Costly Attention and Retirement *

Jamie Hentall MacCuish†

Job Market Paper

Click here for latest version

Abstract

Most people are mistaken about their pension provision implying significant informational frictions. This paper introduces such frictions, a cost of attention to an uncertain pension policy, into a life-cycle model of retirement. Resulting endogenous mistaken beliefs help explain why labour market exits concentrate at official retirement ages despite weak incentives to do so, and the UK female state pension age (SPA) reform provides policy variation. Solving this dynamic rational inattention model with endogenous heterogeneous beliefs represents a significant methodological contribution in itself. Costly attention improves model predicted employment response to the SPA whilst explaining patterns in the belief data. An extension addresses another puzzle, the low take-up of actuarially advantageous deferral options.

KEYWORDS: Rational inattention, Labour supply, Retirement, Pension provision, Learning

JEL CLASSIFICATION: D14, D83, D91, E21, J26, H55

---

*I thank Fabien Postel-Vinay and Eric French for detailed discussions and encouragement. I also thank James Banks, Richard Blundell, Uta Bolt, Jonathan Cribb, Martin Cripps, Aureo de Paula, Mariacristina De Nardi, Rory McGee, Cormac O'Dea, Ran Spiegler, Imran Rasul, Morten Ravn, and participants at a number of seminars for their helpful comments. At different times funding for this work has been received from Grant Inequality and the insurance value of transfers across the life cycle (ES/P001831/1) and ESRC studentship (ES/P000592/1), and this funding is gratefully acknowledged

†University College London and Institute for Fiscal Studies, correspondence to jamie.maccuish.15@ucl.ac.uk

1 Introduction

The ubiquity of mistaken beliefs obscures their deep incompatibility with standard complete information models, and mistaken beliefs about pensions are notably common. Most people are confused about pension policy. Yet these widespread incorrect beliefs about simple infrequently-changed financially important policies strongly indicate incomplete information resulting from informational frictions.

Ignoring informational frictions prevents us from understanding policy uncertainty's impact on people's decisions as we miss the subjective component of uncertainty. People are not only unsure how policy may change; often, they do not know current rules, as mistaken beliefs demonstrate. Conversely, mistaken beliefs are easier to rationalise if we simultaneously acknowledge government policy is objectively uncertain. Governments can and do reform policies making mistakes about them understandable. How retirement is impacted by this interplay between objective policy uncertainty and subjective mistaken beliefs is the focus of this paper.

Specifically, it asks whether costly attention and objective policy uncertainty solves the excess employment sensitivity puzzle whilst explaining observed mistaken beliefs. Many benefit systems offer very weak incentives to retire precisely at pension eligibility ages, yet exits from employment concentrate at them. This is the excess employment sensitivity puzzle, documented in multiple countries (e.g. Behaghel and Blau, 2012; Seibold, 2021). Mistaken beliefs increase the shock received upon reaching these ages because much pension policy uncertainty is resolved upon reaching eligibility; this in turn, may help explain the larger-than-expected employment reaction.

To investigate this, I use the recent reform to the UK female State Pension Age (SPA) to separately identify the SPA's effects on employment from ageing's. 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. So, the SPA provides a motive to retire only to the liquidity-constrained, the inability to borrow against pension income preventing them substituting intertemporally, but implausibly rich women exit employment at the SPA for this explanation. Costly attention offers a data-consistent barrier to intertemporal substitution as it penalises gathering the information needed to do so.

This paper first documents the pertinent facts, concerning mistaken beliefs and excess employment sensitivity, and then builds a model with informational frictions, in the form of costly attention, that accounts for these facts. This model incorporates costly attention, modelled using rational inattention (e.g. Sims, 2003), to an uncertain pension policy into a dynamic life-cycle model of retirement (e.g. French, 2005), thus allowing for endogenous mistaken beliefs that help explain retirement choices.

In incorporating rational inattention to an uncertain pension policy into a dynamic life-cycle model of retirement, this paper is the first, to the best of my knowledge, to solve a dynamic rational inattention model with endogenous heterogeneous beliefs. Endogenising beliefs proves key to explaining retirement choices and their relationship to mistaken beliefs but greatly complicates the model. Weaving together recent theoretical results (Steiner et al., 2017; Caplin and

Dean, 2015; Armenter et al., 2019) into a workable solution methodology for rich structural rational inattention models with endogenous beliefs is a contribution of this paper.

The English Longitudinal Study of Ageing (ELSA), a panel survey of older individuals, provides the data to study mistaken beliefs, self-reported and true SPAs both being observed and their impact on employment. Women subject to the reform are substantially mistaken about their SPA, most being off by over a year at age 58, and their mistakes are predictive of their employment response at the SPA. Moreover, the direction of this relationship implies a role for selection into SPA knowledge, hinting at endogenous learning.

I estimate the model using two-stage simulated method of moments targetting asset and labour supply profiles from ELSA and find policy uncertainty and costly attention significantly improve, compared to a complete information baseline, the model predicted employment response to the SPA. I use the SPA belief data to identify the cost of attention parameter, an empirical first in the rational inattention literature. Ignoring costly attention overstates the SPA's effectiveness as a policy tool to increase old age employment by up to 27%. Costly attention increases the employment response *at* the SPA by intertemporally shifting the employment response informed agents make over their life but decreases the overall employment response.

To investigate excess employment sensitivity, I abstract away from the benefit claiming decision to avoid another puzzle: 87% claim the state pension immediately upon reaching it despite an actuarially advantageous 10.4% p.a. adjustment from deferring. An extension addresses this deferral puzzle by introducing: a claiming decision, policy uncertainty over the adjustment from delayed claiming, and a cost of learning about this uncertainty adjustment. Together these create a new incentive to claim as claiming removes the attention cost; the result is an increase in the proportion claiming early helping to explain this deferral puzzle.

The rest of the paper is structured as follows. Section 2 reviews the literature. Section 3 outlines the institutional context and the data and carries out descriptive and reduced-form analysis to document the excess employment sensitivity puzzle and its relation to mistaken beliefs. Section 4 presents the model, starting with a standard model of complete information and then building in objective uncertainty about pension policy and a cost of attention to this uncertain policy. As solving the model presents novel difficulties and finding a solution method for dynamic rational inattention models with endogenous heterogeneous beliefs represents a contribution, section 5 discusses the solution of the model. Section 6 discusses estimation and section 7 model fit and implications. Section 8 presents the extension addressed at the deferral puzzle. Section 9 concludes.

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 expense 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 from this literature I include 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, that paper solving 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 endogenous heterogeneous beliefs that results 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 predicts 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 as 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. [1] 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 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 knowledge of current pension rules. Manski (2004) documents precisely one such study, finding 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. I find similar patterns to Crawford and Tetlow (2010) and Amin-Smith and Crawford (2018), prevalent mistaken beliefs predictive of labour supply, and also document 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, most papers using belief data to identify objects individuals hold private information about and maintain the assumption of complete information.

[1] Note that the insights from these models were not found to be incorrect. For example, medicare eligibility does seem to significantly impact employment. Rather the post-reform data did not support these models completely explaining the excess employment sensitivity puzzle.

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.

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. A reform to the female SPA is used to separate the effects of ageing from this response to the SPA, and it is detailed in section in 3.1 highlighting aspects that make it illuminating of this excess employment sensitivity puzzle. Section 3.2 discusses the data. Sections 3.3 - 3.4 provide descriptive and reduced form analysis, section 3.3 documenting the excess employment sensitivity puzzle, and section 3.4 documenting erroneous beliefs about the SPA as well as their relationship to employment sensitivity to SPA.

3.1 Institutional Context

The State Pension Age (SPA) is the earliest age at which retirement benefits, known as the state pension, can be claimed in the UK. In other words, it is the Early Retirement Age of the UK pension system. The UK does not have a Normal or Full Retirement age, so 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 implicit full retirement age. [2]

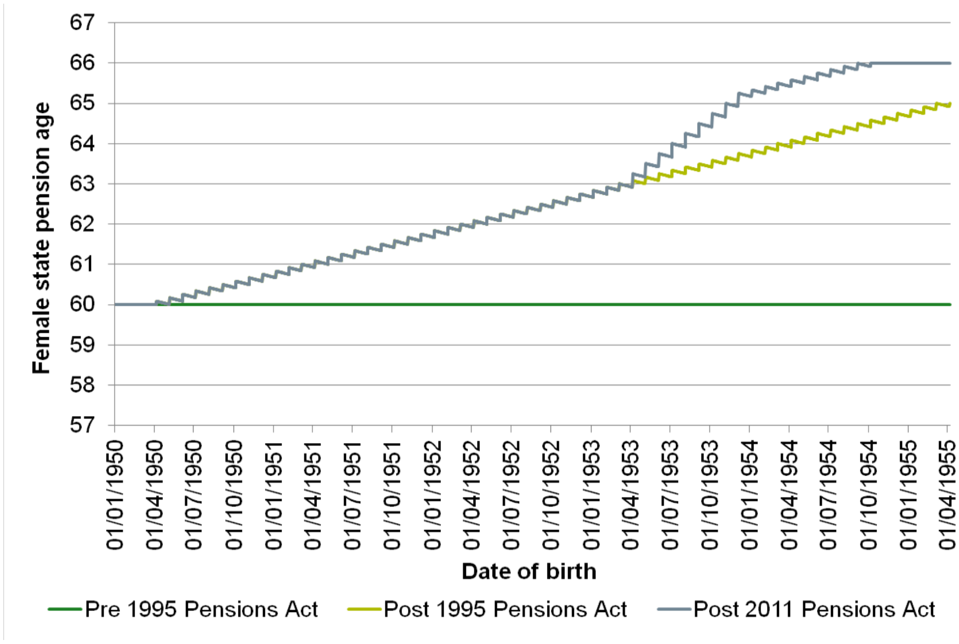
The UK State Pension came into force in 1948, with the SPA set at 65 for men and 60 for women. This remained unchanged until the Pensions Act 1995 legislated for the female SPA to gradually rise from 60 to 65, one month every two months, over the ten years from April 2010. The Pensions Act 2011 accelerated 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 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 retirement age are ruled out. Firstly, firms cannot force employees to retire solely based on age: this would be classed as age discrimination under UK law[3]. 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

[2]Despite an extremely generous actuarial adjustment, deferral was rare, leading to a deferral puzzle, discussion of which is deferred to an extension addressing this puzzle in section 8.

[3]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 A

Figure 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).

not provide major tax incentives to exit the labour market at the SPA. Unlike the US system, there is no earnings test[4], and although the state pension is taxable income, a component of income tax, called National Insurance contributions, is removed upon reaching the SPA[5].

These three facts imply the state pension is essentially an anticipatable increase in non-labour income with the SPA its eligibility age. The reform increased this eligibility age without changing the benefit level and, as such, is an anticipatable decrease in non-labour income. As the reform was announced in 1995 and began in 2010, this 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 [6].

[4]An earnings test is a feature of some social security systems that reduces benefits for those working whilst claiming retirement benefits. Those unfamiliar with it need not worry as it is not a feature of the UK system; its absence is only mentioned to reassure those familiar with systems, including an earnings test.

[5]Cribb et al. (2013) estimate changes to an individual's participation tax rate at SPA and find this does not predict the labour supply response to the SPA.

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

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 3.3 focuses on ruling out this explanation.

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[7] dataset that strikes the best balance along these two aspects, 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 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.

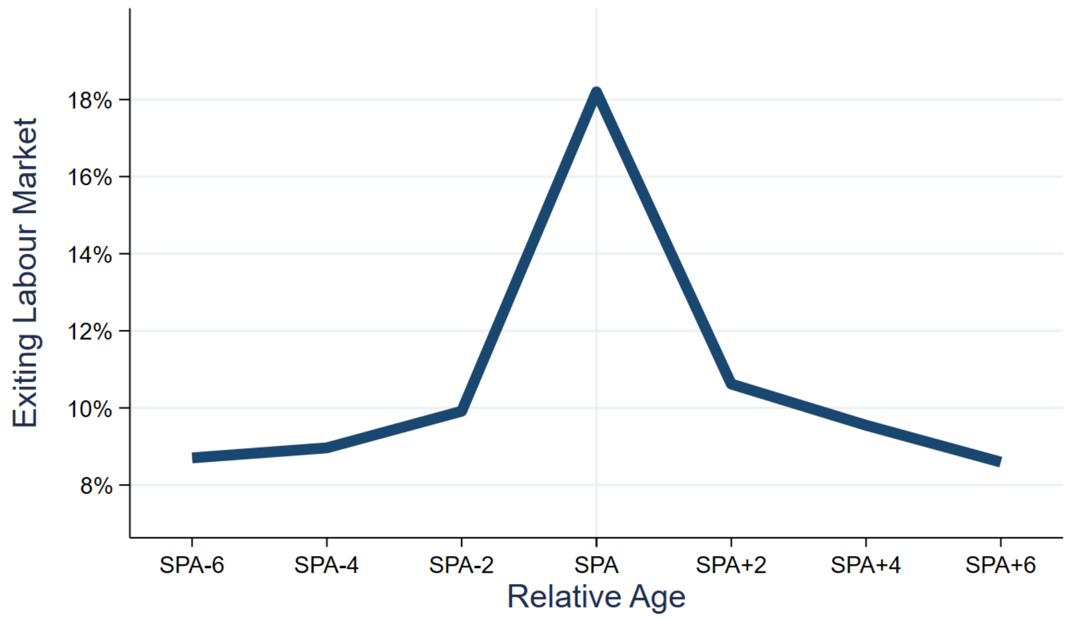
ELSA waves 1 (2002/03) through to 7 (2014/15) capture those affected by the 1995 pension age reform reaching SPA; hence I take the sample used for analysis and estimation from these waves. As this paper is concerned with the reform to the female SPA, males are dropped from the sample, except when estimating 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. Having earlier waves is important to control for pre-trends and increases power when estimating inputs to the structural model. 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 exogenous processes.

