Costly Attention and Retirement *

Jamie Hentall MacCuish[†]

Job Market Paper

Click here for latest version

Abstract

Most people are mistaken about aspects of their pensions, suggesting informational frictions. This paper models these frictions. Exploiting reforms to the UK female State Pension Age (SPA), I solve and estimate a dynamic lifecycle model of retirement with rationally inattentive household learning about an uncertain SPA. Inattention leads to endogenous mistaken beliefs, which help explain a puzzle: labour market exits concentrate at official retirement ages despite weak economic incentives. Costly attention improves the predicted employment response to the SPA, explaining 30%-74% of the shortfall, whilst explaining patterns in the belief data. Though key to explaining this puzzle, costly attention introduces a high-dimensional state (beliefs) and a high-dimensional choice (the learning strategy). Solving the model represents a novel contribution and to achieve this I develop a general-purpose solution method. Costly attention attenuates the SPA's effectiveness as a policy lever to increase old-age labour force participation by up to 27% because less informed agents are less responsive to policy changes. Rational inattention can also explain 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. For their helpful comments, I also thank James Banks, Richard Blundell, Uta Bolt, Jonathan Cribb, Martin Cripps, Aureo de Paula, Mariacristina De Nardi, Rory McGee, Attila Lindner, Cormac O'Dea, Ran Spiegler, Imran Rasul, Morten Ravn, and participants at the NBER Summer Insitute Behavioral Macro Session, IIPF Annual Congress, RES Junior Symposium, CESifo Public Area Conference, Young Economist Symposium, NETSPAR Pensions Workshop, Econometric Society European Meeting, SED Anual Meeting, CERGE-EI, ENTER, Toulouse, Cambridge, UCL, and the IFS. For thoughtful discusion, I thank Andrew Caplin, Dylan Moore, Leanne Nam, Nicolas Fernandez-Arias, Jim Been, Patrizia Alexandra Massner, and Sebastian Seitz. 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

Most people are confused about pensions. One example is they frequently mistake the age from which they can receive pension benefits by multiple years, as seen in Figure 3 . Mistaken pension beliefs are so common they may seem unsurprising. They are, however, incompatible with standard complete information models. Widespread mistaken beliefs about financially important policies suggest incomplete information resulting from information frictions or cognitive limitations.

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

Specifically, I embed rationally inattentive households in a lifecycle model which generates mistaken beliefs and helps explain the excess employment sensitivity puzzles. This puzzle, documented in multiple countries [1], is that people exit employment at pension-eligibility ages whilst benefit systems offer only weak incentives to stop working precisely then. Mistaken beliefs increase the wealth and uncertainty shocks received from the resolution of pension uncertainty upon reaching these eligibility ages. These increased shocks help explain the excessive employment reaction. Costly attention stands out from alternative explanations of the puzzle by also explaining stated beliefs.

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

Endogeneity of beliefs drives the relationship between retirement and mistaken beliefs but complicates the model by introducing a state (beliefs) and a choice (learning strategy) that are both high-dimensional. I weave together recent theoretical results [2]into a general-purpose solution method for dynamic rational inattention model with endogenous heterogeneous beliefs that overcomes these complications thanks to sparsity, amongst other properties. To the best of my knowledge, solving a structural rational inattention model with endogenous heterogeneous beliefs is a first.

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

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

[2] These papers are Steiner et al. (2017), Caplin and Dean (2015), and Armenter et al. (2019)

barrier to intertemporal substitution.

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

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

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

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

The paper is structured as follows. Section 2 reviews literature. Section 3 outlines institutional context and data and presents descriptive and reduced-form analysis. Section 4 presents the model, starting with a standard model of complete information and then building in objective pension policy uncertainty and a cost of attention to this uncertain policy. Section 5 discusses the solution method. Section 6 discusses estimation and Section 7 model fit and implications. Section 8 presents the extension addressing 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 expenses onto these innovations. Much of this literature is US-focused, and some of its concerns are not relevant in the UK context I study (e.g. medical insurance). The key features I include from this literature are uncertainty, borrowing constraints, and individual heterogeneity, and the most similar paper is O'Dea (2018) who estimates a model of males in the UK.

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

Excess employment sensitivity Employment being more sensitive to statutory pension ages than standard models 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 and full retirement ages remained unchanged until 2004, when the full retirement age increased, providing the variation to estimate its impact on employment. Larger effects were detected than predicted by standard models (Mastrobuoni, 2009) and part of the age 65 spike followed the full retirement age despite medicare eligibility remaining at 65 (Behaghel and Blau, 2012), undermining medicare eligibility as its sole cause. [3] 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 people's knowledge of current pension rules. Manski (2004) documents precisely one such study, that finds much individual uncertainty about their benefits is explained by a lack of understanding of current social security formula. Rohwedder and Kleinjans (2006) study the dynamics of forecast errors and find they shrink as individuals approach retirement, providing evidence of learning. Crawford and Tetlow (2010) look at self-reported SPAs and find large errors common; Amin-Smith and Crawford (2018) document these mistakes are predictive of the employment response to the SPA. This paper finds similar patterns to Crawford and Tetlow (2010) and Amin-Smith and Crawford (2018), prevalent mistaken beliefs predictive of

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

labour supply, together with a similar pattern of learning to that found by Rohwedder and Kleinjans (2006). I use these patterns to identify attention which represents a novel use of belief data, since most papers use belief data to identify parameters which individuals hold private information about whilst maintaining the assumption of complete information.

Wider Literature Policy uncertainty plays an important role in this paper, and so it relates to others investigating policy uncertainty, such as Baker et al. (2016). Of particular relevance, Luttmer and Samwick (2018) measure the welfare cost of individuals' perceived uncertainty about their social security benefits.

3 Institutional Context, Data, and Analysis

Explaining the puzzlingly large employment response to the UK state pension age (SPA) is the goal of this paper. It identifies this response to the SPA using a reform to the female SPA which Section 3.1 outlines while explaining what makes it illuminating of excess employment sensitivity. 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 the SPA.

3.1 Institutional Context

The UK State Pension Age (SPA) is the earliest age at which retirement benefits, known as the state pension, can be claimed. In other words, it is the Early Retirement Age of the UK pension system. The SPA is the sole focal age of the state pension system. Deferral of receipt does increase the generosity of the benefit; however, during the period considered, this was without a cap on the deferral duration and so did not imply an effective Full Retirement Age. [4]

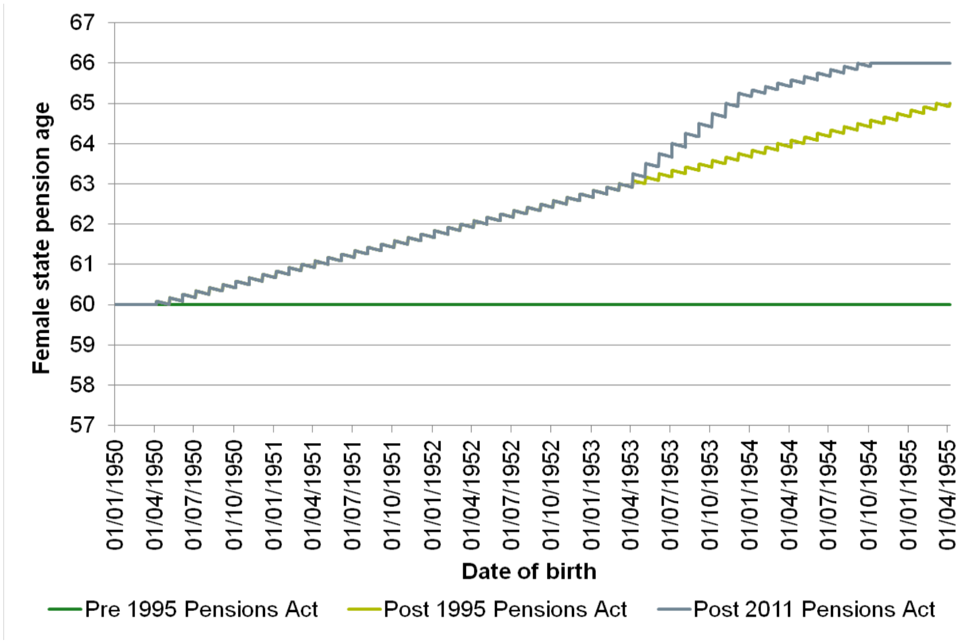
The UK State Pension came into force in 1948, with a SPA of 65 for men and 60 for women. This remained unchanged until the Pensions Act 1995 legislated for the female SPA to gradually rise from 60 to 65, one month every two months, over the ten years from April 2010. The Pensions Act 2011 accelerated the rate of change of the female SPA from April 2016 so that it equalises with men's by November 2018. It additionally legislated an increase to both the male and female SPA to 66 years, phased in between December 2018 to October 2020. Figure 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 illegal age discrimination under UK law[5]. 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

[4] Despite a generous actuarial adjustment, deferral was rare, implying another puzzle, discussion of which is deferred to Section 8.

