-cohorts at the SPA.

2.E Additional Results

2.E.1 First Stage Estimates

Model Types A woman is classed as having a high education if she has more than the compulsory schooling required for her generation. She is classed as married if she is married or cohabiting, as the legal arrangements are less important than the household formation for the questions considered in this paper. As mentioned in the main text, I abstract away from separation in the model. To get around the fact that separation occurs in the data, if a woman is ever observed as married, her household is classified as such in all periods. The reason to classify her as married rather than single is that a divorced or widowed woman will likely receive some form of alimony or widows pension and so she is more accurately modelled as married according to the model. This leads to the following proportion of types: 34% married and low education, 11% single and low education, 44% married and high education, 11% single and high education.

Initial conditions Initial assets a_{55} and average earning $AIME_{55}$ are set from the type-specific empirical joint distribution, some summary statistics of which are presented in Table 2.E.1. Understandable for women of this generation married women have weaker labour market attachment and so lower $AIME_{55}$ but higher household assets a_{55}. Higher education increases both variables.

Labour market conditions The type-specific transition probabilities, estimated with individuals classified as unemployed when they claim unemployment benefits, are shown in Table 2.E.2.

The parameters of the stochastic component of the wage process (persistence and the variance of innovation, measurement error, and initial draw) are shown in

Table 2.E.1: Summary Statistics of Initial Conditions (£)

Type	Variable	Mean	SD
Married, Low Education	Initial Assets	76,226	163,320
	Initial AIME	4,889	2,915
Single, Low Education	Initial Assets	13,231	30,471
	Initial AIME	6,015	4,334
Married, High Education	Initial Assets	148,440	218,143
	Initial AIME	9,358	6,264
Single, High Education	Initial Assets	97,495	186,362
	Initial AIME	10,663	6,676
...total	Initial Assets	102,680	189,801
	Initial AIME	7,618	5,199

Notes: Means and standard deviations of the initial distribution of assets and average lifetime earnings.

Table 2.E.2: Type Specific Unemployment Transition Probabilities

Type	Transition	Probability(%)
Married, Low Education	From employment to unemployment	2.37
	From unemployment to employment	57.75
Single, Low Education	From employment to unemployment	3.20
	From unemployment to employment	27.03
Married, High Education	From employment to unemployment	1.72
	From unemployment to employment	71.08
Single, High Education	From employment to unemployment	3.25
	From unemployment to employment	37.78

Notes: Unemployment and reemployment transition probabilities.

Table 2.E.3: Parameters of the stochastic component of the wages

Type	ρ_w	σ_ε	σ_μ	$\sigma_{\varepsilon,55}$
Married, Low Education	0.911	0.039	0.249	0.266
Single, Low Education	0.901	0.042	0.255	0.178
Married, High Education	0.945	0.035	0.351	0.322
Single, High Education	0.974	0.025	0.358	0.224

Notes: Estimates of the persistence of wages and the variance of their transitory and persistent components as well as initial distribution.

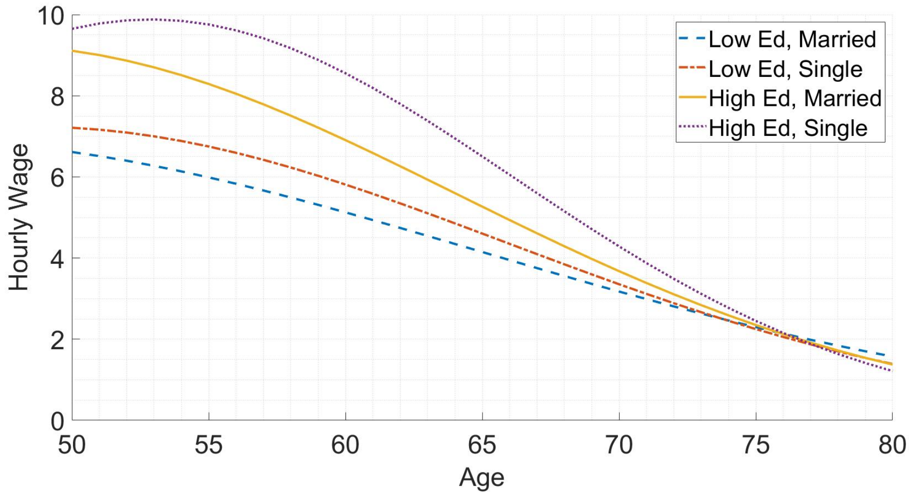
Table 2.E.3

The deterministic component of wages generates the wage profiles in Figure 2.E.1. Spousal Income is shown in Figure 2.E.2.

Social Insurance As mentioned in the main text, much larger differences in State Pension income are observed between married and single women than between high and low education. Amongst State pension claimers, high education women have mean State pension income of £92.52 and low education women £87.11, whereas single women have a State Pension income of £112.50 and married women £80.89. Hence to maximise power whilst capturing the key difference, I restrict heterogeneity in the State Pension process to be between married and single women only. The resulting functions of average lifetime earnings are shown in Figure 2.E.3.

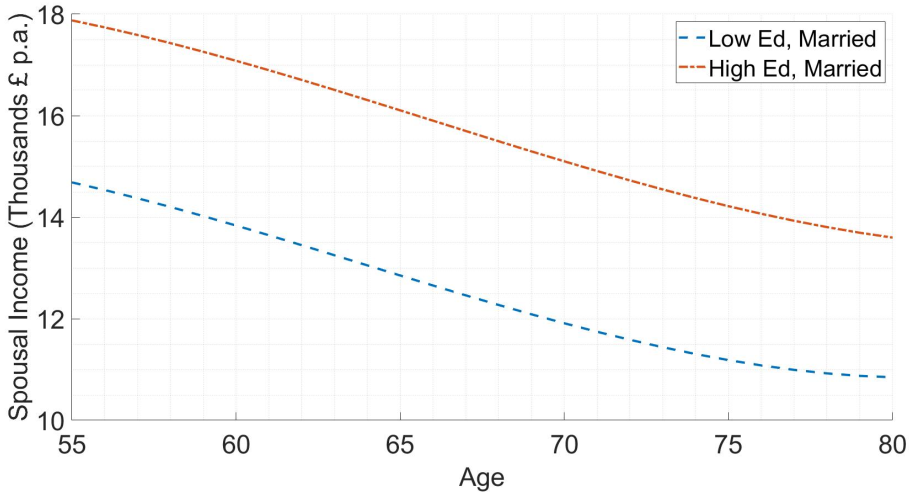
Conversely, differences in private pension income are smaller between married and single women than between high and low education. Amongst those reporting non-zero private pension income, high education women have mean private pension income of £118.50 and low education women £61.42, whereas single women have State Pension income of £100.78 and married women £94.24. Hence to maximise power whilst capturing the key difference, I restrict heterogeneity in the private pension process to be between high and low-education women only. The resulting functions of average lifetime earnings are shown in Figure 2.E.4.

Figure 2.E.1: Wage Profiles



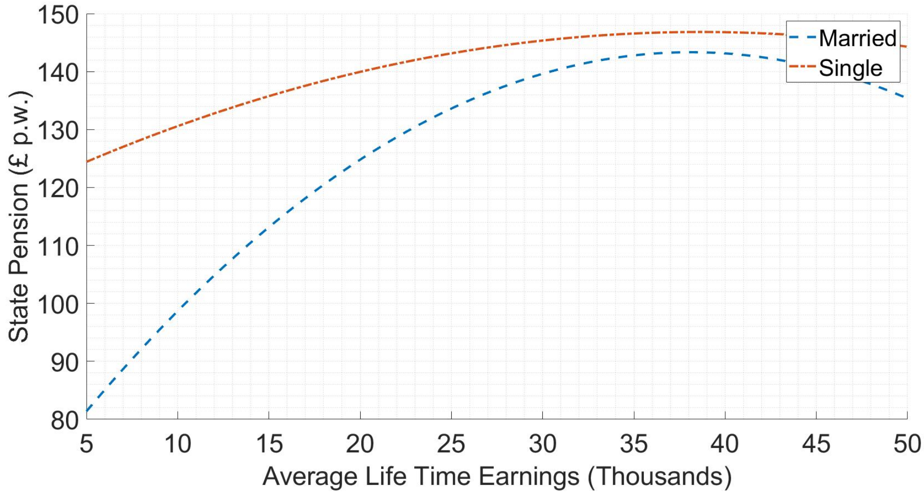
Notes: The deterministic component of female hourly wages for the four model types plotted against female age.

Figure 2.E.2: Spousal Income



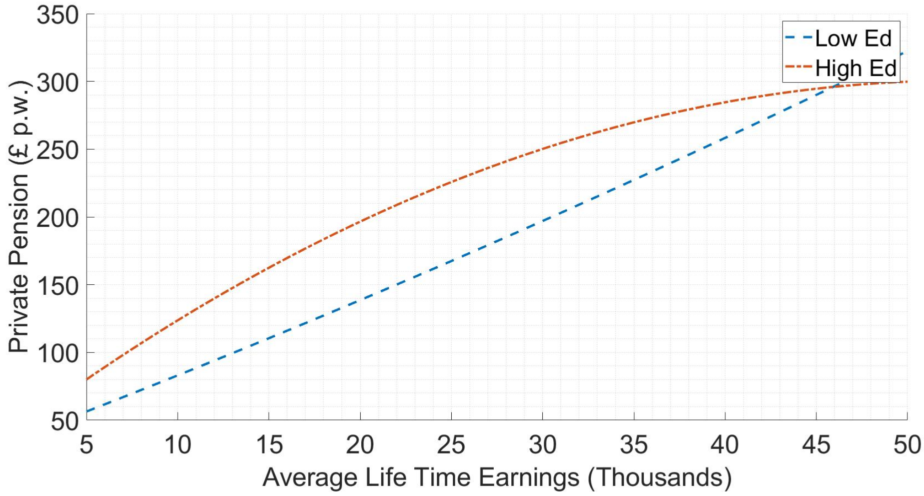
Notes: Spousal income plotted against female age.

Figure 2.E.3: State Pension as Function of Average Earnings



Notes: State Pension income as a function of average lifetime earnings (AIME) for married and single women.

Figure 2.E.4: Private Pension as Function of Average Earnings



Notes: Private Pension income as a function of average lifetime earnings (AIME) for high and low education women.

Table 2.E.4: Regression Analysis of the Determinants of Learning

	Const.	Wlth.	AIME	Wage	t = 56	t = 57	t = 58	t = 59
Coeff.	0.18	-2.9e-07	2.9e-06	9.7e-08	-6.5e-02	-7.8e-02	-9.8e-02	-8.3e-02

Notes: Regression coefficient where the dependent variable is bits of information acquired

2.E.2 Model Fit

As mentioned in the main text, although the different model specifications have different predictions about the labour supply response to the dynamic SPA, the static profiles are not very sensitive to model specifications. All versions are able to match the static profiles. Figures 2.E.5-2.E.8 show the employment and asset profiles for the baseline version and the version with rational inattention with the parameter estimates of Table 2.4.

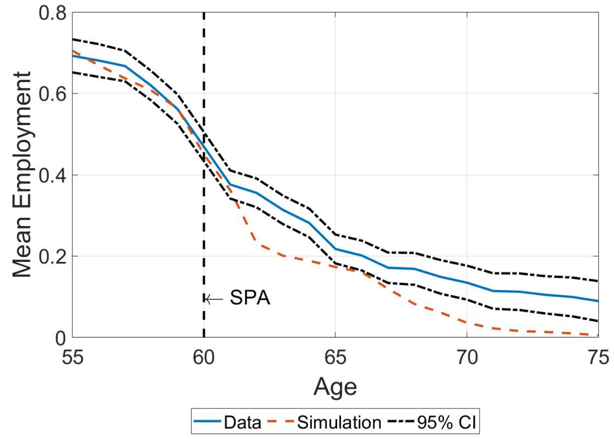
2.E.3 Results Tables

2.E.4 Robustness (Targeting Other Moments)

By design, the estimation procedure used in this paper does not directly target the excess employment sensitivity puzzle. This allows the model to match savings and labour supply across a range of ages and then to investigate how well a model that matches these profiles can explain excess employment sensitivity. However, it may leave some wondering how well the model could match employment response to the SPA if this was directly targeted. This section shows that the baseline model can match the treatment effect in the whole population at the cost of greatly exaggerating liquidity constraints but cannot match the treatment effect of those with above median assets.

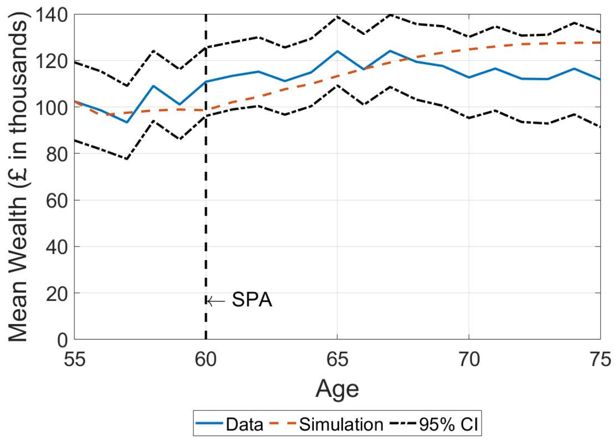
Table 2.E.5 shows that the model nearly perfectly replicates the treatment effect in the whole population but falls dismally short of the one seen in those with above median wealth. Figures 2.E.9-2.E.10 show that this is achieved at the cost of massively exaggerating how much households run down their assets and hence the importance of borrowing constraints.

Figure 2.E.5: Employment Profile Baseline



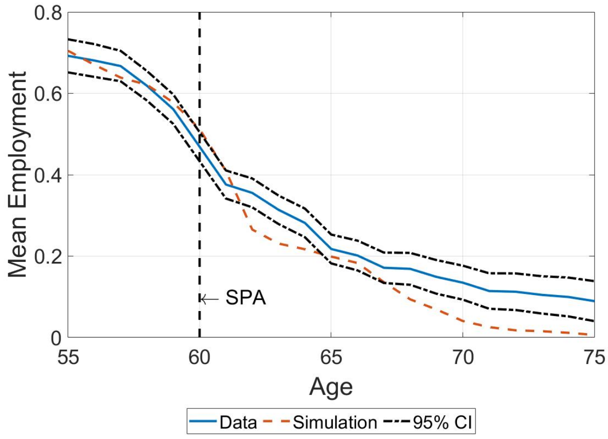
Notes: Model fit to targeted labour supply profile. The empirical profile is for the pre-reform SPA cohort with a SPA of 60. The model was simulated with an unchanging SPA of 60, mimicking the conditions faced by this cohort.

Figure 2.E.6: Asset Profile Baseline



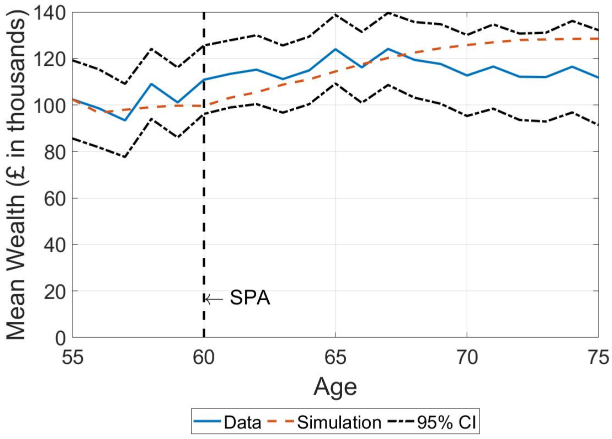
Notes: Model fit to targeted asset profile. The empirical profile is for the pre-reform SPA cohort with a SPA of 60. The model was simulated with an unchanging SPA of 60, mimicking the conditions faced by this cohort.

Figure 2.E.7: Employment Profile Model with Rational Inattention



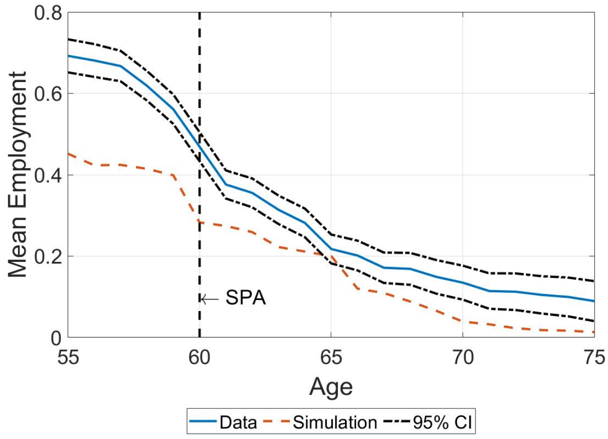
Notes: Model fit to targeted labour supply profile. The empirical profile is for the pre-reform SPA cohort with a SPA of 60. The model was simulated with an unchanging SPA of 60, mimicking the conditions faced by this cohort.

Figure 2.E.8: Asset Profile Model with Rational Inattention



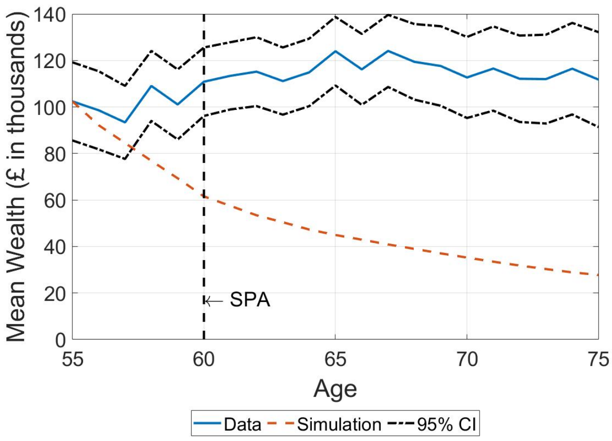
Notes: Model fit to targeted asset profile. The empirical profile is for the pre-reform SPA cohort with a SPA of 60. The model was simulated with an unchanging SPA of 60, mimicking the conditions faced by this cohort.

Figure 2.E.9: Employment Profile Baseline when Targeting Treatment Effects



Notes: Model fit to targeted labour supply profile. The empirical profile is for the pre-reform SPA cohort with a SPA of 60. The model was simulated with an unchanging SPA of 60, mimicking the conditions faced by this cohort.

Figure 2.E.10: Asset Profile Model Baseline when Targeting Treatment Effects



Notes: Model fit to targeted asset profile. The empirical profile is for the pre-reform SPA cohort with a SPA of 60. The model was simulated with an unchanging SPA of 60, mimicking the conditions faced by this cohort.

Table 2.E.5: Effect of SPA on Employment: Heterogeneity by Wealth

	Model	Data
Treatment Effect on employment of being below SPA		
Whole Population	0.078	0.080
Assets >Median(£29,000)	0.007	0.061

Chapter 3

Disentangling Risk and Intertemporal Preferences with Costly Information Acquisition

3.1 Introduction

Economists consider risk aversion synonymous with curvature, with respect to consumption, of utility or a risk-aggregator. Costly information acquisition creates a reason to be averse to risk even in the absence of curvature: the utility cost of reducing uncertainty. Aversion to risk unrelated to curvature raises the prospect of breaking the inverse reciprocals relation between relative risk aversion and the elasticity of intertemporal substitution imposed by standard time-additive expected utility. Other approaches to breaking this relation[1] do not consider this direct utility cost of risk but introduce parameters separately controlling within and across period curvature. For two-period two-state-of-the-world models, this paper shows that costly learning about reducible risk separates risk and intertemporal preferences. It separates them by introducing this additional motive to dislike risk while leaving intertemporal preference largely unchanged. This counterfactual inverse reciprocals relation is at the root of equity premium and the risk-free rate puzzles. In an application, I show that costly learning can solve these puzzles.

[1] e.g. Epstein and Zin (1989) and Weil (1990) or Dillenberger et al. (2020)

To investigate costly attention's impact on risk and intertemporal preference, I compare a rationally inattentive agent who can learn about reducible uncertainty with a standard uninformed agent who cannot. Like-for-like comparisons of their preferences are made before learning can occur. For the standard uninformed agent, the model framework is textbook expected utility. For the rationally inattentive, it is an application of Matějka and McKay (2015) for discrete choice and Jung et al. (2019) for continuous.

Rational inattention implies stochastic choice, and standard definitions of relative risk aversion and the elasticity of intertemporal substitution are ill-defined for stochastic choice as they are predicated on an agent faced with the same decision problem making the same choice. This is an interesting issue, although, for this paper, it more resembles a frustrating tangent. Section 3.3 considers an analyst who misattributes deviations from deterministic choice to measurement error. This generates a stochastic-choice adapted definition of risk aversion: a lottery is prefered to a sure-thing if the agent selects the lottery with a probability greater than one-half. This reflects in a stylised fashion approaches to dealing with the inference of parameters from noisy data.

Sections 3.4 and 3.5 apply the stochastic-choice adjusted definitions to the choices of the rationally inattentive and standard uninformed agents. Sections 3.4 compares their risk aversion by analysing their choices between binary lotteries and a sure-things when the rationally inattentive agent can learn about the outcome of the lottery by paying a utility cost. Propositions, 4 proves that the skewness of the lottery alone determines whether introducing costly attention increases or decreases risk aversion. Neither the cost of attention nor the curvature with respect to consumption affects the direction of change, but they do affect the magnitude. Section 3.5 compares their intertemporal substitution choices when able to self-insure against future learnable income risk using a risk-free in a two-period model. Proposition 5 proves that costly attention does not generally affect intertemporal preference.

Section 3.6 combines these results to show how costly information acquisition separates risk and intertemporal preferences. Previous work (Luo, 2010; Luo and Young, 2016) investigating the ability of rational inattention to explain the equity premium puzzle only considered zero skew uncertainty, so it missed rational inattention's ability to disentangle risk and time preferences. Section 3.6 concludes by calibrating a simple example to the US economy to demonstrate the potential of this mechanism to explain the equity premium and the risk-free rate puzzles.

The model explaining these puzzles assumes risk is learnable, whereas clearly, part of future stock return risk is not learnable. However, the financial literacy literature (e.g. Lusardi et al., 2015) has documented many facts about baseline probabilities of stock returns that consumers have not internalised. Hence, a component of individual uncertainty is reducible through learning. This paper investigates the implication of costly learning about reducible risk, so it abstracts away from the irreducible component.

Two related papers critique risk-aversion in stochastic choice models because the curvature of the utility function alone no longer predicts preferences over lotteries (Wilcox, 2011; Apesteguia and Ballester, 2018). Unlike them, I argue this is a desirable feature. Firstly, because it is an intuitive feature when costly information provides a separate reason to dislike risk. Secondly, because it means these models can offer independent predictions for empirically distinct objects.

Deriving results for tractable binary lotteries allows for illustrations that costly attention can separate risk and time preferences. This simplicity could raise doubts about broader validity. An online appendix, using results from Steiner et al. (2017), provides suggestive evidence for an equivalence between dynamic discrete choice rational inattention models and Hansen and Sargent (1995) risk-sensitive preferences, which Bommier et al. (2017) show are the only Krepes-Porteus recursive preference to separate risk and time preferences whilst preserving monotonicity with respect to first-order stochastic dominance. Hence, these desirable properties should extend to models of costly information acquisition, and rational inattention may provide an appealing alternative to some popular recursive preferences.

3.2 Decision Maker's Perspective: Model Framework

This section presents the model framework of a rationally inattentive agent who can engage in costly learning and a standard uninformed agent who cannot learn. The framework presented here nests specific models used later to analyse the two agents' risk and intertemporal preferences.

3.2.1 The Two Agents: Rationally Inattentive and Standard Uninformed

The standard uninformed agent controls a choice variable d subject to constraint $d \in \mathscr{C} \subset \mathbb{R}$, closed and bounded. She faces uncertainty about some outcome $z \sim Z$, with a known distribution; has utility $U(d,z)$ that is continuous with respect to the standard topology induced on $\mathscr{C}$; and chooses d to maximise expected utility. Her choices solve:

$$\max_{d \in \mathscr{C}} E[U(d,z)].$$

$U(d,z)$ may, or may not, be constructed from a discounted sum of within-period flow-utility functions, $U(d,z) = \sum_{t=1}^{T} \beta_t u(,d,z)$, but if $T > 1$ all uncertainty is resolved after the first decision is taken. This essentially static framework is complex enough to furnish examples that disentangle the risk and intertemporal preferences.

The rationally inattentive agent is identical, except she can receive a signal $x \sim X$ about the uncertain outcome Z by paying a utility cost. She chooses the distribution of the signal $f_{X|Z}(x|z)$, but her utility function is extended with an additive utility cost, higher for more informative signals. She makes her decision conditional on the draw from the noisy signal $d(x) \in \mathscr{C}$. Both agents have a correct prior.

The cost of information is assumed to be directly proportional to the mutual information between the signal X and the uncertain outcome Z: the expected reduction in uncertainty, measured by entropy, about Z from learning X. Hence, her decision problem becomes:

$$\max_{f_{X|Z}(.|.),d(x)\in\mathscr{C}} E[U(d,z)+\lambda I(X,Z)] \tag{3.1}$$

where,

$$I(X,Z) = H(Z) - E[H(Z|X)]$$

and H is the entropy function $H(X) = E[-\log(f_X)]$.

The choice of the signal process makes the rationally inattentive agent's actions stochastic. This paper focuses on the unconditional distribution of actions q ($d \sim q \in \Delta(\mathscr{C})$). This is because, as the standard uninformed receives no signal, only the unconditional distribution of actions q allows for like-for-like comparisons between the two agents. Additionally, after seeing the signal, the rationally inattentive agent is a standard agent with different beliefs, but unchanged preferences. So, nothing new is learned from studying her ex-post preferences.

Finding this unconditional distribution q is made possible by Matějka and McKay (2015) and Jung et al. (2019). I summarise the results from these papers used here:

Result 2. *The actions of the rationally inattentive agents that solve* (3.1) *have an unconditional distribution* $q(d)$ *that solves* (3.2).

$$\max_{q\in\Delta(\mathscr{C})} E_Z\left[\log\Big(E_q[\exp(u(d,z)/\lambda)]\Big)\right] \tag{3.2}$$

3.3 Analyst's Perspective: Inferring Stochastic Choice Agents' Preferences

The textbook definition of preference, including risk and intertemporal preference, assumes that each time an agent faces a choice between A and B, she chooses identically. This does not describe a stochastic choice agent. Hence, I adapt the preference definition to make the like-for-like unconditional comparison between the two agents argued for in Section 3.2. This section proposes the following definition: A is prefered to B when the probability of choosing A exceeds $\frac{1}{2}$. *The reader who is happy with this definition without further justification loses little by skipping the rest of this section.*

The model justifying this definition is of an analyst who attributes all deviations from deterministic behaviour to measurement error. This captures, in a stylised way, approaches in the literature to inference with noisy data.

3.3.1 Model of the Analyst

An analyst observes stochastic choice data. That is repeated decisions $d \in \mathscr{C}$ from a single decision-maker for different decision problems $D = (\mathscr{C}, X(d))$, where $\mathscr{C}$ is the observed choice set and X is the observed outcome resulting from a decision:

Definition 3.3.1 (Stochastic choice data)**.** *This is a collection of decision problems $\{D_i = (\mathscr{C}_i, X_i(d)) | i \in I\}$, each observed with a frequency given by measure $\mu_i \in \Delta(I)$, and a related set of stochastic choice functions $Q = \{q_{D_i} \in \Delta(\mathscr{C})\}$.*

I first apply this definition to a decision maker who chooses between a lottery and a sure-thing amount. I then apply it to a decision maker who chooses savings in a risk-free asset given different interest rates. The first application captures preferences over risk and the second over intertemporal substitution. To illustrate the use of the definition, consider a saver facing different interest rates. Each different interest rate r defines a different decision problem $D_r = (\mathscr{C}, X_r(d))$. A given interest rate constitutes a fraction μ_r of all observations, and for each r, the full distribution of choices is observed q_{D_r}.

Any non-degenerate q_D is inconsistent with deterministic choice. When faced with stochastic choice data, most research does not abandon deterministic choice concepts, nor is it clear that doing so is desirable given the prevalence of measurement error and unobserved heterogeneity. I consider the measures of risk aversion and preference for intertemporal substitution an analyst would arrive at if she treated all deviation from deterministic behaviour as measurement error.

3.3.2 Risk Aversion

The Arrow-Pratt measure can be calculated for the rationally inattentive agents, but it does not predict her preference over lotteries. The certainty equivalent more directly relates to why we care about definitions of risk aversion: they encapsulate preferences between lotteries. Additionally, certainty equivalent is observable, whereas the Arrow-Partt measure is in terms of unobservables. Hence, I define risk aversion in terms of the certainty equivalent.

Setup The analyst has access to stochastic choice data. She observes the probability of accepting a lottery $q_D(x)$ when the agent is offered multiple different sure-thing alternatives x and the relative frequencies with which each x is observed as captured by measure μ_x. There is a lower amount x_l below which $q_D(x) = 0$ and a higher amount x_h above which $q_D(x) = 0$. The analyst believes there is a certainty equivalent amount $\pi \in (x_l, x_h)$ below which the sure-thing is rejected and above which it is accepted. She attributes the stochasticity of the agent's choices to measurement error that switches the binary outcome (accept, reject) with probability ϕ. If $d = 1$ represents accepting the lottery, she has a misspecified model of the DGP of $q(x)$ such that the observation j from $q(x)$ is given by $d_{j,x} = \mathbb{1}[x > \pi] + \varepsilon_j(\mathbb{1}[x < \pi] - \mathbb{1}[x > \pi])$ where $\varepsilon \sim Bernoulli(\phi)$. She wants to infer π.

Proposition 3. *$\hat{\pi}$ such that $q_D(\hat{\pi}) = \frac{1}{2}$ is the maximum likelihood estimator of π*

Proof. This analyst has a misspecified model where the j^{th} observed choice at surething offer x is distributed:

$$d_{j,x} = \mathbb{1}[x > \pi] + \varepsilon_j(\mathbb{1}[x < \pi] - \mathbb{1}[x > \pi]), \tag{3.3}$$

where $\varepsilon \sim Bernouli(\phi)$ and $d = 1$ represent accpeting the lottery and $d = 0$ rejecting.