3.3 Excess Employment Sensitivity

Employment being more sensitive to benefit system retirement ages than implied by incentives is an empirical regularity documented in multiple countries (see section 2 for a discussion of the literature). This section presents evidence of this excess employment sensitivity to the UK SPA. As liquidity constraints are, in essence, the only standard complete information mechanism available to generate this sensitivity to the SPA (see section 3.1), particular attention is given to demonstrating that liquidity constraints alone cannot explain the puzzle.

[7] Technically, ELSA only covers England and Wales.

Figure 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.

Figure 2 captures the fundamentals of the excess employment sensitivity puzzle. It plots the pooled average fraction exiting employment at an age from the SPA. A large spike in exits at the SPA is observed. By adjusting the SPA at the monthly cohort level, the UK female SPA reform allows more careful separation of the employment response to the SPA from the effects of ageing than merely plotting pooled averages.

To do this, I build on the work of Cribb et al. (2016), who use this reform to identify the 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 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[8] leading to the following specification:

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

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

Table 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 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.

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. This treatment is administered to all, but variation in the duration of treatment is induced by the reform. I test the parallel trends assumption by interacting the fixed effects and the Wald test fails to reject the null that these interactions are zero (p = 0.95).

Column 1 of Table 1 presents the results of estimating equation 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[9].

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)[10]. This generates a cut-off of £29,000. I impose this restriction of being in a household with above £29,000 in 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. The objective of the median split 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 1 controls for quarterly age and cohort fixed effects, the control for an individual is someone born in the same year and quarter but a few months older so no longer under the SPA. Thanks to this narrow time window, it is easier to argue against liquidity constraints: households having more than £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 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

[9] I additionally test the parallel trends assumption with placebo test and these return null results see table 2.

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

1% level.

Column 3 of table 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 A contains a much more detailed empirical analysis. In it, I consider 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 liquid constraints to explain away the treatment effect. Appendix A also considers whether any of these factors, neglected for brevity in this section like health, private 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 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 average treatment effects in line with those estimated in this section. Hence, I conclude it is reasonable to give a causal interpretation to the treatment effects estimated in this section.

However, what follows does not rest on the causal nature of these estimates; I use these regressions as an untargeted auxiliary model to my structural life-cycle models. As such, what is important is the model's ability to replicate the key facts, not whether the treatment effects estimates replicated are unbiased causal estimates. Of course, what follows does depend on the reader finding these results puzzling, at least as far as standard complete information models are concerned. The results of the placebo test in table 2, add weight to the claim there is a puzzle. These show the results of including indicators of being one year over and one year under the SPA in equation 1, and as can be seen, unlike the indicator of being below the SPA, these coefficients are tiny and not significant at any reasonable level. Hence it seems that the results in this section are detecting something specific about the SPA, which is puzzling for those with significant liquid wealth.

3.4 Mistaken Beliefs and Employment Sensitivity

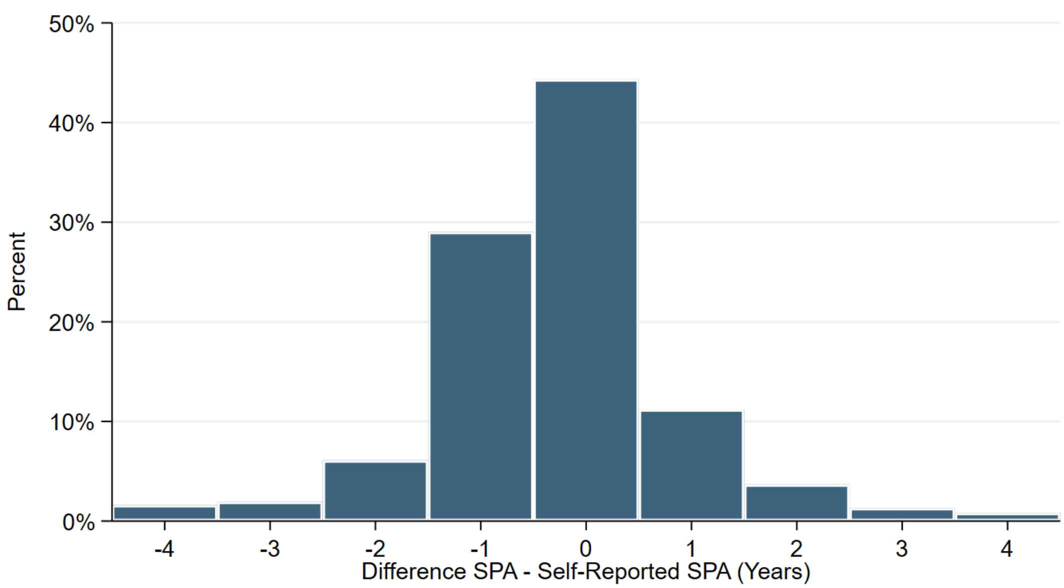
Mistaken beliefs about one's pension provision are so common that few find their existence surprising. Yet, they are difficult to reconcile with frictionless information, for surely this is a topic the individual is incentivised to know about. This section documents these mistaken beliefs, specifically mistakes about the SPA, and how they relate to the excess employment sensitivity documented in section 3.3.

Table 2: Placebo Tests

One Year Below SPA	**0.037**
s.e	(0.0321)
p=	.910
One Year Above SPA	**-0.003**
s.e	(0.0204)
p=	.898
Obs.	7,947
Indv.	3,889

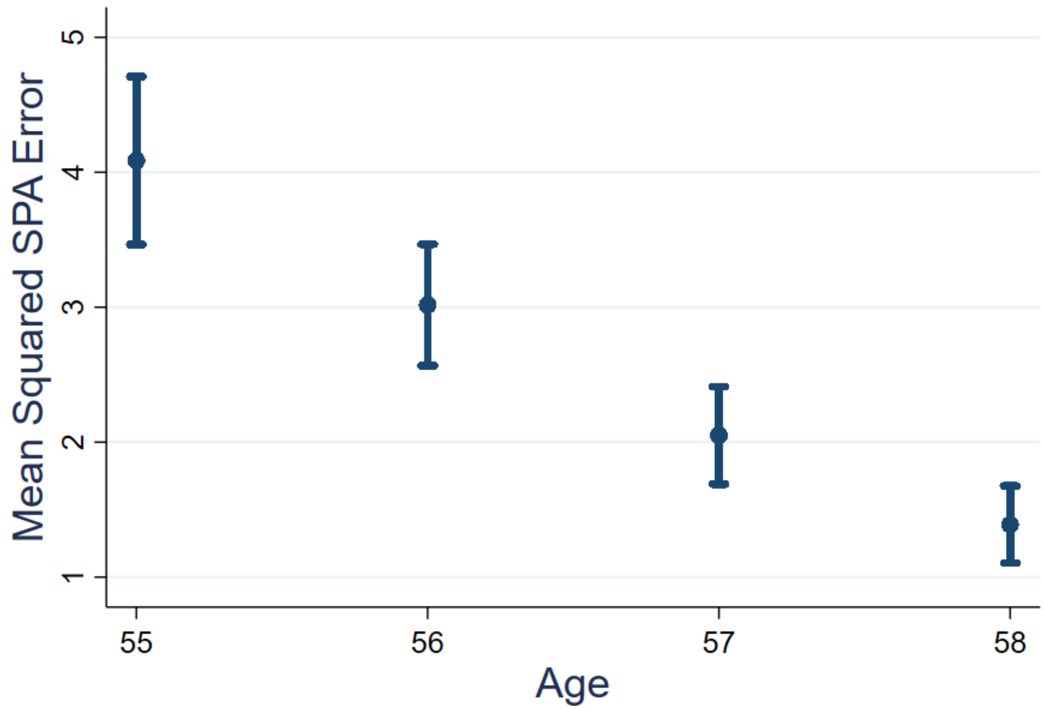
Notes: As a placebo test for a violated parallel trends assumption coefficient on the controls for being one year above and one year below SPA are shown. These are the coefficients from the baseline specification in column (1) of table 1

Figure 3: Mistaken SPA Beliefs of Women Subject to the Reform at Age 58



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

Figure 4: Mean Squared Error in Self-reported SPA



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

The SPA being such a simple aspect of the benefit system, confusion about it is both puzzling and simple to demonstrate. The SPA is a deterministic 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 less than perfect knowledge of one's SPA. Figure 3 shows this difference between the true and reported SPA of 58-year-old women subject to the reform. Although the largest group are those who know their SPA to within a year, this contains those mistaken by a margin of months and still leaves over 40% who are out by a year or more. Striking evidence of the prevalence of mistaken pension beliefs in the UK. Appendix A shows that self-reports cluster around the true SPA for each cohort; just the sort of pattern that emerges from a model of costly information acquisition. [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 4 that plots, against age, the mean squared error in self-reported SPAs. A clear declining age profile can be seen, indicating errors shrink as these women age towards their SPA. This declining mean squared error is the key moment used by the model to identify the cost of attention. This use of belief data to directly estimate the cost of attention is a novel contribution to the

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

Table 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 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 1 results from the question about SPA knowledge only being introduced in wave 3 and only being asked to individuals under 60.

rational inattention literature that adds empirical validity.

So, mistaken beliefs are a feature of reality, but if they were a feature unrelated to the excess sensitivity puzzle, then models attempting to explain this puzzle could safely ignore them. This is not, however, the case. Table 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 1 the size of the error in self-reported SPA in the last wave before reaching 60, after which this question is no longer asked, 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 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 4.

The excess employment sensitivity puzzle is only puzzling for standard models of complete information, deviating

[12] Appendix 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 predictive rather than another explanation in which beliefs are proxy for unobserved heterogeneity.

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 3.

In sum, mistaken beliefs about the SPA are prevalent amongst women subject to this reform, and mistaken beliefs are predictive of the size of the labour supply response to the SPA. So, they are not an empirical regularity we should ignore when trying to understand the excess employment sensitivity puzzle.

4 Model

This section presents the model: section 4.1 a baseline standard complete information model, capturing the relevant features of the UK retirement context, and section 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.

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 labour supply, consumption, 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 life time 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 bequest through a

[13] This is average yearly earnings, despite this to keep notation in line with the literautre I use the abreviation Average Indexed Monthly Earnings,

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 vectors of model state is $X_t = (a_t, w_t, AIME_t, ue_t, t)$[14] and below I detail how they impact the model.

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 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:

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

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 $\varepsilon_1 \sim N(0, \sigma_{\varepsilon,55}^2)$. The quadratic form of the deterministic component of wages captures the observed hump-shaped profile and is common in the literature.

which is the variable US social secuirty depends on.

[14] Types are sometimes included amongst the state variable. Here I exclude them on the technicality that they do not change and so are not needed to capture the state of the model. Hence, they are more accurately described as parameters.

The wage can be conceptualised as being equal to some underlying productivity that the women maintain during unemployment spells. Thus the unemployment status of the women ue_t evolves according to a conditional Markov process, where the probability of unemployment is dependent on current productivity w_t and the type. From age 80, the woman no longer has the choice of working; 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:

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

This specification average out and abstract 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 understand wider labour supply behaviours of older individuals (e.g. Casanova, 2010), is second order at best when considering labour supply responses to the SPA. Since spousal pension benefits are not modelled separately $y^{(k)}(t)$ amalgamates his labour and non-labour income into a single variable. 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 once she reaches the *SPA* and a private pension once she reaches the eligibility age $PPA^{(k)}$. This abstracts away from the benefit claiming decision for two reasons, both briefly 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 most of the period considered. This behaviour 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 8, but deferring it until then gives this baseline model a fair chance of addressing the excess sensitivity puzzle.

Average earning evolves until the woman reaches her private pension age $PPA^{(k)}$, at which point it is frozen. Both the state and private pensions are quadratic in average lifetime earnings $AIME_t$, until attaining their maximum, at which point they are capped. Until being capped, the pensions functions have the following forms

$$S^{(k)}(AIME_t) = sp_{k0} + sp_{k1}AIME_t - sp_{k2}AIME_t^2 \tag{4}$$

$$P^{(k)}(AIME_t) = pp_{k0} + pp_{k1}AIME_t - pp_{k2}AIME_t^2 \tag{5}$$

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 that depends on past as well as current regulations, which cannot be exactly captured without detailed administrative data (see Bozio et al., 2010, for details). 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 would have counted towards their state pension entitlement for some of the women in the sample. 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. Moreover, the model starts after defined benefit savings can be accessed without penalty.