[5] 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[6], and although the state pension is taxable income, a component of income tax, called National Insurance contributions, is removed upon reaching the SPA[7].

These three facts imply the state pension is essentially an anticipatable increase in non-labour income with the SPA its eligibility age. As the reform was announced in 1995 and began in 2010, the income change was anticipatable with a horizon of at least 15 years. Hence, the puzzle is not that employment responds to the SPA reform but that the response concentrates at the SPA when so much forward notice was given. In a standard life-cycle model, with complete information and forward-looking agents, labour supply responses do not concentrate at anticipatable income changes unless liquidity constraints prevent agents from smoothing intertemporally. So, these three features remove incentives to exit the labour market at the SPA for all but the liquidity constrained because the inability to borrow against future pension income forces these people to wait for this anticipatable additional income to decrease labour supply [8]. 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.

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

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

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

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[9] dataset that strikes the best balance between these two desideratum, 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. Earlier waves are important to control for pre-trends and to increase power when estimating model inputs. The oldest women affected by the reform were born on 6 April 1950. Having older cohorts is important as a control group and also informative when estimating model inputs.

3.3 Excess Employment Sensitivity

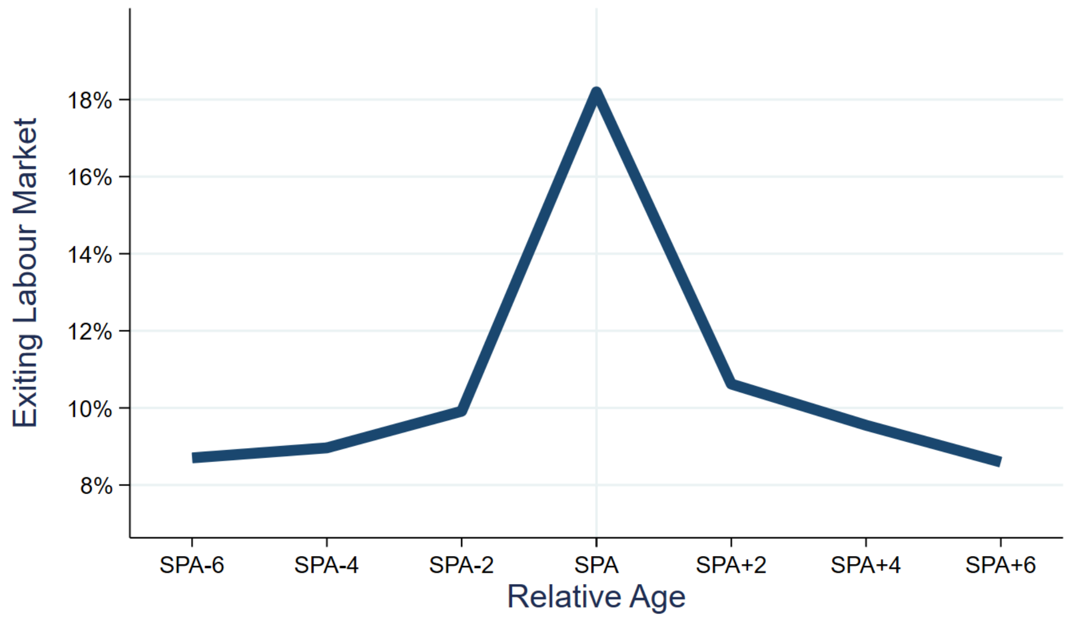
Employment being more sensitive to official retirement ages than implied by incentives is a puzzle documented in multiple countries (see Section 2). This section presents evidence of this puzzle in relation to the UK SPA. Liquidity constraints being essentially the only standard complete information mechanism that could generate this sensitivity to the SPA (see Section 3.1), I focus on demonstrating that liquidity constraints alone do not explain the puzzle.

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 female SPA reform allows more careful identification of the employment response to the SPA.

To do this, I build on Cribb et al. (2016), who use this reform to identify the labour supply response to the SPA and find it significant. They argue against constraints driving their results because, whilst homeowners are less likely to

[9] Technically, ELSA (Banks et al., 2021) 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.

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[10] leading to the following specification:

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

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

Column 1 of Table 1 presents the results of estimating equation 1. I find a 0.080 increase in the probability of being

[10]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.

in work from being below the SPA significant at the 0.1% level.

To address the question of whether liquidity constraints can explain this treatment effect, I restrict to the subsample of women from households with above median assets and repeat the analysis. Specifically, I restrict to those with above median non-housing non-business wealth (NHNBW) in the wave before they reached their SPA, as this is when the resources to smooth labour supply affect their reaction to the SPA. [11] This generates a cut-off of £29,000. The objective of this cut-off is to restrict to a group whose retirement choices are unlikely to be affected by the liquidity constraint. Given the SPA was reformed in monthly increments, and equation 1 controls for quarterly age and cohort fixed effects, an individual's control is someone born in the same year and quarter but a few months older past the SPA. This narrow time window makes arguing against liquidity constraints easier: women with over £29,000 in NHNBW seem unlikely to need to wait 1-3 months for the state pension to stop working. The results are in column 2 of Table 1. For this subpopulation, we find a treatment effect of 0.061, similar in size to results for the whole population and significant at the 1% level.

Column 3 of Table 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 robustness including restricting to more liquid assets categories and different functional forms such as dropping controls (to address bad control concerns) and having the labour supply response to the SPA vary continuously with assets. All of these specifications lead to the conclusion that, although assets matter for the labour supply

[11] NHNBW 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.

Table 2: Placebo Tests

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

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

response to the SPA, the effect is not strong enough for liquidity constraints to explain away the response. Appendix A also considers whether 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 also of average treatment effects in line with those estimated in this section. I conclude it is reasonable to give a causal interpretation to the treatment effects estimated in this section.

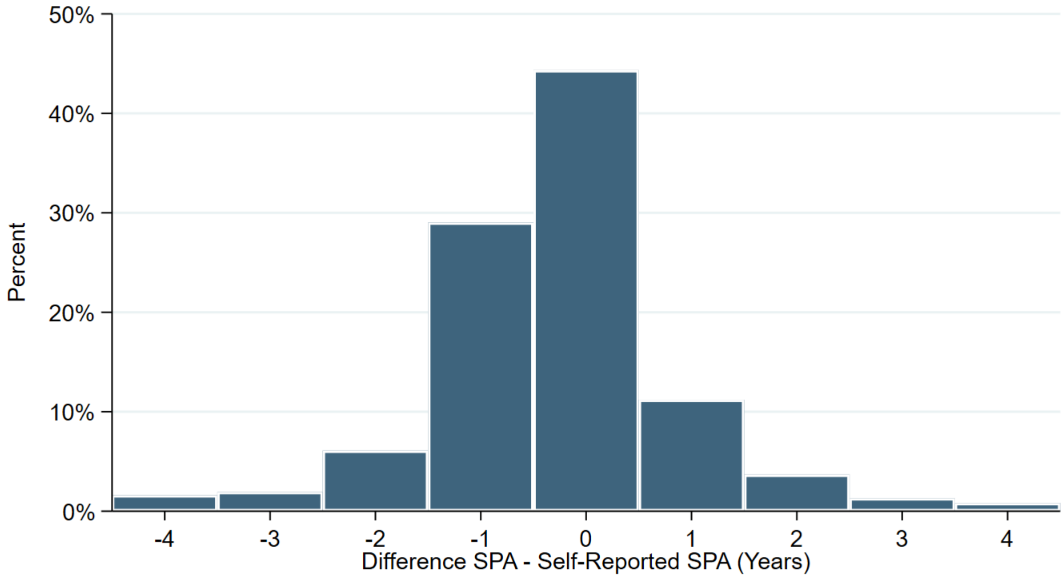
What follows, however, does not rest on the causal nature of these estimates. I use these regression results as an untargeted auxiliary model to a structural model, so what is important is the model's ability to replicate them not whether they are causal. What follows does depend on the reader finding these results puzzling, at least as far as standard complete information models are concerned. The placebo test results in Table 2, support the idea something is puzzling about the SPA. It contains the results of dropping observations over SPA and replacing the treatment in equation 1 with indicators of being one or two years under SPA; unlike the treatment, these coefficients are negative and insignificant. So the results in this section are detecting something specific to the SPA, which is puzzling for those with significant liquid wealth.