The likelihood contribution in her misspecified model from observing d when the sure-thing x is offered is:

$$f(d, x; \phi, \pi) = \phi^{y(d,x,\pi)}(1-\phi)^{(1-y(d,x,\pi))},$$

where $y(d,x,\pi) = \mathbb{1}[x > \pi] + d - 2d\mathbb{1}[x > \pi]$. Then given the distribution of the sample μ_x, the log-likelihood function is:

$$l(\phi,\pi;q_D(x),\mu_x) = \int_{x_l}^{x_h} \sum_{d=0}^{1} q_{D=d}(x) \log\left(f(d,x;\phi,\pi)\right) d\mu_x$$

which gives,

$$\begin{aligned} l(\phi,\pi;q_D(x),\mu_x) = \\ &\log(\phi)\Big(\int_{x_l}^{x_h} \sum_{d=0}^{1} q_{D=d}(x) y(d,x,\pi) d\mu_x\Big) + \\ &\log(1-\phi)\Big(1 - \int_{x_l}^{x_h} \sum_{d=0}^{1} q_{D=d}(x) y(d,x,\pi) d\mu_x\Big). \end{aligned}$$

Treating π as known and maximising w.r.t ϕ gives MLE:

$$\hat{\phi}(\pi) = \int_{x_l}^{x_h} \sum_{d=0}^{1} q_{D=d}(x) y(d,x,\pi) d\mu_x,$$

and the FOC shows:

$$0 = \frac{\partial l}{\partial \phi} = \frac{\int_{x_l}^{x_h} \sum_{d=0}^{1} q_{D=d}(x) y(d,x,\pi) d\mu_x}{\phi} - \frac{1 - \int_{x_l}^{x_h} \sum_{d=0}^{1} q_{D=d}(x) y(d,x,\pi) d\mu_x}{1-\phi}.$$

Define $Y(x;\pi) = \sum_{d=0}^{1} q_{D=d}(x) y(d,x,\pi)$ and substitute $\hat{\phi}(\pi)$ into the log-likelihood function to find the total MLE of π, $\hat{\pi}$, gives:

$$\begin{aligned} l(\phi,\pi;q_D(x),\mu_x) = \\ &\log\Big(\int_{x_l}^{x_h} Y(x;\pi) d\mu_x\Big)\Big(\int_{x_l}^{x_h} Y(x;\pi) d\mu_x\Big) + \\ &\log\Big(1 - \int_{x_l}^{x_h} Y(x;\pi) d\mu_x\Big)\Big(1 - \int_{x_l}^{x_h} Y(x;\pi) d\mu_x\Big). \end{aligned}$$

This is maximised at $\int_{x_l}^{x_h} Y(x;\pi) d\mu_x = 0$ and $\int_{x_l}^{x_h} Y(x;\pi) d\mu_x = 1$. Since each contribution, $y(d,x,\pi)$, is 0 or 1, these values represent the upper and lower attainable bounds of $\int_{x_l}^{x_h} Y(x;\pi) d\mu_x$, and the log-likelihood is maximised by getting closer to these bounds. The FOC to find the extremum is:

$$0 = \frac{\partial}{\partial \pi} \int_{x_l}^{x_h} Y(x;\pi) d\mu_x = \frac{\partial}{\partial \pi} \Big(\int_{x_l}^{\pi} q_{D=1}(x) d\mu_x + \int_{\pi}^{x_h} 1 - q_{D=1}(x) d\mu_x \Big)$$

giving,

$$0 = q_{D=1}(\pi)\mu(\pi) - (1 - q_{D=1}(\pi))\mu(\pi)$$

implying $q_{D=1}(\pi) = \frac{1}{2}$, giving the result.

□

3.3.3 Elasticity of Intertemporal Substitution

Like risk aversion, the elasticity of intertemporal substitution has multiple definitions presupposing deterministic actions. Its definition as the semi-elasticity of consumption growth with respect to the interest rate is in terms of observables, so it is the one the analyst uses to estimate the elasticity (ρ):

$$\rho = \frac{d(\log(\frac{c_{t+1}}{c_t}))}{dr}$$

where r_t is the real interest rate and c_t consumption at time t. To estimate a single elasticity, the analyst assumes isoelastic preferences. With isoelastic preferences, a common method of estimating the elasticity of intertemporal substitution is the linear regression:

$$\log(\frac{c_{t+1}}{c_t}) = \rho r_t + k + \varepsilon_t, \tag{3.4}$$

where k is a constant incorporating the individual's discount factor and any fixed effects, and ε_t is an error term. She assumes all deviations from deterministic actions result from measurement error. However, as regression is robust to measurement error in the dependent variable, where actions enter, and the independent variable r_t is observed without error, no adjustment for measurement error is needed. Hence, the analyst estimates ρ by running OLS on the observations of an individual's consumption and saving decision at the different interest rates in the stochastic choice dataset.

3.4 Risk Aversion

This section compares a standard uninformed agent's risk preferences with a rationally inattentive agent's by combining the model framework (Section 3.2) and the definition of risk aversion arrived at by the analyst (Section 3.3).

3.4.1 Model

An agent with wealth, w, is faced with the choice between a binary lottery $z \sim Z$ and some sure-thing amount.[2] The goal is to find the agent's certainty equivalent π. For a standard uninformed agent with strictly increasing consumption utility is $u(.)$, her certainty equivalent π solves:

$$u(w-\pi) = E_Z[u(w+z)].$$

The rationally inattentive agent can acquire information about the state of the world by paying a utility directly proportional to the mutual information between Z and the signal with constant of proportionality λ. Choosing between a sure-thing χ and the lottery Z, her decision problem is:

$$\max_{d(x)\in\{0,1\}, f_{X|Z}(x|z)\in\Delta} E[u(w+d.z-(1-d)\chi)-\lambda I(X,Z)].$$

The analyst, attributing deviation from deterministic action to measurement error, infers an agent is indifferent between a sure-thing and a lottery when she chooses both with equal probability. Therefore the rationally inattentive agent's certainty equivalent is the sure-thing π that makes her take each option equally.

3.4.2 Solution

As the choice between sure-thing and lottery is a discrete choice, we use the result from Matějka and McKay (2015) to deduce that the probability q of accepting the sure-thing solves:

$$\max_{q\in[0,1]} E_Z[\log(q\exp(u(w-\pi)/\lambda)+(1-q)\exp(u(w+z)/\lambda))].$$

[2] An online appendix derives the condition for a generic lottery.

Ignoring boundary conditions, the FOC with respect to q is:

$$E_Z[\frac{\exp(u(w-\pi)/\lambda)-\exp(u(w+z)/\lambda)}{q\exp(u(w-\pi)/\lambda)+(1-q)\exp(u(w+z)/\lambda)}]=0 \tag{3.5}$$

As we are solving for the certainty equivalent amount, we only need to consider $q = 1/2$ (see Section 3.3), and so:

$$E_Z[\frac{\exp(u(w-\pi)/\lambda)-\exp(u(w+z)\lambda)}{\exp(u(w-\pi)/\lambda)+\exp(u(w+z)/\lambda)}]=0. \tag{3.6}$$

To solve for the utility level that makes the agent choose the lottery and the sure-thing with equal probability, we define the certainty equivalent utility m as the utility resulting from consuming the certainty equivalent amount $m = \exp(u(w-\pi)/\lambda)$. Label the probability of the good state of the world π_G and that of the bad π_B, and label the exponentiated utility in the good state of the world $V_G = \exp(u(w+z_g)/\lambda)$ and in the bad $V_B = \exp(u(w+z_b)/\lambda)$. Then the equation becomes:

$$\pi_G\frac{V_G-m}{V_G+m}+\pi_B\frac{V_B-m}{V_B+m}=0.$$

This is a solvable algebraic expression. Selecting the positive root, which the discriminant shows exists, leads to the following expression for m:

$$m=\frac{2E[V]-(V_G+V_B)+\sqrt{(V_G+V_B-2E[V])^2+4V_GV_B}}{2}. \tag{3.7}$$

This characterises the certainty equivalent of the rationally inattentive agents. If the agent were unable to learn, she would be a standard uninformed agent, so $m = \exp(E[U]) = \exp(\pi_G u_G + \pi_B u_B) = V_G^{\pi G} V_B^{\pi_B}$. Hence, the term:

$$\frac{2E[V]-(V_G+V_B)+\sqrt{(V_G+V_B-2E[V])^2+4V_GV_B}}{2}-V_G^{\pi_G}V_B^{\pi_B} \tag{3.8}$$

tells us how much compensating utility the rationally inattentive agent requires compared with the standard uninformed agent to play the lottery[3].

[3]This normalises the standard uninformed agent's utility function by $1/\lambda$, but this is innocuous.

3.4.3 Analysis

From (3.8), we want to infer qualitative information about certainty equivalent amounts, not to make interpersonal utility comparisons. Considering the two agents as alternative versions of a single person with the same utility from consumption $u(.)$, we can infer the direction of change in risk preferences from introducing costly attention without positing a form for $u(.)$. Since $u(.)$ is increasing, larger utility differences mean a larger certainty equivalent amount. Analysing (3.8) leads to the following proposition.

Proposition 4. *The rationally inattentive agent demands a smaller certainty equivalent than the standard uninformed agent when $\pi_g \in (0, \frac{1}{2})$, a larger certainty equivalent when $\pi_g \in (\frac{1}{2}, 1)$, and the same certainty equivalent when $\pi_g \in \{0, \frac{1}{2}, 1\}$.*

Proof. When $\pi_g = \frac{1}{2}$, $2E[V] = (V_G + V_B)$ and so:

$$m = \frac{\sqrt{4V_G V_B}}{2} = \sqrt{V_G V_B} = V_G^{\pi G} V_B^{\pi_B},$$

hence (3.8) equals zero. When $\pi_g = 1$:

$$\begin{aligned} m &= \frac{2V_G - (V_G + V_B) + \sqrt{(V_G + V_B - 2V_G)^2 + 4V_G V_B}}{2} \\ &= \frac{V_G - V_B + \sqrt{V_B^2 + V_G^2 + 2V_G V_B}}{2} = \frac{2V_G}{2}, \end{aligned}$$

hence (3.8) becomes $V_G - V_G = 0$. By symmetrry when $\pi_g = 0$ (3.8) becomes $V_B - V_B = 0$.

Label (3.8) as a function of π_G, $f(\pi_G)$:

$$f(\pi_G) = \frac{2E[V] - (V_G + V_B) + \sqrt{(V_G + V_B - 2E[V])^2 + 4V_G V_B}}{2} - V_G^{\pi_G} V_B^{1-\pi_G}.$$

When $V_G > V_B$, f is an elementary function without singularity in $[0, 1]$ and so is analytic on the domain. It is zero at $0, \frac{1}{2}, 1$, so we need to check it is negative on $(0, \frac{1}{2})$ and positive on $(\frac{1}{2}, 1)$.

$f(\pi_G)$ can be decomposed into the sum of a positive and negative function:

$$f(\pi_G) = \underbrace{E[V] - V_G^{\pi_G} V_B^{1-\pi_G}}_{g(\pi_G)} + \underbrace{\frac{\sqrt{((1-2\pi_G)V_G + (2\pi_G - 1)V_B)^2 + 4V_GV_B} - (V_G + V_B)}{2}}_{h(\pi_G)}.$$

As the difference between the arithmetic and geometric means, $g(\pi_G)$ is positive over [0,1]. To see $h(\pi_G)$ is negative rearrange it:

$$h(\pi_G) = \frac{\sqrt{(1-4\pi_G+4\pi_G^2)V_G^2 + (1-4\pi_G+4\pi_G^2)V_B^2 - 2(1-4\pi_G+4\pi_G^2)V_GV_B + 4V_GV_B} - (V_G+V_B)}{2}$$

$$\begin{aligned} \Rightarrow h(\pi_G) = & \frac{\sqrt{(V_G+V_B)^2 - 4(\pi_G - \pi_G^2)(V_G - V_B)^2} - (V_G+V_B)}{2} \\ \leq & \frac{\sqrt{(V_G+V_B)^2} - (V_G+V_B)}{2} = 0. \end{aligned}$$

It is easy to check: $g(0) = h(0) = 0$, $g(1) = h(1) = 0$, both are unimodal ($g(\pi_G)$ hump shamped and $h(\pi_G)$ inverse hump shaped), and $h(\pi_G)$ is minimmised at $\pi_G = \frac{1}{2}$ (because $\frac{1}{2}$ maximising $\pi_G - \pi_G^2$). The FOC maximises $g(.)$ is:

$$0 = \frac{dg}{d\pi_G} = V_G - V_B - V_G^{\pi_G^\star} V_B^{1-\pi_G^\star}(\log(V_G) - \log(V_B))$$

$$\Rightarrow \pi_G^\star = \frac{\log(V_G - V_B) - \log(V_B \log(\frac{V_G}{V_G}))}{\log(\frac{V_G}{V_G})}.$$

For fixed V_B $\pi_G^\star$ is monotonic in V_G, converges to $\frac{1}{2}$ as $V_G \downarrow V_B$ and to 1 as $V_G \to \infty$. Since $V_G > V_B$ it follows $\pi_G^\star > \frac{1}{2}$.

We know that $f(\frac{1}{2}) = 0$, at which point h attains its minimum and starts increasing, and g is still increasing. Therefore f is postive over some $(\frac{1}{2}, \frac{1}{2} + \delta_1)$. Moreover, as h smoothly attains its minimum, there exists some region $(\frac{1}{2} - \delta_2, \frac{1}{2})$ such that $-\frac{dh}{d\pi_G} < \frac{dh}{d\pi_G}$ and so moving away from one-half into this region removes more from positive g than it does from negative h, and since $f(\frac{1}{2}) = 0$ it follows that

f is negative over this region.

Next:

$$\frac{df}{d\pi_G} = \frac{(V_G - V_B)^2(2\pi_G - 1)}{\sqrt{(V_G + V_B)^2 - 4(\pi_G - \pi_G^2)(V_G - V_B)^2}} + (V_G - V_B) - (\frac{V_G}{V_B})^{\pi_G} V_B \log(\frac{V_G}{V_B}),$$

$$\Rightarrow \frac{df}{d\pi_G}(1) = \frac{(V_G - V_B)^2}{(V_G + V_B)} + (V_G - V_B) - V_G \log(\frac{V_G}{V_B}).$$

For given V_G, this quantity is strictly increasing in V_B, and so it is maximised when V_B is maximised. $V_B < V_G$ and if we set it to its upper bound $V_B = V_G$ we get $\frac{df}{d\pi_G}(1) = 0$. This is an untenable upper bound, and it is strictly increasing in V_B for any V_B, therefore $\frac{df}{d\pi_G}(1) < 0$.

Similarly,

$$\frac{df}{d\pi_G}(0) = -\frac{(V_G - V_B)^2}{(V_G + V_B)} + (V_G - V_B) - V_B \log(\frac{V_G}{V_B})$$

Given V_B, this quantity is strictly decreasing in V_G, and so it is maximised with V_G minimised. $V_G > V_B$ and if we set it to it's lowever bound $V_G = V_B$ we get $\frac{df}{d\pi_G}(0) = 0$. This is an unattainable upper bound and it is strictly decreasing in V_G for any V_B, therefore $\frac{df}{d\pi_G}(0) < 0$

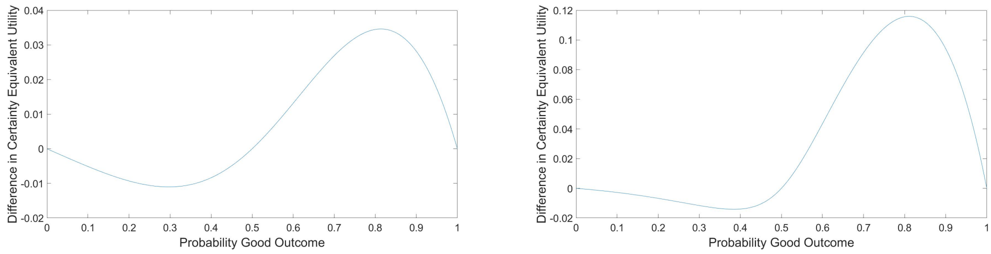
As $f(0) = f(\frac{1}{2}) = f(1) = 0$; f starts and ends negative over $(0, \frac{1}{2})$; and starts and ends positive $(\frac{1}{2}, 1)$, f has at least two turning points implying at least one point where $\frac{d^2 f}{d\pi_G^2} = 0$. If f is not negative (positive) over $(0, \frac{1}{2})$ $((\frac{1}{2}, 1))$, this would require at least two other turning point in the derivative of f , hence two more points where $\frac{d^2 f}{d\pi_G^2} = 0$. However,

$$0 = \frac{d^2 f}{d\pi_G^2} = \frac{8(V_G - V_B)^2 V_G V_B}{((V_G + V_B)^2 - 4(\pi_G - \pi_G^2)(V_G - V_B)^2)^{\frac{3}{2}}} - (\frac{V_G}{V_B})^{\pi}_{G} V_B \log^2(\frac{V_G}{V_B})$$

$$\Rightarrow \pi_G \log(\frac{V_G}{V_B}) + \frac{3}{2}\log((V_G + V_B)^2 - 4(\pi_G - \pi_G^2)(V_G - V_B)^2)$$
$$= \log(8(V_G - V_B)^2 V_G \log^{-2}(\frac{V_G}{V_B}))$$

The RHS of this is constant; the LHS is the sum of a positive linear increasing term

Figure 3.4.1: Comparison of Certainty Equivalent Utility



and a negative inverse hump-shaped term. Therefore the LHS is either monotonic or has one turning point; in either case, it has at most two solutions.

□

Proposition 4 shows that whether the rationally inattentive agents display more or less risk aversion than the standard uninformed agent depends solely on the nature of uncertainty faced (its skewness) - not on the cost of attention or any other preference parameters.

The left panel of Figure 3.4.1 illustrates Proposition 4. As proved, the rationally inattentive agent's certainty equivalent (utility) is lower when the probability of the good outcome is less than one-half and higher when it is greater. The right panel of Figure 3.4.1 shows that when the cost of attention λ is decreased, the shape of the curve is changed, becoming more skewed around $\frac{1}{2}$ and the size of the maximal difference in certainty equivalent increases. If the consumption utility function is unchanged, this represents an increase in the difference of certainty equivalent amounts. Therefore the size of the rationally inattentive agent's certainty equivalent is determined by the curvature of her utility $u(.)$ over consumption, her cost of attention λ, and the nature of uncertainty.

3.5 Elasticity of Intertemporal Substitution

This section investigates the implication of costly attention for the Elasticity of Intertemporal Substitution using self-insurance decisions in a dynamic model.

3.5.1 Model

There are two periods in the model, after which the agent dies with certainty, receiving a terminal value of 0. In period 1, the agent has known and certain income y_1. In period 2, the agent's income is either y_{2b} or y_{2g} occurring with probabilities π_b and π_g where $y_{2b} < y_{2g}$ and $y_1 > y_{2b}$. In period 1, the agent can save in a risk-free asset a with a gross rate of return R. The agent gets utility from consumption and discounts the future at rate β. The agent's consumption flow utility function $u(.) \in \mathscr{C}^2(\mathbb{R}_{>0})$, is strictly increasing, convex, and satisfies an Inada conditions $\lim_{c \to 0} u'(c) = 0$.

The rationally inattentive agent can learn about y_2 by paying an additively separable utility cost, proportional to the mutual information between the signal and y_2 with constant of proportionality λ. Her maximisation problem is:

$$\max_{a(x)\in[0,1],f_{X|Z}(x|z)\in\Delta} E[u(y_1 - a) + \beta u(Ra + y_2)) - \lambda I(X,Z)].$$

I impose a no-borrowing condition ($a > 0$), but the results hold for any borrowing constraint that prevents Ponzi schemes.

3.5.2 Solution

Self-insurance in a risk-free asset is a continuous choice. Jung et al. (2019) prove that, faced with a continuous choice, the rationally inattentive agent simplifies by only considering a finite subset, and we can learn about how she simplifies from the exponentiated-utility space curve:

$$\mathscr{K} = (\exp(u(y_1 - a) + \beta u(Ra + y_{2b}))^{1/\lambda}, \exp(u(y_1 - a) + \beta u(Ra + y_{2g}))^{1/\lambda}) \quad \forall a \in [0, y_1],$$

because the marginal probability distribution q of a is found by solving:

$$\max_{q\in\Delta([0,y_1])} \{\pi_b \log(E_q[q(a)\exp(u(y_1 - a) + \beta u(Ra + y_{2b}))^{1/\lambda}]) + \pi_g \log(E_q[q(a)\exp(u(y_1 - a) + \beta u(Ra + y_{2g}))^{1/\lambda}])\}. \tag{3.9}$$

Figure 3.5.1: Feasible Set Exponentiated Utility Space

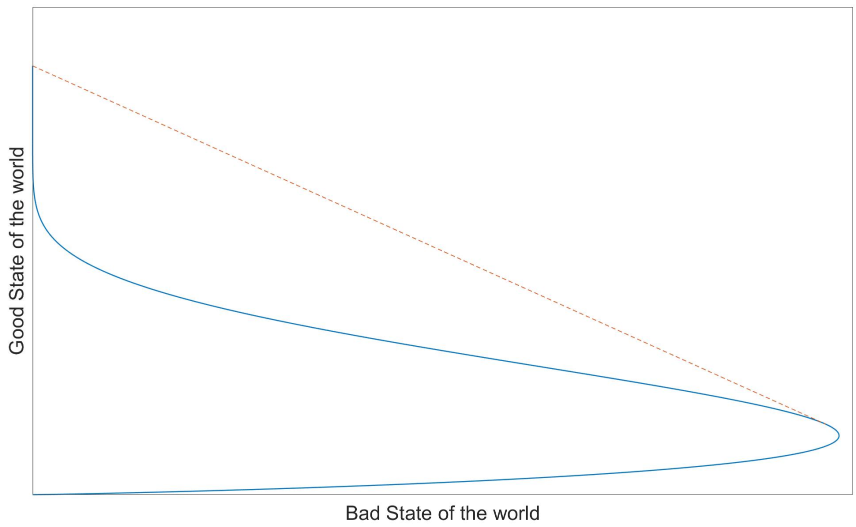
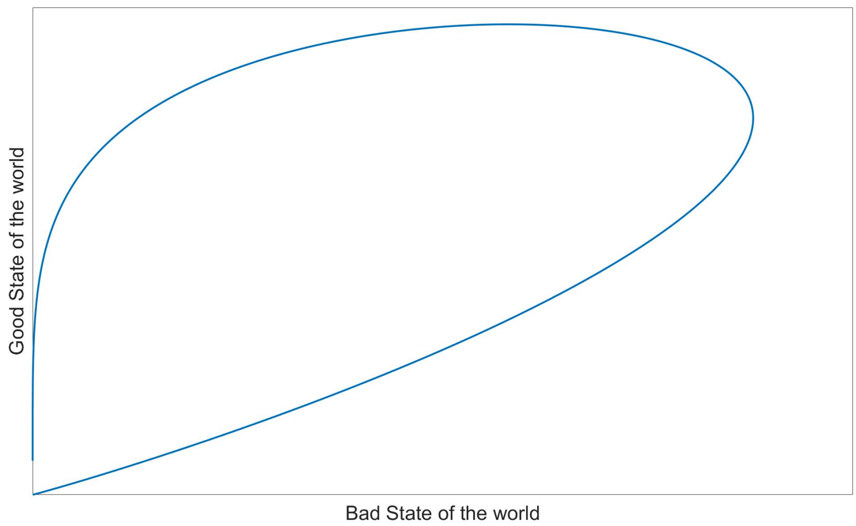


Figure 3.5.2: Feasible Set Exponentiated Utility Space



3.5.3 Analyis

By choosing $q \in \Delta([0, y_1])$, the rationally inattentive agent can attain in expectation any point in the convex hull of $\mathscr{K}$. Since there are two states of the world, at most, two actions are needed to achieve this. Therefore she either gathers no information and behaves like the standard uninformed agent or randomises[4] over two actions. Since she aims to maximise the log-sum-exp objective in (3.9) which is a strictly increasing function in both its arguments, she only randomises if the upper convex hull of $\mathscr{K}$ lies strictly above $\mathscr{K}$. Otherwise, she chooses the same utility-maximising point as the standard uninformed agent. Figure 3.5.1 shows an example where the upper convex hull lies strictly above $\mathscr{K}$, so it offers possibilities to increase net utility. Figure 3.5.2 shows an example where it does not.

Part 1 of Proposition 5 confirms that if the cost of information is sufficiently high, the agent gathers no information, and if sufficiently low enough, she gathers some. Part 2 shows that the rationally inattentive agent has the same preference for intertemporal substitution as the standard uniformed agent unless selecting the borrowing constraint is optimal in the good state of the world. It holds because, conditional on the signal, the rationally inattentive agent is a standard agent with different beliefs. So if her utility is $u(c) = \frac{c^{1-\gamma}}{1-\gamma}$, the analyst observations could be generated by multiple standard agents with the same γ. If saving nothing in the good state of the world is not optimal, the agent is on her Euler Equation, and the elasticity of intertemporal substitution is unchanged by this change in beliefs. If saving nothing in the good state of the world is optimal, rational inattention can increase the frequency with which the agent selects the borrowing constraint, pushing up the elasticity of intertemporal substitution because she can reduce the precautionary saving motive through learning.

Proposition 5. *1. If the cost of attention is high enough, the rationally inattentive agent has degenerate unconditional choice distribution q; if it is low enough, q assigns positive probability to two choices.*

[4] I use randomise to describe non-degenerate action distributions. This differs from usage in the game-theory literature because the randomisation device is partially informative.

2. *If $u(c) = \frac{c^{1-\gamma}}{1-\gamma}$, the elasticity of intertemporal substitution inferred by the analyst $\hat{\rho}$ is $\hat{\rho} = \gamma^{-1}$, unless when $y_2 = y_{2g}$ with probability 1, it is optimal to save nothing in which case $\hat{\rho} \leq \gamma^{-1}$*

Proof. When the upper envelope of $\mathscr{K}$ is concave, nothing is gained by randomising over two savings levels, and the rationally inattentive agent makes the same choice as the standard uninformed agent; thus, her elasticity of intertemporal substitution is the same $\hat{\rho} = \gamma^{-1}$, proving part 2 when $\mathscr{K}$ is concave.

So, we need to identify when the upper envelope is convex and a non-degenerate solution q is possible. Doing this leads to a proof of part 1.

Let $V_i(a) = \exp(U_i(a)/\lambda)$ where $U_i(a) = u(y_1 - a) + \beta u(Ra + y_{2i})$ for $i \in B, G$.

To analyse the shape of $\mathscr{K}$ let's follow the curve traced out in exponentiated utility space by:

$$f(a) = (V_B(a), V_G(a)),$$

starting at the end $a = y_1$, and decreasing a. To facilitate discussion, associate the exponentiated utility in the bad state of the world with the x-axis and in the good with the y–axis.

The Inada conditions mean $f(a)$ is strictly increasing away from $f(y_1)$ in both dimensions because the marginal gains from increasing consumption today at the expense of tomorrow are infinite in both states of the world. Eventually, we reach the optimal saving level in the bad state of the world $a_B^\star$. If $a_B^\star = 0$, there are no gains from randomising and $q(a_B^\star) = 1$, so the propsition is true. So assume $a_B^\star = 0$ is interior. From $a_B^\star = 0$, $V_B(a)$ starts decreasing as we continue to decrease a (its derivtive turns postive). Hence from this point, $f(a)$ doubles back on itself as its payoff in the good state of the world continues to increase but decreases in the bad. No point between $(a_B^\star, y_1]$ can be on the upper convex hull because their supporting lines (tangent) lie below another point on $\mathscr{K}$, $a_B^\star + \varepsilon$.

Continuing to decrease a, $f(a)$ remains aligned along the $x = -y$ plane until it reaches the optimal saving in the good state of the world $a_G^\star$, which may be at the borrowing constraint $a = 0$ or an interior point. If $a_G^\star$ is interior, then as we continue to decrease a past $a_G^\star$, the payoff in the good state of the world starts to decrease, and so $f(a)$ begins to move downward aligned with the $x = y$ plane. No point between $[0, a^\star_G)$ can be on the upper convex hull because their supporting lines (tangent) lie below another point on $\mathscr{K}$, $a^\star_G - \varepsilon$.

Therefore, only points in $[a^\star_G, a^\star_B]$ can be in the upper convex hull, but they will only be so if they lie on a convex portion of the curve (i.e. all points $[a^\star_G, a^\star_B]$ are on the upper envelope, so we need to check convexity).

We implicictly defferntiate f to find $\frac{dV_G}{dV_B}$ and $\frac{d^2V_G}{dV_B^2}$:

$$\frac{dV_G}{dV_B} = \frac{U'_G \exp(U_G/\lambda)}{U'_B \exp(U_B/\lambda)},$$

$$\frac{d^2V_G}{dV_B^2} = \frac{\lambda^{-1}\exp((U_G+U_B)/\lambda)\Big(U'_B(U''_G+\lambda^{-1}(U'_G)^2) - U'_G(U''_B+\lambda^{-1}(U'_B)^2)\Big)}{(U'_B\exp(U_B/\lambda))^3}.$$

U_B is increasing over $[a^\star_G, a^\star_B]$ and so the sign of$\frac{d^2V_G}{dV_B^2}$ is compltely determined by:

$$\Delta(a) := \Big(U'_B(U''_G+\lambda^{-1}(U'_G)^2) - U'_G(U''_B+\lambda^{-1}(U'_B)^2)\Big).$$

U_B and U_G are both concave so U''_B and U''_G are both negative. As U'_B is postive and U'_G negative over $[a^\star_G, a^\star_B]$ is follows that $\forall a \in (a^\star_G, a^\star_B)$:

$$\lim_{\lambda\to\infty} \Delta(a) = (U'_B U''_G - U'_G U''_B)\Big) < 0,$$

$$\lim_{\lambda\to 0} \Delta(a) = \lim_{\lambda\to 0} \Big(U'_B\lambda^{-1}(U'_G)^2) - U'_G\lambda^{-1}(U'_B)^2\Big) > 0.$$

So when the cost of attention is sufficiently large, the upper envelope of $\mathscr{K}$ is concave, so the rationally inattentive agent behaves like the standard uninformed agent. When the cost of attention is sufficiently small, she gathers information and takes two actions with non-zero probability. This completes the proof of part 1.