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

$$Y^{(k)}(t,ue_t,AIME_t) = y^{(k)}(t) + b\mathbb{1}[ue_t = 1] + \mathbb{1}[t \geq SPA]S^{(k)}(AIME_t) + \mathbb{1}[t \geq PPA^{(k)}]P^{(k)}(AIME_t) \tag{6}$$

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

$$V_t^{(k)}(X_t) = \max_{c_t,l_t,a_{t+1}} \{u^{(k)}(c_t,l_t) + \beta(s_t^{(k)}(E[V_{t+1}^{(k)}(X_{t+1})|X_t] + (1-s_t^{(k)})T(a_{t+1}))\} \tag{7}$$

Subject to a budget constraint, a borrowing constraint, and a labour supply constraint:

$$c_t + (1+r)^{-1}a_{t+1} = a_t + w_t(1-l_t) + Y^{(k)}(t,ue_t,AIME_t) \tag{8}$$

$$a_{t+1} \geq 0 \tag{9}$$

$$ue_t(1-l_t) = 0 \tag{10}$$

4.2 Two Additions: Policy Uncertainty and Costly Attention

This section introduces two additions to the model of complete information presented above. Firstly, section 4.2.1 introduces objective policy uncertainty in the form of a stochastic SPA, capturing the observed variability of SPAs over the life-cycle resulting from pension reform. Secondly, section 4.2.2 introduces costly attention to this stochastic SPA, modelled with a disutility cost for more precise information. This allows the model to capture individual uncertainty about government policy in the form of incorrect beliefs about the SPA, and the implications of these beliefs for behaviour.

Since these two changes represent a novel approach section 4.2.3 rounds off with a discussion.

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 first entered the labour force.

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:

$$SPA_{t+1} = \min(SPA_t + e_t, 67) \tag{11}$$

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.

As modelling policy uncertainty in this way represents innovation, a word about interpretation is prudent. In the model, the variable SPA_t represents the current best available information about the age the women will reach the SPA, and as such, the data analogue is the SPA the government is currently announcing for the women'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.

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 observed selection into being informed about your SPA and the pattern of learning.

Directly observed vs learnable states: To make the exposition of this new feature of the model, 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

[15]This is the same collection of variable 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

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. That the other stochastic state variables, w_t and ue_t, are directly observed can be interpreted as these variables being more salient. I focus on costly attention related to the state pension policy rather than any of the other myriad burdens on people's attention because this is the uncertainty that is resolved upon reaching the SPA. Hence, it may help explain why people respond as they do to the SPA, the focus of this paper.

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 and then conditional on the signal draw choosing actions.[16] Provided they pay the utility cost of information, the choice of signal is completely 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 it's 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 signals draws observed z^t. Specifically, the posterior is formed as:

$$Pr(spa|z_t) = \frac{f_t(z_t|spa)\pi_t(spa)}{Pr(z_t)} \tag{12}$$

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 11, to this posterior.

Entropy and mutual information: Mutual information, rigorously defined below, is a concept from information theory. It is the expected reduction in uncertainty about one variable that results from learning another as measured by the entropy. Entropy is, in turn, a measure of uncertainty that captures the least space[17] needed to transmit or store the information contained in a random variable.

Definition 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 $H(X|Y) = E_Y[H(X|Y = y)]$.*

Definition 4.2 (Mutual Information)**.** *The mutual information between $X \sim P_X(x)$ and $Y \sim P_Y(y)$ is the expected reduction*

[16] This is not a substantive modelling assumption but simplifies the exposition. As the household is rationally inattentive, they are dynamically consistent, and they would not deviate from their choices given the opportunity at a later moment. This is analogous to agents committing to actions under complete markets or under certain contract setups.

[17] If the logarithm in the definition is taken with respect to base 2 then entropy measure this space in bits, but the base of the logarithm is not important as the change of base formula guarantees that changing the base will only change the unit of measure. One application that may help intuition is by using these concepts; a computer is able to compress a file.

in uncertainty about X from learning Y (equally about Y from learning X) as measured by the entropy: $I(X,Y) = H(X) - H(X|Y)$.

Utility: Hence utility incorporating information costs 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 constnat of proptionality λ is the cost of attention parameter, and given the above defintions 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. This means that the household always knows that if they are not in receipt of the woman's state pension benefits, she is below her SPA. This avoids any issue in the budget constraint with households not knowing the limits on what they can spend. That arriving at SPA_t in the model provides a positive informational shock reflecting 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: Bring this together the full set of states for the model is $(X_t, SPA_t, \underline{\pi_t}) = (a_t, w_t, AIME_t, ue_t, t, SP$ and the Bellman equation for the model is:

$$V_t^{(k)}(X_t, SPA_t, \underline{\pi_t}) = \max_{d_t, \underline{f_t}} E\Big[u^{(k)}(d_t, \underline{f_t}, \underline{\pi_t}) + \beta\big(s_t^{(k)} V_{t+1}^{(k)}(X_{t+1}, SPA_{t+1}, \underline{\pi_{t+1}}) + (1 - s_t^{(k)}) T(a_{t+1})\big)\Big] \tag{13}$$

subject to the same constraints 8 - 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, depend 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 5.

4.2.3 Discussion of Costly Attention to the Stochastic SPA

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

Expected Entropy Reduction Attention Cost: It is hopefully clear why a cost of information acquisition is included: to accommodate mistaken beliefs which predict employment responses to the SPA. The reasons for the functional form may be less clear. As utility costs of information are uncommon in the life-cycle literature, outlining the motivation for selecting this form may be instructive, and here I offer three motivating arguments.

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.[18] 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.

The second desirable feature is that it can endogenously generate certain rules-of-thumb or heuristics observed sufficiently often to be given names as a behavioural bias. We could treat these simplifying rules-of-thumb, or heuristics, as pre-ordained behavioural rules people blindly follow but, one, this does not explain why the particular rule-of-thumb and, two, it ignores the fact that people change rule-of-thumb as circumstances change. Models with hard-coded behavioural biases suppress a central insight of economics: that people respond to incentives. Endogenously replicating observed heuristics with a cost of attention avoids these pitfalls because, one, it explains why a given heuristics is used and, two, allows an agent to change heuristics in response to incentives. The first two examples come from Kõszegi and Matějka (2020) who show this cost of information can lead to both mental budgeting (total consumption on a category being fixed and composition responding to shocks) and naive diversification (within category consumption composition being fixed and total on categories responding) depending on the circumstance. Another example comes from Caplin et al. (2019) who show it can lead to consideration sets: ignoring many options to focus on a subset of alternatives.

Thirdly, strong a priori reasons to think that a cost of cognition should depend on entropy reduction 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 that 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 process 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 along this dimension is a reasonable approximation when information processing is our focus. [19]

[18] 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.

[19] 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

Interpreting the cost of attention: Costly information is modelled abstractly and so open to various interpretations but to guide the reader, I suggest a couple: 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 multifaceted, and people are confused about most of its 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 8 I look at an extension in which the household also learns about an uncertain actuarially 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. Hence, it captures more than just the hassle costs. 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 conceptualisation of the cost of information acquisition consistent with both data and model.

Interpreting the choice of signal: As our SPA is a number we can look up, this choice of a signal function may be difficult to conceptualise. The first thing to note is that looking up, perfectly remembering, and assimilating into one's action is not an information acquisition strategy that is excluded by the choice of a signal function conception: it corresponds to choosing a perfectly informative signal function. Carefully reading relevant regulations is not, in reality, the only way people learn about government policy in general or the state pension in particular. For example, people learn about how pension reforms affect them from other people and news outlets. In both examples, there is a random component, whether there is a newspaper story or other people talk about pensions, and a component that is a choice, whether you keep reading or ask follow-up questions. This is analogous to the choice of a noisy signal function in that it is partly a choice and partly stochastic, and so this choice captures much about the messy real-world learning process.

5 Model Solution

By introducing a high dimensional state $\underline{\pi_t}$ and a high dimensional choice $\underline{f_t}$, rational inattention has complicated the model to the extent that solving it represents a novel contribution. To achieve this I weave together recent theoretical results into a consistent solution methodology for dynamic rational inattention models with endogenous heterogeneous beliefs, like the one presented above. Section 5.2 explains how this is done, both to communicate the methodological

(for example Frank (2013) or Carhart-Harris et al. (2014)).

innovations and to give some intuition as to how the model is solved. First I provide some other details relevant to solving the model of this paper, but not relevant to solving generic dynamic models of rational inattention.

5.1 Details Specific to this Model

Not every period in the model with rational inattention is complicated by the presence of $\underline{\pi_t}$ and $\underline{f_t}$; only before the realisation of the SPA do they matter. Upon reaching the SPA the true value is revealed and so beliefs ($\underline{\pi_t}$) and learning ($\underline{f_t}$) about the SPA are not relevant. Periods after this can be solved, like the baseline and the model with only policy uncertainty, using standard solution methods. That is using dynamic programming, specifically backward induction where the within period utility maximization problem is solved as a discrete choice problem using search to find the optimal action.

It is instructive to work through the transition from simple post-SPA periods to the complicated pre-SPA. We can solve the model with rational inattention by standard backward induction until we hit age 66. We can proceed as far as age 67 using standard methods because, as the state pension age process is bounded above, the woman receives her state pension with probability 1 from that point on. At age 66, because she knows the underlying data generating process just not the current value of SPA_t, if she is not in receipt of her state pension she knows her SPA is 67 with certainty. So at 66, SPA_t becomes a state variable, because whether or not the woman receives her state pension affects utility, but $\underline{\pi_t}$ is still not relevant, because she is perfectly informed. At age 65 whether or not $\underline{\pi_t}$ is a state depends on the value of SPA_t, and the same is true for all periods 60-65. If $t \geq SPA_t$ then the woman receives the state pension benefit; she knows the value of SPA_t, so $\underline{\pi_t}$ is degenerate and not a state, and she does not need to make an information acquisition choice. Hence, rational inattention is not relevant if $t \geq SPA_t$ and the period can be solved by simply searching for the optimal choice. For ages 55-59, rational inattention is always relevant as SPA_t is always above 60. Hence, because age 65 is the first period for which, when $t < SPA_t$, the true value of the SPA_t cannot simply be inferred (as it could be 66 or 67), age 65 is the first period in which the information acquisition choice is non-trivial and beliefs matter.

The solution of within period problems, when rational inattention matters, because $t < SPA_t$, is outlined immediately below in section 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.

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 $\underline{\pi_t}$ and a high dimensional choice $\underline{f_t}$. This section presents my solution methodology. I use the model of retirement decision, presented earlier, to explain the methodology, but it has wider applicability: 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 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 nigh on 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, and 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. That implies that 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 5.2.1 gives an outline of the proof of the main results from Steiner et al. (2017). Then section 5.2.2 will take the results from section 5.2.1 and present my solution method.

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] Merely citing their result would not provide any intuition. For this reason, and because an understanding of these results is needed to understand the solution methodology, in this section, I present an outline of their proof (see appendix B or the original paper for details).

Steiner et al. (2017) extend Matějka and McKay (2015)[21] to a dynamic setting and most of what is explained here applies equally to static problems. I explain what is relevant from Steiner et al. (2017) to my model using my model as a lens through which to explain their results.

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

$$\overline{V}_{t+1}^{(k)}(d_t, X_t, SPA_t, \underline{\pi_t}) = E\left[s_t^{(k)} V_{t+1}^{(k)}(X_{t+1}, SPA_{t+1}, \underline{\pi_{t+1}}) + (1 - s_t^{(k)}) T(a_{t+1}) \middle| d_t, X_t, SPA_t, \underline{\pi_t}\right] \tag{14}$$

[20] My framework is not quite a direct application of Steiner et al. (2017) but represents a slight extension. Their paper does not allow for endogenous states whilst my model has endogenous states; however, since these endogenous states are observed without information friction and independent of SPA_t this does not violate their key assumptions that actions do not affect the distribution of future unobserved states. For completeness, I present the details of my extension to Steiner et al. (2017) in appendix B, and replicate all proofs I rely on from their paper in my framework. This extension presents a framework that nests the original work and also covers the model in this paper.

[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 taken over the uncertainty in X_{t+1} and SPA_{t+1} and section 5.2.2 explains how to solve for $\underline{\pi_{t+1}}$, then the Bellman equation 13 simplifies to:

$$V_t^{(k)}(X_t, SPA_t, \underline{\pi_t}) = \max_{d_t, \underline{f_t}} E\left[u^{(k)}(d_t, \underline{f_t}, \underline{\pi_t}) + \beta \overline{V}_{t+1}^{(k)}(d_t, X_t, SPA_t, \underline{\pi_t})\right] \tag{15}$$

Steiner et al. (2017) show that the solution to this model has action 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:[22]

$$p_t(d|spa) = \frac{\exp\left(n^{(k)} \frac{((c/n^{(k)})^{\nu} l^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \log(q_t(d)) + \beta \overline{V}_{t+1}^{(k)}(d, X_t, SPA_t, \underline{\pi_t})\right)\Big)}{\sum_{d' \in \mathscr{C}} \exp\left(n^{(k)} \frac{((c'/n^{(k)})^{\nu} l'^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \log(q_t(d')) + \beta \overline{V}_{t+1}^{(k)}(d', X_t, SPA_t, \underline{\pi_t})\right)\Big)} \tag{16}$$

$$\max_{\underline{q_t}} \sum_{spa} \pi_t(spa) \log\left(\sum_{d' \in \mathscr{C}} q_t(d) \exp\left(n^{(k)} \frac{((c/n^{(k)})^{\nu} l^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \beta \overline{V}_{t+1}^{(k)}(d, X_t, SPA_t, \underline{\pi_t})\right)\right) \tag{17}$$

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 which makes 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. In fact, 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 name of the signal seen, re-labelling with the name of choice they should make gives $\underline{f_t}$. 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. From this, it follows that we can re-write the agent's decision problem as:

$$V_t^{(k)}(X_t, SPA_t, \underline{\pi_t}) = \max_{\underline{p_t}} E\left[n^{(k)} \frac{((c/n^{(k)})^{\nu} l^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} - I(\underline{p_t}; \underline{\pi_t}) + \beta \overline{V}_{t+1}^{(k)}(d, X_t, SPA_t, \underline{\pi_t})\right] \tag{18}$$

[22] This is the logit-like result originally from Matějka and McKay (2015) and extended to the dynamic case by Steiner et al. (2017). For a discussion of its advantages vis-a-vis a traditional logit arising from utility shocks, plus a rigorous proof, I direct the reader to the original Matějka and McKay (2015) paper.