3.4 Mistaken Beliefs and Employment Sensitivity

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

The SPA being such a simple facet of the benefit system, confusion about it is both puzzling and simple to demonstrate. It is an exact function of date of birth, recorded in ELSA, and from wave 3, women under 60 are asked what their state pension age is. Any discrepancy demonstrates imperfect knowledge of one's SPA. Figure 3 shows this difference between

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



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

the true and reported SPA of 58-year-old women subject to the reform. The largest group are those who know their SPA to within a year, although this contains many mistaken by a margin of months. It also leaves over 50% who are out by a year or more, striking evidence of the prevalence of mistaken pension beliefs in the UK. Appendix A shows that self-reports cluster around the true SPA of each cohort; just the sort of pattern that emerges from a model of costly attention. [12].

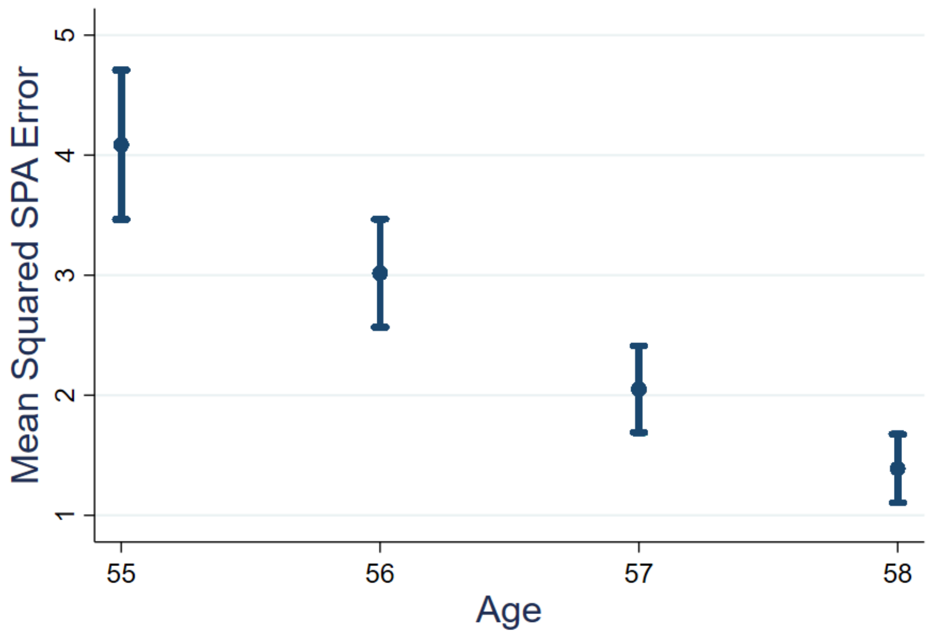
Another prediction of costly information is learning because acquired knowledge is retained, and the marginal value of knowing your SPA increases as you approach it. This prediction is supported by the data as seen in Figure 4 that plots, against age, the mean squared error in self-reported SPAs. A declining age profile can be seen, indicating errors shrink as these women age towards their SPA. The model uses this declining mean squared error as a moment to identify the cost of attention which is a novel contribution to the rational inattention literature that adds empirical validity.

Next we need to ask if mistaken beliefs impact employment's sensitivity to the SPA. Table 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 the question is no longer asked at 60 and an interaction between this error and the indicator of being below the SPA. The interaction is significant and negative indicating that, on average, for each additional year the individual is out by in their SPA self-report, the labour supply response decreases by 6.2 percentage points. [13]

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

[13] 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 causal.

Figure 4: Mean Squared Error in Self-reported SPA



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

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 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.

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.

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 consumption, labour supply, and savings. Households face risk over i) whether they get an employment offer, ii) the wage associated with any offer, and iii) mortality. The households receive non-labour income from state and private pensions after the relevant eligibility age for each.

In more detail, households are divided into four types indexed by k, based on the high or low education status of the female and the presence or absence of a partner. Households choose how much to consume c_t, how much to invest in a risk-free asset a_t with return r, and, if not involuntarily unemployed, how much of the women's time endowment (normalised to 1) to devote to wage labour $1-l_t$ (full-time, part-time or none at all) at a wage offer w_t that evolves stochastically. Unemployment ue_t, where $ue_t = 0$ indicates employment (presence of a wage offer) and $ue_t = 1$ unemployment (the absence), also evolves stochastically. The partner's labour supply is inelastic, and so his behaviour is treated as deterministic. The wife receives the state pension once she reaches the *SPA*, a parameter varied to mimic the UK reform, and a private pension once she reaches the type-specific eligibility age $PPA^{(k)}$. Both pension, $S^{(k)}(.)$ the state pension and $P^{(k)}(.)$ the private pension, are treated as type-specific functions of average life time earning $AIME_t$ ($AIME_{t+1} = \frac{(1-l_{t+1})w_{t+1}+AIME_t t}{t+1}$) [14] . From age 60, the women face a probability s_t^k of surviving the period. Finally, households value bequests through a warm glow bequest function (De Nardi, 2004; French, 2005). Only one birth cohort is modelled at a time, and periods are indexed by age of the women t. Therefore, the full vectors of model state is $X_t = (a_t, w_t, AIME_t, ue_t, t)$.

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

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

where θ determines the intensity of the bequest motive, and K determines the curvature of the bequest function and hence the extent to which bequests are luxury goods. The functional form surrounding $a_t + K$ is the utility from consumption of a household (see below), so approximately captures the utility a descendant would gain from these assets, and hence altruism as a motive for the warm-glow as well as 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 $\varepsilon_1 \sim N(0, \sigma^2_{\varepsilon,55})$. The quadratic form of the deterministic component of wages captures the observed

[14]This is average yearly earnings, to keep notation in line with the literature I use the abreviation Average Indexed Monthly Earnings, which is the variable US social secuirty depends on.

hump-shaped profile and is common in the literature.

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

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

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

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

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

The wife receives the state pension as soon as she reaches the *SPA* which abstracts away from the benefit claiming decision. This is done for two reasons, both touched upon earlier. Firstly, over 85% of people claim the state pension at the SPA, so, in terms of accuracy, little is lost by this simplification. Secondly, this small fraction deferring receipt of the state pension occurs despite deferral having been actuarially advantageous during the period considered. This presents another puzzle to standard models of complete information as they generally imply acceptance of actuarially advantageous offers. This benefit claiming puzzle is taken up in Section 8, but deferring it until then gives this baseline model a chance of addressing the excess sensitivity puzzle.

Lifetime average earning ($AIME_t$) evolves until the woman reaches the age she starts to recieve her $PPA^{(k)}$, at which point it is frozen.[15] Both the state and private pensions are quadratic in $AIME_t$, 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 dependent on past and current regulations, and cannot be exactly captured without detailed administrative data (see Bozio et al., 2010). This functional

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

form captures the dependence of the state pension on working history without getting into these difficulties. Being type-specific allows $S^{(k)}(.)$ to capture indirect influences of education and marital status on the state pension; for example, being a stay-at-home mum counted towards state pension entitlement but only after a reform was enacted. Every private pension scheme is different, but the dependence of $P^{(k)}(.)$ on $AIME_t$ reflects the dependence of most defined benefit schemes on lifetime earnings. This functional form less accurately reflects the structure of defined contribution systems, which are essentially saving accounts, but saving for retirement is captured in the model with the risk-free asset and the models starts after the stautory defined contribution eligibility age beyond which they can be accessed without penalty.