For part 2, conditional on a given signal, the rationally inattentive agent is a standard-utility maximiser with different beliefs. Since the rationally attentive agent's curvature of utility over consumption is unchanged, what the analyst observes is equivalent to multiple standard agents with different beliefs but the same utility curvature. When R is varied, this shifts both beliefs and savings of the rationally inattentive agent, but if $a_G^\star$ is interior, whatever combination of beliefs and savings she ends up at these will be interior and so satisfy the Euler equation. Therefore all observations of the rationally inattentive (a,R) agent lie on a curve that implies the same elasticity of intertemporal substitution as the standard uninformed agent.

If $a_G^\star = 0$, then it may form one of the two points of support, and, as with standard agents when they are at the borrowing constraint, this point will not respond to decreases in the interest rate implying a lower elasticity of intertemporal substitution. Although the standard uniformed agent can also be at the borrowing constraint, she only chooses this if it is the unique expected-utility maximising, whereas the rationally inattentive agent may randomise over this point of support as long as $a_G^\star = 0$. So, if $a_G^\star = 0$, rational inattention may increase observations at the borrowing constraint and hence the elasticity of intertemporal substitution.

□

3.6 Combining Risk and Intertemporal Preferences

This section combines the results on relative risk aversion and the Elasticity of intertemporal substitution to show that rational inattention can break the usual inverse reciprocity between them.

If the agent in Section 3.5 had the option to buy an insurance contract against the income risk instead of the option to self-insure with a risk-free asset, the model would be an application of the risk model from Section 3.4. Thus, changing the agent's choice allows for a comparison of the impact of rational inattention on risk and intertemporal preferences.

Proposition 5 shows that, unless borrowing constraints bind in the good state of the world, the elasticity of intertemporal substitution is unaffected by rational inattention. Risk preference, however, is explained by Proposition 4 and is different from textbook risk preferences except for a handful of edge cases that depend solely on the nature of uncertainty. Hence the reciprocal coupling.

3.6.1 Implication for Finance Puzzles

The difficulty in separating intertemporal and risk preference is central to the equity premium puzzle and the risk-free rate puzzle.

3.6.1.1 Equity Premium Puzzle

The equity premium puzzle (Mehra and Prescott, 1985) is that generating the large observed excess returns of stock over bonds requires unrealistic levels of risk aversion given the low risk of stocks and their poor insurance values against income risk. A large literature attempting to explain it exists (see, Kocherlakota, 1996); however, the aim here is not to evaluate the merits of other explanations but to document how rational inattention can explain this puzzle by separating intertemporal and risk preference. As a full portfolio selection model with consumption growth uncertainty is beyond the analytically solvable models analysed in this paper, I provide suggestive evidence using a simpler model.

Model The model of Section 3.4 was presented as a model of insurance purchase but can be conceived of as a choice between a risky and risk-free asset. It provides a stylised illustration of the equity premium puzzle considering only the extensive margin choice between stocks and bonds.

Consider a version of the model of Section 3.4.3 in which the agent has a choice between investing all wealth w in a risk-free asset with returns r^f or investing a fixed fraction α in a risky asset with uncertain return r having a binary distribution with support $\{r_b, r_g\}$, $r_b < r_g$. We can find the risk-free rate r^f that would make the standard uninformed agent indifferent to the lottery by solving:

$$\frac{((1-\alpha)r^f w + \alpha r^f w)^{1-\gamma}}{1-\gamma} = E[\frac{((1-\alpha)r^f w + \alpha r w)^{1-\gamma}}{1-\gamma}].$$

Since, in a representative agent economy, the agent must be indifferent between assets for both to exist in equilibrium, checking this indifference condition gives a simplified way of investigating the equity premium puzzle using this binary model. Section 3.4 shows that to be indifferent a rationally inattentive agent requires a risk-free rate that solves:

$$\exp\frac{((1-\alpha)r^f w+\alpha r^f w)^{1-\gamma}}{\lambda(1-\gamma)} = \frac{2E[V]-(V_G+V_B)\pm\sqrt{(V_G+V_B-2E[V])^2+4V_GV_B}}{2},$$

where $V_i = \frac{((1-\alpha)r^f w+\alpha r_i w)^{1-\gamma}}{\lambda(1-\gamma)}$ for $i \in \{G,B\}$.

Calibration As a binary distribution has three free parameters, it can be calibrated to match the first three moments of US stock return data. I calibrate skewness to the mean firm-level returns documented in Albuquerque (2012)[5], giving a standardised third central moment of 0.531. I take a mean stock return of 0.081, and its standard deviation of 0.156 from Campbel (2003). Fraction of wealth in the stock market I take from Luo (2010) as $\alpha = 0.22$ and, as all wealth gets consumed in this static model, I set w to the conservative value of \$20,000.

Results With this calibration, the standard uninformed agent needs a utility curvature of $\gamma = 34.3$ for indifference between the risk-free and risky asset. With the curvature of utility at the relatively low value of $\gamma = 2$, the rationally inattentive can be made indifferent between the risk-free and the risky return by lowering her cost of information acquisition.

3.6.1.2 Risk-free Rate Puzzle

The risk-free rate puzzle is that the relatively rapid consumption growth over the life cycle, given the low returns on safe assets, implies an implausibly high elasticity of intertemporal substitution. In standard models, where inverse reciprocity holds, this puts a limit on explaining the equity premium puzzle by simply increasing risk-aversion.

Taking the calibrated model of Section 3.6.1.1, we can generate indifference between holding the risky and risk-free assets with a much lower curvature of the utility $\gamma = 0.5$. This is the type of value typically associated with solving the risk-free rate puzzle. Hence, the ability of rational inattention to separate these parameters offers a potentially simple solution to these two puzzles jointly: set the curva-

[5]The online appendix discusses using aggregate skewness.

ture of utility to solve the risk-free rate puzzle and the cost of attention to match the equity premium.

3.7 Conclusion

This paper has shown how rational inattention can separate the elasticity of intertemporal substitution from relative risk aversion within a time-additive expected-utility framework. This is because costly attention creates an additional reason to dislike learnable risk, namely the cost of reducing uncertainty. Calibrating simple models to stylised facts of the US economy suggests that this ability to disentangle risk and intertemporal preferences may help explain the equity premium and risk-free rate puzzles. This paper use entropy-based cost of attention, but these results may extend to other methods of modelling costly attention (e.g. Gabaix, 2014; Caplin et al., 2022). The parallels between the rationally inattentive agents' choice between information gathering and self-insurance and the choice between self-insurance and self-protection (e.g. Ehrlich and Becker, 1972) suggest a promising avenue of future research.

Chapter 4

Intergenerational Altruism and Transfers of Time and Money: A Life Cycle Perspective

4.1 Introduction

The intergenerational persistence in education, earnings, and wealth is well documented[1], yet the mechanisms behind it are less well understood. To better understand what drives this persistence, this paper estimates a dynastic model that includes three key mechanisms that link generations: i) parental time investments during childhood and adolescence that aid child development; ii) parental aid for education; and iii) cash gifts in the form of inter-vivos transfers and bequests.[2]

We use data from the National Child Development Survey (NCDS), which is an ongoing panel of the entire population of Britain born in a particular week in 1958. The data set contains multiple measures of parental time investments and cognitive skill in childhood, as well as educational outcomes and earnings over the life cycle. We use these data to estimate child skill production functions where parental time

[1]For evidence on intergenerational correlations, see Blanden et al. (2022), Hertz et al. (2008) for education; Solon (1992), Dearden et al. (1997), Mazumder (2005), Chetty et al. (2014) for earnings; and Charles and Hurst (2003) for wealth.

[2]For evidence on parental time investments during childhood and adolescence and their impact on child development see Cunha et al. (2006), Heckman and Mosso (2014), for parental aid for education see Belley and Lochner (2007), Abbott et al. (2019); and for cash gifts in the form of inter-vivos transfers and bequests see Castaneda et al. (2003), De Nardi (2004).

investments affect cognitive skill, which together with education determines wages. We then embed these production functions into a dynastic model in which altruistic couples choose consumption, labor supply, as well as transfers to their children of time, education, and cash.

The model is used to a) compare the relative importance of these types of transfers in explaining lifetime inequality, b) decompose the sources of lifetime income risk starting from birth to marriage, and c) evaluate the role of educational subsidies on educational decisions and subsequent lifetime inequality.

Our model contains five distinct mechanisms which can generate persistence in outcomes across generations. The first three mechanisms generate a positive correlation between the earnings of an individual and the earnings of their parents. These are, first, the borrowing constraint, which limits the ability of low income families to send their children to college. The second mechanism is that we allow parental productivity in investing in children to be correlated with productivity in the labor market. The estimated relationship is positive, which implies that the time investments more educated parents make in their children are more productive than those made by parents with less education. Third, we allow for a dynamic complementarity between early and late time investments and between time investments and educational invesments. While we find only modest complementarity between time investments in early childhood (0 to 7 years) and mid to late childhood (7 to 11 and 11 to 16 years), the complementarity between cognitive skill and years of education is much larger. This generates heterogenous returns to education and amplifies the effects of the first two channels. The fourth channel – positive assortative matching – generates persistence in household earnings over and above that observed between parents and their children. The final mechanism – cash transfers from parents to children – allows for a persistence in income and consumption over and above that seen for earnings. To the best of our knowledge, this is the first paper to include all of the above channels.

The estimated model implies an intergenerational elasticity of wages of 0.24, close to estimates for our cohort of interest in Dearden et al. (1997). The model also replicates the fact (documented by Guryan et al. (2008) and observed in our data) that parents with more education spend more time with their children.

We have three key findings. First, as noted above, we find modest dynamic complementarity between early time investments in children and later time investments. However, we find substantial complementarities between terminal childhood cognitive skill (measured at age 16) and years of education in wages. Among men with college education, a one standard deviation increase in cognitive skill at 16 leads to an additional 19% in wages. Among low education men, this premium is only 9%. As a result, high skill individuals are more likely to select into education than their low skill counterparts. This dynamic complementarity, in combination with borrowing constraints, is a key mechanism that perpetuates income inequality across generations.[3] High income households, who have more resources to send their children to college, have higher returns to investing in their child's cognitive skill than their low income counterparts; thus they invest more in their children.

Second, who one is born to and who one marries are central for explaining life's outcomes. We find that 30% of the variance of men's and 13% of the variance of women's lifetime wages can already be explained by characteristics of their parents, before the individual is even born. By the time individuals are 23, the shares rise to 65% and 45% for men and women, respectively. By modeling marriage and the behavior of both members of couples, we can assess not only the variability of individual but also household income. The characteristics of one's spouse are an important source of uncertainty in lifetime income prior to marriage, especially for women who on average earn less than their spouses. Resolution of this uncertainty explains almost half of the variability in *household* lifetime income for women.

Third, we evaluate the impact of a higher education subsidy on intergenerational persistence. We show that college subsidies are effective in reducing intergenerational persistence, but *only* if they are announced early, i.e. if parents are

[3]The interplay between borrowing constraints and investments in child human capital has also been studied in detail by Caucutt and Lochner (2020). Our paper is complementary to theirs in that we estimate crucial parameters of this mechanism - the wage equation and skill production - directly using a single data set, rather than having to calibrate them using multiple data sets (and different cohorts).

given sufficient time to adjust time investments in childhood. If parents cannot adjust these investments, the policy mostly benefits those who would have sent their children to college even in the absence of the subsidy. In this sense, unannounced transfers mostly provide a lump sum transfer to high income households which, if anything, increases income persistence across generations. If pre-announced, the increased returns to parental investments causes households to invest more in their children. This interplay between pre-announced policies, dynamic complementarity, and parental responses lead to increases in earnings, including for lower income households who would not have sent their children to college in the absence of the subsidy.

This paper relates to a number of different strands of the existing literature, including work measuring the drivers of inequality and intergenerational correlations in economic outcomes, the large literature seeking to understand child production functions and work on parental altruism and bequest motives. The most closely related papers, however, are those focused on the costs of and returns to parental investments in children. The four papers closest to ours are Caucutt and Lochner (2020), Lee and Seshadri (2019) and Daruich (2018) and Yum (2022). Each of those papers, like ours, contains a dynastic model in which parents can give time, education and money to their children. All four papers find that early life investments are key for understanding the intergenerational correlation of income. We build on the contributions of these papers in three ways. The first is that those papers lack data that links investments at young ages to earnings at older ages. As a result, they have to calibrate key parts of the model, while we are able to estimate the human capital production technology and show directly how early life investments and the resulting human capital impact later life earnings. Using the same sample throughout our analysis enables us to measure parental transfers, cognitive skill, and later life wage and other outcomes for one group of people in a single setting. Thus, we can, for example, use the same cognitive skill measures for the estimation of the human capital production function, as well as for the wage equation.

The second is that we explicitly model the behavior of both men and women before and after they are matched into couples. This allows us to show the quantitatively important role that assortative matching plays in amplifying the role of parental transfers in generating persistence in outcomes at the household level.

Finally, the focus of our paper is different. Caucutt and Lochner (2020) focus on identifying the role of market imperfections in rationalizing observed levels of parental investments. The aim of Lee and Seshadri (2019) is to simultaneously rationalize intergenerational persistence in outcomes and cross-sectional inequality in outcomes. Daruich (2018) focuses on the macroeconomic effects of large-scale policy interventions. Yum (2022) focuses on the role of heterogeneity in time investments. Our primary focus, facilitated by our data on each of the three parental inputs for our cohort of interest, is to quantitatively evaluate the role played by each, both for individuals and for households.

Other closely related papers include Del Boca et al. (2014) and Gayle et al. (2018), both of which develop models in which parents choose how much time to allocate to the labor market, leisure and investment in children. Neither paper, however, incorporates household savings decisions, and hence the trade-off between time investments in children now and cash investments later in life. Abbott et al. (2019) focuses on the interaction between parental investments, state subsidies and education decisions, but abstract from the role of parents in influencing skill prior to the age of 16. Castaneda et al. (2003) and De Nardi (2004) build overlapping-generations models of wealth inequality that includes both intergenerational correlation in human capital and bequests, but neither attempts to model the processes underpinning the correlation in earnings across generations. Bolt et al. (2021) use the same data as in this paper and mediation analysis to show that the mechanisms we consider in this paper are the key ones for explaining the persistence of income across generations. However, they do not allow for behavioral responses, and so cannot consider counterfactuals.

The rest of this paper proceeds as follows. Section 4.2 describes the data, and documents descriptive statistics on skill, education and the different types of parental transfers. Section 4.3 lays out the dynastic model used in the paper. Section

4 outlines our two step estimation approach. Section 5 then presents results from the first step estimation, whereas Section 6 presents identification arguments and results from the second step estimation. Section 7 presents results from counterfactuals and Section 4.8 concludes.

4.2 Data and Descriptive Statistics

The key data source for this paper is the National Child Development Study (NCDS). The NCDS follows the lives of all people born in Britain in one particular week of March 1958. The initial survey at birth has been followed by subsequent follow-up surveys at the ages of 7, 11, 16, 23, 33, 42, 46, 50 and 55.[4] During childhood, the data includes information on a number of cognitive skill measures, measures of parental time investments (discussed in more detail below) and parental income. Later waves of the study record educational outcomes, demographic characteristics, earnings and hours of work. For the descriptive analysis in this section, we focus on those individuals for whom we observe both their father's educational attainment (age left school) and their own educational qualifications by the age of 33. This leaves us with a sample of 9,436 individuals.

As the NCDS currently does not have data on the inheritances received or expected, we supplement it using data on individuals drawn from similar birth cohorts in the English Longitudinal Study of Ageing (ELSA). ELSA is a biennial survey of a representative sample of the 50-plus population in England, similar in form and purpose to the Health and Retirement Study (HRS) in the US. The 2012-13 wave of ELSA recorded lifetime histories of gift and inheritance receipt which we can use to augment our description of the divergence in lifetime economic outcomes by parental background. We use data on ELSA members who are born in the 1950s, which gives us a sample of 3,001.[5]

Lastly, to convert the investment measures observed into units of time, we use

[4]The age-46 survey is not used in any of the subsequent analysis as it was a more limited telephone-only interview.

[5]The next wave of the NCDS, which is currently in the field, is currently planned to collect information on lifetime inheritance receipt. We hope to use these new data in later versions of this work.

the UK Time Use Survey (UKTUS), which has detailed measures of time spent in educational investments in the child. We describe these measures in the notes of Table 4.2.2 and in greater detail in Appendix 4.C.

The rest of this section documents inequalities in the three types of parental transfers we are interested in (time investments, educational investments, and cash transfers), as well as subsequent outcomes (skill, lifetime income). Throughout the paper we use low, medium and high to describe education groups – these correspond to having only compulsory levels of education, having some post-compulsory education and having some college respectively.[6] In the US context this would correspond roughly to high school dropout, high school graduate, and some college.

4.2.1 Transfer Type 1: Parental Time Investments

The NCDS has detailed measures of parental time investments received during childhood. The full set of measures we use to estimate the impact of parental time on cognition are listed in Table 4.2.1.[7] These measures come from different sources – some are from surveys of parents, others from surveys of teachers. Here we highlight some of the key features in the data.

The first panel of Table 4.2.2 documents paternal education gradients for some of the investment measures we use. Whilst 52% of high educated fathers read to their age 7 child each week, only 33% of low educated fathers do so. The gradient is even more pronounced for the teacher's assessment of the parents' interest in the child's education: when the child is 7, 66% of high educated fathers are judged by the child's teacher to be 'very interested' in their child's education but only 20% of low education fathers are. While mothers are assessed as having greater interest in their child's education than fathers, there are large differences according to education group (75% of the highest education group are very interested, compared to 33% in the lowest education group).

[6]For this age group of fathers, compulsory education roughly corresponds to leaving school at age 14, post-compulsory means leaving school between ages 15 and 18, and some college means staying at school until at least age 19.

[7]While some of these measures are potentially costly in terms of money as well as time, we focus on the time cost which the previous literature has found to be the key determinant of child cognition (e.g., Del Boca et al. (2014)).

Table 4.2.1: List of all measures used

Skill measures	*Investment measures*
Age 0:	
Birthweight	Teacher's assessment of parents' interest in education
Gestation	Outings with child (mother and father)
	Read to child (mother and father)
	Father's involvement in upbringing
	Parental involvement in child's schooling
Age 7:	
Reading score	Teacher's assessment of parents' interest in education
Math score	Outings with child (mother and father)
Drawing score	Father's involvement in upbringing
Copying design score	Parents' ambitions for child's educational attainment
	Parental involvement in child's schooling
	Library membership of parents
Age 11:	
Reading score	Teacher's assessment of parents' interest in education
Math score	Involvement of parents in child's schooling
Copying design score	Parents' ambitions regarding child's educational attainment
Age 16:	
Reading score	
Math score	

Notes: All investment measures are retrospective, so age 0 investments are measured at age 7, age 7 investments are measured at age 11, age 11 investments are measured at age 16.

4.2.2 Transfer Type 2: Educational Investments

Panel 2 of Table 4.2.2 shows that there is a substantial intergenerational correlation in educational attainment between fathers and their children. Having a high-educated father makes it much more likely that a child will end up with high education. 46% of the children of high educated fathers also end up with high education, compared to only 13% of those whose fathers have low education.

4.2.3 Transfer 3: Inter-vivos Transfers and Bequests

The third panel of Table 4.2.2 documents the receipt of inter-vivos transfers and bequests as reported in ELSA by father's education. The table shows significant differences in the receipt of inter-vivos transfers depending on parental education. Only 6% of individuals from low education families report having received a transfer worth more than £1,000, compared to 20% from high educated families. Moreover, conditional on receipt of a gift, the average value for the two groups differs by about £18,400.

Table 4.2.2: Transfers and outcomes by father's education

			Father's education			
	Avg	SD	low	med	high	F-test
Parental Investments						
Mother reads each week 7	0.49	0.50	0.46	0.56	0.67	0.00
Father reads each week 7	0.36	0.48	0.33	0.44	0.52	0.00
Mother outings most weeks 11	0.54	0.50	0.53	0.61	0.59	0.00
Father outings most weeks 11	0.51	0.50	0.50	0.58	0.56	0.00
Father very interested in educ 7	0.26	0.44	0.20	0.43	0.66	0.00
Mother very interested in educ 7	0.39	0.49	0.33	0.58	0.75	0.00
Father very interested in educ 11	0.31	0.46	0.23	0.52	0.73	0.00
Mother very interested in educ 11	0.39	0.49	0.33	0.59	0.76	0.00
Father very interested in educ 16	0.36	0.48	0.28	0.57	0.80	0.00
Mother very interested in educ 16	0.38	0.49	0.32	0.59	0.78	0.00
Time spent with child [UKTUS]*	9.06	10.05	8.35	8.91	9.87	.52
Child Education						
Fraction low education	0.25	0.43	0.30	0.10	0.02	0.00
Fraction high education	0.16	0.37	0.13	0.31	0.46	0.00
Cash Transfers						
Inter-vivos transfers (>£1000)	0.07	0.26	0.06	0.10	0.20	0.06
Gift value (among recipients only)	39,400	104,600	30,600	77,900	49,100	0.72
Fraction receiving inheritance	0.39	0.49	0.36	0.58	0.54	0.00
Inheritance value (among recipients)	88,200	114,700	75,600	122,400	174,300	0.00
Child Skills						
Reading 7	0.00	1.00	-0.09	0.33	0.58	0.00
Reading 11	0.00	1.00	-0.13	0.46	0.90	0.00
Reading 16	0.00	1.00	-0.11	0.47	0.77	0.00
Maths 7	0.00	1.00	-0.08	0.26	0.54	0.00
Maths 11	0.00	1.00	-0.13	0.48	0.91	0.00
Maths 16	0.00	1.00	-0.14	0.48	0.99	0.00
Lifetime Earnings in £1,000						
Men	1,347	352	1,289	1,533	1,740	0.00
Women	925	239	879	1,048	1,197	0.00

Notes: For different types of transfers and outcomes, Table 4.2.2 shows: Mean, standard deviation, mean conditional on each paternal education group (low, medium, high). 75% of fathers are low education, 20% are middle education, and 5% are high education. **P-values* for an *F-test* of the difference in the mean between the low and high father's education group. **Sum of father's and mother's time spent on the following activities spent with the child in UKTUS data: teaching the child, reading/playing/talking with child, travel escorting to/from education.

Differences in inheritance receipt by parental background are also significant.[8] 54% of those with high educated fathers have received an inheritance, compared to 36% of those with low-educated fathers, and among those who have received an inheritance, those with high educated fathers have received more than twice as much on average (£174,300 compared to £75,600). The net result is that those with high educated fathers inherit £66,000 more than those with low-educated fathers.

4.2.4 Outcome 1: Skill

The fourth panel of Table 4.2.2 shows the average reading and math scores of children at ages 7, 11, and 16, by father's education. As one might expect, children whose fathers have a higher level of education have higher skill levels; at the age of 7, the reading score of children of low educated fathers is 0.09 standard deviations below average, whereas it is 0.58 above average for children of high educated fathers. This gap in reading scores widens with age: by the time the children are 16, reading scores of children of low educated fathers is 0.11 standard deviations below average, whereas they are 0.77 above average for children of high educated fathers. Similar patterns are found for math scores.

4.2.5 Outcome 2: Lifetime Earnings

Finally, we can see that children of more educated fathers have higher lifetime earnings. The gap in lifetime earnings between men with high educated fathers versus those with low educated fathers is £451k. For women, the difference is £318k.

To summarize, we find that children from more highly educated fathers tend to receive more of each of the three kinds of transfers, and they end up with higher skills, as well as lifetime income. In the following, we present a model bringing together these different types of transfers to explain how these operate in generating the intergenerational persistence in outcomes that we observe.

[8] Sample statistics are calculated for those who lost both parents when interviewed, which is 75% of our ELSA sample.

4.3 Model

This section describes a dynastic model of consumption and labor supply in which parents can make different types of transfers to their children. Figure 4.3.1 illustrates the model's timeline. During childhood, parental time investments in children and educational choices affect the evolution of the child's cognitive skill (which we refer to as skill below) and their educational attainment. Upon reaching age 23, they are matched in couples, possibly receive transfers of cash from their parents and begin adult life. They then have their own children, choose consumption, labor supply, and how much to invest in their own children, with implications for their children's future outcomes.

The NCDS interviews respondents every four to seven years from the age of 0 to 55. To be consistent with the data, each of our model periods will cover the time between interviews (and each period will be of different length). Each individual has a life cycle of 20 model periods which can be broken down into four phases.

1. Childhood has periods $t = 1,2,3,4$ which corresponds to ages 0-6, 7-10, 11-15, 16-22. During childhood the individual accumulates human capital and education but does not make decisions.

2. Young Adult consists of one period at $t = 5$ corresponding to ages 23-25. The individual receives a parental cash transfer (which is potentially 0), is matched into a couple and begins making labor supply and savings decisions.

3. Parenthood has five periods $t = 6,7,8,9,10$, corresponding to ages 26-32, 33-36, 37-41, 42-48, 49-54. The couple have identical twin children at the start of the 'Parenthood' phase. In addition to making labor supply and savings decisions, the couple decide how much to invest in their childrens' human capital and education. At the end of this period they have an opportunity to transfer wealth to their children who in turn are matched into couples.

4. Late adult phase consists of 10 regularly-spaced periods corresponding to ages 55-59, ..., 100-104. The household separates from their children and makes their own saving and consumption decisions.

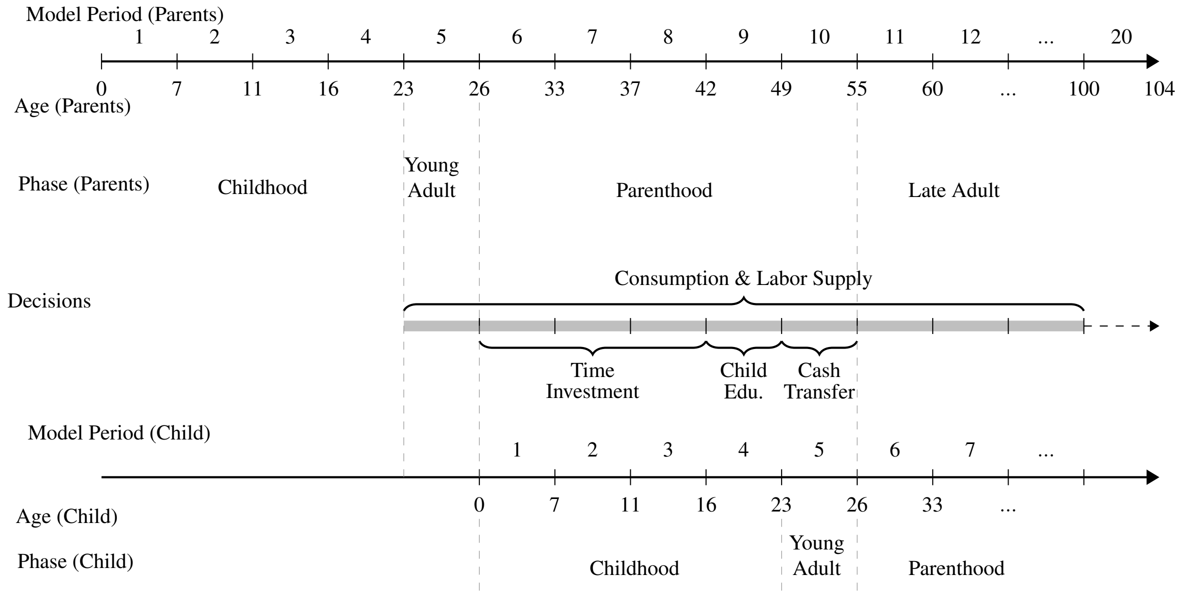


Figure 4.3.1: The life cycle of an individual

In outlining the dynastic model we describe below a life cycle decision problem of a single generation. All generations are, of course, linked; each couple has children. These children, in turn, will form couples who have children, too. To index generations we use t to denote the age (in model periods) of the generation we consider and a prime to denote their childrens' variables. For example, in the model period when adults are aged t, their children are aged t'.[9]

We now provide formal details of the model.

4.3.1 Preferences

The utility of each member of the couple $g \in \{m, f\}$ (male and female respectively) depends on their consumption ($c_{g,t}$) and leisure ($l_{g,t}$):

$$u_g(c_{g,t}, l_{g,t}) = \frac{(c_{g,t}^{\nu_g} l_{g,t}^{(1-\nu_g)})^{1-\gamma}}{1-\gamma}$$

We allow preferences for consumption and leisure to vary with gender. Households equally weight the sum of male and female utility. The household utility function is multiplied by a factor n_t which represents the number of equivalized adults in a

[9]Children are born five model periods after their parents, therefore they are aged $t' = 1$ in model periods when the parent is model-aged $t = 6$.

household in time t (scaled so that for a childless couple $n_t = 1$).

$$u(c_{m,t}, c_{f,t}, l_{m,t}, l_{f,t}, n_t) = n_t \left(u_m(c_{m,t}, l_{m,t}) + u_f(c_{f,t}, l_{f,t}) \right)$$

Total household consumption is split between children, who receive a fraction $\frac{n_t - 1}{n_t}$, and adults who get a share $\frac{1}{n_t}$. The quantity of leisure is:

$$l_{g,t} = T - (\theta ti_{g,t} + hrs_{g,t}) \tag{4.1}$$

where T is a time endowment, $ti_{g,t}$ is time investment hours in children, $hrs_{g,t}$ is work hours, and $l_{g,t}$ is leisure time. $1 - \theta$ is the share of time with the child that represents leisure to the parent: if θ =0 then time with children is pure leisure for the parent, whereas if θ =1 then time with children generates no leisure value.

The annual discount factor is β. The model period length aligns with the differences in time between interviews and so the discount factor between model period varies. Thus the discount rate between t and $t+1$ is $\beta_{t+1} = \beta^{\tau_t}$, where τ_t is the length of model period t.[10]

Each generation is altruistic regarding the utility of their offspring (and future generations). In addition to the time discounting of their children's future utility (which they discount at the same rate they discount their own future utility), they additionally discount it with an intergenerational altruism parameter (λ).

4.3.2 Demographics

All individuals are matched probabilistically into couples, conditional on education. The probability that a man of education ed_m gets married to a women with education ed_f is given by $Q_m(ed_m, ed_f)$. The matching probabilities for females are $Q_f(ed_f, ed_m)$. The draw of spousal skills and initial wealth is therefore drawn from a distribution that depends on one's own education.