As the problem is treated as discrete choice there exists some 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:

$$\max_{\underline{p_t}} \sum_{spa} \pi_t(spa) \sum_{i=1}^{N} p_t(d_i|spa) \Big(n^{(k)} \frac{((c_i/n^{(k)})^{\nu} l_i^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} - I(\underline{p_t}; \underline{\pi_t}) + \beta \overline{V}_{t+1}^{(k)}(d_i, X_t, SPA_t, \underline{\pi_t}) \Big) \tag{19}$$

and from the symmetry of mutual information: [23]

$$I(\underline{p_t}; \underline{\pi_t}) = \sum_{spa} \pi_t(spa) \left(\sum_{d} p_t(d|spa) \log(p_t(d|spa)) \right) - \sum_{d} q_t(d) \log(q_t(d)) \tag{20}$$

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

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

Substituting 20 into 19, rearranging, and collapsing the repeated sums gives:

$$\max_{\underline{p_t}} \sum_{spa} \pi_t(spa) \sum_{i=1}^{N} \Big(n^{(k)} \frac{((c_i/n^{(k)})^{\nu} l_i^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \log(q_t(d_i)) - \log(p_t(d_i|spa_i)) + \beta \overline{V}_{t+1}^{(k)}(d_i, X_t, SPA_t, \underline{\pi_t}) \Big) \tag{21}$$

Taking $\underline{q_t}$ as given, optimality with respect to any $p_t(d|spa)$ requires the following FOC, derived from differentiating 21, be satisfied [24]

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

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:

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

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

[23] We have been thinking of mutual information as the expected reduction in entropy about the state of the world from learning the signal, or equivalently, what action to take. However, that is mathematically equivalent to the expected reduction in entropy about the action from learning the state of the world, which is what is expressed above.

[24] The eagle-eyed reader may have noted that this treats the continuation value as fixed. Showing that "one can ignore the dependence of continuation values on beliefs and treat them simply as functions of histories" was a major achievement of Steiner et al. (2017) that I abstract from here to explain the intuition behind the results. I will touch again on this point briefly in section 5.2.2, but for a proper treatment please refer to the original paper.

nuisance terms which gives the solution for $\underline{p_t}$:

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

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 16 into 21 and noting that the logarithm of the numerator in 16 cancels all other terms in 21 leaving only the summation from the denominator. So $\underline{q_t}$ can be found by solving:

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

5.2.2 Solution Method

Being the first to solve a dynamic rational inattention model with endogenous heterogeneous beliefs, this paper requires a new solution methodology. At its core the solution methodology is to solve 17 for $\underline{q_t}$ and subsitute the solution into 16 to solve for $\underline{p_t}$. This basic description conceals two major hurdles, and this section explains how they were overcome leading up to a description of the algorithm used.

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 16 and 17 into the Bayesian updating formula 12 gives

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

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 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 to the model in order to know next period's beliefs given the chosen action and hence know the effective conditional continuation values:

$$\overline{V}_{t+1}^{(k)}(d_t, X_t, SPA_t, \underline{\pi_t}) = E\left[s_t^{(k)}V_{t+1}^{(k)}(X_{t+1}, SPA_{t+1}, \underline{\pi_{t+1}}) + (1-s_t^{(k)})T(a_{t+1})\middle| d_t, X_t, SPA_t, \underline{\pi_t}\right] \tag{22}$$

Steiner et al. (2017) dodge 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.[25] 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. [26] Redundant information in the state space is problematic because the curse of dimensionality means this is often the binding constraints in 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 17 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. [27]

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 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 17, but moreover, if after removing the actions that will never be taken 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 realisation of the SPA. Hence, all actions that are strictly dominated across all realisation 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 realisation 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 actions that are strictly dominated only takes into account the ordinal characteristics of utility and not the cardinal aspect of inter-personal expected utility. Using the necessary and sufficient condition from Caplin et al. (2019),

[25] This allowed them to show we can ignore the dependence of continuation values on beliefs, because "the solution can be interpreted as an equilibrium of a common interest game played by multiple players. The player in each period observes the history but not the choice rule used in the past. In equilibrium, each player forms beliefs according to the others' equilibrium strategies."

[26] In the original paper past actions mattered not only because they impacted beliefs but the authors' allowed the possibility of past action impacting current utility. This creates a potential reason why two histories leading to the same belief might represent different states in the original paper. This is not a possibility in this paper.

[27] Although, I have not proved this is a contraction mapping the fixed point iteration was always found to converge and generally in relatively few iterations.

it is easily shown that if a there exists a decision $d^\star = (c^\star, l^\star)$ which satisfies

$$\sum_{spa} \pi_t(spa) \frac{\exp\left(n^{(k)} \frac{((c^\star/n^{(k)})^\nu l^{\star 1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \beta \overline{V}^{(k)}_{t+1}(d^\star, X_t, spa, \underline{\pi_t})\right)}{\exp\left(n^{(k)} \frac{((c/n^{(k)})^\nu l^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \beta \overline{V}^{(k)}_{t+1}(d, X_t, spa, \underline{\pi_t})\right)} \geq 1 \tag{23}$$

for all other decisions $d = (c, l)$ then it is the only action taken $q(d^\star) = 1$. [28] Unlike dropping strictly dominated alternative, which reduces the dimensionality and so makes solving equation 17 easier, checking equation 23 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 23 depends on the problem faced and how frequently it shows 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 17, 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 a action d that satisfies 23 **then**

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

            else

                Solve 17 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

[28] For the reader who does not want to reference Caplin et al. (2019) equation 23 can be derived from the boundary condition in equation 17 and this is done in appendix C

end while

end if

Substitute $\underline{q_t}$ into 16 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 C.

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.

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 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 as data generation process:

$$\log(w_{j,t}) = \delta_{k0} + \delta_{k1}t + \delta_{k2}t^2 + \varepsilon_{j,t} + \mu_{j,t} \tag{24}$$

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.

Unemployment Transition Matrix I estimate type-specific transition probabilities in and out of unemployment using self-declared employment status: the transition probabilities between employed and unemployed.

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 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 D.

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 D.

Table 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 targetting the pre-reform labour supply and assets profiles.

Due to the novel nature of the cost of attention parameter in this 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 targetting 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 targetted. This comes at the cost of not using all information to identify λ.

7 Results

Section 7.1 presents the goodness of fit to targetted moments and the model's ability to replicate the key empirical facts regarding excess employment sensitivity, mistaken beliefs and the relationship between them. Section 7.2 presents implications of the model about patterns of information acquisition and the welfare cost of costly attention. Section 7.3 concludes with model policy predictions.

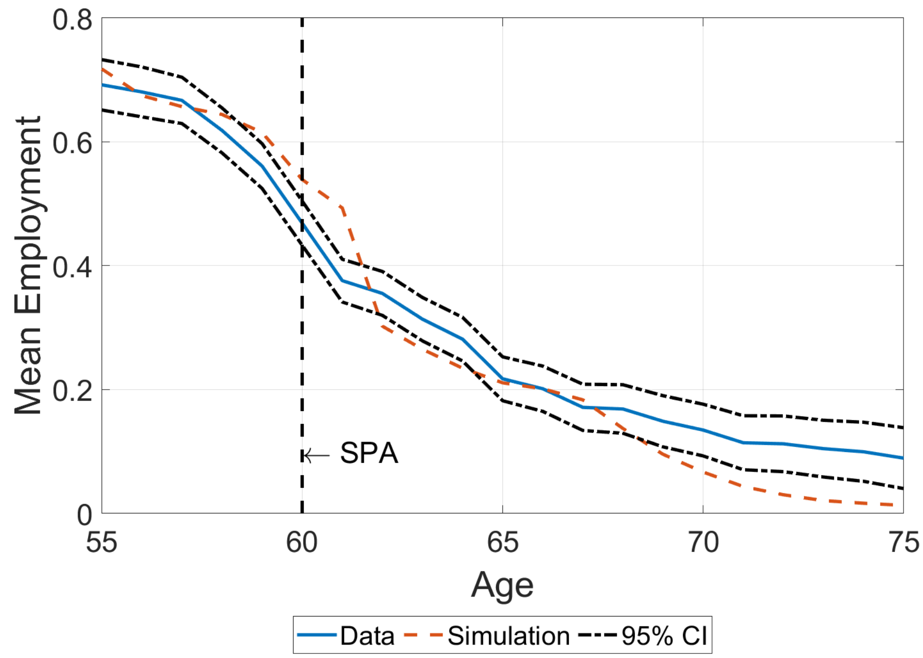
7.1 Model Evaluation

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 can be found in appendix E.1.

Figures 5 and 6 shows the baseline model fit to pre-reform employment and asset profile with $SPA = 60$. Table 4 contains the estimated parameter. The graphs for the versions with policy uncertainty and policy uncertainty and rational inattention can be found in appendix D but are practically indistinguishable. In their response to the dynamic SPA as it changes, however, these model versions clearly distinguish themselves.

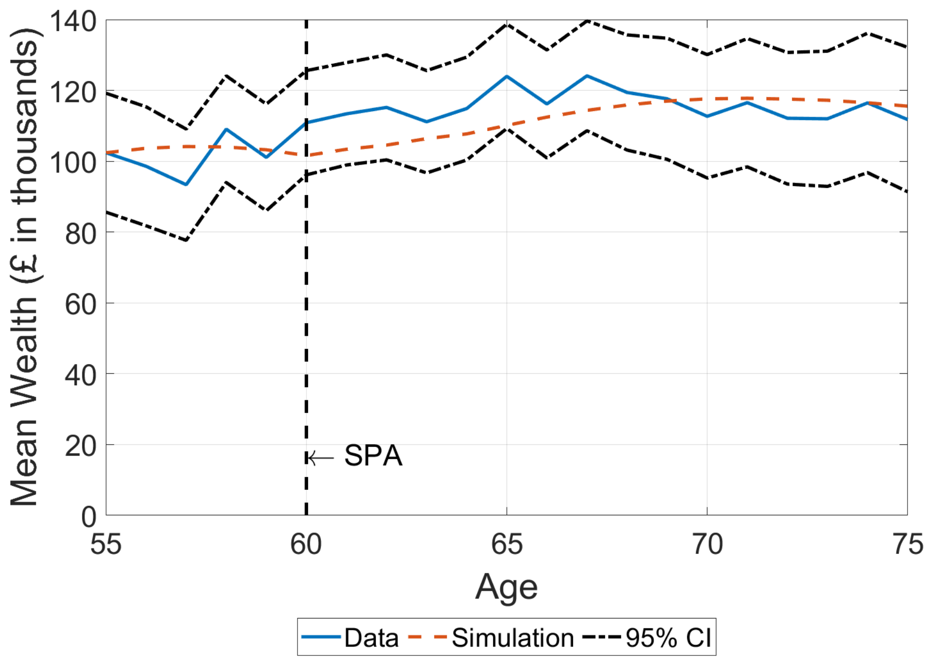
To investigate this response to the SPA, I simulate the model with $SPA = 60$, $SPA = 61$, and $SPA = 62$, as these are the SPAs reached in ELSA waves 1-7. Then, I repeated the regression analysis from section 3.3 on the simulated data using an adaptation of equation 1 to the limitations of the model. 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

Figure 5: Employment Profile



Notes: Model fit to targetted 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 6: Asset Profile



Notes: Model fit to targetted 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 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	0.018	0.023	0.052	0.038	0.080
[95% C.I]					[0.044,.0116]
Assets >Median(£29,000)	0.026	0.029	0.052	0.031	0.061
[95% C.I]					[0.018,.0103]
	Reduction in MSE of SPA Self-Reports				
MSE Reduction 55-58	-	-	-0.85	1.53	1.69
[95% C.I]					[0.31,3.36]
Coefficient	Treatment Effect Heterogeneity by SPA Error				
Interaction	-	-	-0.001	-0.022	-0.066
[95% C.I]					[-0.094,-034]

Notes: The columns show results from three separate costs of attention. The top panel shows labour supply response across the wealth distribution as per table 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.

the treatment indicator, a full set of age and cohort fixed effects (no date here as date is not distinct from age in the model) and the controls from the empirical specification having counter-parts in the model. As in section 3.3, I then repeated this on the subsample with above median empirical assets (£29,000) in the period before their SPA. The top panel of table 5 contains the results, column 5 replicating the empirical estimates found in columns table 1 and 2 of table 1 . 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 motivates policy uncertainty and costly attention, and to see their impacts separately, I introduce them sequentially. Column 2 introduces policy uncertainty, and 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 closer to the lower treatment effect in the richer subpopulation. Section 3.3 shows how the treatment effect in this richer subpopulation is most puzzling ex-ante, and appendix C shows if we directly target these treatment effects, the baseline can match that of the whole population by pushing borrowing constraints to extremes, but not of the subpopulation. With these parameters estimated with static profiles, however, it struggles most with the aggregate.

Column 3 introduces costly attention to the stochastic SPA, with a cost of attention of $\lambda = 3 \times 10^{-6}$. The treatment effect in both the whole population and those with above median assets move significantly toward the data falling in the confidence intervals.