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

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

Household maximisation problem and value functions The Bellman equation 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 the following budget constraint, borrowing constraint, and labour supply constraint:

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

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

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

4.2 Two Additions: Policy Uncertainty and Costly Attention

This section introduces two additions to the complete information model. Firstly, Section 4.2.1 introduces objective policy uncertainty in the form of a stochastic SPA, capturing the observed variation of SPAs over the life-cycle resulting from pension reform. Secondly, Section 4.2.2 introduces costly attention to this stochastic SPA, in the form of a disutility for more precise information, allowing the model to capture mistaken beliefs. As these additions represent innovation, Section 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 began working life.

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

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

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

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

4.2.2 Costly Attention (Rational Inattention)

The second addition is the cost of information acquisition about the stochastic SPA. This allows the model to capture the fact that people are mistaken about their SPA and that these mistaken beliefs are the results of an endogenous learning process. As such, it creates a potential for the model to replicate the patterns of learning documented in Section 3.4.

Directly observed vs learnable states: To make the exposition of rational inattention to the SPA as clear as possible, I introduce two notational simplifications. I group decisions into a single variable $d_t = (c_t, l_t, a_{t+1})$ and all states other than the SPA into a single state variable $X_t = (a_t, w_t, AIME_t, ue_t, t)$.[16] The stochastic SPA SPA_t is separated because, unlike other state variables, it is not directly observed by the household. Instead, the household must pay a utility cost to receive more precise information about the SPA, as outlined below. The other stochastic state variables, w_t and ue_t being directly observed can be interpreted as these variables being more salient. Rather than any of the other myriad burdens on people's attention I focus on costly attention to the state pension policy because this is the uncertainty that is resolved upon reaching the SPA and hence may help explain why people respond as they do to the SPA.

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

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

information about the SPA to acquire, its choice can be thought of as a two-step process: first choosing a signal distribution and then conditional on the signal draw choosing actions. Although subject to a utility cost of information, the choice of signal is unconstrained; the household is free to learn about SPA_t however they want. More precisely, a household with non-hidden states X_t and $\underline{\pi_t}$ is free to choose any conditional distribution function $\underline{f_t}[X_t, \underline{\pi_t}](z|SPA_t)$ for its signal $z_t \sim Z_t$ given the value of the hidden state SPA_t.

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

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

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: The cost of attention is directly proportional to the mutual information, defined below, between signal and SPA. Mutual information is the expected reduction in uncertainty, as measured by the entropy, about one variable resulting from learning the value of another. Entropy, in this information theoretic sense, is a measure of uncertainty that captures the least space[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 in uncertainty, as measured by entropy, about X from learning Y (equally about Y from learning X) : $I(X,Y) = H(X) - H(X|Y)$.*

Utility: Incorporating information costs, utility takes the form:

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

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

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

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

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

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

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

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, depends upon the full distribution of signals. This means that the continuation value is not known until the solution is known; this problem will be taken up in Section 5.

4.2.3 Discussion of Costly Attention to the Stochastic SPA

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

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

Firstly, although this functional form is not widely used in life-cycle models, this is because most life-cycle models ignore costly information acquisition, not because any other functional form is widely used. In fact, a cost of information acquisition that is directly proportional to the mutual information is among the most common in the costly information literature leading to two important advantages.[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.

[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.

Secondly, it endogenously generates certain rules-of-thumb or heuristics observed sufficiently often to be christened as a behavioural bias. We could treat these simplifying rules-of-thumb, or heuristics, as pre-ordained behavioural rules people blindly follow. This has major disadvantages: one, this does not explain why the particular rule-of-thumb and, two, it ignores the fact that people change rule-of-thumb as circumstances change. Hard-coded behavioural biases suppress a central insight of economics: people respond to incentives. Endogenising observed heuristics with a cost of attention avoids these pitfalls because, one, it explains why a given heuristics is used and, two, allows an agent to change heuristics in response to incentives. Two examples come from Kõszegi and Matějka (2020) who show this cost of attention generates both mental budgeting (quantity allocated to a category being fixed and composition changing) and naive diversification (composition being fixed and quantity allocated changing) depending on the circumstance. A third example comes from Caplin et al. (2019) who show it leads to consideration sets: ignoring many options to focus on a subset.

Thirdly, strong a priori reasons to think that a cost of cognition should depend on entropy exist. The information-theoretic concept of entropy was developed to explain how computers process information and gives a lower bound on the efficient transmission and storage of information. The computational theory of mind Mcculloch and Pitts (1943) holds the human mind is a computer. This is controversial and well outside the scope of this paper, but even its most stringent opponent would agree the brain performs some tasks like a computer, with information processing a primary candidate. So, if the brain 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 in this dimension is reasonable when information processing is our focus. [19]

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

In the broader interpretation learning about the SPA can be taken as illustrative of learning about the state pension system in general. The pension system is multifaceted, and people are confused about most of these facets. The model concentrates all costs of information acquisition onto tracking one aspect of the pension benefit system, the SPA. So the model may also capture learning about these other facets and the resolution of uncertainty about them. Hence, it is possible to think of this cost of learning about the SPA as a cost of learning about pension policy more generally, and I believe the reader taking this perspective can equally draw interesting lessons from this model. In Section 8 I look at an extension in which the household also learns about an uncertain actuarial adjustment to deferred claiming.

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

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

SPA in decision-making. Therefore, including the cognitive cost of remembering and assimilating information as well as any hassle cost is the minimum data and model consistent conceptualisation.

Interpreting the choice of signal: The choice of a signal function to learn about the SPA may be difficult to conceptualise. The SPA is a number we can just look up which seems simpler than choosing a signal function. However, looking it up is a learning strategy encompassed by the choice of a signal function conception, corresponding to choosing a perfectly informative signal function.[20] In reality, carefully reading relevant regulations is not the main way people learn about government policy in general or the state pension in particular: people learn from other people or news outlets. In both examples, there is a random component, what stories newspapers run and what other people talk about, and a choice component, whether you keep reading or ask follow-up questions. This is analogous to the choice of a signal function in that it is partly a choice and partly stochastic, and so it captures much about the messy real-world learning process.

5 Model Solution

By introducing a high dimensional state $\underline{\pi_t}$ (beliefs) and a high dimensional choice $\underline{f_t}$ (signal), rational inattention has complicated the model to the extent that solving it is a contribution. To achieve this I weave together recent theoretical results into a consistent solution method for dynamic rational inattention models with endogenous heterogeneous beliefs, like the one presented above. Section 5.2 explains how this is done, both to communicate the methodological innovations and to provide intuition of the model solution. First I explain details specific to solving the model of this paper.

5.1 Details Specific to this Model

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

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

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

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

The solution of within period problems, when rational inattention matters, because $t < SPA_t$, is explained immediately below in Section 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 $\underline{\pi_t}$ (beliefs distribution) and a high dimensional choice $\underline{f_t}$ (signal distribution). This section presents my solution method. I use the model of retirement decision from this paper to explain the method, but it applies to any dynamic rational inattention models with endogenous heterogeneous beliefs.

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

The rest of this section precedes as follows. Firstly, Section 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. [21] To provide some intuition, and because an understanding of these results is needed to understand the solution methodology, in this section, I present an outline of their proof using my model as a lens through which to explain their results. Steiner et al. (2017) extend Matějka and McKay (2015)[22] to a dynamic setting and so most of what is explained here applies equally to static problems.