At age 26, a pair of identical twins is born to the couple. In order to match the average fertility for this sample, which is close to two, yet still maintain com-

[10] In addition, to account for varying period length and within period discounting we weight each period's utility by $\sum_{q=0}^{\tau_t} \beta^q = \frac{1-\beta^{\tau_t+1}}{1-\beta}$.

putational tractability, we follow Abbott et al. (2019) and assume that the twins are faced with identical sequences of shocks.

Mortality is stochastic - the probability of survival of a couple (we assume that both members of a couple die in the same year) to period $t+1$ conditional on survival to period t is given by s_{t+1}. We assume households face mortality risk after the age of 50 and that death occurs by the age of 105 at the latest.

4.3.3 Human Capital

This section describes the production function for skill and education from birth to age 23. During this part of the life cycle, parental time investments do not directly impact the contemporaneous utility of their children, but leads (in expectation) to the children having higher wages, more able spouses and more able childrens' children, all of which matters to the altruistic parent.

4.3.3.1 Child Skill Production Function

Between birth and age 16, children's skill updates each period according to the production function:

$$h'_{t'+1} = \gamma_{1,t'} h'_{t'} + \gamma_{2,t'} ti_{t'} + \gamma_{3,t'} ti_{t'} \cdot h'_{t'} + \gamma_{4,t'} ed_m + \gamma_{5,t'} ed_f + u'_{h,t'+1} \tag{4.2}$$

where $h'_{t'}$ represents children's skill when the children are age t'. Children's skill depends on their parents' level of education, the sum of the time investments ($ti_{t'} = ti_{m,t'} + ti_{f,t'}$) those parents make, past skill, and a shock ($u'_{h,t+1}$). Skill evolves until period 4 (age of 16), after which it does not change.

We allow education of the parents, ed_m and ed_f, to impact skill to capture the idea that high skill individuals who are productive in the labor market may also be productive at producing skills in their children. This is a mechanism that features prominently in several recent studies of the labor market (e.g., Lee and Seshadri (2019)).

Children's initial skill at birth $h'_{t'=1}$ is a function of their parents' level of education and a shock:

$$h'_{t'=1} = \gamma_{4,0} ed_m + \gamma_{5,0} ed_f + u'_{h,0}. \tag{4.3}$$

4.3.3.2 Education

When children are age 16, parents choose the education level of their children. There was compulsory education to age 16 for our sample members. Thus we model the decision to send children to school until age 16, age 18 (completing secondary education) or 21 (completing undergraduate education). Because there were no tuition fees for the cohort we study, we model the cost of education as forgone labor income when at school.

4.3.3.3 Wages

The wage rate evolves according to a process that has a deterministic component which varies with age and whether the individual works part-time or fulltime, and a stochastic component:

$$\ln w_t = \delta_0 + \delta_1 t + \delta_2 t^2 + \delta_3 t^3 + \delta_4 PT_t + v_t \tag{4.4}$$

where PT_t is a dummy for working part-time. To capture the impact of skill on lifetime wages, we model the initial wage draw in period 5 (age 23) as a function of final skill (h) and a shock: subsequent values follow a random walk

$$v_t = \begin{cases} \delta_5 h + \eta_t, & \eta_t \sim N(0, \sigma^2_{\eta 4}) \text{ if } t = 5 \\ v_{t-1} + \eta_t, & \eta_t \sim N(0, \sigma^2_{\eta}) \text{ if } t > 5 \end{cases}. \tag{4.5}$$

Skill impacts the age 23 wage shock v_5 and thus impacts wages at all ages because v_t is modeled as having a unit root. Thus we do not need to keep track of skill after turning age 23, but instead we keep track of wages as a state variable, which includes v_t and thus final skill. While the associated subscripts are suppressed above, each of $\{\delta_0, \delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \sigma_\eta, \sigma_{\eta 4}\}$ varies by gender (g) and education (ed). This flexibility means that we allow skill to impact wages through its relationship with education δ_5. As we show below this flexibility is important as the returns to skill are higher for the highly educated.

4.3.4 Budget Constraints

Constraints Households face an intertemporal budget constraint and a borrowing constraint:

$$a_{t+1} = (1+r_t)(a_t + y_t - (c_{m,t} + c_{f,t}) - x_t) \quad (4.6)$$

$$a_{t+1} \geq 0 \quad (4.7)$$

where a_t is the household wealth, y_t is household income and x_t is a cash transfer to children that can only be made when the members of the couple are 49 and their children are 23 (and so $x_t = 0$ in all other periods). The gross interest rate $(1+r_t)$ is equal to $(1+r)^{\tau_t}$ where r is an annual interest rate and τ_t is the length in years of model period t.

Earnings and household income Earnings are equal to hours worked (hrs) multiplied by the wage rate, for example: $e_{f,t} = hrs_{f,t} w_{f,t}$. Household net-of-tax income is

$$y_t = \tau(e_{m,t}, e_{f,t}, e'_t, t) \quad (4.8)$$

where $\tau(.)$ is a function which returns net-of-tax income and $e_{m,t}$ and $e_{f,t}$ are male and female earnings respectively. Before children turn 16 their earnings (e'_t) are 0. Upon turning age 16, if children leave education, the work full time at the median wage given their age and gender for the rest of the model period. Their parents are still the decision-maker in this period and any income the children earn is part of the parental household income.

4.3.5 Decision Problem

4.3.5.1 Decision Problem in the Young Adult Phase

An individual becomes an active decision maker at age 23 when they are already formed into a household as part of a childless couple. As such $t = 5$ is the first model period with a decision problem to solve.

Choices Each period during this phase, couples choose consumption $(c_{m,t}, c_{f,t})$ and hours of work of each parent $(hrs_{m,t}, hrs_{f,t})$ where $hrs_{g,t} \in \{0, 20, 40, 50\}$ hours per week. The resulting vector of decision variables is $\mathbf{d_t} = (c_{mt}, c_{f,t}, hrs_{m,t}, hrs_{f,t})$.

Uncertainty Couples face uncertainty over the innovation to each of their wages next period $\{\eta_{m,t}, \eta_{f,t}\}$ and the initial skill level of their future children $u'_{h,0'}$.

State variables The vector of state variables $(\mathbf{X}_t)$ during young adulthood is $\mathbf{X}_t = \{t, a_t, w_{m,t}, w_{f,t}, ed_m, ed_f\}$ where t is age, a_t is assets, $w_{g,t}, ed_{g,t}$ are the wages and education of each parent for $\{g \in m, f\}$.

Value function The value function for the young adult phase is given below in expression (4.9):

$$V_t(\mathbf{X}_t) = \max_{\mathbf{d_t}} \left\{ u(c_{mt}, c_{f,t}, hrs_{m,t}, hrs_{f,t}, n_t) + \beta_{t+1} \mathbb{E}_t [V_{t+1}(\mathbf{X}_{t+1})] \right\} \quad (4.9)$$

subject to the intertemporal budget constraint in equation (4.6) and the borrowing constraint in equation (4.7) where the expectation operator is over the innovation to the wage of each of spouse $(\eta_{m,t}, \eta_{f,t})$ and the initial skill of the child $(u'_{h,0'})$.

4.3.5.2 Decision Problem in the Parenthood Phase: Before Children Reach Young Adulthood

Choices Households make decisions on behalf of both the adults and children within the household each period. They choose consumption and hours of work of each parent $(c_{m,t},\ c_{f,t}, hrs_{m,t},\ hrs_{f,t})$, time investments in children of each parent $(ti_{m,t}$ and $ti_{f,t})$ until their child turns 16, and childrens' education ed' in the period the children turn 16. The resulting vector of decision variables is $\mathbf{d_t} = (c_{m,t}, c_{f,t}, hrs_{m,t}, hrs_{f,t}, ti_{m,t}, ti_{f,t}, ed')$.

Uncertainty Couples face uncertainty over the innovation to each of their wages $\{\eta_{m,t}, \eta_{f,t}\}$ and the innovations to the childrens' skills $(u'_{h,t})$.

State variables The set of state variables in this phase is $\mathbf{X}_t = \{t, a_t, w_{m,t}, w_{f,t}, ed_m, ed_f, g', h'_t\}$, which is the same as in the independent adult phase plus the childrens' gender (g')

and their skill level (h'_t).

Value function The household's value function and the constraints are the same as in equation (4.9), except adapted to have the sets of choices, uncertainty and states described immediately above.

4.3.5.3 Decision Problem in the Parenthood Phase: When Children Become Young Adults

Couples can make a final decision that affects their dependent children when age 49 (with children age 23). Their children then become fully independent, get married, and begin making their own decisions.

Choices During this phase couples choose consumption ($c_{m,t}, c_{f,t}$), hours of work for each parent ($hrs_{m,t}, hrs_{f,t}$), and a cash gift (x_t) which is split equally between their two children. The resulting vector of decision variables is $\mathbf{d_t} = (c_{m,t}, c_{f,t}, hrs_{m,t}, hrs_{f,t}, x_t)$.

Uncertainty Couples face two distinct types of uncertainty. The first is uncertainty over the characteristics of their children as they start adulthood. The dimensions of uncertainty here are the childrens' initial wage draw and the attributes of their future spouse (his/her skill, education level, assets, and initial wage draw). The second dimension of uncertainty is with respect to their own circumstances next year – that is their next period wage draws.

State variables The set of state variables in this phase is the same as in the parenthood phase plus childrens' education (ed').

Value function The decision problem in this final period of parenthood is:

$$V_t(\mathbf{X}_t) = \max_{\mathbf{d_t}} \left\{ u(c_{m,t}, c_{f,t}, l_{m,t}, l_{f,t}, n_t) + 2\lambda \mathbb{E}_t[V'_{t'}(\mathbf{X}'_{t'})] + \beta_{t+1}\mathbb{E}_t[V_{t+1}(\mathbf{X}_{t+1})] \right\} \tag{4.10}$$

subject to equations (4.6) and (4.7). Note that there are two continuation value functions here. The first is the expected value of the couple to which the (soon to be independent) children of the parents will belong to, and the expectation operator is over the children's initial wage draw and their future spouse's attributes. The altruistic parents take this into account in making their decisions. This continuation utility is discounted by the altruism parameter (λ) and the integration is with respect to the children's initial wage draw and the characteristics of their spouse (these shocks are realized after the parents make their decisions). We have assumed that parents have two identical children and therefore we multiply this continuation value by 2. The second continuation value function is the future expected utility that the parents will enjoy in the next period (when they will enter the late adult phase). This expectation operator is with respect to next period's wage draws, which are stochastic, and discounted by β_{t+1}, the time discount factor.

4.3.5.4 Decision Problem in the Late Adult phase

At this stage the children have entered their own parenthood phase and the parent couple enters a late adult phase.

Choices Households make labor supply and consumption/saving decisions only ($\mathbf{d_t} = (c_{mt}, c_{f,t}, hrs_{m,t}, hrs_{f,t})$).

Uncertainty There is uncertainty over next period's wage draws and survival s_t (we assume both members of the couple die in the same period).

State variables The vector of state variables is $\mathbf{X}_t = \{t, a, w_m, w_f, ed_m, ed_f\}$. The skill level and education of the (now-grown-up) children are no longer state variables.

Value function Given the definitions of choices, states, and uncertainty for the late life phase the value function and the constraints take the same form as for the young adult phase (expression (4.9)).

$$V_t(\mathbf{X}_t) = \max_{\mathbf{d_t}} \left\{ u(c_{mt}, c_{f,t}, hrs_{m,t}, hrs_{f,t}, n_t) + \beta_{t+1} s_{t+1} \mathbb{E}[V_{t+1}(\mathbf{X}_{t+1})] \right\}$$

subject to equations (4.6) and (4.7).

4.4 Estimation

We adopt a two-step strategy to estimate the model. In the first step, we estimate or calibrate those parameters that, given our assumptions, can be cleanly identified outside our model. In particular, we estimate the skill function, the wage process, marital sorting process, and mortality rates. In addition, we also estimate the initial conditions (of the joint distribution of education, skill level, gender, and parental transfers received at age 23) directly from the data. We calibrate the interest rate, parameters of the tax code (taken from Taxben, a tax-benefit microsimulation model developed by the Institute for Fiscal Studies (Waters (2017))), and household equivalence scale parameter.

In the second step we estimate the remaining parameters using the method of simulated moments and correct for selection bias in the wage equation.

4.4.1 Estimating the Human Capital Production Function

Estimating the latent factor production function We have multiple noisy measures of children's latent skill ($h'_{t'}$) and parental investment ($inv_{t'}$) in our NCDS data. Following the recent literature (Agostinelli and Wiswall (2022)), we estimate a human capital production function where latent skill is a function of previous period's (latent) skill level and investments, parental education, and a shock:

$$h'_{t'+1} = \alpha_{1,t'} h'_{t'} + \alpha_{2,t'} inv_{t'} + \alpha_{3,t'} inv_{t'} \cdot h'_{t'} + \alpha_{4,t'} ed^m + \alpha_{5,t'} ed^f + u'_{h,t'} \qquad (4.11)$$

We explicitly account for measurement error in the latent factors using a GMM implementation of the methods in (Agostinelli and Wiswall (2022)). Following the literature (Cunha and Heckman (2008), Cunha et al. (2010)), we assume independence of measurement errors, allowing us to use all possible combinations of (noisy) input measures to instrument for one another using a system GMM approach described in Appendix 4.D.

Converting latent investments to time Equation (4.11) gives us the coefficient of a unit of latent investment on a unit of latent skills. However, latent skills and latent investments do not have a natural scale. We normalize the scale of the skill measure via the wage equations (4.4 and 4.5), which we discuss in Appendix 4.G below.

We anchor latent parental investments to hours of investment time, as this is the relevant object in the model. To anchor the latent investments estimated using the NCDS to time, we use another data set that contains information on hours of time spent with children – the UK Time Use Survey (UKTUS). We assume time investments with children impact latent investments according to:

$$inv_{t'} = \kappa_{0,t'} + \kappa_{1,t'}(ti_{m,t'} + ti_{f,t'}) \tag{4.12}$$

where $\kappa_{1,t'}$ is the hours-to-latent investments conversion parameter which determines the productivity of time investments and $\kappa_{0,t'}$ is a constant that ensures we match mean time investments. We allow the κ parameters to vary by age, to reflect that parental time investments, and the productivity of those investments, varies by age.

The parameters $\kappa_{0,t'}$ and $\kappa_{1,t'}$ are estimated using MSM by matching age 16 skill by father's education in the NCDS data and time investments by parental education in the UKTUS data. We discuss the estimation and identification of $\kappa_{0,t'}$ and $\kappa_{1,t'}$ in Section 6.

With the parameters $\kappa_{0,t'}$ and $\kappa_{1,t'}$ in hand, we substitute equation (4.12) into equation (4.11) as follows, where the second line is (4.2) which is the production function we use in our dynamic programming model:

$$\begin{aligned} h'_{t'+1} &= \alpha_{1,t'}h'_{t'} + \alpha_{2,t'}(\kappa_{0,t'} + \kappa_{1,t'}ti_{t'}) + && (4.13)\\ &\quad \alpha_{3,t'}(\kappa_{0,t'} + \kappa_{1,t'}ti_{t'}) \cdot h'_{t'} + \alpha_{4,t'}ed^m + \alpha_{5,t'}ed^f + u'_{h,t'} && (4.14)\\ &= \gamma_{0,t'} + \gamma_{1,t'}h'_{t'} + \gamma_{2,t'}ti_{t'} + \gamma_{3,t'}ti_t \cdot h'_{t'} + \gamma_{4,t'}ed^m + \gamma_{5,t'}ed^f + u'_{h,t'} \end{aligned}$$

where $\gamma_{0,t'} = \alpha_{2,t'}\kappa_{0,t'}, \quad \gamma_{1,t'} = (\alpha_{3,t'}\kappa_{0,t'} + \alpha_{1,t'}),$
$\gamma_{2,t'} = \alpha_{2,t'}\kappa_{1,t'}, \quad \gamma_{3,t'} = \alpha_{3,t'}\kappa_{1,t'}, \quad \gamma_{4,t'} = \alpha_{4,t'}, \quad \gamma_{5,t'} = \alpha_{5,t'}.$

4.4.2 Identification and Estimation of the Wage Equation

We estimate the wage equation laid out in equations (4.4) and (4.5), but allow for i.i.d. measurement error in wages u_t. Using those equations and noting that $v_t = \delta_5 h_4 + \sum_{k=5}^{t} \eta_k$ yields:

$$\ln w_t^* = \ln w_t + u_t = \delta_0 + \delta_1 t + \delta_2 t^2 + \delta_3 t^3 + \delta_4 PT_t + \delta_5 h_4 + \sum_{k=5}^{t} \eta_k + u_t \quad (4.15)$$

for each gender and education group.

In our procedure we must address three issues. First, wages are measured with error u_t. Second, the skill level h_4 is measured with error. Third, we only observe the wage for those who work, which is a selected sample.

We can address some problems of selectivity using our panel data. To address the issue of composition bias (the issue of differential labor force entry and exit by lifetime wages), we use a fixed effects estimator. Given our assumption of a unit root in $v_t = \delta_5 h_4 + \sum_{k=5}^{t} \eta_k$, which we estimate to be close to the truth, we can allow v_5 (the first shock to wages) to be correlated with other observables, and estimate the model using fixed effects. In particular, we estimate δ_1, δ_2, δ_3, δ_4 and an individual fixed effects using a fixed effects estimator:

$$\ln w_t^* = \delta_1 t + \delta_2 t^2 + \delta_3 t^3 + \delta_4 PT_t + FE + \xi_t$$

where $FE = \delta_0 + \delta_5 h_4 + \eta_5$ is a person specific fixed effect capturing the time invariant factors and $\xi_t = \sum_{k=6}^{t} \eta_k + u_t$ is a residual. We then use a methodology similar to that described in section 4.4.1 to estimate δ_5 where we use multiple noisy measures of skills to instrument for each other. We then estimate the variances of the wage shocks $(\sigma_{\eta_5}^2, \sigma_{\eta}^2)$ and the variance of the measurement error (σ_u^2) using an error components procedure.

The above procedure addresses problems of measurement error in skill as well as selection based on permanent differences in productivity but not selection based on wage shocks. We control for this last aspect of selection bias by finding the wage profile that, when fed into our model, generates the same estimated profile (i.e., the same δ parameters from equation (4.15)) that we estimated in the data. Because the simulated profiles are computed using only the wages of those simulated agents that work, the simulated profiles should be biased for the same reasons they are in the data. We find this bias-adjusted wage profile using the iterative procedure described in French (2005). See Appendix 4.G for details.

4.4.3 Method of Simulated Moments

We estimate the rest of the model's parameters (discount factor, consumption weight for both spouses, risk aversion, altruism weight, share of time with the child that represents leisure to the parent, the hours-to-latent investments conversions):

$$\Delta = (\beta, \nu_f, \nu_m, \gamma, \lambda, \theta, \{\kappa_{0,t'}, \kappa_{1,t'}\}_{\{t'=1,2,3\}})$$

with the method of simulated moments (MSM), taking as given the parameters that were estimated in the first step. In particular, we find the parameter values that allow simulated life-cycle decision profiles to "best match" (as measured by a GMM criterion function) the profiles from the data.

Because we wish to understand the drivers of parental labor supply and time investments, we match employment choices for both spouses and also household time spent with children, by parents' age and education. Because we wish to understand the drivers of education and money transfers, we also match educational decisions, as well as cash transfers to children when the children are older. Because we wish to understand how households discount the future, we match wealth data. Finally, to understand the relationship between time and latent investments, we match observed hours spent with children and their observed skill level. In particular, the moment conditions that comprise our estimator are given by

1. Employment rates, by age, gender, and education, from the NCDS data (30 moments)

2. Fraction in full time work conditional on being employed, by age, gender, and education, from the NCDS data (30 moments)

3. Mean annual time spent with children, by child's age and parent's gender and education, from the UKTUS data (18 moments)

4. Mean age at which individuals left fulltime education by fathers' education level from the NCDS data (3 moments)

5. Mean lifetime receipt of inter-vivos transfers, from ELSA (1 moment)

6. Median wealth at 60 from ELSA (1 moment)

7. Mean skill at age 16 by father's education, from the UKTUS data (3 moments)

We observe hours and investment choices of individuals in the NCDS, and thus match data for these individuals for the following years: 1981, 1991, 2000, 2008, and 2013 when they were 23, 33, 42, 50 and 55.

The mechanics of our MSM approach are as follows. We simulate life cycle histories of shocks to skill level, wages, partnering and childrens' gender and skills for a large number of artificial individuals over multiple generations. Each individual is endowed with a gender and a value of the age-23 education, wealth, and partner characteristics drawn from the empirical distribution from the NCDS data. The initial stochastic component of wages v_5 is drawn from a parametric distribution estimated on the NCDS data (see section 4.4.2).

Next, using value function iteration, we solve the model numerically. We solve backwards through time, embedding a backwards recursion over each life cycle of multiple generations. Our solution concept involves finding a fixed point in decisions rules over generations. Using these decision rules, in combination with simulated endowments and the trajectories of shocks, we simulate the profiles of behavior for a large number of artificial households, each composed of a man and woman. The behaviors that we can simulate are those that our modelled agents decide: assets, work hours and time investments, child's educational choices, and inter-vivos transfers. We use the resulting profiles to construct moment conditions, and evaluate the match using our GMM criterion function. We search over the parameter space for the values that minimize this criterion. Appendix 4.I contains a detailed description of our moment conditions, the weighting matrix in our GMM criterion function, and the asymptotic distribution of our parameter estimates and Appendix 4.H gives details of our computational procedures.

4.5 First Step Estimation Results

In this section we describe results from our first-step estimation that we use as inputs for our structural model. We present estimates of the effect of parental time investments on children's skill, and how that skill in turn affects subsequent education and adult earnings. This exploits a key advantage of our data - that we measure for the same individuals their parents' investments, their level of skill and the value of that skill in the labor market.

4.5.1 The Determinants of Skill

In Section 4.2 we documented that children of high educated parents do better in cognitive tests, and that the skill gaps between children of high and low educated parents grow over time. Combining multiple test scores to create a measure of skills, we estimate a skill production function using the methods described briefly in Section 4.4.1 and in more detail in the appendix.

We estimate equation (4.2) for skills at ages 7, 11, and 16. The time investments entering the equation are those corresponding to ages 0-6, 7-10, and 11-16, respectively. Estimates are presented in Table 4.5.1 (Appendix 4.E gives estimates of the initial skill draw). To ease interpretation, we normalize our skill and time measures to have unit variance in every period.

We estimate age 7 skill as a function of age 0 skill, age 0 time investments, the interaction of skill and time investments, and mother's and father's education. It shows that time investments have a significant effect on skill, even after conditioning on background characteristics and initial skill. Evaluated at mean skill, a one standard deviation increase in time investments at age 0-6 raises age-7 skill by approximately 0.15 standard deviations, a one standard deviation increase in time investments at age 7-10 raises age-11 skill by 0.10 standard deviations, and a one standard deviation increase in time investments at age 11 raises age-16 skill by 0.13 standard deviations. Skill levels are very persistent, especially at older ages, implying a high level of self-productivity.

Interestingly, the interaction between skills and investments is negative for age 7 and 16, but positive for age 11. This implies that whilst at young ages, invest-

Table 4.5.1: Determinants of skills.

	Age 7	Age 11	Age 16
Lagged Skill	0.154	0.739	0.939
	[0.057, 0.251]	[0.696, 0.834]	[0.918, 0.993]
Investment	0.146	0.097	0.131
	[0.113, 0.171]	[0.079, 0.116]	[0.093, 0.161]
Lagged Skill × Investment	-0.021	0.040	-0.038
	[-0.067, 0.010]	[0.027, 0.068]	[-0.066, -0.009]
Mum: Medium Education	0.448	0.181	0.027
	[0.347, 0.552]	[0.109, 0.235]	[-0.026, 0.075]
Mum: High Eduction	0.593	0.414	-0.088
	[0.388, 0.776]	[0.292, 0.571]	[-0.242, 0.055]
Dad: Medium Education	0.472	0.262	0.056
	[0.252, 0.611]	[0.179, 0.321]	[0.002, 0.115]
Dad: High Education	0.401	0.460	0.107
	[0.313, 0.495]	[0.290, 0.548]	[0.010, 0.218]
Skill shock: $Var(u_{ab',t'})$	0.031	0.067	0.026

Notes: GMM estimates. Confidence intervals are bootstrapped using 100 replications. For the production function at age 7, we use skill measured at age 7 as a function of skill at age 0, time investments measured at age 7 (and referring to investments at age 0-6). For the production function at age 11, we use skill measured at age 11 as a function of skill at age 7, time investments measured at age 11 (and referring to investments at age 7-10). For the production function at age 16, we use skill measured at age 16 as a function of skill at age 11, time investments measured at age 16 (and referring to investments at age 11-15).

ments are more productive for low-skilled children, at older ages, productivity is higher for the higher-skilled ones. The positive and statistically significant coefficients on the age 11 interaction terms indicates that the skill production function exhibits dynamic complementarity at this stage of childhood (as found by Cunha et al. (2010)). However, at all ages the extent of complementarity or substitutability is modest. For example, for those with age 7 skill levels one standard deviation below (above) mean, a one standard deviation in investment delivers a 0.097-0.040=0.057 (0.097+0.040=0.137) increase in age 11 skills levels.

While the richness of our data allows us to account for measurement error in skills and investments, we do not believe our setting allows for credible exclusion restrictions that would allow us to account for the potential endogeneity of investments. The literature has not yet come to a consensus as to whether potential endogeneity would lead us to over- or understate the returns to investments. Attanasio et al. (2020) and Attanasio et al. (2020) find that failure to account for endogeneity leads to an understatement of the returns to investments in all periods, whereas Cunha et al. (2010) find that it leads to an overstatement of the returns for older children[11].

We find that parental education strongly impacts future skills, providing empirical support for a key mechanism for perpetuating inequality across generations. High education parents are effective in producing human capital in their children (as also shown in some of the papers cited in Heckman and Mosso (2014) and is assumed in Becker et al. (2018) and Lee and Seshadri (2019)) in addition to having more resources to afford college. The high productivity of high education parents means that all else equal, their children will be of higher skills. As we will show below, skills and years of education are highly complementary in the production of wages. The combination of these features of human capital production gives high education parents yet another incentive to send their children to higher education.

These results are robust to the inclusion of a number of other covariates into the equation, such as parental age and number of children in the household.

4.5.2 The Effect of Skills and Education on Wages

Our approach allows us to better understand whether differences in wages across individuals represents differences in skills versus shocks. In this section we give our estimates of the wage process shown in Section 4.3.3.3 for each gender and education group.

We allow the impact of skills on wages to depend on education to capture the possibility that returns to skills are greater for the more educated. Table 4.5.2 shows estimates of this impact (δ_5) for each gender and education group. These estimates

[11] More generally, Nicoletti and Tonei (2020) find that parents tend to compensate for low cognitive skills, whereas e.g. Aizer and Cunha (2012) find that parents reinforce children's skills.

show the log-point increase in wages associated with a one standard deviation increase in age-16 skill for each education and gender group. The extent of complementarity is similar to that estimated in Delaney (2019) and Daruich (2018), and is implicit in much of the literature on match quality (e.g., Arcidiacono (2005)) and college preparedness in educational choice (e.g., Blandin and Herrington (2018)).

Table 4.5.2: Log-point change in wages for a 1 SD increase in skill, by education level

	Male	Female
Low	0.084 (0.025)	0.078 (0.024)
Middle	0.167 (0.019)	0.103(0.018)
High	0.205 (0.027)	0.127(0.027)

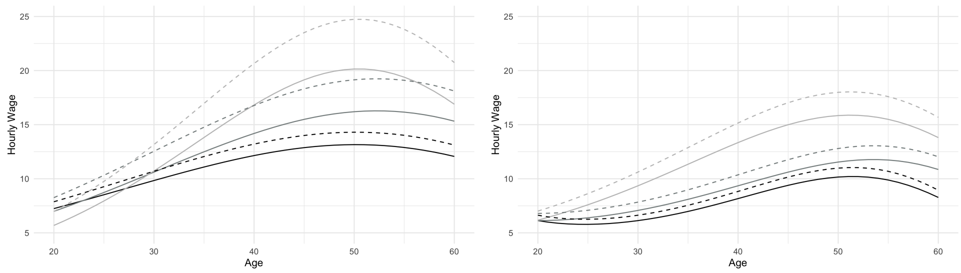
Notes: Cluster bootstrapped standard errors in parentheses (500 repetitions).

The table shows that, as one would expect, age-16 skill has a significant positive impact on wages conditional on education for all groups. Perhaps most interestingly, it shows evidence of complementarity between education and skill in the labor market, particularly for men. While low education men see only a 0.08 log-point increase in hourly wages for every additional standard deviation of skill, high education men (with some college education) see an average increase of 0.21 log-points in hourly wages for every additional standard deviation of skill. High educated women also receive greater returns to skill than low or middle educated women, although the gradient is more modest relative to that of men.

Figure 4.5.1 shows wage profiles by age, education and gender for full time workers with average skills and also skill levels that are one standard deviation above average. Men and those with high education have higher wages and faster wage growth.

As we show below this dynamic complementarity between skill and education has implications both for optimal time investments in children, and also for optimal educational decisions. Because of forward looking behavior, households who are more likely to invest in the education of their child have a stronger incentive to invest time in producing skills in their children. Furthermore, those with high skill have an incentive to select into high education.

Figure 4.5.1: Wages, by age, education and gender



(a) Male **(b)** Female

Note: Wages measured in 2014 pounds. Wage profiles have been corrected for selection and solid lines are evaluated at mean skill, dashed lines are evaluated at mean skills plus one standard deviation.

Turning to the variance of innovations to wages (σ_η^2), Table 4.5.3 shows that the estimated variance ranges from 0.0024 to 0.0048 implying that a one standard deviation of an innovation in the wage is 5-7% of wages, depending on the group. These estimates are similar to other papers in the literature (e.g. French (2005), Blundell et al. (2016)). Furthermore, we find evidence that the variance of wage innovations is increasing with education, implying that education is a risky investment.

Interestingly, we estimate the variance of the initial wage shock $\sigma_{\eta_5}^2$ to be small for all groups. While there is significant cross sectional variation in wages, even early in life, we estimate that most of that variation is explainable by our latent skill measure and measurement error in wages.