SPA self-reports in ELSA, however, 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, as this is the cohort simulated during estimation. The middle panel of table 5 shows these numbers. Column 4 shows that for a 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 mistaken beliefs predict the labour supply response to the SPA was an impetus to investigating the role of informational frictions in this excess sensitivity puzzle; a natural question is whether the model can replicate this relationship. Table 3 in section 3 shows those who are better informed of their SPA in their late 50s have a smaller labour supply response upon reaching their SPA in their 60s. Two countervailing forces exist in the model linking the degree of SPA knowledge to the labour supply response to the SPA. Firstly, SPA knowledge is endogenous, implying those whose actions depend least on the SPA acquire the least information about it. This mechanism pushes in the direction of the empirical finding of a negative relationship between SPA knowledge and labour supply response. Conversely, if we compare to ex-ante equivalent households where, by luck of the draw, one ended up worse informed than the other, then the worse informed household will receive a larger shock upon discovering their SPA and so have a larger reaction. Hence, whether the model generates a positive or negative relationship between the degree of SPA knowledge and the labour supply response to the SPA depends on which dominates. The bottom panel of table 5 shows this number. The first column in the bottom panel of table 5 shows that for the values of λ considered, we get a slightly negative relationship or no relationship.

The fact that the model replicates the key facts from the data indicates that it has the key mechanism required to explain the data. A limitation is that it requires different values of λ to replicate different facts. This indicates the level of some incentives may be misaligned. Unidimensional policy uncertainty over the SPA is a massive simplification: in reality, many more aspects of pension policy are uncertain. As such introducing uncertainty and learning about another dimension of the state pension system seems a natural next step to make progress on this limitation; section 8 takes up this challenge.

7.2 Model Implications

Size of informational frictions λ is not an easily having 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. Hence, to make λ interpretable, I first convert to consumption using marginal consumption expressing λ in consumption per bit. Expressing λ in this way exaggerates the cost of attention because uncertainty and opportunities to learn are more limited in this, and all, models than in reality,

Table 6: Regression Analysis of the Determinants of Learning

Intercept	0.18
Assets	-2.898e-07
AIME	2.826e-06
Labour Income	9.655e-08
Age = 56	-6.458e-02
Age = 57	-7.781e-02
Age = 58	-9.796e-02
Age = 59	-8.282e-02

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

and so each bit represent a greater proportion of total learnable uncertainty. To account for the limited total amount of information, I express λ as the utility cost of going from completely uninformed to perfectly informed about the current SPA_t.

Table 7 shows summary statistics of the distribution of consumption equivalent attention costs for the whole population and by education level for both values of λ, $\lambda = 1.3 \times 10^{-7}$ and $\lambda = 3 \times 10^{-6}$. For $\lambda = 1.3 \times 10^{-7}$, the consumption costs are low, especially given these numbers represent the maximum cost of removing uncertainty (they are the cost of going from completely uninformed to completely informed). However, there is substantial heterogeneity arising from the differing levels of wealth of the households. Higher education households have higher attention costs in consumption equivalent units because, as they are richer and have lower marginal utility, the same utility attention cost needs more consumption units to equate to it. For $\lambda = 3 \times 10^{-6}$, a similar qualitatively picture arises, but the costs are correspondingly higher.

This gives some idea of the potential welfare gains available from reducing this informational friction, for example, by sending more 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 and material and cannot exceed £1.

Who learns and when The model allows us to investigate just what determines who learns and when. Table 6 presents the partially correlates from a pooled regression of the number of bits of information acquired on the states of the model and indicators of age. This is done over the age range 55-59 for a cohort with a SPA of 60.

That the total available information in the model is so small makes interpreting magnitudes difficult, but the sign of the regression coefficients convey information. On average, a positive amount of information is gathered, those with more alternative resources (assets and labour income) gather less information, and those who have higher future pensions (AIME) gather more. On average most information is acquired at age 55, declining until age 58 and then jumping back up at age 59, immediately before the SPA.

Table 7: Summary Statistics of Attention Cost Converted to Consumption (£)

λ	Education Groups	Mean	SD	Median	5% Percentile	95% Percentile
1.3×10^{-7}	Low	29	44	22	8	66
	High	72	241	41	12	179
	...both	52	183	31	10	134
3×10^{-6}	Low	654	1025	495	158	1486
	High	1617	5314	952	268	4131
	...both	1189	4047	692	184	3796

Notes: Distribution of consumption equivalent amounts to utility cost of attention for two different costs of attention.

7.3 Model Predictions

7.3.1 Optimal Savings

Retirement preparedness is a concern, particularly in countries in which private savings accounts form a growing fraction of retirement income. The academic literature on whether households under or over save for retirement is inconclusive (e.g. Scholz et al., 2006; Crawford and O'Dea, 2020). As this paper takes the age 55 asset distribution as given, its ability to answer this question is limited. It does, however, offer novel predictions about the impact of informational frictions on agents' ability to prepare for retirement. To investigate this, I compare the savings of the rationally inattentive household to the friction-free benchmark (i.e. the model with only policy uncertainty).

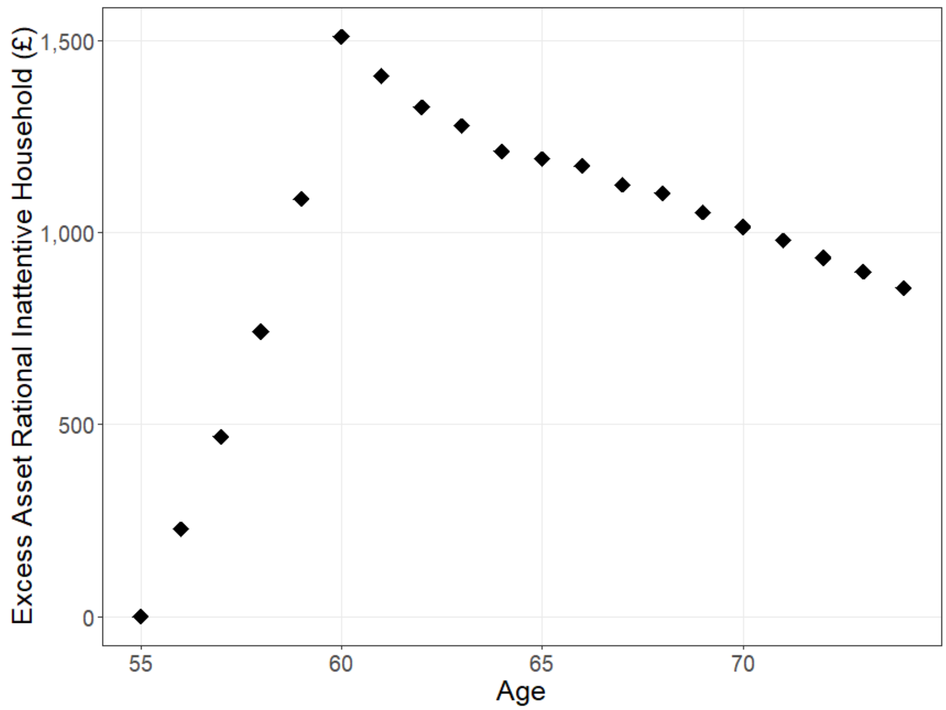
Figure 7 shows this excess saving when $\lambda = 3 \times 10^{-6}$, and a clear pattern of excess saving for retirement can be seen. As the cost of attention is so high rather than learning, the SPA households insure against policy uncertainty. For these simulations, the SPA was kept constant at 60 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 8 show this excess saving when $\lambda = 1.3 \times 10^{-7}$. The mistakes are, unsurprisingly, much smaller, and there is no longer such a clear pattern. Averaging across households, they slightly undersave for retirement and then increases their assets later. This additional saving later in life results from lower accumulated lifetime earnings meaning lower pension income.

7.3.2 Increasing Old Age Participation with the Pension Ages

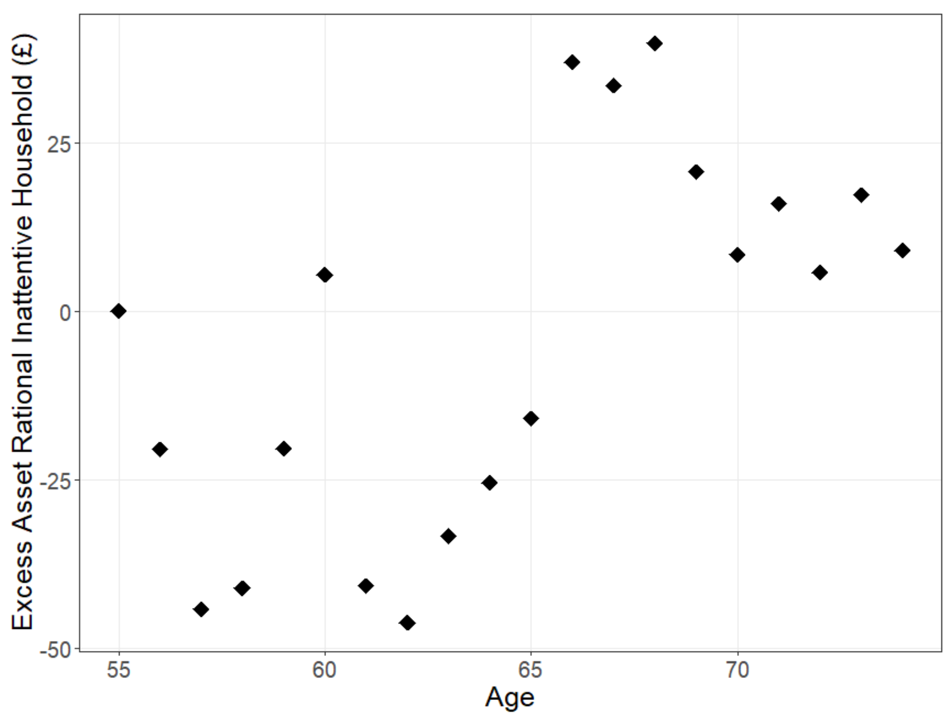
The rising old-age dependency ratio due to ageing populations have made increasing the labour force participation of older individuals a policy objective of governments around the world (OECD, 2000), and statutory retirement ages are often seen as a key tool to achieve this goal. Rational inattention increased the responsiveness of employment at the SPA, so it would be natural to conclude it makes the SPA a more powerful lever to address this issue. This is not necessarily

Figure 7: Excess Saving when $\lambda = 3 \times 10^{-6}$



Notes: Excess saving relative to model with only policy uncertainty when $\lambda = 3 \times 10^{-6}$ and the true SPA was 60 throughout the simulations.

Figure 8: Excess Saving when $\lambda = 1.3 \times 10^{-7}$



Notes: Excess saving relative to model with only policy uncertainty when $\lambda = 1.3 \times 10^{-7}$ and the true SPA was 60 throughout the simulations.

the case.

Figure 9 which plots, against age, the mean employment change resulting from increasing the SPA from 60 to 66 in the model with $\lambda = 1.3 \times 10^{-7}$ and the model with policy uncertainty alone. Because the rationally inattentive household is less aware of the change in SPA, in the lead up to the SPA (ages 55 to 65), they increase labour supply less than the fully aware household. Over this period, in the model without informational frictions, average employment increases by 0.41 years when the SPA goes from 60 to 65. With costly attention, it increases by 0.32 years, and so ignoring this cost leads to an overestimation of the increase in employment of 23%.

Post the new SPA of 66, the rationally inattentive agent works more to make up for missed opportunities, and so this model predicts a larger increase in labour supply than the frictionless version. However, generally declining employment and imperfect intertemporal substitutability mean that the differences are more than in the pre-SPA period. Over the working life (55-79), the rationally inattentive version predicts an additional 0.42 years of employment, the version with only policy uncertainty 0.49 additional years, a difference of 15%.

Costly attention generates a smaller overall increase in employment but a larger employment response at the SPA because much of the bunching at the SPA represents an intertemporal shifting of employment. The completely informed household immediately internalises changes to the SPA, increasing labour supply when the woman is in her 50s and continuing throughout the pre-SPA period. The rationally inattentive household only partially responds until, when much closer to the SPA, they realise the need to make up for lost time.

This forward shifting of employment to just before the SPA under costly attention registers as the barely detectable crossing at age 65 in figure 9 but can be seen clearly in figure 10 which shows the impact of increasing the SPA from 60 to 62. Here the larger increases at the SPA in the rationally inattentive version dominates the smaller increase in employment earlier in life, the rationally inattentive version predicting a marginally larger mean employment increase (0.149 vs 0.146 years). The reason is that by age 65, the rationally inattentive agent knows the SPA is either 66 or 67, and so the informational shock, and hence the employment response, is smaller than if she discovered her SPA in her early 60s when a range of SPAs were possible. Although this results from an arbitrary upper limit to the SPA, it provides an interesting policy insight: if the government's ability to increase the SPA declines as the SPA gets higher, including in more gradual ways than a limit, then the employment response to the SPA will decline with subsequent reforms.

Rational inattention only partially explains excess employment sensitivity (see table 5), casting doubt on the model predicted labour supply responses. Simultaneously intriguing evidence framing effects, or norms, might be important to the employment response to the pension ages (e.g. Seibold, 2021; Gruber et al., 2022) should be considered alongside the evidence presented in this paper for the importance of mistaken beliefs. It seems natural to think that both mechanisms may be at play, and a simple way to introduce a norm to stop working at the SPA is a fraction of passive decision makers in the style of Chetty et al. (2014) or Lalive et al. (2017). The fraction of passive agents I introduce is also naive, not anticipating their passive retirement. I compare the prediction of a model that exclusively uses passive agents to explain the excess employment sensitivity puzzle with a model that combines costly attention and passivity. 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% to replicate the data, the information and wealth shocks at SPA closing the gap.