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

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

where expectations are over X_{t+1} and SPA_{t+1} and Section 5.2.2 describes finding $\underline{\pi_{t+1}}$, the Bellman equation 13 becomes:

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

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

$$p_t(d|spa) = \frac{\exp\left(n^{(k)} \frac{((c/n^{(k)})^{\nu} l^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \log(q_t(d)) + \beta \overline{V}_{t+1}^{(k)}(d, X_t, SPA_t, \underline{\pi_t})\right)\right)}{\sum_{d' \in \mathscr{C}} \exp\left(n^{(k)} \frac{((c'/n^{(k)})^{\nu} l'^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \log(q_t(d')) + \beta \overline{V}_{t+1}^{(k)}(d', X_t, SPA_t, \underline{\pi_t})\right)\right)}, \tag{14}$$

$$\max_{\underline{q_t}} \sum_{spa} \pi_t(spa) \log\left(\sum_{d \in \mathscr{C}} q_t(d) \exp\left(n^{(k)} \frac{((c/n^{(k)})^{\nu} l^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \beta \overline{V}_{t+1}^{(k)}(d, X_t, SPA_t, \underline{\pi_t})\right)\right). \tag{15}$$

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

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

The next step is a revelation principle type argument. As the household is rational and pays a utility cost for information they will not select any extraneous information. All information has a cost $\lambda I(\underline{f_t}; \underline{\pi_t})$, but only information that leads

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

[22] 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.

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

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

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

$$\max_{\underline{p_t}} \sum_{spa} \pi_t(spa) \sum_{i=1}^{N} p_t(d_i|spa) \Big(n^{(k)} \frac{((c_i/n^{(k)})^{\nu} l_i^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} - I(\underline{p_t};\underline{\pi_t}) + \beta \overline{V}_{t+1}^{(k)}(d_i, X_t, SPA_t, \underline{\pi_t})\Big) \tag{16}$$

and from the symmetry of mutual information: [23]

$$I(\underline{p_t};\underline{\pi_t}) = \sum_{spa} \pi_t(spa) \left(\sum_{d} p_t(d|spa) \log(p_t(d|spa))\right) - \sum_{d} q_t(d) \log(q_t(d)) \tag{17}$$

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

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

Substituting 17 into 16, rearranging, and collapsing the repeated sums gives:

$$\max_{\underline{p_t}} \sum_{spa} \pi_t(spa) \sum_{i=1}^{N} \Big(n^{(k)} \frac{((c_i/n^{(k)})^{\nu} l_i^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \log(q_t(d_i)) - \log(p_t(d_i|spa_i)) + \beta \overline{V}_{t+1}^{(k)}(d_i, X_t, SPA_t, \underline{\pi_t})\Big). \tag{18}$$

Taking $\underline{q_t}$ as given, optimality with respect to any $p_t(d|spa)$ requires the following FOC, derived from differentiating 18, 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,

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

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

$\sum_{d\in\mathscr{C}} p_t(d|spa)=1$. Rearranging gives:

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

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

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

This derivation assumed $\underline{q_t}$ was given, but as $\underline{q_t}$ is the marginal to conditional $\underline{p_t}$ it is also chosen. The form of $\underline{q_t}$ can be found from substituting 14 into 18 and noting that the logarithm of the numerator in 14 cancels all other terms in 18 leaving only the summation from the denominator. So $\underline{q_t}$ can be found by solving:

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

5.2.2 Solution Method

Being the first to solve a dynamic rational inattention model with endogenous heterogeneous beliefs, requires a new solution method. At its core the solution method is to solve 15 for $\underline{q_t}$ and subsitute the solution into 14 to get $\underline{p_t}$. This basic description conceals two major hurdles which this section explains culminating in a description of the algorithm.

The first major difficulty is that next period's beliefs given actions are not known until the full probability distribution of actions is known. This is because we do not know how strong a signal of a given SPA an action is unless we know how likely they were to take that action given other possible SPAs. It follows that next period's effective conditional value function $\overline{V}_{t+1}$ is not known, even when the next period's value function V_{t+1} is known, because we do not know the beliefs tomorrow that will result from an action today. Substituting the results of 14 and 15 into the Bayesian updating formula 12 gives:

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

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 ($\underline{q_t}$) to know next period's beliefs given choices and hence to know the effective conditional continuation values:

$$\overline{V}^{(k)}_{t+1}(d_t,X_t,SPA_t,\underline{\pi_t})=E\left[s^{(k)}_t V^{(k)}_{t+1}(X_{t+1},SPA_{t+1},\underline{\pi_{t+1}})+(1-s^{(k)}_t)T(a_{t+1})\middle|d_t,X_t,SPA_t,\underline{\pi_t}\right]\tag{19}$$

Steiner et al. (2017) evade this difficulty by removing the beliefs from the state space and replacing them with the full history of actions. They can do this because, given initial beliefs, the full history of signals, or equivalently actions, perfectly predicts the beliefs in period t. This is an inspired move for a theory paper and is a key step in extending Matějka and McKay (2015) to the dynamic case.[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. Redundant information in the state space is problematic because the curse of dimensionality means this is often the binding constraints to producing rich models. What moves this here from problematic to a non-starter is that this redundant information grows exponentially with the number of periods.

Hence, I rely on the theoretical results of Steiner et al. (2017) that used the history of action state-space representation, but in practice, I use the more compact belief state-space representation for the actual computational work. To get around the issue that I need $\underline{q_t}$ to know $\overline{V}_{t+1}$ I use a simple guess-and-verify fixed-point strategy. First I guess a value $\underline{\tilde{q}_t}$ and solve the fixed point iteration for the effective conditional continuation value defined by substituting 22 into 23. Then given $\overline{V}_{t+1}$ I solve 15 for $\underline{q_t}$. If resulting $\underline{q_t}$ is sufficiently close to $\underline{\tilde{q}_t}$, I accept this solution otherwise I replace $\underline{\tilde{q}_t}$ with $\underline{q_t}$ and repeat. [26]

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 15. Moreover, if after removing these actions we are left with a single action, then we have solved the problem without further calculation.

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

Removing strictly dominated actions only uses the ordinal information encoded in the utility. Expected utility implies that utility encodes cardinal information as well, which can be exploited using the necessary and sufficient condition from

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

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

Caplin et al. (2019). It is easily shown (see appendix B.2) that if a there exists a decision $d^\star = (c^\star, l^\star)$ which satisfies:

$$\sum_{spa} \pi_t(spa) \frac{\exp\left(n^{(k)} \frac{((c^\star/n^{(k)})^\nu l^{\star 1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \beta \overline{V}^{(k)}_{t+1}(d^\star, X_t, spa, \underline{\pi_t})\right)}{\exp\left(n^{(k)} \frac{((c/n^{(k)})^\nu l^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \beta \overline{V}^{(k)}_{t+1}(d, X_t, spa, \underline{\pi_t})\right)} \geq 1, \tag{20}$$

for all other decisions $d = (c, l)$ then it is the only action taken $q(d^\star) = 1$. Unlike dropping strictly dominated alternative, which reduces the dimensionality and so makes solving equation 15 easier, checking equation 20 is only advantageous when the optimal behaviour is to take the same action in all realisations of SPA_t. As such the benefit of checking condition 20 depends on the problem faced and how frequently it reveals the optimal solution without needing to solve an optimisation. For the retirement model in this paper, it was found useful.

Finally, when sparsity does not provide a shortcut to a solution I employ sequential quadratic programming to solve 15, an approach to static rational inattention problems suggested by Armenter et al. (2019). Hence, bringing this together, a high-level summary of the solution algorithm is:

Remove d from $\mathscr{C}$ that are strictly dominated across all possible combinations of SPA_t and $\underline{\pi_{t+1}}$
if $|\mathscr{C}| = 1$ **then**
    Set $\underline{q_t}$ to degenerate distribution at $d \in |\mathscr{C}|$
else
    Set initial value of $\underline{\tilde{q}_t}$ and Error > Tolerance
    while Error > Tolerance **do**
        Solve for $\overline{V}_{t+1}$ given $\underline{\tilde{q}_t}$
        Remove d from $\mathscr{C}$ that are strictly dominated across all possible SPA_t given $\underline{\pi_{t+1}}$
        if $|\mathscr{C}| = 1$ **then**
            Set Error = 0 < Tolerance and $\underline{q_t}$ to degenerate distribution at $d \in |\mathscr{C}|$
        else
            if there is an action d that satisfies 20 **then**
                Set Error = 0 < Tolerance and $\underline{q_t}$ to degenerate distribution at d
            else
                Solve 15 using sequential quadratic programming for $\underline{q_t}$
                Set Error to distance between $\underline{q_t}$ and $\underline{\tilde{q}_t}$
                Update $\underline{\tilde{q}_t} = \underline{q_t}$
            end if
        end if
    end while
end if

Substitute $\underline{q_t}$ into 14 to solve for $\underline{p_t}$.