Table 4.5.3: Variance of innovations to wages, by education level

		Men	
	Low	Middle	High
σ_η^2	0.0024	0.0038	0.0045
	(0.0006)	(0.0006)	(0.001)
		Women	
	Low	Middle	High
σ_η^2	0.0020	0.0034	0.0048
	(0.0003)	(0.0004)	(0.0006)

Note: σ_η^2 is the variance of the annual innovation to wages. Bootstrapped standard errors in parentheses.

In our formulation, wage shocks have an autocovariance of one: wages are a random walk with drift. This implies skills have a permanent effect on wages. To test this restriction we also estimated versions of the wage process where we allowed the autocovariance to be less than one. However, we found little evidence against this restriction and thus use the more parsimonious formulation.

4.5.3 Marital Matching Probabilities

Table 4.5.4 shows the distribution of marriages, conditional on education, that we observe in the NCDS data. It also shows the share of men and women in each educational group. An important incentive for education is that it increases the probability of marrying another high education, high wage person. Table 4.5.4 shows evidence of assortative mating, as shown by the high share of all matches that are along the diagonal on the table: 12% of all marriages are between couples who are both low educated, 38% are between those who are middle educated and 4% among those who are highly educated.

Table 4.5.4: Marital matching probabilities, by education

	Low education male	Medium education male	High education male	Share of females in education group
Low education female	0.12	0.19	0.02	0.33
Medium education female	0.13	0.38	0.05	0.56
High education female	0.01	0.07	0.04	0.12
Share of males in education group	0.26	0.64	0.11	

Notes: The numbers represent cell proportions, which are the percentage of all marriages involving a particular match, i.e. these frequencies sum to one. NCDS data, marriages at age 23

4.5.4 Other Calibrations

Other parameters set outside the model are the interest rate r, parameters of the tax system τ, the household equivalence scale (n_t), time endowment T, and survival probabilities s_t.

The interest rate is set to 4.69%, following Jordà et al. (2019). To model taxes, we use IFS TAXBEN which is a microsimulation model which calculates both taxes and benefits of each family member as a function of their income and other detailed characteristics. We then calculate taxes and benefits (including state pensions) for our sample members at each point in their life, and estimate a three-parameter tax system which varies across three different phases of life: young without children (ages 23-25), working adult (ages 26-64), pension age (age 65, onwards). This three parameter tax system has the following functional form: $y_t = d_{0,t} + d_{1,t}(e_{m,t} + e_{f,t} + e_{f,t} + e'_t)^{d_{2,t}}$. We set the time endowment to $T = 16$ available hours per day $\times$ 7 days per week $\times$ 52 weeks per year=5,824 hours per year. We use the modified OECD equivalence scale and set $n_t = 1.4$ for couples with children. Survival probabilities are calculated using national life tables from the Office for National Statistics.

4.6 Second Step Results, Identification, and Model Fit

We now present the estimated structural parameters, how they are identified and the model's fit. Table 4.6.1 presents estimates from the structural model.

4.6.1 Utility Function Estimates and Identification

Table 4.6.1: Estimated structural parameters.

Parameter	Estimate
β: discount factor	0.985
	(0.0001)
ν_f: consumption weight, female	0.454
	(0.0002)
ν_m: consumption weight, male	0.433
	(0.0003)
γ: risk aversion	3.462
	(0.0078)
λ: altruism parameter	0.313
	(0.0010)
θ: time cost of investment	0.041
	(0.0002)
$\kappa_{1,1}$: latent investments per hour, ages 0-6	0.175
	(0.0006)
$\kappa_{1,2}$: latent investments per hour, ages 7-10	0.153
	(0.0010)
$\kappa_{1,3}$: latent investments per hour, ages 11-15	0.224
	(0.0009)
Coefficient of relative risk aversion, consumption*	2.092

Notes: Standard errors: in parentheses below estimated parameters. NA: parameters fixed for a given estimation.
* Average coefficient of relative risk aversion, consumption, averaged over men and women. Calculated as
$-(1/2)[(\nu_m(1-\gamma)-1)+(\nu_f(1-\gamma)-1)]$. β is an annual value.

The parameter γ is the coefficient of relative risk aversion (or the inverse of the intertemporal elasticity) for the consumption-leisure aggregate. It is the key parameter for understanding both the coefficient of relative risk aversion for consumption and for understanding the willingness to intertemporally substitute consumption and labor supply. The coefficient of relative risk aversion for consumption is 2.09 averaging over men and women,[12] which is similar to previous estimates that rely on

[12] We measure the individual's coefficient of relative risk aversion using the formula $-\frac{(\partial^2 u_t/\partial c_{g,t}^2)c_{g,t}}{(\partial u_t/\partial c_{g,t})} = -(\nu_g(1-\gamma)-1)$, and so the average is $-(1/2)[(\nu_m(1-\gamma)-1)+(\nu_f(1-\gamma)-1)]$. Note that this variable is measured holding labor supply fixed. The coefficient of relative risk aver-

different methodologies (see Browning et al. (1999) for reviews of the estimates).

Identification of the coefficient of relative risk aversion for consumption is similar to Cagetti (2003) and French (2005) who estimate models of buffer stock savings over the life cycle using asset data as we do. Within this framework, a small estimate of the coefficient of relative risk aversion means that individuals save little given their level of assets and their level of uncertainty. If they were more risk averse, they would save more in order to buffer themselves against the risk of bad income shocks in the future. We also obtain identification from labor supply since precautionary motives can explain high employment rates when young, despite the low wages of the young: more risk averse individuals work more hours when young in order to accumulate a buffer stock of assets. Furthermore, since γ is the inverse of the intertemporal elasticity of substitution for utility and thus is key for determining the intertemporal elasticity of labor supply.[13] Wage changes cause substitution from work both into leisure and into time spent with children.

Our estimate of the time discount factor β is equal to 0.985, and is also identified using our wealth data and our data on labor supply over the life cycle, both of which suggest households are relatively patient. First, wealth holdings at age 60 are relatively high given pension benefits and high consumption demands up to this age. Second, young individuals work many hours even though their wage, on average, is low. This is equivalent to stating that young people buy relatively little leisure, even though the price of leisure (their wage) is low. Between ages 35 and 60, people buy more leisure (i.e., work fewer hours) as they age even though their price of leisure (or wage) increases. Therefore, life cycle labor supply profiles provide evidence that individuals are patient. French (2005) also finds that $\beta(1+r) > 1$ when using life cycle labor supply data.

The parameters v_m and v_f are identified by the share of total non-childcare hours devoted to time worked in the market. To see this, note that the the after tax

---

sion for consumption is poorly defined when labor supply is flexible.

[13] Assuming certainty, linear budget sets, and interior conditions, the Frisch elasticity of leisure is $\frac{v_g(1-\gamma)-1}{\gamma}$ and the Frisch elasticity of labor supply is $-\frac{l_{g,t}}{hrs_{g,t}} \times \frac{v_g(1-\gamma)-1}{\gamma}$. However, an advantage of the dynamic programming approach is that it is not necessary to assume certainty, linear budget sets, or interior conditions.

wage is approximately linked to marginal rate of substitution between consumption and leisure as follows:

$$\begin{aligned} w_{g,t}(1-\tau'_{g,t}) &\leq -\frac{\partial u_t}{\partial hrs_{g,t}} / \frac{\partial u}{\partial c_g} \\ &\leq -\frac{1-\nu_{g,t}}{\nu_{g,t}} / \frac{c_{g,t}}{l_{g,t}} \end{aligned} \tag{4.16}$$

which holds with equality when work hours are positive, where $\tau'_{g,t}$ is individual g's marginal tax rate at time t.[14] Inserting the time endowment equation (4.1) into equation (4.16) and making the approximation $c_{g,t} \approx w_{g,t} hrs_{g,t}(1-\tau'_{g,t})$ yields

$$\nu_g \approx \frac{hrs_{g,t}}{T - ti_{g,t}}. \tag{4.17}$$

Thus ν_g is approximately equal to the share of non-childcare hours that is spent at work. We find that this share is somewhat less than 0.5, and thus our estimate of ν_g is modestly less than 0.5 for both men and women.

Our estimate of the weight that the altruistic parents place on the utility of both their children (2λ) is 0.63 which is the middle of the range of estimates reported in the literature. This is higher than estimates by Daruich who estimated it to be 0.48 and Lee & Seshadri who estimate it to be 0.32, and lower than Gayle, Golan, and Soytas whose estimate is 0.80 and Caucutt & Lochner whose estimate is 0.86. These papers model a parent with only one child, whereas in our framework a parent has two children. Thus we multiply by 2 the continuation values of the children.

The parameter λ is identified from two sources. First, households invest in the formal education of their children. The foregone household income from children going to school represents a direct loss of resources to the household. Second, households make cash transfers to their children. We find that cash transfers to children are modest. However, they are the most direct manifestation of altruism. To see this, note from equation (4.10) that in the phase when the child is in their young adult phases ($t = 9$, when the parent is 49 and the child is 23), parents have

[14] This relationship is not exact because of the part time penalty to work hours and the discreteness of the hours choice.

the opportunity to transfer resources, and the following optimality condition holds

$$\frac{\partial u_t}{\partial c_{g,t}} \geq \frac{2\lambda \partial \mathbb{E}_t V'_{t'}(\mathbf{X'}_{t'})}{\partial A'_{t'}} = \frac{2\lambda \mathbb{E}_t \partial u'_t}{\partial c'_{g,t'}}$$

and holds with equality if transfers are positive. The term on the right is the sum (over both children) of the childrens' expected marginal utility of consumption value of assets, which the parents can transfer to the children when the children are age 23. At the time of the transfer the children will be at a low earning time during their life cycles, and will soon have their own children and the time and money expenses of those children. This, and the fact that they are likely to be borrowing constrained, will mean they will have a higher marginal utility of consumption than their parents. In order to rationalize relatively modest transfers to children, λ must be less than 1. Nevertheless, the fact that these transfers are made is perhaps the strongest evidence that $\lambda > 0$ and households are altruistic. Furthermore, in Section 4.7.3 we show that the returns to education average 7.4% per year of education, which is well above the market interest rate of 4.7%. Recall that the returns to education accrue to the child, whereas the return to cash accrues to the parents. The fact that many parents invest little in their childrens' education, but some invest a lot, again provides evidence that λ is less than 1 but is greater than 0.

The parameter θ is identified by the relative productivity of time investments with children. Recall that $1-\theta$ is the share of time with the child that represents leisure to the parent: if $\theta = 1$ then time with children has the same utility cost as work, whereas if $\theta = 0$ then time with children has the same utility benefit as leisure. Thus, if $\theta = 1$ optimal behavior implies that the economic benefit of an additional hour of investment in the child (i.e., the increase in the expected present value of the childrens' lifetime income) will (approximately) equal the economic benefit of an additional hour of work (the parent's wage). Conversely, if $\theta = 0$ then parents will spend time with their children even if it does not affect the childrens' future wages. Appendix 4.K provides a more formal discussion of identification of θ. Because we find that the impact of parents' time on childrens' skill is positive but modest, we estimate θ to be 0.04, meaning that 96% of the time that parents spend with their children is leisure for them. There is little evidence on the magnitude of this parameter. The closest study to ours is Daruich (2018) who uses a specification slightly different than ours, but also finds that time spent with children is largely leisure.

The $\kappa_{1,t'}$ parameters govern the relationship between units of latent investments and units of time. Identification of these parameters comes from that we observe gradients in time investments (from the UKTUS data) by parental education and that we also observe corresponding skill gradients from the NCDS. That is, we observe more educated parents spending more time with their children, as well as a gradient in final skill by parental education. This, together with the production function in first stage pins down how a unit of time maps into a unit of investments. Appendix 4.K contains a more formal derivation.

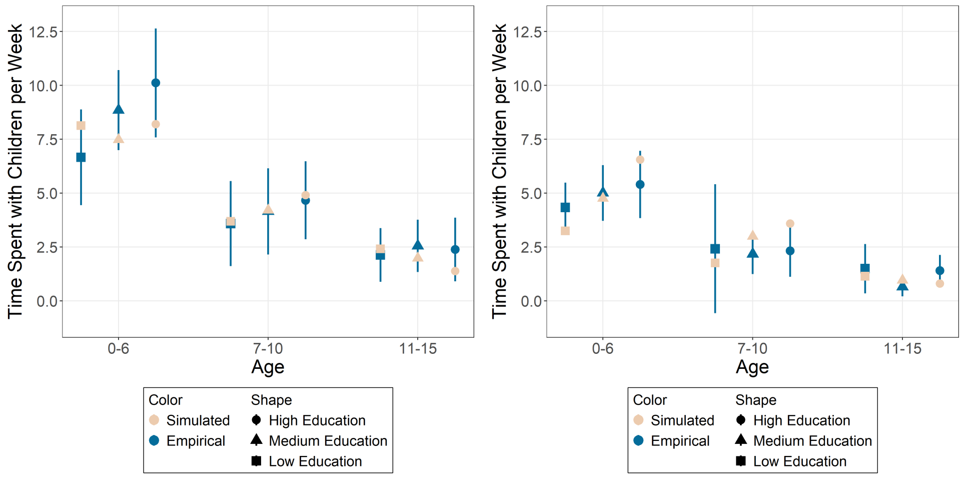
4.6.2 Model Fit

In this section we focus on the moments that are critical for understanding intergenerational altruism: transfers of time, educational investments, and money.

Figure 4.6.1 shows transfers of time from mothers and fathers in the left and right panels, respectively. The model fits three key patterns in the data well. First, time investments decline with age. Second, mothers invest more in their children than fathers. This higher rate of investment reflects the lower wage, and thus the lower opportunity cost of time for women. Third, high education parents invest more time in their children than low education parents. This pattern is driven by a combination of the higher education levels of their children and the complementarity between skills and education in wages that we have estimated.

This higher level of time investments of educated parents, in combination with their greater productivity of these investments, leads to higher skills levels of their children as can be seen in panel (a) of Figure 4.6.2. Our model captures well how higher time investments of the educated lead to higher skill levels of their children. Children of low education fathers have skill levels that are 0.14 standard deviations below average, whereas children born to high education fathers have skill levels that are 0.80 standard deviations above average. Our model matches these patterns well,

Figure 4.6.1: Model fit: parental time with children



(a) Mother's Time with Kids **(b)** Father's Time with Kids

Notes: Measures of educational time investments. Source: UKTUS. See Appendix C.3 for details.

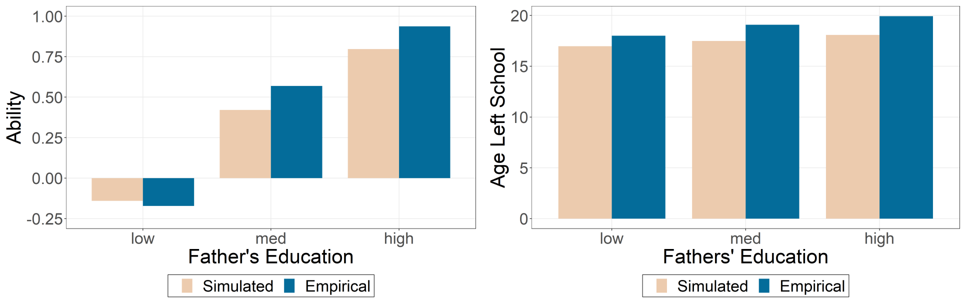
although we slightly overstate the gradient.

Next, panel (b) of Figure 4.6.2 shows children's education, by father's education. Although the model slightly underpredicts educational attainment of children, it captures the gradient of children's education by parent's education. The difference between the average age left school of the children with high educated fathers and those with low educated fathers that our model predicts is 1.12 years, close to the difference of 1.92 years found in the data.

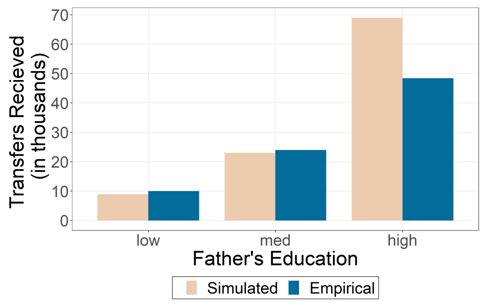
Table 4.6.2 shows that we match well the mean level of financial transfers received and the median level of assets at age 60. These financial transfers include inter-vivos transfers when younger and bequests received when older. These amounts are discounted to age 23: when undiscounted, the amounts are considerably larger. In the data, as in the model, median transfers are 0. Thus we match mean transfers. Figure 4.6.2c shows that in addition to matching well mean transfers to children, we also replicate the untargeted gradient of transfers by father's education. We match well the transfers to children of low and medium educated fathers but over-predict transfers to children of high educated fathers.

Finally, our model can reproduce key labor supply moments of men and women with different education levels as shown in Appendix 4.J. Both female labor force participation and fulltime work conditional on employment are slightly over-predicted in the model. However, the model does well in generating a dip in female participation and fulltime work between ages 33 and 48 (when children are in the household). Moreover, as in the data, the model predicts higher participation rates for more educated women at older ages. For men, the model does well in generating a level of labor supply that is consistent with the data both on the intensive and the extensive margin.

Figure 4.6.2: Model fit: education and skill



(a) Skill level by father's education **(b)** Education choices by father's education



(c) Transfers recieved by father's education

Notes: Empirical education and skills from NCDS data.

Table 4.6.2: Model fit: transfers and assets

	Empirical	Simulated
Mean transfers	£12,900	£12,800
Median Assets	£306,400	£291,700

Notes: Values in 2014 GBP. Mean transfers and median assets calculated using ELSA data. Transfers include inter-vivos transfers and bequests and are discounted to age 23 at the real rate of return. See Appendix 4.C for more details.

4.6.3 Intergenerational Persistence

Although we do not target them directly, our model fits well the intergenerational persistence in economic outcomes that are commonly estimated in other studies. This includes the intergenerational correlation of education and the intergenerational elasticity (IGE) of lifetime earnings and consumption. We estimate the following regression on our simulated data: $y' = a_0 + a_1 y + u$ where y' denotes the child's outcome (e.g., the number of years of schooling or the log of childrens' household earnings) and y the parents' corresponding outcome (e.g. parents' years of schooling or lifetime household earnings).

The model predicted correlation of childrens' and parent's education is 0.23, with a correlation with father's education of 0.19 and mother's of 0.18, which is similar to the estimates presented in Hertz et al. (2007) who report 0.31 for Great Britain. The model predicted intergenerational elasticity of lifetime household earnings is 0.24, which is similar to the estimated values reported in Belfield et al. (2017) and Bolt et al. (2021). A more complete measure of lifetime resources is consumption. The model predicted intergenerational elasticity of consumption is 0.51 which is in line with the findings in Gallipoli et al. (2020) who find an average consumption IGE in the PSID of 0.46, a value that is substantially above their estimates of the elasticity of earnings. Wealth transfers across generations cause consumption to be more persistent than earnings. That our model reproduces key patterns of intergenerational persistence gives us additional confidence in its use for evaluating the drivers of this persistence and policy counterfactuals.

Table 4.6.3: Intergenerational Persistence

Outcome	Model-Implied	Literature
Intergenerational Correlation, Education	0.23	Hertz (2007) $\approx$ 0.3
Intergenerational Elasticity, Earnings	0.24	Dearden et al. (2007), Bolt et al. (2021) $\approx$0.3
Intergenerational Elasticity, Consumption	0.51	Gallipoli et al. (2022) $\approx$0.5 (in the US)

Notes: Intergenerational correlations and elasticities calculated from model simulated data. Earnings and consumption calculated as average over ages 23-65.

4.7 Results

4.7.1 How is Income Risk Resolved over the Life Cycle?

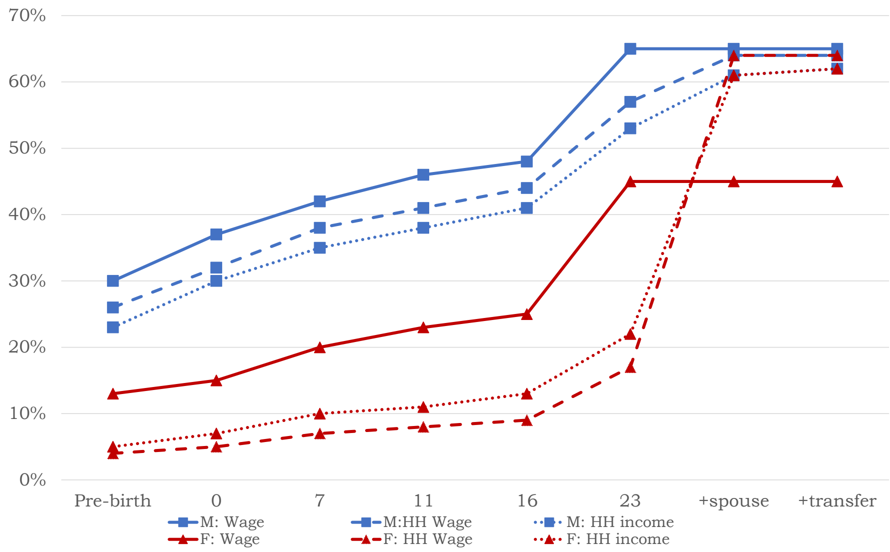
How much of the cross-sectional variance in lifetime income can we predict using information known at different ages? Already before birth, information on the parents can help us predict an individual's lifetime income through predicted future investments that parents will make as well as through the productivity of those investments. As the child is born and grows older, decisions are made and shocks are realized, thus increasing the extent to which lifetime income can be predicted.

We take as given the age-23 joint distribution of the state variables of the NCDS sample members, draw histories of shocks and calculate optimal decisions for both NCDS sample members and their children. This allows us to simulate lifetime outcomes for two generations. Next, we calculate the share of the variance in the childrens' lifetime income that can be predicted by the following variables that are known at each age: parental assets, wages, and education; the child's skill level, gender, education, and wages; and the education and wages of the child's spouse. This approach allows us to decompose the relative importance of (predictable) circumstances and choices; the remainder being explained by shocks. This builds upon the approach in Huggett et al. (2011) who calculate the share of lifetime income known to the individual at age 23 and Lee and Seshadri (2019) who calculate it before birth, at birth, and age 24. By showing the amount of lifetime income variability known at multiple ages, we illustrate how this uncertainty is resolved with age and how it is resolved after marriage.

Our decomposition makes use of the law of total variance: a random variable can be written as the sum of its conditional mean plus the deviation from its conditional mean. As these two components are orthogonal, the total variance equals the sum of the variance in the conditional mean plus the variance around the conditional mean. We then divide the variance in the conditional mean of lifetime income by the total variance of lifetime income. We illustrate how uncertainty about three measures of household resources is resolved over the life cycle in Figure 4.7.1.

The first measure of resources is individual lifetime wages (represented by

Figure 4.7.1: Resolution of uncertainty over the life cycle



Notes: "M: Wages" denotes male individual wages, "M: HH Wages" denotes male household wages, "M: HH income" denotes male household income; "F" for females, analogously. "+spouse" denotes age 23 after being matched into a couple. "+transfers" denotes age 23 after transfers from parents received. This graph shows the share of variance explained by characteristics of both parent and child known at a given age of the child. Wages and income are discounted pre-tax values. Wages are measured as the potential earnings if working full time. Household income is the sum of earnings plus parental transfers received by both the individual and their spouse.

solid lines in the graph) which are calculated as the discounted pre-tax earnings between ages 23 and 64 that an individual would earn if they worked full time in every period. These numbers reflect the difference in *potential* rather than realized earnings, which depend on labor supply choices. The figure shows that 30% and 13% of lifetime wages are known for males and females, respectively, even before they are born. These shares are explained by parents' education (which affects initial skill and the productivity of parental investments) and also household financial resources (which affects the quantity of investments the child receives). As the child ages, new information is realized, both about their own skill and their parents' financial resources. Immediately after birth, initial skill and parental wage shocks are revealed, causing the shares explained to rise to 37% and 15% for males and females, respectively. By the time the children are aged 23, educational choices have been made and their initial wage draw has been realized, causing the shares to rise to 65% and 45%. Thus close to half of lifetime wage variability is realized by age

23. The higher share of lifetime wages that is explainable for men reflects the higher return to skill for men, especially those who obtain a high level of education. From age 23 onwards, the share of wage variance explained does not change anymore, as wages are not affected by spousal characteristics or transfers from parents.

Our second measure of resources is household wages (represented by dashed lines in the graph), the sum of lifetime wages for both spouses. At age 23 individuals marry, resolving uncertainty about age 23-characteristics of the spouse (their wage, education, and parental transfer). Before matching occurs, household wages (the sum of lifetime wages for both individual and spouse) are less explainable than individual wages since matching is not perfectly assortative. Just before marriage, the share of lifetime household wages explained is 57% and 17% for males and females, respectively. The share explained is much lower for women than for men because wages are both lower and less variable for women than men. Marriage explains much of the remaining variability in lifetime household wages, especially for women: the share explained jumps to 64% for both men and women after marriage. Household wages are less explainable than individual wages before marriage, but are more explainable afterwards. That is, before marriage, the characteristics of one's future spouse is an important risk; after marriage one's spouse becomes an important form of insurance, at least for women.

Our third measure of resources is household lifetime income (represented by dotted lines in the graph), which is the sum of realized earnings and parental transfers received by both the individual and their spouse. Household income is about as explainable as the household wage. Transfers explain little of lifetime income, both because transfers are small relative to lifetime earnings and also because the transfers made are highly explainable given all the other variables known.

4.7.2 What Explains Income Inequality?

The previous section shows how uncertainty is resolved over the life cycle as shocks are realized and choices are made. However, it does not show the relative importance of parental choices which lead to intergenerational persistence in outcomes. This section shows the relative roles of different types of parental transfers in contributing to variability in lifetime income. To address the importance of choices relative to other variables, we perform counterfactual experiments where we hold all choices, of both the parents' and childrens' cohorts, constant except that we equalize in turn parental time investments, education, and money transfers. We evaluate individual lifetime wages, household lifetime wages, and household lifetime income for the childrens' cohort and report the proportionate fall in variance that these equalizations would induce.

Table 4.7.1: Fraction of outcome variance for males explained by time investments, education, and skill

Equalize:	Education	Time Investments	Transfers
Individual's wage	8%	13%	-
Wage of household	16%	16%	-
Household's income	14%	11%	4%

Notes: Percentage reduction in variance of variables when equalizing a channel to its model median for education and means otherwise. Wages, earnings, and income are discounted pre-tax values received between ages 23 and 64. Wages here are measured as the potential earnings if working full time. Household income is the sum of earnings plus parental transfers received by both the individual and their spouse. "-" means no change relative to the baseline case.

Table 4.7.1 shows that equalizing time investments received would reduce the variance of individual lifetime wages by 13%. Equalizing educational investments would have a smaller impact on individual lifetime wages, reducing this variance by 8%. Equalizing education reduces the variance of *household* wages more than *individual* wages. This is because equalizing education not only removes the variation in household wages coming from the education of the spouses, but also removes the additional variation across households due to assortative matching. Equalizing time investments also reduces the variance of household wages more than individual wages since in the model those who receive high investments also receive high education, and so equalizing investments also reduces assortative mating. This highlights the interplay between education, the family, and lifetime risks: because highly-educated individuals are more likely to marry other highly-educated individuals, inequality in education contributes more to variability in household wages than it does to individual wages. Assortative matching amplifies inequality.

The final row of Table 4.7.1 considers household income, which includes transfers as well as labor earnings. Equalizing transfers across households would reduce the variability in income less than equalizing time investments or education. If all households received mean transfers, the variance of household income would fall by 4%; equalizing education and time investments would reduce this variance by more. Altonji and Villanueva (2007) and Black et al. (2022) also emphasize the modest role that transfers play in lifetime inequality.

4.7.3 The Returns to Education

In previous sections, we established that 1) education is very persistent across generations and 2) education is the strongest driver of the variance in household lifetime income. Why do different parents make different education choices, and what difference would an education subsidy make? To answer these questions, we first study in detail the returns to education. We then evaluate a counterfactual education policy in the next section.

In our model the return to education is heterogeneous across the population since wages depend on skills, education, and their interaction. We measure the return to education for different groups by exogenously changing the education levels of agents in the model, then calculating the resulting annualized percent change in lifetime wages from age 23-65. As in Section 4.7.1, we use the model and the age-23 joint distribution of initial conditions to simulate the behavior and resulting lifetime wages of two generations. In order to measure the childrens' return to education, we simulate their childrens' lifetime wages twice: first assuming that they receive low education and second assuming they receive high education. We then calculate the annualized percent change in lifetime wages. In column (1) of Table 4.7.2 we assume that the education level is unanticipated: the household's decision rules are thus calculated assuming that the household (erroneously) believes they can choose the child's education level. Thus, in this first experiment, changing education holds constant the decision to invest in child's skills. If everyone received low education, average discounted lifetime wages (i.e., their pre-tax earning if they worked full time) would be £382,000. Conversely, if everyone selected high education, wages would be £563,000, a difference of 47.3%. Given the five year difference in schooling between low and high educated individuals, annualizing this translates into a 8.1% increase in lifetime wages per year of education: education is a lucrative investment. To put the impact of education transfers in context: the increase in lifetime earnings from moving from low to high education (£181,000) is significantly higher than the average cash transfer to children reported in Table 4.6.2 (£13,000).

Section 4.5.2 presented evidence of dynamic complementarity: the returns to education are higher for those with high skill levels. This has two implications for economic behavior that are evident in Table 4.7.2.

First, it provides an incentive for high skill individuals to self-select into education. To measure the extent of this self-selection, we calculate the return to education for two groups: those who, in the baseline case, select high (college) education, and for those who in the baseline case select low (compulsory) education. The bottom panel of Table 4.7.2 shows that the return to education is higher for those who would have selected high education (8.5%, which is the treatment effect on the treated) than the return for those who select low education (7.9%, which is the treatment effect on the untreated). Complementarity between skills and education, in combination with self-selection in the model, explains this result.

Table 4.7.2: Returns to education.