Figure 11 plots the change in employment when the SPA increases from 60 to 66 for the two models with passive agents. As in figure 9 rational inattention predicts a smaller increase in employment from increasing the SPA, but now the smaller share of passive agents in that model makes the difference between models larger. This happens because when the SPA is 60, past that age, a smaller share passively stops working, further shrinking the employment increase in the rationally inattentive model. The smaller share of passive agents also decreases and sometimes reserves the larger employment increase for the rationally inattentive household post the SPA of 66 relative to the frictionless version seen previously. The model with costly attention predicts an additional 0.59 years of mean employment between 55 and 79, in contrast to 0.71 without costly attention, and an additional 0.50 years of mean employment during the prime working ages of 55 to 65, compared to 0.63 without. 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.

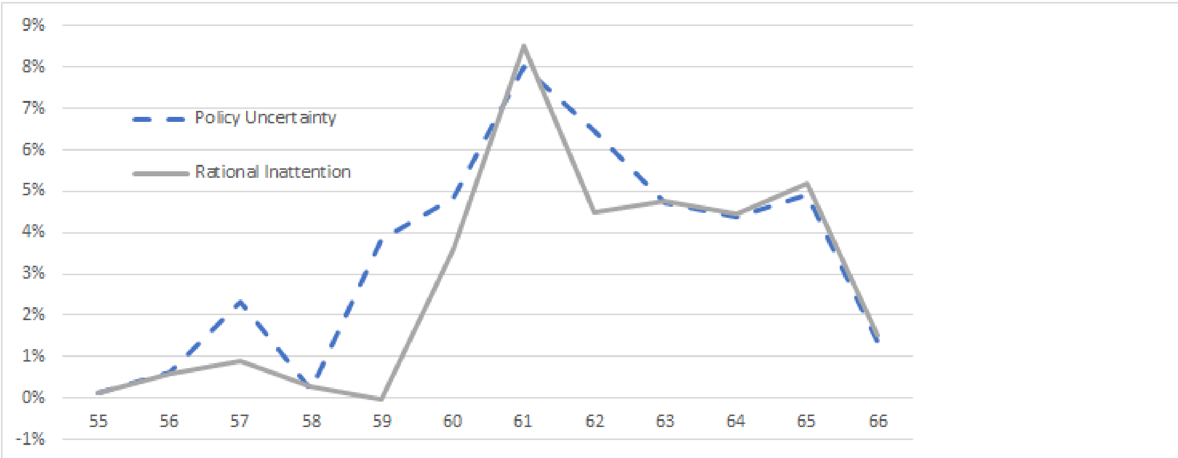
8 Extension

As was seen in section 7 the model requires different costs of attention to replicate different features of the data pointing toward misaligned in the relative levels of incentives even though the model contains mechanisms that can explain each feature in isolation. Since the stochastic state pension age was a simplification of the true extent of policy uncertainty around the state pension, this seems like a natural place to look for this misalignment.

For this reason, I introduce learning and uncertainty about another aspect of the state pension system into the model: the actuarial adjustments to benefits from deferring. Combined with a claiming decision this not only makes the model more realistic helping to align incentives but also helps explain the deferral puzzle, detailed in the section below. Rational inattention speaks directly to this puzzle because the calculation implying actuarially favourable deferral ignores the attention cost of learning the deferral rate, and claiming removes the attention cost of tracking this aspect of the pension system. Thus creating an additional incentive to claim.

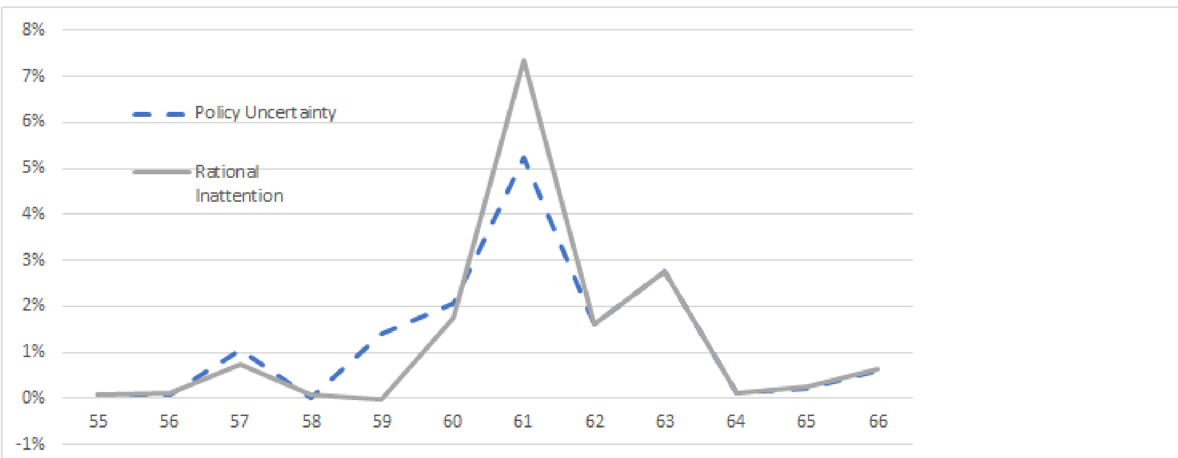
The version of the model presented in section 4.2, does not incorporate such a mechanism for two reasons. Firstly, the model does not include a benefit claiming decision. Secondly, the only source of uncertainty subject to an attention cost is the SPA and once this age is reached the attention cost disappears whether the agent claims or not. Including more sources of uncertainty subject to an attention cost would make the model more realistic. If an additional sources were uncertainty about the deferral rate and a benefit claiming decision was added, then the model would include an incentive not to defer resulting from cognitive costs. Hence this provides an incentive not to defer which is ignored in the claims that deferral is more than actuarially fair. The simplest possible extension with these features is presented in the rest of this section along with some results.

Figure 9: Increase in Employment from Increasing SPA 60 to 66



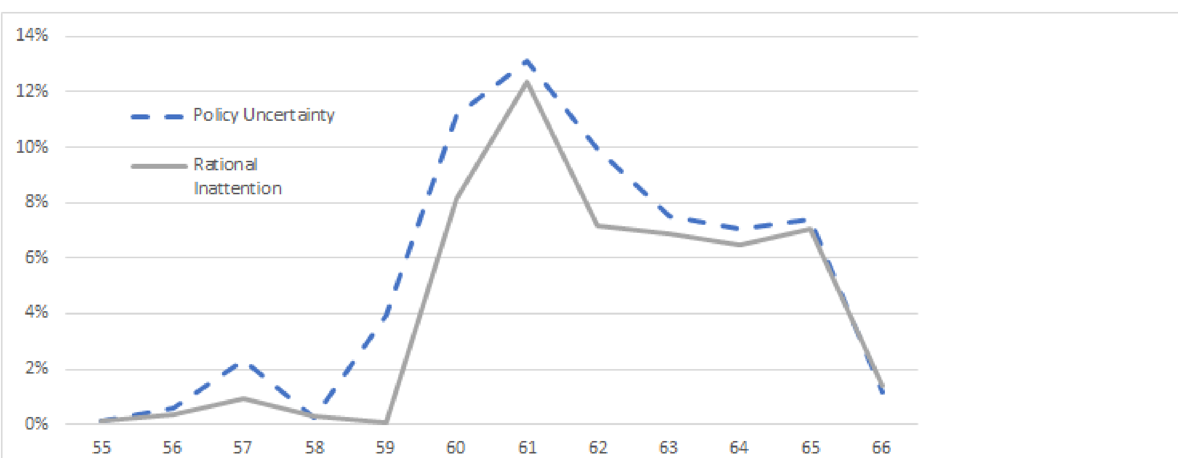
Notes: For each model the difference in employment increase between simulation of households with a female SPA of 60 and those with a female SPA of 66.

Figure 10: Increase in Employment from Increasing SPA 60 to 62



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

Figure 11: Increase in Employment from Increasing SPA 60 to 66 with a Fraction of Passive Agents



Notes: Graph shows the difference between simulation of households that had a female SPA of 60 and those that had a female SPA of 66, where the households with a SPA of 66 initially had a SPA of 65 which is increased when they are 57

8.1 Deferral Puzzle

The deferral puzzle refers to the fact that deferral of state pension benefits was extremely 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 the 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.

Amongst the small group of women, we observe deferring the duration of deferral was low. 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 have 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

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 70 is placed on deferral.

To keep the state space manageable, stochastic deferral adjustment is modelled as iid with two points of support. Having only two points of support limits the growth of the state due to solving the model with different values for the adjustment rate to a factor of two. Having the uncertainty be iid means that beliefs do not enter as a state variable as yesterday's learning is not relevant to today's knowledge; the agent knows the underlying probabilities so these form their prior each period. 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 since 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 stated in absolute rather than percentage terms the ONS time series of state

Table 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.

Table 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	0.109
Assets >Median(£29,000)	0.0903	0.089
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. Top panel shows labour supply response across the wealth distribution as per table 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.

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, a stochastic SPA and actuarial adjustment, is then re-estimated to match the same pre-reform employment and assets profiles with a constant realization of 10.5% for the deferral adjustment, which was the deferral rate these cohorts faced. New parameter estimates are in table 8. For these parameter values, only 6.2% of individuals claim the state pension before the mandatory claiming age of 70, much lower than the 99% plus claiming by that age 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 9, this cost of attention produced a relatively good fit along all dimensions of interest.

9 Conclusion

This paper shows that incorporating one empirical regularity, mistaken beliefs resulting from information frictions, into a model of retirement can help explain other puzzling empirical regularities, in particular, the excess sensitivity of employment to statutory retirement ages. Including rational inattention to an objectively uncertain pension policy significantly improves the model prediction of the labour supply response to the SPA. The mechanism driving this comes from an interplay of objective policy uncertainty and subjective beliefs: arriving at the State Pension Age (SPA) resolves the policy uncertainty and the size of the resulting shock is larger due to people's mistaken beliefs.

In doing so this paper makes other auxiliary contributions. It is the first, to the best of my knowledge, to solve a dynamic rational inattention model with endogenous heterogeneous beliefs. Allowing for the large choice and state variables implicit in incorporating endogenous heterogeneous beliefs presents computational challenges and weaving together recent theatrical results into a consistent solution methodology is one of these contributions. Doing so is not just an exercise in pushing the limits of computation, however, as the fact that mistaken beliefs are endogenously selected is key to explaining the relationship between these mistakes and the labour supply response to the State Pension Age (SPA). People who are most misinformed about their SPA have the smallest labour supply response upon reaching the SPA because the SPA is not relevant to their actions and so they choose not to learn about it.

By including an explicit model of belief formation this paper takes an approach to the beliefs preferences identification problem that avoids loading all explanations onto preferences by making the same sort of functional form assumptions about beliefs that are routinely made about preferences. This paper then uses beliefs data to pin down the cost of attention.

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 attention cost which can be avoided by claiming; hence this assertion omits an incentive not to defer.

References

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.

Armenter, R., M. Müller-Itten, and Z. R. Stangebye (2019). Rational Inattention and the Ignorance Equivalent.

Baker, S., N. Bloom, and S. Davis (2016). Measuring Economic Policy Uncertainty. *The Quarterly Journal of Economics 131*(4), 1593–1636.

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.

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.

Bernheim, D. (1988). *Social Security Benefits: An Empirical Study of Expectations and Realizations*.

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.

Burtless, G. (1986). Social Security, Unanticipated Benefit Increases, and the Timing of Retirement. *Review of Economic Studies 53*(5), 781–805.

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.

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.

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.

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.

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.

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.

Gabaix, X. (2019). Behavioral inattention. In *Handbook of Behavioral Economics: Applications and Foundations 1*, Volume 2, pp. 261–343. Elsevier.

Gruber, J., O. Kanninen, and T. Ravaska (2022). Relabeling, retirement and regret. *Journal of Public Economics 211*, 104677.

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.

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.

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.

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.

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.

Mcculloch, W. S. and W. Pitts (1943). A logical calculus of the ideas immanent in nervous activity. Technical report.

O'Dea, C. (2018). Insurance , Efficiency and Design of Public Pensions . *WP* (1), 1–63.

OECD (2000). *Reforms for an Ageing Society*.

Porcher, C. (2020). Migration with Costly Information.

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.

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.

Steiner, J., C. Stewart, and F. Matějka (2017). Rational Inattention Dynamics: Inertia and Delay in Decision-Making. *Econometrica 85*(2), 521–553.

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.

A Additional Empirical Details

A.1 Additional Instituional Details

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 conincided with a compolusary retirement age. The equality Act (2010) went further and banned all compolusary retirement ages with a handful of specific exceptions knows 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 no strictly relevant to the emprics. However, some background and anecdote about them may ilustrate how strict UK age descrimintion law is as regards focring people to retire. Seldon v Clarkson, Wright and Jakes (2012) clarfied exactly when EJRA are justified. It laid out three criteria an EJRA must meet: one, the reason jsutifiying the EJRA must by objective of a public interest (e.g. intergenerational fairness) not just of the firm; two, this objectve must be consistent with the social policy aims of the state; and, three, an EJRA must be a proportionate means to acheive this objective.