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

$$\log(w_{j,t}) = \delta_{k0} + \delta_{k1}t + \delta_{k2}t^2 + \varepsilon_{j,t} + \mu_{j,t} \tag{21}$$

for individual j, of type k, in period t, where period t is indexed by female age and type k indicates whether high or low education and single or married. The parameters of the age-dependent deterministic component of the wage process $(\delta_{k0}, \delta_{k1}, \delta_{k2})$ are estimated by type-specific regression. The parameters of the stochastic component of the wage equation

$(\rho_w, \sigma_\varepsilon, \sigma_{\varepsilon,55}, \sigma_\mu,)$ are estimated using a standard approach (e.g. Guvenen, 2009; Low et al., 2010) that chooses values that minimise the distance between the empirical covariance matrix of estimated residuals and the theoretical variance covariance matrix of $\varepsilon_t + \mu_{j,t}$.

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

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

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

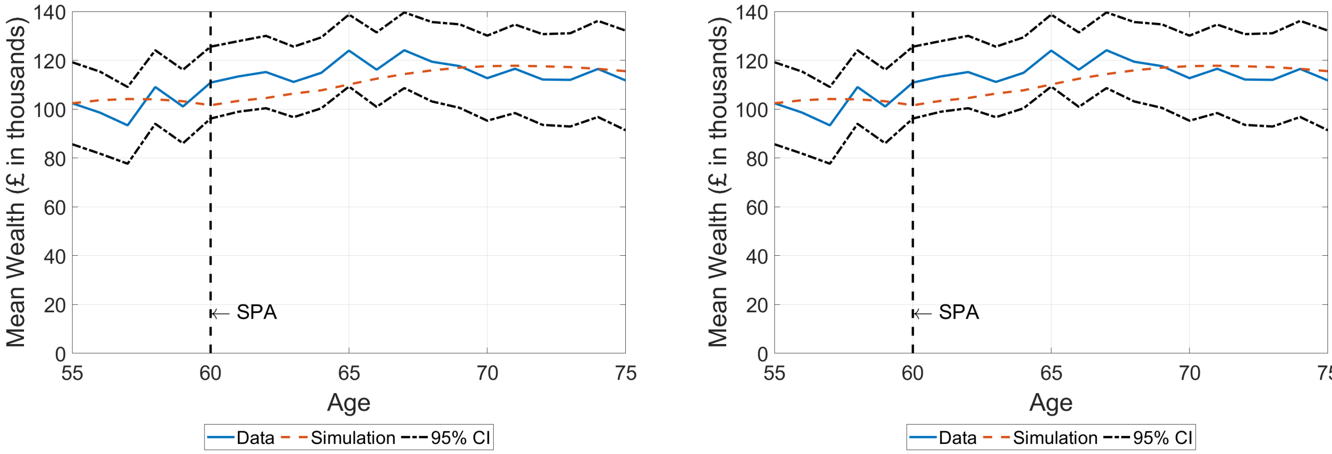
6.2 Second Stage

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.

Due to the novel nature of the cost of attention parameter in the lifecycle literature, I investigated a range of values for λ alongside attempts to identify it from the reduction in self-reported SPA mean squared error between 55 and 58. Estimation of λ is done separately from targetting the other moments and holding the values of the other parameters

Figure 5: Fit to Targetted Profiles



(a) Employment Profile

(b) Asset Profile

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

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. Finally, 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 are in appendix E.1.

The model with policy uncertainty attains a good fit to pre-reform employment and asset profile with $SPA = 60$ as shown in Figures 5a and 5b, and Table 4 contains the estimated parameter. The graphs for the baseline and the version with policy uncertainty together with rational inattention are very similar and are in appendix E.2. However, in the response generated to the dynamic SPA, these model versions clearly distinguish themselves.

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

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.

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 the treatment indicator, a full set of age and cohort fixed effects (not date as it is not distinct from age in the model) and the controls from the empirical specification having counter-parts in the model. As in Section 3.3, I repeat 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 repeats the empirical estimates of these treatment effects found in columns 1 and 2 of Table 1, hence the difference between column 5 and column 1 shows the baseline model struggles to match both the aggregate response to SPA and the correlation of this response with wealth.

This failure of the baseline reflects the excess employment sensitivity puzzle that led to investigations of policy uncertainty and costly attention. To examine their impacts separately, I introduce them sequentially. Column 2 shows policy uncertainty alone makes little to no difference. This is because the level of objective policy uncertainty is low; we observe changes to the SPA arrive infrequently. Both this version and the baseline fall short of matching the treatment effect in both the whole population and the population with above median assets at SPA but are closer to the lower treatment effect in the richer subpopulation. [27]

Column 3 shows the result with costly attention to the stochastic SPA, with a relatively arbitrary cost of attention of $\lambda = 3 \times 10^{-6}$ that fits these treatments' effect tolerably well.[28] The treatment effect in both the whole population and those with above median assets move significantly toward the data, falling in the confidence intervals.[29]

SPA self-reports in ELSA offer an opportunity to improve on this arbitrary value of λ, as they offer clear and direct identifying variation. Exploiting this, I identify λ from the reduction in mean squared error in self-reported SPAs between ages 55 and 58 for the cohort with a SPA of 60, the cohort simulated during estimation. The middle panel of Table 5 shows these numbers. Column 4 shows that for a smaller cost of attention $\lambda = 1.3 \times 10^{-7}$, the degree of learning, captured by the reduction in mean squared error, is well matched. For this lower value of λ, the fit for the employment response to

[27] Section 3.3 shows the response by the rich is puzzling ex-ante, Appendix C shows, if we directly target treatment effects, the baseline matches that of the whole population but not of the rich. With these parameters estimated, however, the baseline struggles most with the aggregate.

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

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

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	0.019	0.023	0.052	0.037	0.080
[95% C.I]					[0.044,.0116]
Assets >Median(£29,000)	0.027	0.032	0.052	0.040	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 two 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 SPA is worse, although it still improves on the fit of the baseline and the model with only policy uncertainty. For the larger value of λ in column 3, we see that knowledge of the SPA actually gets worse between age 55 and 58, indicating that if any learning is happening, it is getting drowned out by drift from agents updating with the known laws of motion.

That people better informed of their SPA in their late 50s have a smaller labour supply response at their SPA in their 60s (see Table 3) was an impetus to investigating the role of endogenous information in this excess sensitivity puzzle. A natural question is whether the model replicates this relationship. Two countervailing model forces exist linking SPA knowledge to the labour supply response at the SPA. On the one hand, SPA knowledge is endogenous, implying those whose actions depend least on the SPA acquire the least information about it. On the other, comparing two ex-ante equivalent households where, by luck, one ended up worse informed than the other, the worse informed household receives a larger shock upon discovering their SPA and so has a larger reaction. Which dominates determines whether the model generates a positive or negative relationship between SPA knowledge and the labour supply response at the SPA. The bottom panel of Table 5 shows a negative relationship for the smaller λ and no relationship for the larger.

The fact that the model replicates the key facts from the data indicates it has the mechanism required to explain the data. A limitation is that it requires different values of λ to replicate different facts. Two explanations spring to mind: the levels of some incentives may be misaligned, or other mechanisms may contribute to the excessive employment response. Unidimensional policy uncertainty is a massive simplification, many aspects of pension policy are uncertain, which may explain misalignment of incentives. Section 8 enriches pension policy uncertainty, but data limitations make the extension necessarily more speculative. Section 7.3.2 introduces a norm to retire at SPA alongside costly attention.

Table 6: Summary Statistics of Attention Cost Converted to Compensating Assets (£)

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

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

7.2 Model Implications

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

Table 6 shows summary statistics of the distribution of compensating assets for $\lambda = 1.3 \times 10^{-7}$ and $\lambda = 3 \times 10^{-6}$. For $\lambda = 1.3 \times 10^{-7}$, the cost range from £0.32 at the 5th percentile to £37.74 at the 95th with a mean of £15.37, but there is substantial heterogeneity. The correlations with assets are negative, indicating that information frictions impose the highest cost on the poorest members of society. For $\lambda = 3 \times 10^{-6}$, qualitatively a similar picture arises, but the costs are higher. These costs give an upper bound on welfare gains from reducing uncertainty by, for example, sending information letters. The gains are modest, but so are the marginal costs. Once the fixed administrative and technological costs have been paid, marginal costs can include little more than postage which costs £0.68.