	Unanticipated (1)	Anticipated (2)
	A: Full Sample	
Lifetime earnings, if low education	382,000	362,000
Lifetime earnings, if high education	563,000	595,000
Return	47.3	64.4
Annualized return	8.1	10.5
	B: By Baseline Education Choice	
Annualized return		
... among those who selected high education	8.5	11.9
... among those who selected low education	7.9	9.6

Notes: The return R is calculated as percent change in discounted pre-tax lifetime wage earnings between having high and low education. Annualized returns equal $(1+R)^{1/5}-1$. Anticipated means that the household is certain that it will be forced to either have high or low education starting from birth. Unanticipated means that the household believes it will make the optimal educational choice at age 16 and makes time investments given the belief of optimal educational choice, then is forced to either have high or low education.

Second, if parents are forward-looking their investment decisions will depend on the probability their children continue education. As with column (1), column (2) solves the model both for the case where children receive low education and for the case where children receive high education, and reports the resulting return to education. However, in column (2) households are certain of their childrens' future education level. This means households can change their time investment and other decisions in response to the education change. When households are certain their children will receive high education, they respond by increasing time investments, since the return to these investments is now higher. Column (2) shows that when allowing for these anticipation effects, the return to education rises from 8.1% to 10.5% once households anticipate this higher level of education. This highlights the importance of pre-announced policies that can deliver higher returns than policies that are not pre-announced. Pre-announcing the policy allows parents to adjust their time investments accordingly.

4.7.4 Evaluating an Education Subsidy

The financing of university education is at the center of current policy debate in multiple countries. Key issues discussed are whether college subsidies mostly benefit high income households and whether they are a good investment for the government.

To address these issues, we evaluate the impact of introducing £10,000 annual grants to those attending university. This translates into a £30,000 subsidy over the course of a three-year university degree. We evaluate the impact of the subsidy on educational attainment, lifetime wages, and the intergenerational persistence of lifetime outcomes. Finally, we calculate the tax revenue this reform generates and thus government surplus. In evaluating the impact of the subsidy, as in the previous section, we make two different assumptions about whether the subsidy was anticipated.

We first assume that the subsidy is unanticipated: the household's decision rules are thus calculated assuming that the household (erroneously) believes that there exists no subsidy for university attendance. Thus, in this first experiment, the subsidy holds constant parental investment decisions. In the second experiment we assume the policy is known from the start of the parent generation's working life. Thus, they can adjust time and education investments in their children.

In both experiments we account for the equilibrium effect of the policy on the marriage market. In particular, the subsidy impacts not only the education of an individual, but also the distribution of educational levels within the economy and therefore the distribution of potential spouses. Thus, the marital matching probabilities change. Although we do not impose an equilibrium matching model, our approach respects marriage market clearing by exploiting historical variation in marriage matching probabilities conditional on education as documented in Appendix 4.M. We then calculate the impact of the reform on children's outcomes and intergenerational persistence.

Table 4.7.3 shows that the reform significantly impacts educational decisions. If the education subsidy were unanticipated, the fraction of children who attend university rises from the baseline value of 0.24 to 0.34. This additional education raises average economy-wide lifetime individual wages if in full time work from £454,000 to £468,300, an increase of £14,300. The gain in lifetime household earnings is more modest, however, because of reduced labor supply. Most households who receive the subsidy would have attended university even without the benefit. For these households, the subsidy is merely an income transfer, creating a wealth effect that reduces labor supply. As a result, tax revenue from the reform rises, but only by £2,400. Because the discounted cost of the reform, averaged over the cohort we consider, is £8,800 (the discounted cost of the subsidy is £25,900, and 34% of all individuals go to university), the net present value of the reform on government revenue is -£6,400.

Two groups of households benefit from the reform. The first group comprises those households who would have sent their children to college, even in the absence of the subsidy. These tend to be high income households with high skill children. For these households, the grant is merely a lump sum transfer. The second group comprises those households who send their children only if they receive the subsidy, i.e. those with relatively high skill children who have a relatively high return to going to university. These tend to be households with parents who have relatively high education and high earnings themselves, and thus, the reform, if anything *increases* the intergenerational persistence of outcomes. For example, the intergenerational correlation in education increases from 0.23 to 0.27. This increases the intergenerational elasticity of earnings and consumption also.

Table 4.7.3 shows that if the education subsidy is anticipated, its impact is significantly larger than if unanticipated. The fraction of children who attend university rises from the baseline value of 0.24 to 0.58, significantly larger than when the subsidy is unanticipated. These results are consistent with Caucutt and Lochner (2020), who also find that the impact of education subsidies on education are more than twice as large as large if they are anticipated.[15] The larger gain in schooling

[15] Dynarski (2003)'s estimates imply a 1.3 year increase in completed education from an equally size subsidy for a reform that was likely partly, but not fully anticipated. Her estimated impacts are between the anticipated and unanticipated effects predicted by our model.

Table 4.7.3: Impact of Education Subsidy

Outcome	Baseline	Education subsidy: Unanticipated	Anticipated
Years of education	17.44	17.83	18.89
Share with college education	0.24	0.34	0.58
Discounted lifetime individual wages	454,000	468,300	525,200
Discounted household individual earnings	807,200	811,100	879,500
Mean skill	0.42	0.42	0.59
Mean skill (low ed father)	-0.09	-0.09	0.19
Mean skill (med ed father)	0.50	0.50	0.67
Mean skill (high ed father)	0.97	0.97	0.96
Average discounted cost	-	£8,800	£12,800
Average discounted additional revenue	-	£2,400	£33,500
Average surplus	-	-£6,300	£20,700
Intergenerational correlation, education	0.23	0.27	0.18
Intergenerational elasticity, earnings	0.24	0.35	0.18
Intergenerational elasticity, consumption	0.51	0.69	0.42

Notes: Discounted values are discounted to when children of NCDS sample members are age 18. Average discounted cost is the discounted cost of the college subsidy, averaged across all members of the cohort. Average discounted additional revenue is the discounted additional taxes paid over the life cycle, averaged across all members of the cohort. Discounted individual earnings equals $\frac{\text{discounted household earnings}}{2}$. Average surplus is the difference between average discounted cost and revenue.

when the subsidy is anticipated is a direct result of the complementarity between skill and education in the wage equation. When the education subsidy is anticipated, parents are more likely to send their child to university, which raises returns to early life investments. Thus, the change in incentives for college-going also changes incentives for early life investments.

These induced increases in parental investments raise age 16 skill by 0.17 standard deviations. The jumps in skill are especially large among households with less educated fathers. Because children born to high education fathers planned to attend college in the absence of the subsidy, education decisions are largely unaffected for this group. In contrast, the subsidy has much larger impacts on college going and thus early life investments to children born to lower education fathers. These parental investments raise age 16 skill by 0.28 standard deviations for children born to low education fathers. As a result, anticipated subsidies reduce the intergenerational persistence of education, earnings, and consumption.

This subsidy raises lifetime wages if in full time work from £454,000 to

£525,200, an increase of £71,200. Impacts on household earnings are similarly large. The impact on earnings is much larger for the anticipated than unanticipated case, for three reasons. First, the impact on educational attainment is larger. Second, the anticipated reform increases age 16 skill. Third, in the anticipated case, most households who receive the subsidy would not have sent their children to college in the absence of the subsidy. Recall that for households who would have sent their children to college in the absence of the subsidy, lifetime wages are unchanged and thus the subsidy is a lump sum transfer, reducing labor supply. But for households who choose to send their children to college as a result of the anticipated subsidy, resulting higher lifetime wages incentivizes longer working hours.

As a result, tax revenue from the reform rises by £33,500. Because the cost of the reform is £12,800, the average surplus from the reform is £20,700. Put differently, unlike in the unanticipated case, the subsidy more than pays for itself.[16]

4.8 Conclusion

This paper estimates a dynastic model of parental altruism where parents can invest in their children through time, educational expenditures, and transfers of cash. We estimate human capital production functions and the effect of skill on wages using data from a cohort of children born in 1958, thus presenting the results of a model which is the first estimated using data which links early life investments received by individuals to their earnings over their whole of the life cycle. In addition, we model the investment decisions of two parents, allowing us to consider the role of assortative mating on the intergenerational persistence of outcomes. Our model is able to replicate realistic patterns of intergenerational persistence in wages, earnings, wealth and consumption.

We find that approximately one fifth of the variance of lifetime household income can already be explained by characteristics of the parents before individuals are born. The predictability of earnings pre-birth is due to both the direct effects of

[16]Note that in this calculation we are comparing costs of the subsidy itself, and not the full costs and benefits of government provided university education. University tuition was free for members of this cohort, but imposed a cost to the government to pay for staff pay and other costs of universities. See Fu et al. (2019) for an assessment of these additional costs.

parental characteristics on individual's skills, and also due to increased investments of higher educated parents. The share explained rises to over 60% after marriage. Prior to marriage, the characteristics of one's future spouse is an important risk; after marriage one's spouse becomes an important form of insurance.

We find that an important mechanism generating intergenerational persistence is the dynamic complementarity between time and educational investments – the returns to education are higher for high skill individuals. Borrowing constraints prevent low-income families from investing in education, and this dynamic complementarity reduces the incentive for low-income families to invest in children's skill earlier in life. This has consequences for the design of policies that aim to reduce intergenerational persistence, such as education subsidies. We find that if such policies are announced early, parents increase early life investments, leading to a higher return to the policy. In contrast, if such policies are introduced unexpectedly, they can even *increase* the intergenerational persistence in outcomes.

Appendix

4.A Parameter definitions

Table 4.A.1 summarises the parameters that enter the model and which are introduced in the body of the paper (excluding the appendices).

Table 4.A.1: Parameter definitions

	Preference Parameters		**State variables**
β	Discount factor, annual	$g \in \{m, f\}$	Gender
β_{t+1}	Discount factor, between model periods	t	Model period
ν_g	Consumption weight in utility function	ed	Educational Attainment
λ	Intergenerational altruism parameter	a_t	Wealth
$1-\theta$	Share of investment time perceived as leisure	$w_{g,t}$	Wage
$\kappa_{0,t}, \kappa_{1,t}$	Time to investment conversion parameters		
	Labor market		**Household choices**
y_t	Household income	$c_{g,t}$	Consumption
$\tau(.)$	Net-of-tax income function	$l_{g,t}$	Leisure
$e_{g,t}$	Earnings	$hrs_{g,t}$	Work hours
η_t	Wage innovation	$ti_{g,t}$	Time investment in children
σ^2_η	Variance of wage innovation	x_t	Cash transfer ($t = 10$)
δ_j	Wage profile parameters		
	Skill		**Utility function and arguments**
$h'_{t'}$	Child's skill at t'	$u()$	Single period utility function
γ_j	Skill production parameters	$V_t(\mathbf{X_t})$	Value function
u_h	Stochastic skill component	$\mathbf{X_t}$	Vector of all state variables
		n_t	Number of equiv. adults in hh
	Assets	T	Time endowment
$(1+r_t)$	Gross interest rate, between model periods	$\mathbf{d_t}$	Vector of decision variables
r	Annual interest rate		
			Other
	Measurement Systems	τ	Length (years) of period t
ω	Vector of child skill and time investment	$Q_g()$	Marriage probability function
		s_{t+1}	Survival rate across period t

4.B Time Periods, States, Choices and Uncertainty

Table 4.B.1 lists all model time periods, parents' and chilrens' age in those time periods, the state variables, choice variables, and sources of uncertainty during those time periods.

Table 4.B.1: Model time periods, and states, choices and sources of uncertainty during those time periods

Time Periods																				
Model period	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Parent generation's age	0	7	11	16	23	26	33	37	42	49	55	60	65	70	75	80	85	90	95	100
Child generation's age						0	7	11	16	23										
Parent generation's datasets																				
NCDS					x		x		x	x	x									
Time use survey						x	x	x												
ELSA										x										
Child generation's datasets																				
NCDS						x	x	x												
Parent generation's states																				
Assets					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Wage of male and female					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Education of male and female					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Children's gender						x	x	x	x	x										
Children's skill						x	x	x	x	x										
Children's education										x										
Parent generation's choices																				
Work hours of male and female					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Time spent with children, male and female						x	x	x												
Consumption, male and female					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Cash transfer to children										x										
Education of children									x											
Parent generation's uncertainty																				
Wage shock of male and female					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Initial skill of children						x														
Skill shock to children							x	x	x											
Children's partner										x										
Children's initial wage										x										
Mortality										x	x	x	x	x	x	x	x	x	x	x

Notes: Between periods 1 and 4 the parent generation makes no choices, and in this sense has no state variables or uncertainty.

4.C Data

We use data from the NCDS, ELSA, and UKTUS in our analysis, and use sample selection rules which are consistent across the three data sets. The sample selection rules are described in more detail below.

4.C.1 NCDS

Our main data set is the National Child Development Survey (NCDS) which started with 18,558 individuals born in one week in March 1958. We use the NCDS Data in three different ways: First, for estimating the skill production functions. Second, for estimating the income process. And third, to derive moment conditions on marital matching, education shares, employment rates, the fraction of full-time work. We explain the samples used for these three purposes in more detail below.

Production function estimation: For the production function estimation, we require individuals to have a full set of observations on all skill measures, investment measures between the ages of 0-16, parental education, and parental income (see table 4.2.1 for a full list of measures). This reduces the original sample of 48,644 observations to 11,596 observations across the four waves considered.

Income process: For the estimation of the income process, we consider the waves collected at ages 23, 33, 42, 50, and 55, leaving out age 46 due to low-quality data. This leads to a total of 54,352 observations in adulthood. Of these, we drop all self-employed people (5,932 excluded), those who are unmarried after age 23 (7,602 excluded), those for who we only have one wage observation (9,909 excluded) leaving us with 30,909 observations. We trim wages at the top and bottom 1% for each sex and education group.

Moments: For the moments, we exclude all self-employed people (5,932 excluded), and those who are unmarried after age 23 (7,602 excluded), leaving us with a total number of observations of 40,818.

4.C.2 ELSA

We use the ELSA data both for asset data at age 60 which we use in our moment conditions and also for the gift and inheritance data which we use in our moment condition. ELSA is a biannual survey of those 50 and older, starting in 2002. We use data up through 2016.

Although NCDS sample members are asked about assets at age 50, these data are considered to be of low quality because the data omit housing wealth; thus we use ELSA instead. For our wealth measure, we use the sum of housing wealth including second homes, savings, investments including stocks and bonds, trusts, business wealth, and physical wealth such as land, after financial debt and mortgage debt has been subtracted.

For the asset moment condition at age 60 we begin with 2746 respondents who are age 60 at the time of the survey. We drop members of cohorts not born before 1950 (which excludes 1604 observations), unmarried people (which excludes 239 observations), and the self-employed (which excludes 132 observations). Finally, we have 23 individuals who live in the same household as another ELSA member. In order to not double count these households, we exclude one observation from these multi-respondent households, resulting in 748 individuals remaining.

ELSA has high quality data on gifts and inheritances in wave 6 (collected in 2012-2013). In this wave respondents were asked to recall receipt of inheritances and substantial gifts (defined as those worth over £1,000 at 2013 prices) over their entire lifetimes. Respondents are asked age of receipt and value for three largest gifts and three largest inheritances.[17] From our original sample of 10,601 in 2012, we drop members of cohorts not born between 1950-1959 (which excludes 7,223 observations), singles (921 excluded), and self-employed (328 excluded), resulting in 2,129 individuals remaining. Of these, we only keep individuals for whom both parents had died by the time of the survey (1107 individuals) and for who we have information on the father's education resulting in a final sample of 984 individuals.

Table 4.C.1 compares education shares and median net weekly earnings in both NCDS and ELSA. The ELSA sample has modestly higher education and lower earnings, but overall the samples match quite well.

[17]Only 3.6% of all individuals have three or more large inheritances or bequests (Crawford 2014), so the restriction is unlikely to signficantly affect our results.

Table 4.C.1: Sample comparison: NCDS and ELSA

	Education shares			
	Male		Female	
	NCDS	ELSA	NCDS	ELSA
Low	16%	20%	22%	26%
Medium	49%	38%	49%	40%
High	35%	43%	29%	34%
	Median net weekly earnings in £			
	Male		Female	
	NCDS	ELSA	NCDS	ELSA
Low	399	315	223	171
Medium	479	383	266	221
High	665	519	399	358

Notes: In NCDS, low education includes no educational qualification or CSE2-5, Medium education includes O-level or A-level, High education includes higher qualifications or a degree. In ELSA, low education includes no educational qualification or CSE, Medium education includes O-level or A-level, High education includes higher qualifications below a degree or a degree. Earnings are median net weekly earnings in £2013.

4.C.3 UKTUS

Using the measures of parental investments in the NCDS we can construct a latent time investment index. However, the NCDS does not directly measure hours of investment time. For measuring hours of investment time we use UKTUS data from 2000-2001. Respondents use a time diary to record activities of their day in 144 x 10-minute time slots for one weekday and one weekend day. In each of these slots the respondent records their main ("main activity for each ten minute slot") and secondary activities ("most important activity you were doing at the same time"), as well as who it was carried out with. We have diaries for both parents and the children, but use only the parent diaries.

We construct our measure of time spent with children by summing up across both parents the ten minute time slots during which an investment activity with a child takes place either as a main or a secondary activity. Although we know the number of children and the age of each child within the household, we do not know the precise age of the child that received the investment; we assume this to be the youngest child. We include all of the following activities as time spent with the child when constructing the investment measure: teaching the child, reading/playing/talking with child, travel escorting to/from education.

Our original sample includes 11,053 diary entries. We keep only married individuals with a child $\leq$ 15 yrs (which excludes 6,694 observations), drop households with more than 2 adults (797 excluded), keep those for whom we have diary information on both parents for both a weekend day and a weekday (506 excluded), and keep only 2 child families (1,660 excluded), leaving us with 1,396 remaining observations: (349 households with 4 entries (weekend, weekday for mum, dad)).

4.D Estimation of the Skill Production, Parental Investment, and Wage Functions

4.D.1 Production Function

The production function for skills that we estimate is as specified in equation (4.11) in the main text:

$$h'_{t'+1} = \alpha_{1,t'} h'_{t'} + \alpha_{2,t'} inv_{t'} + \alpha_{3,t'} inv_{t'} \cdot h_{t'} + \alpha_{4,t'} ed^m + \alpha_{5,t'} ed^f + u'_{h,t'} \qquad (4.18)$$

where $u'_{h,t'}$ is independent of all other right hand side variables.

4.D.2 Measurement

We do not observe children's skills (h'), or parental investments (inv) directly. However we observe $j = \{1, ..., J_{\omega,t}\}$ error-ridden measurements of each. These measurements have arbitrary scale and location. That is for each $\omega \in \{h', inv\}$ we observe:

$$Z_{\omega,t,j} = \mu_{\omega,t,j} + \lambda_{\omega,t,j} \omega_t + \varepsilon_{\omega,t,j} \qquad (4.19)$$

All other variables are assumed to be measured without error.

4.D.3 Assumptions on Measurement Errors and Shocks

Measurement errors are assumed to be independent across measures and across time. Measurement errors are also assumed to be independent of the latent variables (skill and investment), the structural shocks, and parental education $(u'_{h,t'}, ed_f, ed_m)$.

4.D.4 Normalizations

As mentioned above, skills and investments do not have a fixed location or scale which is why we need to normalize them. In the first period, we normalize the mean of the latent factors to be zero which fixes the location of the latent factors. In all other periods, the mean of the latent factor for skills h_t is allowed to be different from zero although the mean of investment is assumed 0 in all periods. Moreover, for each period, we set the scale parameter $\lambda_{\omega,t,1} = 1$ for one normalizing measure $Z_{\omega,t,1}$.

Agostinelli and Wiswall (2016) have shown that renormalization of the scale parameter $\lambda_{\omega,t,1} = 1$ can lead to biases in the estimation of coefficients in the case of overidentification of the production function coefficients when assuming that $\alpha_{1,t'} + \alpha_{2,t'} + \alpha_{3,t'} = 1$ in equation (4.18). This is not the case in our estimation as we do not assume $\alpha_{1,t'} + \alpha_{2,t'} + \alpha_{3,t'} = 1$ when estimating equation (4.18).

4.D.5 Intial Conditions Assumptions

Children are born in period $t' = 1$. The mean of h'_1, ed_f, ed_m and inv_1 are 0 by normalization and without loss of generality. h'_1 depends on parents' education and is normally distributed conditional on parents' education.

4.D.6 Estimation

1. **Scale parameters (λs) and variance of latent factors** . Here we estimate the parameters for the measurement equations for the child skill and investment latent variables. Using equation (4.19) we can derive the variance of each of the latent factors:

$$Cov(Z_{\omega,t,j}, Z_{\omega,t,j^*}) = \lambda_{\omega,t,j}\lambda_{\omega,t,j^*}\, Var(\omega_t) \tag{4.20}$$

   We have normalised $\lambda_{h,t,1} = \lambda_{inv,t,1} = 1$ to set the scale of h_t and of inv_t. For each other measure $j \neq 1$, and for $\omega \in \{h', inv\}$, using equation (4.20) we can show that:

$$\lambda_{\omega,t,j} = \frac{Cov(Z_{\omega,t,j}, Z_{\omega,t,j^*})}{Cov(Z_{\omega,t,1}, Z_{\omega,t,j^*})} \tag{4.21}$$

The model defined in equation (4.21) is overidentified if we have more than three measures since there are many different combinations of j and j^* that can be used here ($j^* \neq j$). We use GMM with an identity weighting matrix to estimate the λs where the moments are all the combinations of measures possible using equation (4.21). With these estimates of the λs in hand, we then estimate $Var(\omega_t)$ using equation (4.20). This equation is also overidentified with more than three measures, and again we estimate this using GMM.[18]

2. **Location parameters (μs) in measurement equations** At the child's birth ($t' = 1$), we normalize the mean of h'_1 and inv_1 to zero. Therefore:

$$\mu_{h',1,j} = \mathbb{E}[Z_{h',1,j}], \quad \mu_{inv,1,j} = \mathbb{E}[Z_{inv,1,j}] \tag{4.22}$$

3. **Calculation for next step** For each measure, we need to calculate a residualized measure of each Z for $\omega_t \in \{h_t, inv\}$:

$$\tilde{Z}_{\omega,t,j} = \frac{Z_{\omega,t,j} - \mu_{\omega,t,j}}{\lambda_{\omega,t,j}} \tag{4.23}$$

This will be used below in Step 1. Note that:

$$\omega_t = \tilde{Z}_{\omega,t,j} - \underbrace{\frac{\varepsilon_{\omega,t,j}}{\lambda_{\omega,t,j}}}_{\equiv \tilde{\varepsilon}_{\omega,t,j}} \tag{4.24}$$

It gives skill (or investment) plus an error rescaled to match scale of the skill (which is also the scale of skill measure 1).

[18]Note that at age 0 (period t'= 1) and age 16 (period t'= 4), we only have 2 measures of skill, respectively. To identify $\lambda_{h',4,j}$, we use covariances across time. For example, we use $Cov(Z_{h',3,j}, Z_{h',4,j}) = \lambda_{h',3,j}\lambda_{h',4,j}Cov(h'_3, h'_4)$ and $Cov(Z_{h',3,j}, Z_{h',4,j*}) = \lambda_{h',3,j}\lambda_{h',4,j*}Cov(h'_3, h'_4)$, thus $\frac{Cov(Z_{h',3,j}, Z_{h',4,j})}{Cov(Z_{h',3,j}, Z_{h',4,j*})} = \frac{\lambda_{h',3,j}\lambda_{h',4,j}Cov(h'_3, h'_4)}{\lambda_{h',3,j}\lambda_{h',4,j*}Cov(h'_3, h'_4)}$. For the normalizing measure $Z_{h',4,j*}$, $\lambda_{h',4,j*} = 1$, so this becomes $\frac{Cov(Z_{h',3,j}, Z_{h',4,j})}{Cov(Z_{h',3,j}, Z_{h',4,j*})} = \lambda_{h',4,j}$.

4. **Estimate latent skill production technology**

We will only describe the estimation of the production technology, as the estimation of the investment equation is analogous. Recall the production function:

$$h'_{t'+1} = \alpha_{1,t'} h'_{t'} + \alpha_{2,t'} inv_{t'} + \alpha_{3,t'} inv_{t'} \cdot h_{t'} + \alpha_{4,t'} ed_m + \alpha_{5,t'} ed_f + u'_{h,t'}$$

and using equation (4.24) note that we can rewrite the above equation as:

$$\begin{aligned} \frac{Z_{h',t'+1,j} - \mu_{h',t'+1,j} - \varepsilon_{h',t'+1,j}}{\lambda_{h',t'+1,j}} = \; & \alpha_{1,t'}(\tilde{Z}_{h',t',j} - \tilde{\varepsilon}_{h',t',j}) + \\ & \alpha_{2,t'}(\tilde{Z}_{inv,t',j} - \tilde{\varepsilon}_{inv,t',j}) + \\ & \alpha_{3,t'}(\tilde{Z}_{inv,t',j} - \tilde{\varepsilon}_{inv,t',j}) \cdot (\tilde{Z}_{h',t',j} - \tilde{\varepsilon}_{h',t',j}) + \\ & \alpha_{4,t'} ed_m + \alpha_{5,t'} ed_f + \\ & u'_{h,t'} \end{aligned} \quad (4.25)$$

or

$$\begin{aligned} \frac{Z_{h',t'+1,j} - \mu_{h',t'+1,j}}{\lambda_{h',t'+1,j}} = \; & \alpha_{1,t'} \tilde{Z}_{h',t',j} + \\ & \alpha_{2,t'} \tilde{Z}_{inv,t',j} + \\ & \alpha_{3,t'} \tilde{Z}_{inv,t',j} \cdot \tilde{Z}_{h',t',j} + \\ & \alpha_{4,t'} ed_m + \alpha_{5,t'} ed_f + \\ & \Bigg(u'_{h,t'} - \tilde{\varepsilon}_{h',t',j} - \tilde{\varepsilon}_{inv,t',j} - \tilde{\varepsilon}_{inv,t',j} \cdot \tilde{\varepsilon}_{h',t',j} + \frac{\varepsilon_{h',t'+1,j}}{\lambda_{h',t'+1,j}} \\ & - \alpha_{3,t'}(\tilde{Z}_{inv,t',j} \tilde{\varepsilon}_{h',t',j} + \tilde{Z}_{h',t',j} \tilde{\varepsilon}_{inv,t',j}) \Bigg). \end{aligned} \quad (4.26)$$

OLS is inconsistent here, as $\tilde{Z}_{h',t',j}$ and $\tilde{\varepsilon}_{h',t',j}$ are correlated. We resolve this issue by instrumenting for $\tilde{Z}_{h',t',j}$ using the other measures of skill $\tilde{Z}_{h',t',j^*}$ in that period.

Recall that we only normalized the location of factors in the first period, but have not done so for the subsequent periods (in this case $\mu_{h',t'+1,j}$). We estimate the location parameter for each measure j by estimating equation (4.26) using only output measure j on the left hand side. The intercept then identifies $\mu_{h',t'+1,j}$.

We estimate all location parameters (the μs) and the α parameters jointly using equation (4.26) by using a system GMM with diagonal weights. By using the system GMM we make efficient use of all available measures.

5. **Estimate the variance of the production function shocks**

The variance of the structural skills shock can be obtained using residuals from equation (4.26), where $\pi_{h',t,j} \equiv \left(u'_{h,t'} - \tilde{\varepsilon}_{h',t',j} - \tilde{\varepsilon}_{inv,t',j} - \tilde{\varepsilon}_{inv,t',j} \cdot \tilde{\varepsilon}_{h',t',j} + \frac{\varepsilon_{h',t'+1,j}}{\lambda_{h',t'+1,j}} - \alpha_{3,t'} (\tilde{Z}_{inv,t',j}\tilde{\varepsilon}_{h',t',j} + \tilde{Z}_{h',t',j}\tilde{\varepsilon}_{inv,t,j}) \right)$:

$$Cov\left(\frac{\pi_{h',t',j}}{\lambda_{h',t',j}}, \tilde{Z}_{h',t',j^*}\right) = \sigma^2_{h',t',j} \text{ for } j \neq j^*$$

which is true since the measurement errors $\tilde{\varepsilon}$ are uncorrelated across measures and so $Cov(\tilde{\varepsilon}_{h',t',j}, \tilde{Z}_{h',t',j^*}) = 0$ if $j \neq j^*$. As before, these covariances are overidentified, so we estimate these variances using GMM where the variance covariance matrix of the $\widehat{\sigma^2_{h',t',j}}$s is estimated using the bootstrap.

4.E Initial Skill

Initial skill at birth is a function of mother's education level, father's education level, and a shock. Using minimum distance methods, we estimate initial skill (after adjusting for their different scales) as a function of parental education dummies. We then estimate the variance of the shock analogously to Step 5 in the previous section. Table 4.E.1 shows the results of the minimum distance, and the variance of the initial skill shock.

Table 4.E.1: Initial skill regression

	Coefficient	SE
Mother's education		
Medium	0.092	(0.041)
High	0.079	(0.103)
Father's education		
Medium	0.066	(0.044)
High	-0.007	(0.081)
Constant	-0.038	(0.022)
Variance of shock	0.880	

4.F Signal to Noise Ratios

Note that using equation (4.19) the variance of measure $Z_{\omega,t,j} = (\lambda^2_{\omega,t,j})Var(\omega_t) + Var(\varepsilon_{\omega,t,j})$, where $(\lambda^2_{\omega,t,j})Var(\omega_t)$ comes from the variability in the signal in the measure and $Var(\varepsilon_{\omega,t,j})$ represents measurement error, or "noise". The signal to noise ratios for measure $Z_{\omega,t,j}$ is calculated in the following way:

$$s_{\omega,t,j} = \frac{(\lambda^2_{\omega,t,j})Var(\omega_t)}{(\lambda^2_{\omega,t,j})Var(\omega_t) + Var(\varepsilon_{\omega,t,j})}$$

Intuitively, this is the variance of the latent factor (signal) to the variance of the measure (signal+noise) and thus describes the information content of each measure.

Table 4.F.1 presents signal to noise ratios for skill. At birth, birthweight is the most informative measure. At age 7, reading, maths, coping, and drawing scores are all roughly equally informative. At ages 11 and 16, maths scores become the most informative.

Table 4.F.1: Signal to noise ratios: Skill measures

Age 0		*Age 7*		*Age 11*		*Age 16*	
birthweight	0.862	read	0.385	read	0.555	read	0.570
gestation	0.140	maths	0.335	maths	0.942	maths	0.713
		copy	0.259	copy	0.104		
		draw	0.281				

Note: At ages 0 and 16, we only have 2 measures of skill.