The plantif 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 firm two of the most discussed EJRA are at the UK top Universities: Oxford and Cambridge. Other UK universities appear to have removed compulososry reitement requirements ages, and intrestingly Oxford recent lost an employment tribunal that judged that there EJRA was not justified. Ewart v University of Oxford (2019) found that although the objective of Oxfords EJRA (intergenerational fairness) was valid an EJRA was not a propotionate way to acheive this due to limited demonstrated effectiveness weighed against it clearly determental impacts. Hopefully, this goes someway to illustrate that UK law treats forced retirement very seriously as age descrimination and that the few exepctions made are precisely that: exceptional.

A.3 Excess Employment Sensitivity

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 10 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

Table 10: 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 1 interacting a continuous measure of NHNBW with the treatment and all fixed effects and controls.

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.

A.3.2 Ristricted Asset Categorisation

As the gaol in investigating treatment effect heterogenienty by asset holdiings is to understand the role played liquidity constraintys the main text restrictd to NHNBW. However, parts of NHNBW can be illiquid and so in table 11 repeat the analysis but for a more restricted asset category very liquid asset, which is only assets that can reasonably by liquidated in a matter of weeks. As can be seen the restuls are qualitively very similart to those using NHNBW and do not support liquidi constrinat alone explaining away the treatment effect. The treatment effect for those with above median assets is still postive and, although the difference between the two subgroups is now signifcant, column 4 containing the continuous interaction terms shows tha,again, this heterogenienty is to weak for the treatment to be completely explained by liquidit constinarts.

A.3.3 Bad Control Concerns

Bad controls concerns are particaulrly important in the case of DID. Some take the veiw that only time invariant controls should be inculde because controls imply that we are impose parellel 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 qualitively the conclusion draw are not impacted by the presence or otherwise of controls.

Table 12 shows the results of this exercise of dropiing controls. As can be seen the results are very little changed from those with controls.

Table 11: 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 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.

Table 12: 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 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.

A.3.4 Imputation Approach to DID

Using a two-way fixed effects regresion to esimate difference-in-difference model assumes treatment effect hetergoeneinty across time and across units. When timing of treatment induces the variation in treatment, as is the case in this paper, violations of these assumption can lead to estimated treatment effects being nosensiscal combinations of the individual level treatment effect. This issues, and realated issues, have been flagged by a recent wave of literautre, but thankfully this literture also propose solution that relax these assumtpions.

Here I impletement the inputation approach of Borusyak et al. (2021). This approach alows for nver treated, but does not allow for always treated units. To be consitent 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 exspect.

Figure 12 shows the dyanmic treatment effets before and after the SPA. There is no indication of violated parrellel trends of anticpation effects as no of the pre-SPA treatment effects are signifincatly different from zero. Indeed jointly testing for violations of parrellel trends fails to reject the null of parrellel trends ($p = .799$). Conversly 7 of 9 post-SPA treatment effects are individual signicant and we can easilly 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 13 looks at wether these individual treatment effects varry between waves. The treatment affect look quite uniform across waves, although again we can reject this hyptothesis ($p = .137$). However, neither violation of homogeneinty seems serve and geneerally the graphs looks supportative 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 approixamtion.

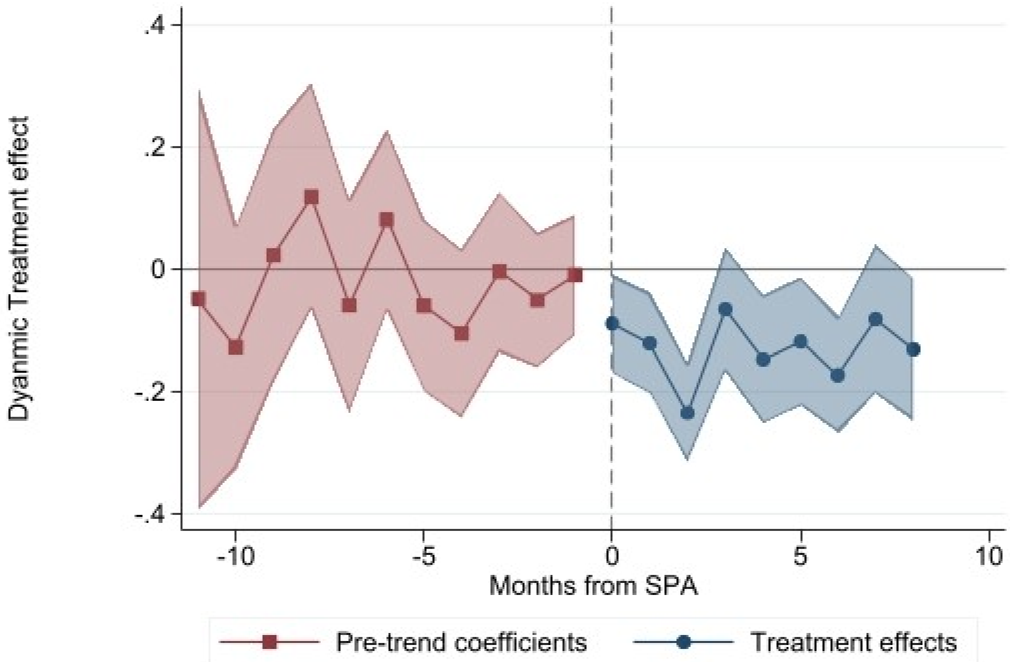
If you are more concerned about the violations of homogneous treatment effects then these results show that even allow for arbitrairy heterogneinty there is simething special happening at the SPA which is diffcult to explain in standard complete information models

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.

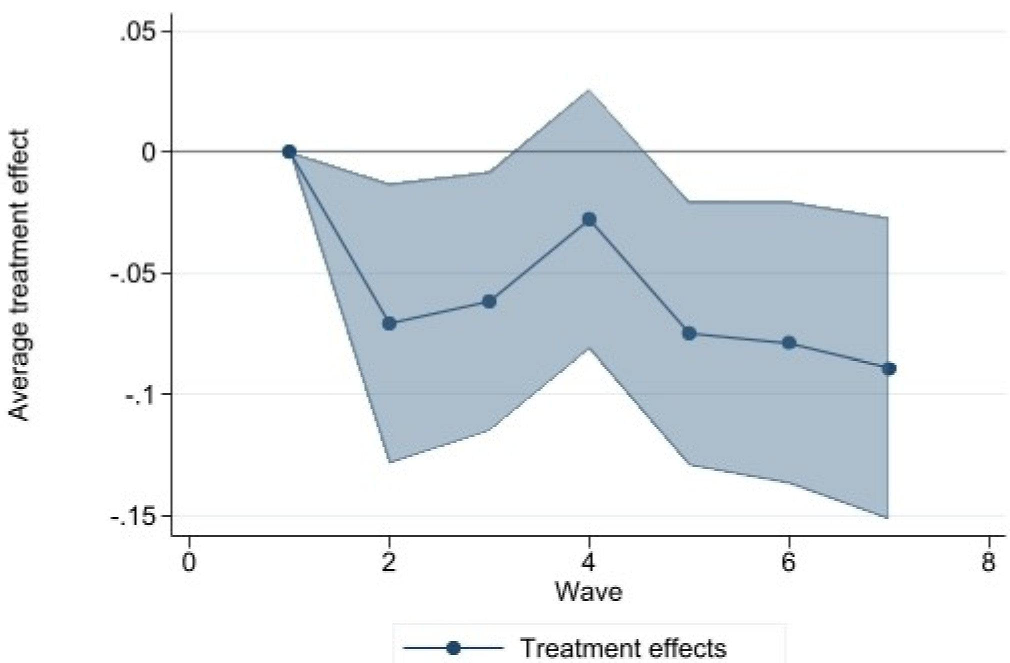
Wealth effects play a important role in determining labour supply, and women who have a later SPA are life-time poorer. The puzzle is not that they have a higher labour supply, the puzzle is that their labour supply response should be concentrated at the SPA, the change in SPA having been announced over 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 1 difference in life-time wealth, including those induced by SPA differences, between year-of-birth cohorts are absorbed by

Figure 12: Dynamic Treatment Effects by Time from SPA



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

Figure 13: Average Treatment Effect by Wave



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

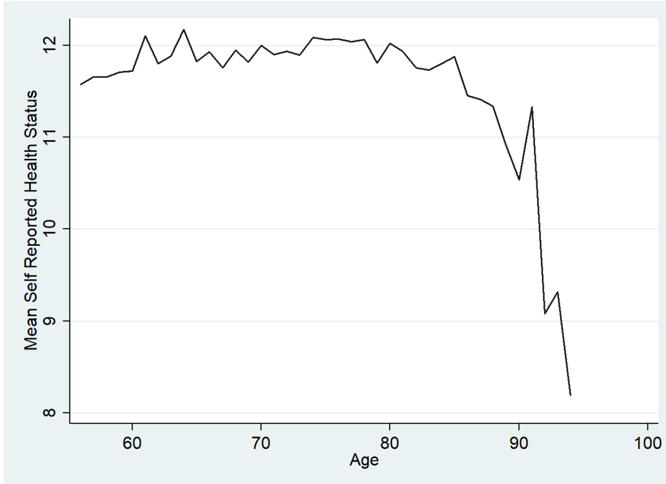


Figure 14: Self Reported Health Profile

the cohort effects. Hence, the only wealth difference the treatment effect will detect are between individual 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 limited impact of wealth on the treatment effect; 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 el al., 2011). However, there is no reason to suspect it interacts with the SPA, and so no reason for it to explain 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 see in figure 14 which shows the age profile of health status. All the same, as it is such an important factor in retirement table 13 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[29].

Finally, the timing of private pension eligibility is important for retirement choices. However, occupational pension schemes are very unlikely to have adjusted their pension ages in line with the female SPA, because private pension do

[29] See appendix A for details.

Table 13: 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

Table 14: 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

not generally offer different eligibility ages to men and women[30], and this reform only changed the female SPA. Still, checking for 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 age are unreliable, as is documented in section 3.4. However, only defined benefit pension system 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 16, and as can be seen, despite the loss of power, the treatment remains present and significant.

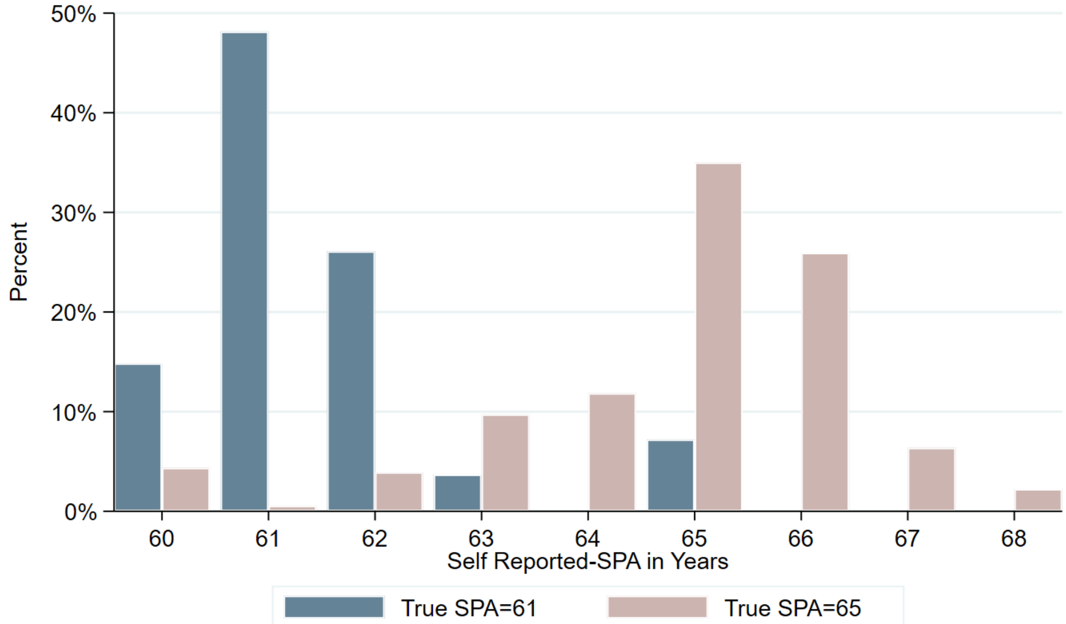
A.4 Mistaken Beliefs and Excess Employment Sensitivity

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 of 65. To get at these distinctions figure 15 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

[30]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 15: 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.

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 16 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 there own SPA to precisely the right month. The main thing we can gleen from this graph that we cannot when the date is binned at a yearly frequency is that the spike every 1 months here show that people display an unsurprising round number bias.