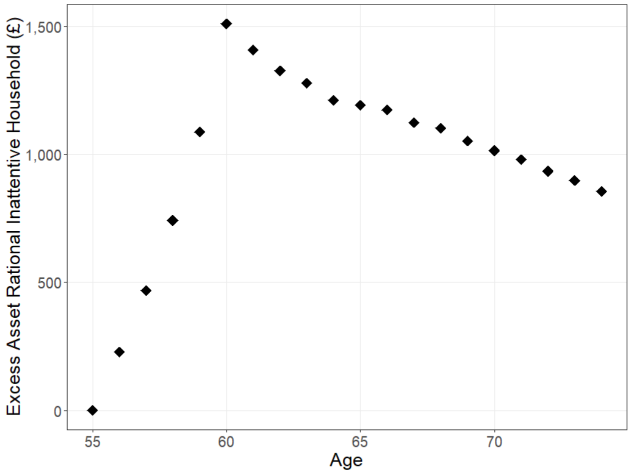
Who learns and when The model allows us to investigate who learns and when. To investigate this, I run a pooled regression of bits of information acquired on the states of the model and indicators of age for women 55-59 with a SPA of 60. The results are unsurprising and, for this reason, are in the appendix, Table 20. On average: a positive amount of information is gathered, those with more alternative resources (assets and labour income) gather less information, and those with higher future pensions (AIME) gather more. Most information is acquired at age 55, declining until age 58 and then jumping back up at age 59, immediately before the SPA.

7.3 Model Predictions

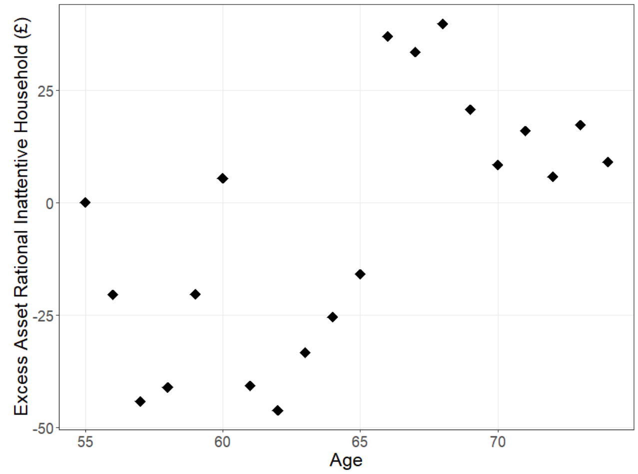
7.3.1 Optimal Savings

Retirement preparedness is a concern, particularly in countries where private savings accounts form a growing fraction of retirement income, but the academic literature fails to reach a consensus on whether households under or over save for

Figure 6: Excess Saving



(a) $\lambda = 3 \times 10^{-6}$



(b) $\lambda = 1.3 \times 10^{-7}$

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

retirement (e.g. Scholz et al., 2006; Crawford and O'Dea, 2020). As this paper takes age 55 assets as given, its ability to answer this question is limited, but comparing the savings of the rationally inattentive household to the frictionless benchmark (i.e. with only policy uncertainty) illuminates how informational frictions impact retirement preparation.

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

7.3.2 Increasing Old Age Participation with the Pension Ages

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

The left panel of Table 7 shows the change in mean employment resulting from increasing the SPA from 60 to a SPA in the range 61-65 for the model with $\lambda = 1.3 \times 10^{-7}$ and with policy uncertainty alone. Averages are shown over prime working years (55-65) and all working life (55-79). Focusing on the SPA increase to 65, with costly attention mean employment increases by 0.33 years over 55-65 and 0.42 years over 55-79; this compares to 0.40 and 0.49 additional

Table 7: Additional Mean Employment from Increasing SPA from 60

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

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

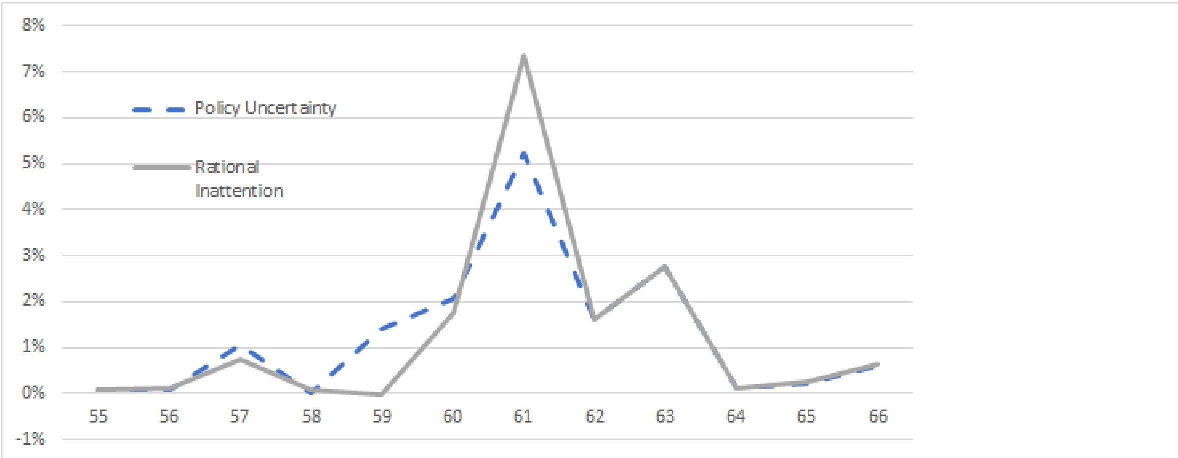
years in the model without informational frictions. Neglecting costly attention overestimates the employment increase by 23% over 55-65 and 15% over 55-79. The reason for a smaller response with costly attention is the rationally inattentive household is less aware of SPA changes, so increases labour supply less in the lead-up to their new SPA. Post their new SPA, the rationally inattentive agent works more to compensate. This reduces but does not overturn the difference over 55-79 because of imperfect intertemporal substitutability and lower employment at older ages.

Costly attention generates a smaller overall increase in employment but a larger response at the SPA because much of the bunching at SPA reflects intertemporal shifting of employment. Fully aware households immediately internalise changes to the SPA, increasing labour supply when the woman is in her 50s. The rationally inattentive household only partially responds until they realise, much closer to the SPA, the need to make up for lost time. Figure 7 illustrates these dynamics for an increase of the SPA to 62.

Costly attention only partially explains excess employment sensitivity (see Table 5), leaving room to doubt these labour supply predictions. Also, evidence framing effects, or norms, affect the employment response to pension ages (e.g. Seibold, 2021; Gruber et al., 2022) should be considered alongside evidence presented in this paper for the importance of mistaken beliefs. A simple way to include a norm to stop working at the SPA, alongside costly attention, is to have a fraction of passive decision makers in the style of Chetty et al. (2014) or Lalive et al. (2017). I introduce a fraction of naive-passive agents, not anticipating their passive retirement, and compare predictions of models with and without costly attention after the inclusion of these naive-passive agents. The model without costly attention requires 11% of women are passive to replicate the employment response to the SPA, whilst with $\lambda = 1.3 \times 10^{-7}$ requires only 8%.

The left panel of Table 7 shows these results. Focusing again on the SPA increase to 65, costly attention again predicts smaller employment increases, but now the difference is larger. When the SPA is 60, passive retirees decrease employment

Figure 7: Increase in Employment from Increasing SPA 60 to 62



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

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

8 Extension: Deferal Puzzle

Section 7 shows that, even though the model can explain each feature of the data in isolation, it requires different costs of attention to replicate different features pointing toward misalignment in incentives. As the stochastic State Pension Age (SPA) understates policy uncertainty about the State Pension, this is a natural place to look for this misalignment.