Table 4.F.2 presents signal to noise ratios for investment. Here we have many

measures of investment. The most informative measures when young are the frequency of father's outings with the child, and both mother's and father's frequency of reading to the child. At older ages, the most informative variable is the teacher's assessment of each parent's interest in the child's education.

Table 4.F.2: Signal to noise ratios: Investment measures

Age 0-6		*Age 7-10*		*Age 11-15*	
mum: interest	0.164	mum: interest	0.356	mum: interest	0.796
mum: outing	0.270	mum: outings	0.235	dad: interest	0.765
mum: read	0.456	dad: outings	0.166	other index	0.344
dad: outing	0.773	dad:interest	0.386	parental ambition	0.221
dad: interest	0.082	dad:role	0.033		
dad:read	0.539	parents initiative	0.206		
dad: large role	0.069	parents ambition uni	0.093		
other index	0.136	parents ambition school	0.249		
		library member	0.253		

Notes: All investment measures are retrospective, so age 0-6 investments are measured at age 7, age 7-10 investments are measured at age 11, age 11-15 investments are measured at age 16.

4.G Accounting for Measurement Error in Skill Levels and Wages

We estimate the wage equation laid out in equations (4.4) and (4.5), but allow for i.i.d. measurement error in wages u_t. Using those equations and noting that $v_t = \delta_5 h_4 + \sum_{k=5}^{t} \eta_k$ yields:

$$\ln w_t^* = \ln w_t + u_t = \delta_0 + \delta_1 t + \delta_2 t^2 + \delta_3 t^3 + \delta_4 PT_t + \delta_5 h_4 + \sum_{k=5}^{t} \eta_k + u_t \qquad (4.27)$$

for each gender and education group.

In our procedure we must address three issues. First, wages are measured with error u_t. Second, skill h_4 is measured with error. Third, we only observe the wage for those who work, which is a selected sample.

We address issues of selectivity by relying on our panel data as much as is possible. To address the issue of composition bias (the issue of whether lifetime high or low wage individuals drop out of the labor market first), we use a fixed effects estimator. Given our assumption of a unit root in $v_t = \delta_5 h_4 + \sum_{k=5}^{t} \eta_k$, which we estimate to be close to the truth (see Appendix 4.G.1 for estimates that relax this assumption and allow v_t follow an AR(1)), we can allow v_5 (the first shock to wages, age 23) to be correlated with other observables, and estimate the model using fixed effects. In particular, the procedure is:

Step 0: From equation (4.27) note that:

$$\ln w_t^* = \delta_1 t + \delta_2 t^2 + \delta_3 t^3 + \delta_4 PT_t + FE + \xi_t$$

where FE is a person specific fixed effect capturing the time invariant factors $\delta_5 h_4 + \eta_5$ and ξ_t is a residual.

Step 1: Estimate δ_1, δ_2, δ_3, δ_4 using fixed effects (FE) regression.

Step 2: Predict the fixed effect:

$$\begin{aligned} \widehat{FE} &\equiv \bar{\ln w_t^*} - \hat{\delta}_1 \bar{t} - \hat{\delta}_2 \bar{t^2} - \hat{\delta}_3 \bar{t^3} - \hat{\delta}_4 \bar{PT}_t \\ &= \delta_0 + \delta_5 h_4 + \eta_5 \\ &= \delta_0 + \delta_5 \tilde{Z}_{h,4,j} + \eta_5 - \delta_5 \tilde{\varepsilon}_{h,4,j} \qquad (4.28) \end{aligned}$$

where the means are over all observations of an individual, e.g., $\bar{t}$ is the mean age of an individual over all years she was observed, and $h_4 = \tilde{Z}_{h,4,j} - \tilde{\varepsilon}_{h,4,j}$ and where $\tilde{Z}_{h,4,j}$ and $\tilde{\varepsilon}_{h,4,j}$ have been defined in equation (4.24). The above equation holds for all measures j. Although the estimated fixed effect, $\widehat{FE}$, is affected by variability in the sequence of wage shocks $\{\eta_t\}_{t=5}^{12}$ and measurement errors $\{u_t\}_{t=5}^{12}$, this merely adds in measurement error on the left hand side variable in equation (4.28). However, measurement error on the right hand side h_4 is more serious: we only have the noisy proxies $\tilde{Z}_{h,4,j}$ which are correlated with $\tilde{\varepsilon}_{h,4,j}$ by construction. We address this problem in the next step.

Step 3: Using GMM, we project the predicted fixed effect ($\widehat{FE}$) on each measure of skill, $\tilde{Z}_{h,4,j}$, and instrument by using the respective other measures, $\tilde{Z}_{h,4,j'}$, to obtain $\hat{\delta}_0$ and $\hat{\delta}_5$. Since we have two measures of skill (reading and math), we have two equations and two instruments. When reading is the skill measure, we instrument for this using math, and vice versa. Our GMM procedure efficiently combines different measures of skill and yields consistent estimates of $\hat{\delta}_0$ and $\hat{\delta}_5$ even in the presence of measurement error in the skill measures.

Step 4: Then use covariances and variances of residuals to calculate shock variances.

Substituting a noisy measure of skill into the wage equation (4.27) yields

$$\ln w_t^* = \delta_0 + \delta_1 t + \delta_2 t^2 + \delta_3 t^3 + \delta_4 PT_t + \delta_5 \tilde{Z}_{h,4,j} + \sum_{k=5}^{t} \eta_k + u_t - \delta_5 \tilde{\varepsilon}_{h,4,j}$$

where we use the fact that $h_4 = \tilde{Z}_{h,4,j} - \tilde{\varepsilon}_{h,4,j}$ as defined in equation (4.24). Next we define a wage residual that will exist for each skill measure:

$$\begin{aligned}\widetilde{\ln w_{t,j}} &\equiv \ln w_t^* - (\delta_0 + \delta_1 t + \delta_2 t^2 + \delta_3 t^3 + \delta_4 PT_t + \delta_5 \tilde{Z}_{h,4,j}) \\ &= \sum_{k=5}^{t} \eta_k + u_t - \delta_5 \tilde{\varepsilon}_{h,4,j}\end{aligned}$$

Note that from the measurement equation (4.19), $Var(\tilde{Z}_{h,4,j}) = Var(h_4) + Var(\tilde{\varepsilon}_{h,4,j})$, where we have previously estimated $Var(h_4)$ using equation (4.20) and $Var(\tilde{Z}_{h,4,j})$ is the variance of the renormalized measures in the data. We can then back out the variance of the measurement error and plug it into the following equation to estimate the parameters of the wage shocks:

$$Cov(\widetilde{\ln w_{t,j}}, \widetilde{\ln w_{t+l,j}}) = Var(\sum_{k=5}^{t} \eta_k) + \delta_5^2 Var(\tilde{\varepsilon}_{h,4,j}) \text{ for } l > 0$$

$$Var(\widetilde{\ln w_{t,j}}) = Var(\sum_{k=5}^{t} \eta_k) + Var(u_t) + \delta_5^2 Var(\tilde{\varepsilon}_{h,4,j})$$

Step 5: correct the δ parameters for selection. The fixed-effects estimator is identified using wage growth for workers. If wage growth rates for workers and non-workers are the same, composition bias problems—the question of whether high wage individuals drop out of the labor market later than low wage individuals—are not a problem. However, if individuals leave the market because of a wage drop, such as from job loss, then wage growth rates for workers will be greater than wage growth for non-workers. This selection problem will bias estimated wage growth upward.

We control for selection bias by finding the wage profile that, when fed into our model, generates the same fixed effects profile as in the estimates using the NCDS data. Because the simulated fixed effect profiles are computed using only the wages of those simulated agents that work, the profiles should be biased upwards for the same reasons they are in the data. We find this bias-adjusted wage profile using the iterative procedure described in French (2005).

4.G.1 Wage shock process estimates without imposing random walk

In section 4.5.2, we impose that wage shocks have an autocovariance of 1. Table 4.G.1 shows the coefficients and standard errors when we relax this assumption and allow the persistence parameter to be different from one. We also report results from an overidentification test statistic. To do this we initially regress log wages on age, education, skill and part time status as before, then estimate the process for the residuals using an error components model where we match the variance covariance matrix of wage residuals. When estimating, we allow for an AR(1) process with homoskedastic (i.e., with age-invariant variances) innovations and a transitory shocks in which we allow for heteroskedasticity. We have 5 periods of data, and thus 15 unique elements of the variance covariance matrix which we treat as moment conditions for each gender/education group. We estimate the variances of the transitory shocks (5 parameters), the initial variance of the AR(1) component, the variance of the AR(1) shocks, and ρ, meaning that we have 8 parameters to estimate and thus 15-8=7 degrees of freedom, meaning that under the null of correct model specification our test statistic should be distributed $\chi^2(7)$. Overall, the model fits the data well and we cannot reject the hypothesis of correct model specification for many groups. Perhaps more importantly, we can see that for all groups except low educated females, we cannot reject the hypothesis that the persistence parameter is 1. Even for this group, the value of $\rho = 0.94$. Thus throughout we assume $\rho = 1$ for all groups.

Table 4.G.1: Estimates for AR(1) process without random walk restriction

		Male	
Education:	Low	Medium	High
ρ	1.034	0.968	1.027
	(0.022)	(0.018)	(0.019)
Test stat:	10.74	12.96	36.60
		Female	
Education:	Low	Medium	High
ρ	0.940	0.985	0.971
	(0.029)	(0.023)	(0.023)
Test stat:	35.38	23.56	19.54

This table shows the persistency parameter for an AR(1) wage shock process when we relax the assumption that the process is a random walk. Bootstrapped standard errors are in parentheses. The rows entitled "Test stat" show the overidentification test statistic.

4.H Computational Details

This Appendix details how we solve for optimal decision rules as well as our simulation procedure. We describe solving for optimal decision rules first.

1. To find optimal decision rules, we solve the model backwards using value function iteration. The state variables of the model are model period, assets, wage rates, education levels, childrens' gender, childrens' skill, and childrens' education. At each model period, we solve the model for 50 grid points for assets, 10 points for wage rates (for each spouse), 3 education levels for each spouse, childrens' gender, childrens' skill (5 points), and childrens' education. Because we assume that the two children are identical, receive identical shocks, and that parents make identical decisions towards the two children, we only need to keep track of the state variables for one child. Our

approach for discretizing wage shocks follows Tauchen (1986). The bounds for the discretisation of the wage process is ± 3 standard deviations. For skills we use Gauss-Hermite procedures to integrate. We use linear interpolation between grid points when on the grid, and use linear extrapolation outside of the grid.

2. Parents can each choose between between 4 levels of working hours (non-employed, part-time, full-time, over-time) and in model period $t = 6,7$ and 8 they can choose between six levels of time spent with children. In all model periods except $t = 10$ we solve for the optimal level of next period assets using golden search. In period $t = 10$ parents may also transfer assets to children: we solve this two-dimensional optimization problem using Nelder-Mead. We back out household consumption from the budget constraint and then solve for individual level consumption from the intra-temporal first order condition, which delivers the share of household consumption going to the male in the household. As this first order condition is a non-linear function we approximate the solution using the first step of a third order Householder algorithm. This allows us to use the information contained in the first three derivatives of the first order condition. We found this method to give fast and accurate solutions to the intra-temporal problem. Details of this are available from the authors.

Next we describe our simulation procedure.

1. Our initial sample of simulated individuals is large, consisting of 50,000 random draws of individuals in the first wave of our data at age 23. Given that we randomly simulate a sample of individuals that is larger than the number of individuals observed in the data, most observations will be drawn multiple times. We take random Monte Carlo draws of education and assets, which are the state variables that we believe are measured accurately and are observed for everyone in the data. For the variables with a large amount of measurement error, or which are not observed for all sample members (i.e., initial

skill of child, and wages of each parent), we exploit the model implied joint distribution of these state variables. We assume child's gender is randomly distributed across the population.

2. Given the optimal decision rules, the initial conditions of the state variables, and the histories of shocks faced by both parents and children, we calculate life histories for savings, consumption, labor supply, time and education investments in children, which then implies histories for childrens' skill, educational attainment. For discrete choice variables (e.g. participation), we evaluate whether the choice is the same at all surrounding grid points. If not, we resolve the household's problem given each of the household's choices (e.g., work and not work), and choose the value that delivers the highest value. If so, we take the implied discrete variable, and if any of the continuous state variables (e.g. assets) is between grid-points, we interpolate to find the implied decision rule.

3. We aggregate the simulated data in the same way we aggregate the observed data and construct moment conditions. We describe these moments in greater detail in Appendix 4.I. Our method of simulated moments procedure delivers the model parameters that minimize a GMM criterion function which we also describe in Appendix 4.I.

4. To search for the parameters that minimize the GMM criterion function, we use the BOBYQUA algorithm developed by Powell (2009). This is a derivative-free algorithm that uses a trust region approach to build quadratic models of the objective function on sub-regions.

4.I Moment Conditions and Asymptotic Distribution of Parameter Estimates

We estimate the parameters of our model in two steps. In the first step, we estimate the vector χ, the set of parameters than can be estimated without explicitly using our model. In the second step, we use the method of simulated moments (MSM) to estimate the remaining parameters, which are contained in the $M \times 1$ parameter vector Δ =(β, ν_f, ν_m, γ, $\lambda, \theta, \{\kappa_{1,t'}\}_{\{t'=1,2,3\}}$). Our estimate, $\hat{\Delta}$, of the "true" parameter vector Δ_0 is the value of Δ that minimizes the (weighted) distance between the life-cycle profiles found in the data and the simulated profiles generated by the model.

We match data from three different sources. For most of our moments we use data from the NCDS. However, the NCDS currently lacks high quality asset and transfer data after age 23, and does not have detailed time use information with children. For the asset and transfer data we also match data from ELSA, and for the information on time with children we also use UKTUS.

From the NCDS we match, for three education groups *ed*, two genders (male and female) g, $T = 5$ different periods: $t \in \{5, 7, 9, 10, 11\}$ (which corresponds to ages 23, 33, 42, 50, 55) the following moment conditions: $3 \times 2 \times T = 6T$ moment conditions: employment rates (forming 6T moment conditions), mean annual work hours of workers ($6T$). In addition, from the NCDS we match age 16 skill and the mean education leaving age, conditional on father's education level (6 moment conditions).

From ELSA we match mean lifetime inter-vivos transfers received (1 moment) and also household median wealth at age 60 (1 moments).

From UKTUS we match mean annual time spent with children, by age of child (ages 0-7, 8-11, 12-16) and gender and education of parent (18 moments).

In the end, we have a total of $J = 86$ moment conditions.

Our approach accounts explicitly for the fact that the data are unbalanced: some individuals leave the sample, and we use multiple datasets, so an individual who belongs in one sample (e.g., NCDS) likely does not belong to another sample (e.g., ELSA or UKTUS). Suppose we have a dataset of I independent individuals that are each observed in up to J separate moment conditions. Let $\varphi(\Delta; \chi_0)$ denote the J-element vector of moment conditions described immediately above, and let $\hat{\varphi}_I(.)$ denote its sample analog. Letting $\widehat{\mathbf{W}}_I$ denote a $J \times J$ weighting matrix, the MSM estimator $\hat{\Delta}$ is given by

$$\underset{\Delta}{\operatorname{argmin}} \frac{I}{1+\tau} \hat{\varphi}_I(\Delta;\chi_0)' \widehat{\mathbf{W}}_I \hat{\varphi}_I(\Delta;\chi_0),$$

where τ is the ratio of the number of observations to the number of simulated observations.

In practice, we estimate χ_0 as well, using the approach described in the main text. Computational concerns, however, compel us to treat χ_0 as known in the analysis that follows. Under regularity conditions stated in Pakes and Pollard (1989) and Duffie and Singleton (1993), the MSM estimator $\hat{\Delta}$ is both consistent and asymptotically normally distributed:

$$\sqrt{I}\left(\hat{\Delta} - \Delta_0\right) \rightsquigarrow N(0, \mathbf{V}),$$

with the variance-covariance matrix $\mathbf{V}$ given by

$$\mathbf{V} = (1+\tau)(\mathbf{D}'\mathbf{W}\mathbf{D})^{-1}\mathbf{D}'\mathbf{W}\mathbf{S}\mathbf{W}\mathbf{D}(\mathbf{D}'\mathbf{W}\mathbf{D})^{-1},$$

where $\mathbf{S}$ is the variance-covariance matrix of the data;

$$\mathbf{D} = \frac{\partial \varphi(\Delta;\chi_0)}{\partial \Delta'}\Big|_{\Delta=\Delta_0} \tag{4.29}$$

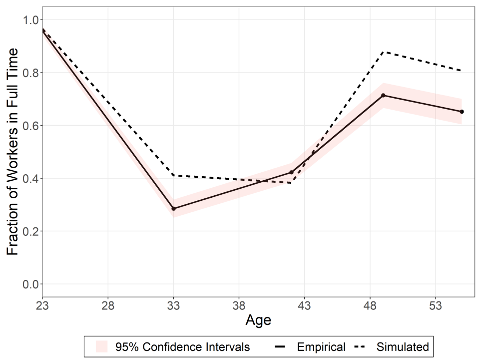
is the $J \times M$ gradient matrix of the population moment vector; and $\mathbf{W} = \operatorname{plim}_{I\to\infty}\{\widehat{\mathbf{W}}_I\}$. The asymptotically efficient weighting matrix arises when $\widehat{\mathbf{W}}_I$ converges to $\mathbf{S}^{-1}$, the inverse of the variance-covariance matrix of the data. When $\mathbf{W} = \mathbf{S}^{-1}$, $\mathbf{V}$ simplifies to $(1+\tau)(\mathbf{D}'\mathbf{S}^{-1}\mathbf{D})^{-1}$.

But even though the optimal weighting matrix is asymptotically efficient, it can be biased in small samples. (See, for example, Altonji and Segal (1996).) We thus use a “diagonal” weighting matrix. This diagonal weighting scheme consists of the inverse of the moments along the diagonal, which will weight more heavily moments with low means so that they too will contribute significantly to the GMM criterion function, regardless of how precisely estimated they are.

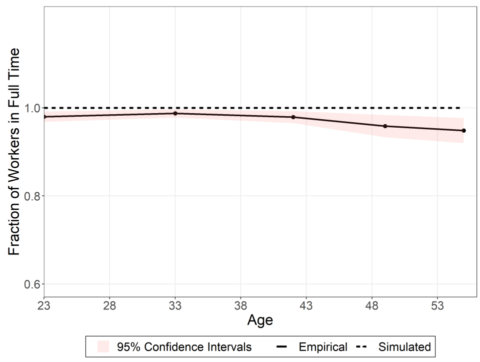
We estimate **D**, **S**, and **W** with their sample analogs. For example, our estimate of **S** is the $J \times J$ estimated variance-covariance matrix of the sample data. One complication in estimating the gradient matrix **D** is that the functions inside the moment condition $\varphi(\Delta;\chi)$ are non-differentiable at certain data points (e.g., for employment). This means that we cannot consistently estimate **D** as the numerical derivative of $\hat{\varphi}_I(.)$. Our asymptotic results therefore do not follow from the standard GMM approach, but rather the approach for non-smooth functions described in Pakes and Pollard (1989), Newey and McFadden (1994), and Powell (1994). When calculating gradients we vary step-sizes, then take the average gradient over the different step-sizes.

4.J Further Details on Model Fit

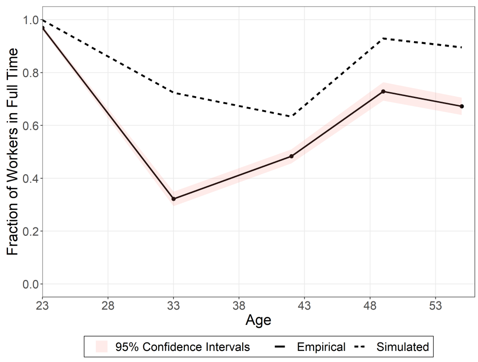
Figure 4.J.1: Model fit: full-time work conditional on employment



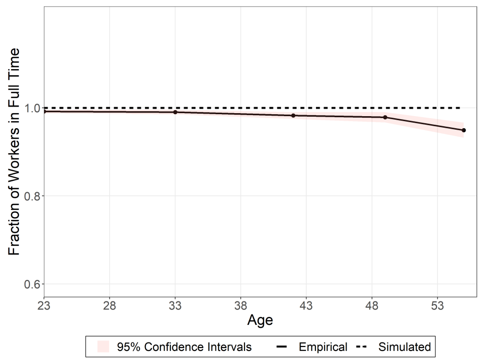
(a) Women: low educated



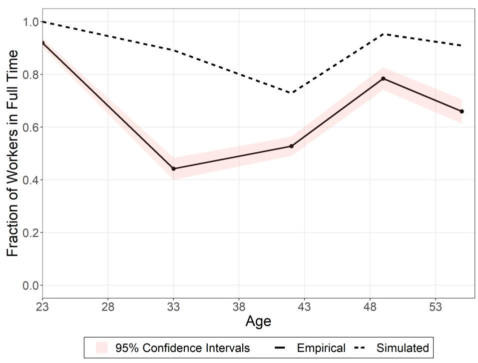
(b) Men: low educated



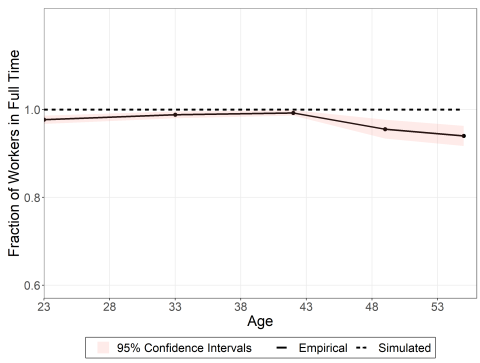
(c) Women: medium educated



(d) Men: medium educated



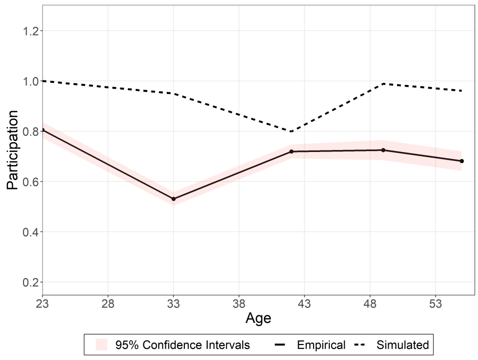
(e) Women: high educated



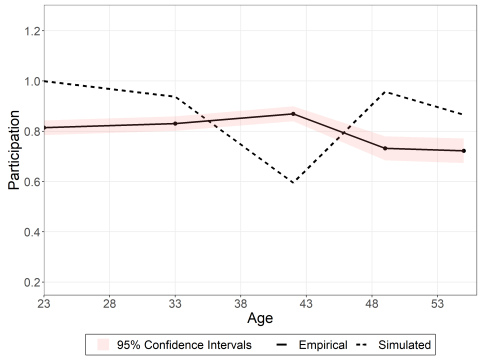
(f) Men: high educated

Notes: Figures show fraction in fulltime work at different ages conditional on being employed for women and men. Empirical data come from NCDS.

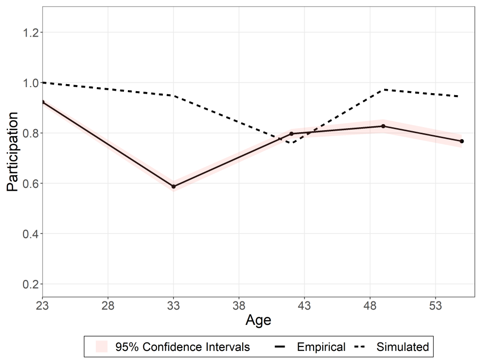
Figure 4.J.2: Model fit: participation



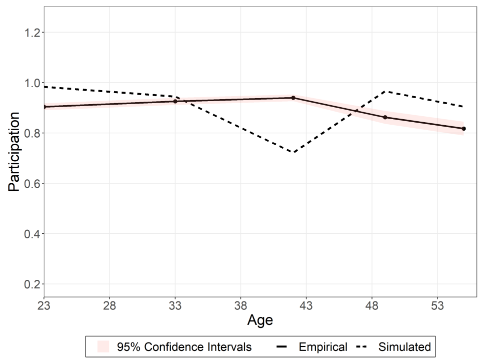
(a) Women: low educated



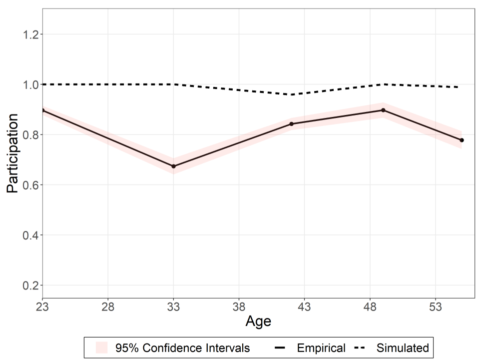
(b) Men: low educated



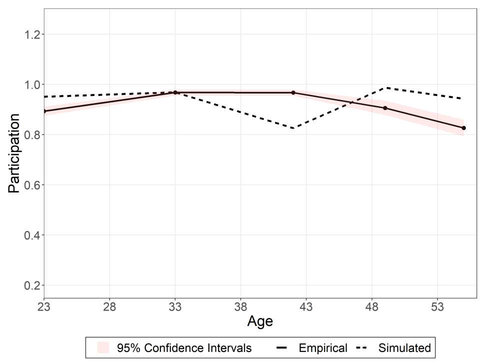
(c) Women: medium educated



(d) Men: medium educated



(e) Women: high educated



(f) Men: high educated

Notes: Figures show fraction of individuals employed at different ages for women and men. Empirical data come from NCDS.

4.K Identification of the time cost of investments θ

To give some intuition regarding the identification of θ, we use a simplified two period version of our dynastic model, where we abstract from couples, uncertainty, and where we assume a linear production function. The household's state variables are: education ed, skill h, and their initial assets a_1. The parent is altruistic towards their child and incorporate their child's value function into their problem, but discounts it by factor λ. Households choose consumption c_t, leisure l_t, time investments ti_t, monetary transfers to their child x_1' and the education of the child ed' which can be dropout (D), high school (HS) or college (C). Each education choice is associated with a price p_k, $k \in \{D, HS, C\}$, which can be interpreted as the price of foregone labor earnings of the child. The child initially has no other assets than the monetary transfer from the parent. We first describe the discrete decision problem of the parent who selects their children's education level. They maximize their value function which nests the child's value function:

$$V(ed,h,a_1) = \max_{ed'=\{D,HS,C\}} \{V_{ed'=D}, V_{ed'=HS}, V_{ed'=C}\} \tag{4.30}$$

where $V_{ed'=k}$ denotes the value of the problem if the parents choose education level k for their child. The above nests the following decision problem over consumption, leisure, time investments and asset transfers:

$$V_{ed'=k}(ed,h,a_1) = \tag{4.31}$$

$$\max_{c_1,l_1,ti_1,x_1',c_2,l_2,ti_2} u(c_1,l_1) + \beta u(c_2,l_2) + \lambda V'(ed'=k,h',a_1') \tag{4.32}$$

subject to:

$$\begin{aligned} &(1+r)^2 a_1 + (1+r) hrs_1 w_1(ab,ed) + hrs_2 w_2(h,ed) = \\ &(1+r)c_1 + c_2 + x_1' + \sum_{ed'=\{D,HS,C\}} p_k \mathbb{1}_{[ed'=k]} \end{aligned} \tag{4.33}$$

$$h' = \alpha_0 + \alpha_1 ti_1 + \alpha_2 ti_2 \tag{4.34}$$

$$T = \theta ti_1 + hrs_1 + l_1 \tag{4.35}$$

$$T = \theta ti_2 + hrs_2 + l_2 \tag{4.36}$$

$$a'_1 = x'_1, x'_1 \geq 0 \tag{4.37}$$

where (4.33) describes the monetary budget constraint over 2 periods, (4.34) shows the human capital production function over two periods where α_1, α_2 are the productivity of time investments for final skill. (4.35) and (4.36) are the time constraints in period 1 and 2, and (4.37) states that initial assets equal the initial parental cash transfer.

Assuming interior conditions for the choice variables $\{c_1, l_1, ti_1, c_2, l_2, ti_2\}$ but allowing the constraint x'_1 to bind we can now rewrite this problem and derive optimality conditions:

$$\begin{aligned} V_{ed'=k}(ed,h,a_1) = \quad & \max_{c_1,l_1,ti_1,x'_1,c_2,l_2,ti_2} u(c_1,l_1) + \beta u(c_2,l_2) + \lambda V'(h',ed',x'_1) \\ & +\mu[(1+r)^2 a_1 + (1+r)hrs_1 w_1(h,ed) + hrs_2 w_2(h,ed) \\ & -(1+r)c_1 - c_2 - x'_1 - p_k \mathbb{1}_{[ed'=k]}] \\ & +\kappa(\alpha_0 + \alpha_1 ti_1 + \alpha_2 ti_2 - h') + \phi(x'_1) \\ & +\zeta_1(\theta ti_1 + hrs_1 + l_1 - T) \\ & +\zeta_2(\theta ti_2 + hrs_2 + l_2 - T) \end{aligned}$$

Euler equation: $\frac{\partial u}{\partial c_1} = \beta \frac{\partial u}{\partial c_2}(1+r)$

FOC wrt ti_1: $\zeta_1 \theta - \kappa \alpha_1 = 0$

FOC wrt h': $\kappa + \lambda \frac{\partial V'}{\partial h'} = 0$

$$\kappa + \lambda \mu'[(1+r)\frac{\partial w'_1(h',ed'=k)}{\partial h'} hrs'_1 + \frac{\partial w'_2(h',ed'=k)}{\partial h'} hrs'_2] = 0$$

FOC wrt l_1:$-\zeta_1 + \frac{\partial u}{\partial l_1} = 0$

FOC wrt l_2:$-\zeta_2 + \beta \frac{\partial u}{\partial l_2} = 0$

FOC wrt hrs_1:$-\zeta_1 + \mu(1+r)w_1(h,ed) = 0$

FOC wrt hrs_2:$-\zeta_2 + \mu w_2(h, ed) = 0$

FOC wrt x'_1:$-\mu + \lambda \frac{\partial V'}{\partial x'_1} = 0$

$$\mu = \lambda \mu'(1+r)^2 + \phi \Rightarrow \mu \geq \lambda \mu'(1+r)^2$$

From this, we can derive the following optimality condition for investments in period 1:

$$w_1(h, ed)\theta \leq \alpha_1 [\frac{1}{(1+r)^2} \frac{\partial w'_{1'}(h', ed')}{\partial h'} hrs'_{1'} + \frac{1}{(1+r)^3} \frac{\partial w_{2'}(h', ed')}{\partial h'} hrs'_{2'}] \quad (4.38)$$

This equation is key to understanding the identification of θ. On the left hand side, we have the marginal cost of investments to the parent which is their wage times θ – the amount of leisure they lose per hour of time spent with the child. On the right hand side, we have the marginal benefit of an hour spent with the child; this is the increase in the present discounted value of the child's future income from the hour of investment. The increase equals the productivity of an hour of time α_1, multiplied by the resulting marginal increase in income over the life cycle to the child. If cash transfers are positive equation (4.38) holds with equality, although if cash transfers are 0 then it is an inequality. Dividing both sides by $w_1(h, ed)$ shows that we can place an upper bound on θ by calculating the present values of the gain in child's lifetime income from one hour of time investment relative to the wage. In the appendix below we perform exactly this calculation.