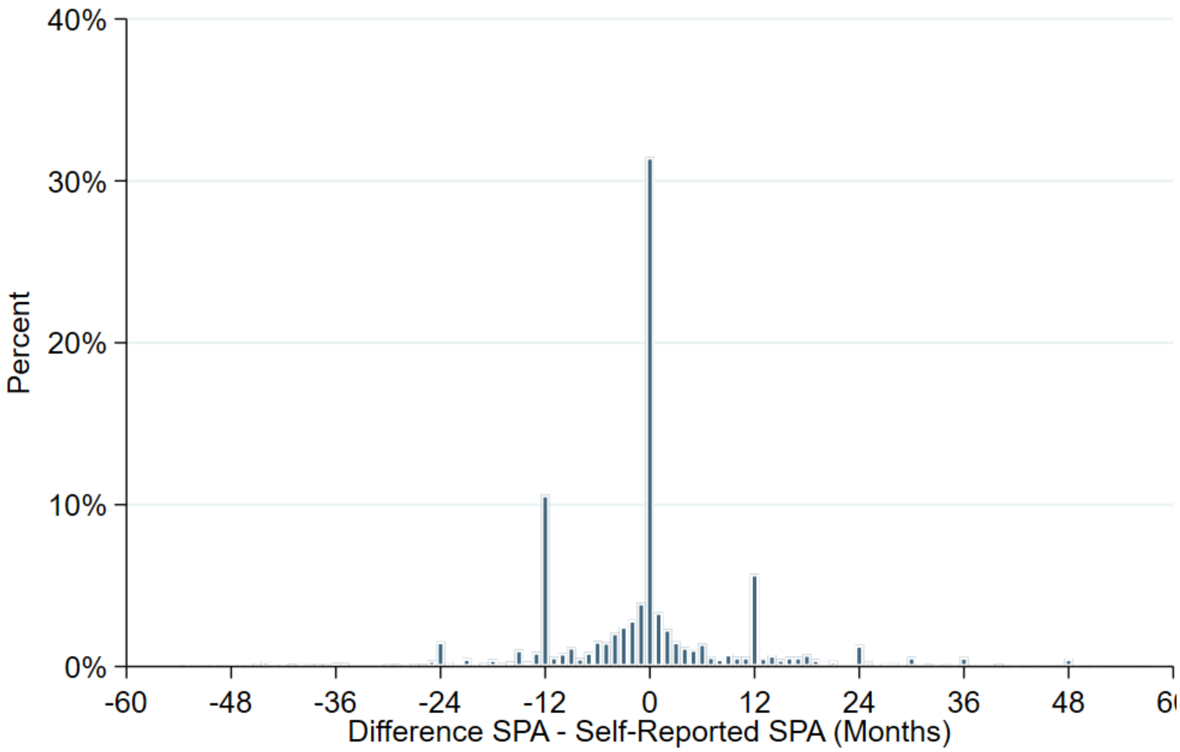
A.6 Treatment Effect Hetrogeneinety by Beliefs

I have interpreted the fact those who are more mistaken about their SPA in there late 50s have a smaller labour supply reaction upon reacing their SPA in thier early 50s as evidence of the importants of beliefs. However, this fact is consitent with beliefs proxying for some unobserved heterogneienty, for example if those who are more mistaken have lower cogntive 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 effect vary with beliefs that are consitent with selection into SPA knowledge driving the results and that are harder to explain with an appeal to unobserved heterogneinty.

Firstly, those whose beliefs imply they will recieve a postive shock (i.e. they overestimate the SPA) are the ones to have the largest labour supply response to it. This can been seen in table 15, it shows how the treatment effect varries between three groups: those who think there SPA is sooner than it is, those who correctly state thier SPA, and those who

Figure 16: 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.

Table 15: Treatment Effect Heteorgneinety 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

Table 16: Treatment Effect Heteorgneinety 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

think it is further away than it is. The last groups in the excluded category and we can see that for this group the treatment effect is postive and signifcant at 0.785. Then for those who underestimate their SPA we see that theyir predicted treatment effect is smaller and loses signifcance when the sum of coffeicient is tested joints ($p = .0999$). Those who know their SPA have a larger response, rather than a smaller one

Secondly, those whose knowledge of the SPA got better had the largest labour supply response to the SPA. This can be seen in table 16 which shows how the treatment effect varries between three groups: those whose SPA self reports gets worse between first and last time they are asked, those whose self reports stays the same, and those whose self reports gets better. It can be seen that those whose knoweldge imporves have the largest labour supply respnse to the SPA.

B Additional Mathematical Details

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

This section presents a extension of Steiner et al. (2017) that makes their results applicable to my model. Thier paper does not allow for any endogenous state variables but here I show their results still hold if there are endognous state but these are completely revealved by the costless signal. The reson is that ... So for comparability In this section I adopt much of the notation of SSM and notation is not related to the rest of the paper.

My variant of their setup is the following. There is a payoff relevant state $\theta_t \in \Theta_t$ according to measure $\pi \in \Delta(\prod_t \theta_t)$ and agents must make a payoff relevant decision from a choice set D_t. Before making a decision, the agent first observes a costless signal $y_t \in Y_t$, $y_t \sim g_t(y_t|\theta^t, y^{t-1})$ and then can choose any costly signal about θ_t on signal space X_t. 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 $z^t = (x^t, y^{t+1})$. It is assumed that $y_{t+1} \perp (x^t, d^t)|(\theta^t, y^t)$. The sets Θ_t, D_t,Y_t, and X_t are finite and that $|D_t| \leq |X_t|$. My setup differs from SSM's in that I adapt the timing assumption so that the costless signal is received before the action is taken each period rather than after it. This change in timing only affects the proof of lemma 1 from SSM's paper and I show below that this results still holds using a slightly different strategy to prove it.

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

$$\max_{f,\sigma} E[\sum_{t=0}^{T} \beta^t (u(\sigma_t(z^{t-1}, x_t), \theta^t) - I(\theta^t, x_t|z^{t-1})] \tag{25}$$

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. For notational convenience, let $\omega^t = (\theta^t, z^{t-1})$ be the current state and the agents 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*

$$\max_{p} E[\sum_{t=0}^{T} \beta^t (u(d^t, \theta^t) - I(\theta^t, d_t|z^{t-1})] \tag{26}$$

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 26 induces a strategy solving the dynamic RI problem.

Proof. We precede in steps.

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

$$E[\sum_{t=1}^{\infty}\beta^t I(\theta^t,\zeta_t|z^{t-1})] = E[\sum_{t=?}^{\infty}\beta^t(H(\theta^t|\zeta^{t-1},y^t) - H(\theta^t|\zeta^t,y^t))] \tag{27}$$

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

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

Also $y_{t+1} \perp (x^t,b^t)|(\theta^t,y^t) \Rightarrow H(y_{t+1}|\theta^t,x^t,y^t) = H(y_{t+1}|\theta^t,y^t) = H(y_{t+1}|\theta^t,b^t,y^t)$, so by 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 27 becomes

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

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, 26-25 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 26$\geq$25 .

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 25. For this strategy generated by this mapping, the probability choice rule makes equation26=equation25

Step 4: If s solves 25 the corresponding PCR p must solve 26, as by step 2 the value from p in 26$\geq s$ in 25, so if p doesn't solve 26 $\exists$ PCR producing greater net lifetime utiltiy than s in 25. But by step 3 this produces a viable solution to 25 with greater net life-time utility contradicting s being a solution to 25.

Step 5: If p solve 26 then by step 3 it produces a viable solution to 25 but then 26$\geq$25 so this strategy must be the optimal solution to 25 $\square$

The remainder of the proof follow as stated in SSM for the case where the choice variables are discrete as is the case in this paper.

It is worth saying a few words about how the model in this paper maps to the class of models in this appendix based on SSM as the correspondence is not obvious. The clearest difference between the SSM setup and the model presented in section 4.2 is that SSM only allow for exogenous states whilst I have an endogenous state in the form of assets a_t. However, since utility can depend upon the entire history of choices and states there is a simple mapping from the endogenous states without history dependent preferences to the world of exogenous states with history dependent preferences. The state in the sense of SSM now only contains the exogenous states $\Theta_t = Supp(SPA) \times Supp(Y_t) \times Supp(AIME_t)$, $(SPA_t, y_t, u_t) = \theta_t \in \Theta_t$ but since $a_t \in d_{t-1}$ and $AIME_t = g(d^{t-1}, \theta^{t-1})$ for the function g that follows from the definition of $AIME_t$ given in section 4.1. Hence, we can re-express the the utility given in terms of section 4.2 states X_t and the current decision $u(d_t, X_t)$ in terms of the history of exogenous state θ^t and the history of decisions $u(d^t, \theta^t)$. And since the SSM agent condition their action on everything useful from $z^{t-1} = (d^{t-1}, y^t)$, they can condition on all states.

Details available upon request.

C Additional Computational Details

C.1 Solving the Models without Costly Attention

The models are solved by backward inducation starting at age 101 when the household dies with certainty. The household problme is consider as a discrete choice problem. This within period descrete choice optimisation problem is solved by grid search, selecting the value that mximises the households utility. States are discreteised 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 household when making their saving choice. This keeps the size of the state space maageable whilst not unduly constraining households and is equivalent to having a finer grid for consumption than for assets. When evavluting continuation values of off-grid values I use linear interpolation of the value function.

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 discrete the distribution I consider all possible combinations moving probabilty masses of a given size between the 8 possible SPAs 60-67. As no amount of Bayesina 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 girdpoint of beliefs and so I imposed a minum probabiltiy 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 concerete, I broke the total probability into four masses that I moved between SPAs to form the grid over beliefs. In the abence of this minum probability of any SPA that would mean the probability masses being moved between SPAs was of a size of 0.25. In period in which there are 8 possible SPAs, because $t < 60$ and the women has not aged past any possible SPA, theses probabiltiy 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 though of as an ordering of the 4 gris points and the breaks between the 8 grid points. As the women sucessively ages past SPAs this shrink 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 $\mathscr{R}^7$ I order these numbers in lexogrpahic ordering which is convient to construct all possible combinations of the probability masses.

High Dimensional Interpolation When the prior with which a household starts next period is off this grid I use k-nearest neightbour inverse distance weighting to carry out the multidimesional 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 burderb and also a source of approximation error. For this reason I initially start using just two nearest grid point 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 BLAH during periods in which and at states at which rational inattention matters because $t < SPA_t$ the central equation that need to be solved to find the households optimal decision is the following:

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

Following approaches to used in the RUM literature I normalise the payoff inside this equation. As first I do this by diving through by the highest payoff in all possible SPAs. However the precense of λ in this equation makes this process of exponentiating utility even more problematic. Data and not computational coonsideration should determine what values of λ we consider, however the fact this parameter appears as a denomnitor in in an exponetiated expression means that as λ gets small the difference between exponetiated payoof gets larger. Since a lower SPA is better than having a later one, the values inside the log assocaited with SPA=60 are larger, and decreasing the cost of attention exagerates these diferences. However, when λ gets small the fact that the exponetiated payoff assocaited with SPA=67 are much smaller than those assocatied with SPA=60 does not mean the former are not important to the optimisation because for very small values log() approaches minus infinity and it's rate of change approaches infinity. So how probabilities are allocated over these outcomes when the expontetiated payoof is very small has vey large implication as to the value of the objective function. Therefore, we cannot ignore vanishingly small exponteiated payoff because they have outsized implications for the logarythmic opbjective function. This fact combined with the very small values of the cost of attention implied by the belief data lead me to very carefully optimise the code with respect to the storgage of very smalle utlity values, rather than just dropping them as could be more happilly done with a more standard obejctive function. To store these smaller values I use quadruple precision flaot points leading to smallest value distingushiable from zero of 10^{-4965}. Hoever, since complier are optimised to conduct double precision operation moving from double to quadrubple precision leads to a much greater that factor of two slow down in runtime. For this reason, I only use quadruple precision when absolutely nescerrary, checking beforehand if normalising payoff leads to an underfolw so that important values would be lost and treated as zero in double precision.

C.3 Simulating

My initial sample of simulated individuals is large, consisting of 50,000 random draws of individuals age 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 life-time earnings, which are the state variables that 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 initialsed to a fixed value mirroring the SPA of the cohort curently being simulated. I make the assumption that SPA answer represent draws from an individuals belief distribution and that everyone start at age 55 with the same beliefs. This allows me to initialise the belief distribtion to the distribution of point estimates seen for SPA self-report in the ELSA data.

Given these initial condition I simultate the choice of the indivudal households using the decision rule found when solving the model and the exogenous process estimate in the first stage. I thien 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 D. The method of simulated moments procedure delivers the model parameters that minimize a GMM criterion function, which we also describe in Appendix 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 routine at promising initial conditions.

D Additional Econometric Details

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 quinitc in NHNBW plus a very rich set of additional controls the include variaibles on: health, education, location, labour market behavirou, housing tenure, cohort, age, wage, and mesure 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 observered heteroscedasticity. To do this I run regression of the non-imputed $AIME_{55}$ values on a quinitc of NHNBW without the controls (because the model does not contain the other variables) and then regress the squared residuels on the same polyomial of NHNBW. Since the imputed $AIME_{55}$ are by construction homoroscedasticity assing a noise term with variance given by this last regression replicates the heteroscedasticity seen in the regression of $AIME_{55}$ on the qunitic of NHNBW.

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 probabilties floowing 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 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 measuremnt error issue.

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 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 targetted profiles to their average value whilst also allowing for the key variation in behvirou between SPA-cohorts at the SPA.

Table 17: Sumarry Statistics of Initial Conditions (£)

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

E Additional Results

E.1 First Stage Estimates

Model Types A women is classed as high education if she has more than the compolusry 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 seperation in the model. To get around the fact that seperation occurs in the data, if a women is ever observed as married her household is classified as such in all periods. The reason to classifiy her as married rather than single is that a divorced or windowed women will likely recieve some form of alumoney or windows pension and so she is more acurately modelled as married according to the model. This leads to the following propoation 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 sumarry statistic of which are presented in table 17. Understandable for women of this generation married women have weaker labour market attachment and so lower $AIME_{55}$ but higher household assets a_{55}. Hiher education increase both variables.

Labour marekt conditions

E.2 Alternative Model Specifications

Details available upon request.