4.K.1 Approximating the PDV of time investments

We estimate θ to be 0.049, implying that 95% of time spent with the child is leisure time. This is surprising, given that some studies have estimated sizeable returns to early life investments such as Perry Pre-School (e.g. García et al. (2020)). Furthermore, many studies have found positive returns to parental investment. To gain some intuition, we conduct a back-of-the envelope calculation using equation (4.38) that takes into account the production function of skill, the opportunity cost of time to parents, and the returns to time investments in the form of higher lifetime earnings of the child.

In the following calculation, we consider a family with a low educated father. We assume that at baseline, the amount of time that the family invests in their child is at the mean in each period and calculate the resulting skill at age 16. We also calculate the child's expected lifetime earnings, assuming that the child is male, has low education, and works 40 hours per week up until age 65.

We then consider the impact of one additional hour per week spent with the child in the first period of life (0-6). The resulting skill increase translates to a wage increase of 0.5% at each age, causing lifetime earnings to increase by £5,079 when not discounted, or by £1,141 when discounting back to age 6 (using an interest rate of r=0.0469).

To calculate the lifetime costs to the parents of increasing investments by 1 hour per week, we assume that parents have an hourly wage of £9.3, which is the average expected wage between ages 26-32 for a low educated male. They thus forgo $£9.3 \times 52$ weeks $\times$ 6 years $= £2,902$ when they increase their investment by 1 hour per week during the first childhood stage 1. Thus, the ratio of the NPV of returns to cost is $\frac{1,141}{2,902} = 0.39$.

4.L Identification of κ

Our structural model maps hours of parental time into future skill. However, the NCDS has latent investments and future skill. Here we show more on the mapping between hours of time and latent investments.

As described in section 4.4.1, we assume hours of parental time spent with children $ti_{m,t'}, ti_{f,t'}$ are converted to latent investment units $inv_{t'}$ according to equation (4.12) which we reproduce here:

$$inv_{t'} = \kappa_{0,t'} + \kappa_{1,t'}(ti_{m,t'} + ti_{f,t'})$$

where $\kappa_{1,t'}$ is the hours-to-latent investments conversion parameter which determines the productivity of an hour of time and $\kappa_{0,t'}$ is a constant. We allow the κ parameters to vary by age. With three investment periods and two parameters in each period, this gives us six parameters to estimate.

To gain intuition regarding the identification of the κ parameters, recall equation (4.13) which shows the relationship between time investments and skill:

$$
\begin{aligned}
h'_{t'+1} &= \alpha_{1,t'}h'_{t'} + \alpha_{2,t'}inv_{t'} + \alpha_{3,t'}inv_{t'} \cdot h_{t'} + \alpha_{4,t'}ed^m + \alpha_{5,t'}ed^f + u'_{h,t'} \\
&= \alpha_{1,t'}h'_{t'} + \alpha_{2,t'}(\kappa_{0,t'} + \kappa_{1,t'}ti_{t'}) + \alpha_{3,t'}(\kappa_{0,t'} \\
&+ \kappa_{1,t'}ti_{t'}) \cdot h'_{t'} + \alpha_{4,t'}ed^m + \alpha_{5,t'}ed^f + u'_{h,t'}
\end{aligned}
$$

The top line is the estimating equation using the NCDS data: the α parameters are estimated using the latent investment, skill, and parental education measures in the NCDS data. The κ parameters are estimated within the dynamic programming model. Identification of the κ_1 parameters comes from the gradients in time investments $ti_{m,t'} + ti_{f,t'}$ (from the UKTUS data) by parental education and the corresponding skill gradients (from the NCDS). All the α parameters and parental education ed^m, ed^f are known. From UKTUS we know that at each age, high education parents spend more time with their children than low education parents and from the NCDS we know that at each age the children of high education parents have higher skill levels, even after controlling for the direct effect of parental education on skill: $\alpha_{4,t'}ed^m + \alpha_{5,t'}ed^f$. κ_1 thus captures how differences in time investments by parental education translate into differences in skill, controlling for parental education. The κ_0 parameters allow us to match mean time investments at different ages, observed in the UKTUS. The means of hours of time is positive, whereas the mean of latent investment is 0.

4.M Updating the matching probabilities in counterfactuals

In Section 4.5.3, we show that marital matching probabilities depend on the education level of the male and the female. These probabilities reflect the prevailing distribution of education levels in the population for the cohort we study. When evaluating the education subsidy in Section 4.7.4, we must account for the fact that, in counterfactual settings, the distribution of education levels in the population may change, which will lead to changes in the marital matching probabilities. We account for this in our counterfactuals by allowing matching probabilities to depend on population education shares.

We estimate these matching probabilities as a function of the distribution of education levels observed in the population using data from the Family Expenditure Survey (FES) and its successor surveys from 1978 to 2017. During this time, there were major changes in the distribution of education, both for men and women. For example, the share of women with high education increased from less than 10% in 1987 to more than 40% in 2017. We use these data to estimate the following ordered logit model where for each gender and education level, we estimate the probability of matching with someone of the other gender with a certain education level, conditional on the distribution of education in the population of both genders. For example, we estimate the probability that an individual of gender g and education level $ed = j$ partners with an individual of education level $ed^P = i \in \{$low, medium, and high educated$\}$ as:

$$\begin{aligned} p_{i,j,g} &= Pr\left(ed^P_{j,g} = i\right) = Pr\left(\kappa_{i-1,j,g} < \mathbf{x}\boldsymbol{\beta}_{j,g} + u \leq \kappa_{i,j,g}\right) \\ &= \frac{1}{1+\exp\left(-\kappa_{i,j,g} + \mathbf{x}\boldsymbol{\beta}_{j,g}\right)} - \frac{1}{1+\exp\left(-\kappa_{i-1,j,g} + \mathbf{x}\boldsymbol{\beta}_{j,g}\right)} \end{aligned} \tag{4.39}$$

where $\mathbf{x}\boldsymbol{\beta}\beta\beta_j = \beta_{1,j} S_{m,low} + \beta_{2,j} S_{m,medium} + \beta_{3,j} S_{f,low} + \beta_{4,j} S_{f,med}$. $S_{g,ed}$ denotes the share in the population who are in gender group g and education group ed and the $\kappa_{i,j}$ parameters are the estimated thresholds for each group. Equation (4.39) is estimated separately for each education level and gender. In our dynastic model, any given policy environment generates population shares $S_{g,ed}$. These can be used with the parameters estimated here to deliver the matching probabilities that characterize the marriage market under the new equilibrium.

4.M.1 Budget Constraints and Income Sources

Constraints Households face an intertemporal budget constraint and a borrowing constraint:

$$a_{t+1} = (1+r_t)(a_t + y_t - (c_{m,t} + c_{f,t}) - x_t) \tag{4.40}$$

$$a_{t+1} \geq 0 \tag{4.41}$$

where a_t is the household wealth, y_t is household income and x_t is a cash transfer to children that can only be made when the members of the couple are 49 and their children are 23 (and so $x_t = 0$ in all other periods). The gross interest rate $(1+r_t)$ is equal to $(1+r)^{\tau_t}$ where r is an annual interest rate and τ_t is the length in years of model period t.

Earnings and household income Household income is given by

$$y_t = \tau(e_{m,t}, e_{f,t}, e'_t, t) \tag{4.42}$$

where $\tau(.)$ is a function which returns net-of-tax income and $e_{m,t}$ and $e_{f,t}$ are male and female earnings respectively. Before age 16 children's earnings (e'_t) are 0. In their last period before the 'Independence' phase of life (age 16), children can participate in the labor market if they are no longer in education. Their parents are still the decision-maker in this period and any income the children earn is part of household resources in that period. When not at school in this period we assume the child works full time. We model the potential income if the child works, and thus the loss of household income if the child receives additional years of education.

Earnings are equal to hours worked (h) multiplied by the wage rate for each spouse: $e_{g,t} = h_{g,t} w_{g,t}$.

Bibliography

Abbott, B., G. Gallipoli, C. Meghir, and G. Violante (2019). Education policy and intergenerational transfers in equilibrium. *Journal of Political Economy*.

Agostinelli, F. and M. Wiswall (2016, July). Identification of dynamic latent factor models: The implications of re-normalization in a model of child development. Working Paper 22441, National Bureau of Economic Research.

Agostinelli, F. and M. Wiswall (2022, July). Estimating the technology of children's skill formation. Working Paper 22442, National Bureau of Economic Research.

Aizer, A. and F. Cunha (2012). The production of human capital: Endowments, investments and fertility. Technical report, National Bureau of Economic Research.

Albuquerque, R. (2012, 5). Skewness in stock returns: Reconciling the evidence on firm versus aggregate returns. *Review of Financial Studies 25*, 1630–1673.

Altonji, J. G. and L. M. Segal (1996). Small sample bias in gmm estimation of covariance structures. *Journal of Business and Economic Statistics 14*(3), 353–366.

Altonji, J. G. and E. Villanueva (2007). The marginal propensity to spend on adult children. *The BE Journal of Economic Analysis & Policy 7*(1).

Amin-Smith, N. and R. Crawford (2018). State pension age increases and the circumstances of older women. In *The Dynamics of Ageing: Evidence from the english longitudinal study of ageing*, pp. 9–39.

Apesteguia, J. and M. A. Ballester (2018). Monotone stochastic choice models: The case of risk and time preferences. *Journal of Political Economy 126*, 74–106.

Arcidiacono, P. (2005). Affirmative action in higher education: How do admission and financial aid rules affect future earnings? *Econometrica 73*(5), 1477–1524.

Armenter, R., M. Müller-Itten, and Z. R. Stangebye (2019). Rational Inattention and the Ignorance Equivalent.

Attanasio, O., S. Cattan, E. Fitzsimons, C. Meghir, and M. Rubio-Codina (2020). Estimating the production function for human capital: results from a randomized controlled trial in colombia. *American Economic Review 110*(1), 48–85.

Attanasio, O., C. Meghir, and E. Nix (2020). Human capital development and parental investment in india. *The Review of Economic Studies 87*(6), 2511–2541.

Baker, S., N. Bloom, and S. Davis (2016). Measuring Economic Policy Uncertainty. *The Quarterly Journal of Economics 131*(4), 1593–1636.

Banks, J., G. Batty, J. Breedvelt, K. Coughlin, R. Crawford, M. Marmot, J. Nazroo, Z. Oldfield, N. Steel, A. Steptoe, M. Wood, and P. Zaninotto (2021). English longitudinal study of ageing: Waves 0-9. 1998-2019 [data collection].

Barr, N. and P. Diamond (2009). Reforming pensions: Principles, analytical errors and policy directions. *International social security review 62*(2), 5–29.

Bartoš, V., M. Bauer, J. Chytilová, and F. Matějka (2016, jun). Attention Discrimination: Theory and Field Experiments With Monitoring Information Acquisition. *American Economic Review 106*(6), 1437–1475.

Becker, G. S., S. D. Kominers, K. M. Murphy, and J. L. Spenkuch (2018). A theory of intergenerational mobility. *Journal of Political Economy 126*(S1), S7–S25.

Behaghel, L. and D. Blau (2012). Framing Social Security Reform : Behavioral Responses to Changes in the Full Retirement Age. *American Economic Journal: Economic Policy 4*(4), 41–67.

Belfield, C., C. Crawford, E. Greaves, P. Gregg, L. Macmillan, et al. (2017). Intergenerational income persistence within families. Technical report.

Belley, P. and L. Lochner (2007). The changing role of family income and ability in determining educational achievement. *Journal of Human capital 1*(1), 37–89.

Bernheim, B. D. and D. Taubinsky (2018). Behavioral public economics. *Handbook of behavioral economics: Applications and Foundations 1 1*, 381–516.

Bernheim, D. (1988). *Social Security Benefits: An Empirical Study of Expectations and Realizations*.

Black, S. E., P. J. Devereux, F. Landaud, and K. G. Salvanes (2022). The (un) importance of inheritance. Technical report, National Bureau of Economic Research.

Blanden, J., M. Doepke, and J. Stuhler (2022). Educational inequality. Technical report, National Bureau of Economic Research.

Blandin, A. and C. Herrington (2018). Family structure, human capital investment, and aggregate college attainment.

Blundell, R., M. Costa Dias, C. Meghir, and J. Shaw (2016). Female labor supply, human capital, and welfare reform. *Econometrica 84*(5), 1705–1753.

Bolt, U., E. French, J. H. Maccuish, and C. O'Dea (2021). The intergenerational elasticity of earnings: Exploring the mechanisms.

Bommier, A., A. Kochov, and F. L. Grand (2017). On monotone recursive preferences. *Econometrica 85*, 1433–1466.

Borusyak, K., X. Jaravel, and J. Spiess (2021). Revisiting event study designs: Robust and efficient estimation. *arXiv preprint arXiv:2108.12419*.

Bozio, A., R. Crawford, and G. Tetlow (2010). The history of state pensions in the UK: 1948 to 2010. Technical report.

Browning, M., L. P. Hansen, and J. J. Heckman (1999). Micro data and general equilibrium models. *Handbook of macroeconomics 1*, 543–633.

Burtless, G. (1986). Social Security, Unanticipated Benefit Increases, and the Timing of Retirement. *Review of Economic Studies 53*(5), 781–805.

Cagetti, M. (2003). Wealth accumulation over the life cycle and precautionary savings. *Journal of Business & Economic Statistics 21*(3), 339–353.

Campbel, J. Y. (2003). Consumption-based asset pricing. *Handbook of the Economics of Finance*, 803–887.

Caplin, A. and M. Dean (2015). Revealed Preference, Rational Inattention, and Costly Information Acquisition. *American Economic Review 105*(7), 2183–2203.

Caplin, A., M. Dean, and J. Leahy (2017). Rationally Inattentive Behavior: Characterizing and Generalizing Shannon Entropy. *WP*, 1–45.

Caplin, A., M. Dean, and J. Leahy (2019). Rational Inattention, Optimal Consideration Sets, and Stochastic Choice. *Review of Economic Studies 86*(3), 1061–1094.

Caplin, A., M. Dean, and J. Leahy (2022). Rationally inattentive behavior: Characterizing and generalizing shannon entropy. *Journal of Political Economy 130*(6), 1676–1715.

Carhart-Harris, R. L., R. Leech, P. J. Hellyer, M. Shanahan, A. Feilding, E. Tagliazucchi, D. R. Chialvo, and D. Nutt (2014, feb). The Entropic Brain: a Theory of Conscious States Informed by Neuroimaging Research With Psychedelic Drugs. *Frontiers in Human Neuroscience 8*(20).

Carroll, C. and A. Samwick (1996). The Nature of Precautionary Wealth. *NBER Working Paper*.

Casanova, M. (2010). Happy Together : A Structural Model of Couples ' Joint Retirement Choices. *Working Paper* (5), 1–53.

Castaneda, A., J. Diaz-Gimenez, and J.-V. Rios-Rull (2003). Accounting for the u.s. earnings and wealth inequality. *Journal of Political Economy 111*(4), 818–857.

Caucutt, E. M. and L. Lochner (2020). Early and late human capital investments, borrowing constraints, and the family. *Journal of Political Economy 128*(3), 1065–1147.

Cesarini, D., E. Lindqvist, M. J. Notowidigdo, and R. Östling (2017). The effect of wealth on individual and household labor supply: evidence from swedish lotteries. *American Economic Review 107*(12), 3917–46.

Charles, K. K. and E. Hurst (2003). The correlation of wealth across generations. *Journal of political Economy 111*(6), 1155–1182.

Chetty, R. (2012). Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply. *Econometrica 80*(3), 969–1018.

Chetty, R. (2015). Behavioral economics and public policy: A pragmatic perspective. *American Economic Review 105*(5), 1–33.

Chetty, R., J. N. Friedman, S. Leth-Petersen, T. H. Nielsen, and T. Olsen (2014). Active vs. passive decisions and crowd-out in retirement savings accounts: Evidence from denmark. *The Quarterly Journal of Economics 129*(3), 1141–1219.

Chetty, R., N. Hendren, P. Kline, and E. Saez (2014). Where is the land of opportunity: The geography of intergenerational mobility in the united states. *Quarterly Journal of Economics 129*(4), 1553–1623.

Crawford, R. and C. O'Dea (2020). Household portfolios and financial preparedness for retirement. *Quantitative Economics 11*(2), 637–670.

Crawford, R. and G. Tetlow (2010). Employment, retirement and pensions. *Financial circumstances, health and well-being of the older population in England: ELSA 2008 (Wave 4)*, 11–75.

Cribb, J., C. Emmerson, and G. Tetlow (2013). Incentives, shocks or signals: labour supply effects of increasing the female state pension age in the UK. *IFS Working Paper W13/03*.

Cribb, J., C. Emmerson, and G. Tetlow (2016). Signals matter? Large retirement responses to limited financial incentives. *Labour Economics 42*, 203–212.

Cunha, F. and J. J. Heckman (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources 43*(4), 738–782.

Cunha, F., J. J. Heckman, L. Lochner, and D. V. Masterov (2006). Interpreting the evidence on life cycle skill formation. *Handbook of the Economics of Education 1*, 697–812.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica 78*(3), 883–931.

Daruich, D. (2018). The macroeconomic consequences of early childhood development policies. *FRB St. Louis Working Paper* (2018-29).

De Nardi, M. (2004). Wealth inequality and intergenerational links. *The Review of Economic Studies 71*(3), 743–768.

De Nardi, M., E. French, and J. B. Jones (2009). Life expectancy and old age savings. *American Economic Review 99*(2), 110–15.

De Nardi, M., E. French, and J. B. Jones (2010). Why Do the Elderly Save? The Role of Medical Expenses. *Journal of Political Economy 118*(1), 39–75.

Dearden, L., S. Machin, and H. Reed (1997). Intergenerational mobility in britain. *The Economic Journal*, 47–66.

Del Boca, D., C. Flinn, and M. Wiswall (2014). Household choices and child development. *The Review of Economic Studies 81*(1), 137.

Delaney, J. (2019). Born unequal? understanding the components of life time inequality. Technical report, ESRI working paper.

Dillenberger, D., D. Gottlieb, and P. Ortoleva (2020). Stochastic impatience and the separation of time and risk preferences.

Duffie, D. and K. J. Singleton (1993). Simulated moments estimation of markov models of asset prices. *Econometrica 61*(4), 929–952.

Dynarski, S. M. (2003). Does aid matter? measuring the effect of student aid on college attendance and completion. *American Economic Review 93*(1), 279–288.

Ehrlich, I. and G. S. Becker (1972). Market insurance, self-insurance, and self-protection. *Journal of Political Economy 80*(4), 623–648.

Epstein, L. G. and S. E. Zin (1989). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica 57*, 937–969.

Fosgerau, M., E. Melo, A. de Palma, and M. Shum (2020, nov). Discrete choice and rational inattention: a general equivalence result. *International Economic Review 61*(4), 1569–1589.

Frank, S. L. (2013, jul). Uncertainty Reduction as a Measure of Cognitive Load in Sentence Comprehension. *Topics in Cognitive Science 5*(3), 475–494.

French, E. (2005). The Effects of Health, Wealth, and Wages on Labour Supply and Retirement Behavior. *Review of Economic Studies 72*(2), 395–427.

French, E. and J. B. Jones (2011). The Effects of Health Insurance and Self-Insurance on Retirement Behavior. *Econometrica 79*(3), 693–732.

Fu, C., S. Ishimaru, and J. Kennan (2019). Government expenditure on the public education system. Technical report, National Bureau of Economic Research.

Fuster, A., R. Perez-Truglia, M. Wiederholt, and B. Zafar (2022). Expectations with endogenous information acquisition: An experimental investigation. *Review of Economics and Statistics 104*(5), 1059–1078.

Gabaix, X. (2014). A sparsity-based model of bounded rationality. *The Quarterly Journal of Economics 129*(4), 1661–1710.

Gabaix, X. (2019). Behavioral inattention. In *Handbook of Behavioral Economics: Applications and Foundations 1*, Volume 2, pp. 261–343. Elsevier.

Gallipoli, G., H. Low, and A. Mitra (2020). Consumption and income inequality across generations.

García, J. L., J. J. Heckman, D. E. Leaf, and M. J. Prados (2020). Quantifying the life-cycle benefits of an influential early-childhood program. *Journal of Political Economy 128*(7), 2502–2541.

Gayle, G.-L., L. Golan, and M. Soytas (2018). What is the source of the intergenerational correlation in earnings. Technical report, Mimeo.

Gruber, J., O. Kanninen, and T. Ravaska (2022). Relabeling, retirement and regret. *Journal of Public Economics 211*, 104677.

Guryan, J., E. Hurst, and M. Kearney (2008). Parental education and parental time with children. *Journal of Economic perspectives 22*(3), 23–46.

Gustman, A. and T. Steinmeier (2001). Imperfect Knowledge, Retirement amd Saving. *NBER Working Paper 3*(5).

Gustman, A. L. and T. L. Steinmeier (1986). A Structural Retirement Model. Technical Report 3.

Gustman, A. L. and T. L. Steinmeier (2005, feb). The social security early entitlement age in a structural model of retirement and wealth. *Journal of Public Economics 89*(2-3), 441–463.

Guvenen, F. (2009). An empirical investigation of labor income processes. *Review of Economic Dynamics 12*(1), 58–79.

Hansen, L. P. and T. J. Sargent (1995, 5). Discounted linear exponential quadratic gaussian control. *IEEE Transactions on Automatic Control 40*, 968–971.

Heckman, J. J. and S. Mosso (2014). The economics of human development and social mobility. *Annu. Rev. Econ. 6*(1), 689–733.

Hertz, T., T. Jayasundera, P. Piraino, S. Selcuk, N. Smith, and A. Verashchagina (2008). The inheritance of educational inequality: International comparisons and fifty-year trends. *The BE Journal of Economic Analysis & Policy 7*(2).

Huggett, M., G. Ventura, and A. Yaron (2011). Sources of lifetime inequality. *American Economic Review 101*(7), 2923–54.

Jordà, Ò., K. Knoll, D. Kuvshinov, M. Schularick, and A. M. Taylor (2019). The rate of return on everything, 1870–2015. *The Quarterly Journal of Economics 134*(3), 1225–1298.

Jung, J., J. H. J. Kim, F. Matějka, and C. Sims (2019, 11). Discrete actions in information-constrained decision problems. *Review of Economic Studies 86*, 2643–2667.

Kõszegi, B. and F. Matějka (2020, may). Choice Simplification: A Theory of Mental Budgeting and Naive Diversification. *The Quarterly Journal of Economics 135*(2), 1153–1207.

Kalyvas, C. and T. Tzouramanis (2017). A survey of skyline query processing. *arXiv preprint arXiv:1704.01788*.

Kocherlakota, N. R. (1996). The equity premium: It's still a puzzle. *Source: Journal of Economic Literature 34*, 42–71.

Koşar, G. and C. O'Dea (2022). Expectations Data in Structural Microeconomic Models. *National Bureau of Economic Research*.

Lalive, R., A. Magesan, and S. Staubli (2017). Raising the Full Retirement Age : Defaults vs Incentives. *NBER Working Paper*, 1–57.

Landais, C., J. Kolsrud, D. Reck, and J. Spinnewijn (2021). Retirement consumption and pension design.

Lee, S. Y. and A. Seshadri (2019). On the intergenerational transmission of economic status. *Journal of Political Economy 127*(2), 000–000.

Lockwood, B. (1991). Information externalities in the labour market and the duration of unemployment. *The Review of Economic Studies 58*(4), 733–753.

Low, H., C. Meghir, and L. Pistaferri (2010, sep). Wage risk and employment risk over the life cycle. *American Economic Review 100*(4), 1432–1467.

Lumsdaine, R. L., J. H. Stock, and D. A. Wise (1996). Why Are Retirement Rates So High at Age 65? In *Advances in the Economics of Aging*, pp. 61–82.

Luo, Y. (2008). Consumption dynamics under information processing constraints. *Review of Economic Dynamics 11*(2), 366–385.

Luo, Y. (2010, 10). Rational inattention, long-run consumption risk, and portfolio choice. *Review of Economic Dynamics 13*, 843–860.

Luo, Y. and E. R. Young (2016). Long-run consumption risk and asset allocation under recursive utility and rational inattention *. *Journal of Money, Credit and Banking 48*, 325–362.

Lusardi, A., P.-C. Michaud, and O. S. Mitchell (2015). Optimal financial knowledge and wealth inequiality. *Journal of Political Economy 125*, 1–48.

Luttmer, E. F. and A. Samwick (2018). The Welfare Cost of Perceived Policy Uncertainty: Evidence from Social Security. *American Economic Review 108*(2), 275–307.

Macaulay, A. (2021). Cyclical Attention to Saving *.

Maćkowiak, B. and M. Wiederholt (2009). Optimal sticky prices under rational inattention. *American Economic Review 99*(3), 769–803.

Maćkowiak, B. and M. Wiederholt (2015). Business cycle dynamics under rational inattention. *Review of Economic Studies 82*(4), 1502–1532.

Manoli, D. and A. Weber (2016, nov). Nonparametric evidence on the effects of financial incentives on retirement decisions. *American Economic Journal: Economic Policy 8*(4), 160–182.

Manski, C. (2004). Measuring expectations. *Econometrica 72*(5), 1329–1376.

Mastrobuoni, G. (2009). Labor supply effects of the recent social security benefit cuts: Empirical estimates using cohort discontinuities. *Journal of public Economics 93*(11-12), 1224–1233.

Matějka, F. and A. McKay (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review 105*(1), 272–298.

Mazumder, B. (2005). Fortunate sons: New estimates of intergenerational mobility in the united states using social security earnings data. *The Review of Economics and Statistics 87*(2), 235–255.

Mcculloch, W. S. and W. Pitts (1943). A logical calculus of the ideas immanent in nervous activity. Technical report.

Mehra, R. and E. C. Prescott (1985). The equity premium a puzzle. *Journal of Monetary Economics 15*, 145–161.

Newey, W. K. and D. L. McFadden (1994). Large sample estimation and hypothesis testing. In R. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics, Volume 4*. Elsevier, Amsterdam.

Nicoletti, C. and V. Tonei (2020). Do parental time investments react to changes in child's skills and health? *European Economic Review 127*, 103491.

O'Dea, C. (2018). Insurance , Efficiency and Design of Public Pensions . *WP* (1), 1–63.

OECD (2000). *Reforms for an Ageing Society*.

Pakes, A. and D. Pollard (1989). Simulation and the aysmptotics of optimization estimators. *Econometrica 57*(5), 1027–1057.

Porcher, C. (2020). Migration with Costly Information.

Powell, J. (1994). Estimation of semiparametric models. In R. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics, Volume 4*. Elsevier, Amsterdam.

Powell, M. J. (2009). The bobyqa algorithm for bound constrained optimization without derivatives. *Cambridge NA Report NA2009/06, University of Cambridge, Cambridge 26*.

Ravid, D. (2020, sep). Ultimatum Bargaining with Rational Inattention. *American Economic Review 110*(9), 2948–2963.

Rohwedder, S. and K. Kleinjans (2006). Dynamics of Individual Information about Social Security. *RAND WP*.

Rust, J. and C. Phelan (1997). How Social Security and Medicare Affect Retirement Behavior In a World of Incomplete Markets. *Econometrica 65*(4), 781.

Schittkowski, K. (2014). Nlpqlp-nonlinear programming with non-monotone and distributed line search.

Scholz, J. K., A. Seshadri, and S. Khitatrakun (2006). Are americans saving "optimally" for retirement? *Journal of political economy 114*(4), 607–643.

Seibold, A. (2021). Reference points for retirement behavior: Evidence from german pension discontinuities. *American Economic Review 11*(4), 1126–1165.

Sims, C. (2003). Implications of rational inattention. *Journal of Monetary Economics 50*(3), 665–690.

Solon, G. (1992). Intergenerational income mobility in the united states. *The American Economic Review*, 393–408.

Steiner, J., C. Stewart, and F. Matějka (2017). Rational Inattention Dynamics: Inertia and Delay in Decision-Making. *Econometrica 85*(2), 521–553.

Taubinsky, D. and A. Rees-Jones (2018). Attention variation and welfare: theory and evidence from a tax salience experiment. *The Review of Economic Studies 85*(4), 2462–2496.

Tauchen, G. (1986). Finite state markov-chain approximations to univariate and vector autoregressions. *Economics Letters 20*(2), 177–181.

van der Klaauw, W. and K. I. Wolpin (2008). Social security and the retirement and savings behavior of low-income households. *Journal of Econometrics 145*(1-2), 21–42.

Waters, T. (2017). Taxben: The ifs tax and benefit microsimulation model. *Institute for Fiscal Studies User Guide, https://www. ifs. org. uk/uploads/publications/docs/taxben_guide. pdf*.

Weil, P. (1990). Nonexpected utility in macroeconomics. *Source: The Quarterly Journal of Economics 105*, 29–42.

Wilcox, N. T. (2011, 5). 'stochastically more risk averse:' a contextual theory of stochastic discrete choice under risk. *Journal of Econometrics 162*, 89–104.

Yum, M. (2022). Parental time investment and intergenerational mobility. *International Economic Review*.