For this reason, I introduce learning and uncertainty about another aspect of the state pension system: the actuarial adjustments to benefits from deferring. Combined with a claiming decision, this not only helps to align incentives by making the model more realistic but also helps explain the deferral puzzle (detailed below). Since the adjustments rate becomes irrelevant upon claiming, rational inattention to this aspect of the pension system speaks directly to this puzzle because calculations implying deferral is actuarially favourable ignore the attention benefits of claiming: claiming removes the need to pay attention to this adjustment rate. The model of Section 4.2 does not incorporate such a mechanism for two reasons. Firstly, it does not include a benefit-claiming decision. Secondly, the only source of uncertainty subject to an attention cost is the SPA, and once it is reached, the uncertainty resolves, irrespective of claiming. The simplest extension that contains this new incentive to claim is presented in the rest of this section, along with some results.

8.1 Deferral Puzzle

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

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

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

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

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.

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

The two points of support are chosen as 10.4% and 5.8%, the actuarial adjustment from 2006 to 2016 and post-2017 respectively. The probability of being offered the higher actuarial adjustment of 10.4% is chosen to match the average actuarial adjustment since 1955, resulting in a probability of 0.415. Deferral rules are taken from Bozio et al. (2010) and

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	0.080
Assets >Median(£29,000)	0.0903	0.061
Age	Variance of SPA Answers	
55	2.985	2.852
58	1.795	1.180
Coefficient	Treatment Effect Heterogeneity by SPA Error	
Treatment Effect	0.0532	0.157
Interaction	-0.0111	-0.023

Notes: Costly attention refers to the model with, additionally, a cost of information acquisition about the stochastic policy. The top panel shows labour supply response across the wealth distribution as per Table 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.

since earlier deferral rules were previously stated in absolute rather than percentage terms, the ONS time series of state pension spending going back to 1955 (https://www.gov.uk/government/publications/benefit-expenditure-and-caseload-tables-2021) is used to work out implied average percentage deferral adjustments.

The model with policy uncertainty, to the stochastic SPA and adjustment rate, is then re-estimated to match the same pre-reform employment and assets profiles with a constant realisation of 10.5% for the deferral adjustment, which was the deferral rate these cohorts faced. Parameter estimates are in Table 8 and, for these values, only 6.2% of individuals claim the state pension before the mandatory claiming age of 70, much lower than the 99% plus claiming seen in the data.

Next, I introduce costly attention with a cost of attention corresponding to approximately £10 of consumption to the median consuming household to be fully informed. This increased the number voluntarily claiming to 22.2%, approximately a fourfold increase on the model without informational frictions, but still short of the rate observed in the data. As can be seen in Table 9, this cost of attention produced a relatively good fit along all dimensions of interest.

9 Conclusion

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

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

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

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

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

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.

Banks, J., G. Batty, J. Breedvelt, K. Coughlin, R. Crawford, M. Marmot, J. Nazroo, Z. Oldfield, N. Steel, A. Steptoe, M. Wood, and P. Zaninotto (2021). English longitudinal study of ageing: Waves 0-9. 1998-2019 [data collection].

Barr, N. and P. Diamond (2009). Reforming pensions: Principles, analytical errors and policy directions. *International social security review 62*(2), 5–29.

Bartoš, V., M. Bauer, J. Chytilová, and F. Matějka (2016, jun). Attention Discrimination: Theory and Field Experiments With Monitoring Information Acquisition. *American Economic Review 106*(6), 1437–1475.

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.

Cesarini, D., E. Lindqvist, M. J. Notowidigdo, and R. Östling (2017). The effect of wealth on individual and household labor supply: evidence from swedish lotteries. *American Economic Review 107*(12), 3917–46.

Chetty, R. (2012). Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply. *Econometrica 80*(3), 969–1018.

Chetty, R., 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.

Kalyvas, C. and T. Tzouramanis (2017). A survey of skyline query processing. *arXiv preprint arXiv:1704.01788*.

Koşar, G. and C. O'Dea (2022). Expectations Data in Structural Microeconomic Models. *National Bureau of Economic Research*.

Lalive, R., A. Magesan, and S. Staubli (2017). Raising the Full Retirement Age : Defaults vs Incentives. *NBER Working Paper*, 1–57.

Landais, C., J. Kolsrud, D. Reck, and J. Spinnewijn (2021). Retirement consumption and pension design.

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.

Powell, M. J. (2009). The bobyqa algorithm for bound constrained optimization without derivatives. *Cambridge NA Report NA2009/06, University of Cambridge, Cambridge 26*.

Ravid, D. (2020, sep). Ultimatum Bargaining with Rational Inattention. *American Economic Review 110*(9), 2948–2963.

Rohwedder, S. and K. Kleinjans (2006). Dynamics of Individual Information about Social Security. *RAND WP*.

Rust, J. and C. Phelan (1997). How Social Security and Medicare Affect Retirement Behavior In a World of Incomplete Markets. *Econometrica 65*(4), 781.

Schittkowski, K. (2014). Nlpqlp-nonlinear programming with non-monotone and distributed line search.

Scholz, J. K., A. Seshadri, and S. Khitatrakun (2006). Are americans saving “optimally” for retirement? *Journal of political economy 114*(4), 607–643.

Seibold, A. (2021). Reference points for retirement behavior: Evidence from german pension discontinuities. *American Economic Review 11*(4), 1126–1165.

Sims, C. (2003). Implications of rational inattention. *Journal of Monetary Economics 50*(3), 665–690.

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 coincided with a compulsory retirement age. The equality Act (2010) went further and banned all compulsory retirement ages with a handful of specific exceptions known as employer justified retirement ages (EJRA).

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

The plaintiff in Seldon v Clarkson, Wright and Jakes (2012) was a partner in a law firm, and it was judged that this EJRA was justified. Documented cases of EJRAs are relative few; apart from partners in law firms, two of the most discussed EJRA are at the UK top Universities: Oxford and Cambridge. Other UK universities appear to have removed compulsory retirement requirements ages, and interestingly Oxford recent lost an employment tribunal that judged that their EJRA was not justified. Ewart v University of Oxford (2019) found that although the objective of Oxford's EJRA (intergenerational fairness) was valid, an EJRA was not a proportionate way to achieve this due to limited demonstrated effectiveness weighed against its clearly detrimental impacts. Hopefully, this goes someway to illustrate that UK law treats forced retirement very seriously as age discrimination and that the few expectations made are precisely that: exceptional.

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

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.

experience a significant treatment effect of a 0.162 increase in her probability of being in work from being below the SPA. NHNBW of £409,000 seems ample to smooth labour supply over the horizon of one to three months. So, although wealth matters for the impact of the SPA on employment, it seems liquidity constraints cannot explain away the effect.

A.3.2 Ristricted Asset Categorisation

As the goal of investigating treatment effect heterogeneity by asset holdings is to understand the role played by liquidity constraints, the main text is restricted to NHNBW. However, parts of NHNBW can be illiquid, and so in Table 11 repeat the analysis but for a more restricted asset category, very liquid asset, which is only assets that can reasonably be liquidated in a matter of weeks. As can be seen, the results are qualitatively very similar to those using NHNBW and do not support liquidi constraint alone explaining away the treatment effect. The treatment effect for those with above median assets is still positive, and although the difference between the two subgroups is now significant, column 4, containing the continuous interaction terms, shows that, again, this heterogeneity is too weak for the treatment to be completely explained by liquidity constraints.

A.3.3 Bad Control Concerns

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

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

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

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

Figure 8 shows the dynamic treatment effects before and after the SPA. There is no indication of violated parallel trends of anticipation effects as none of the pre-SPA treatment effects are significantly different from zero. Indeed jointly testing for violations of parallel trends fails to reject the null of parallel trends ($p = .799$). Conversely, 7 of 9 post-SPA treatment effects are individually significant, and we can easily reject the null ($p = .000$) of them being jointly zero. The graph also doesn't provide much indication that the post-SPA treatment effects differ from each other, although we can reject that hypothesis ($p = .198$).

Figure 9 looks at whether these individual treatment effects vary between waves. The treatment effects look quite uniform across waves, although, again, we can reject this hypothesis ($p = .137$). However, neither violation of homogeneity seems to serve, and generally, the graphs look supportive of the interpretation of a homogenous treatment effect that turns on at the SPA (as assumed in the baseline), although the statistical test show this is only an approximation.

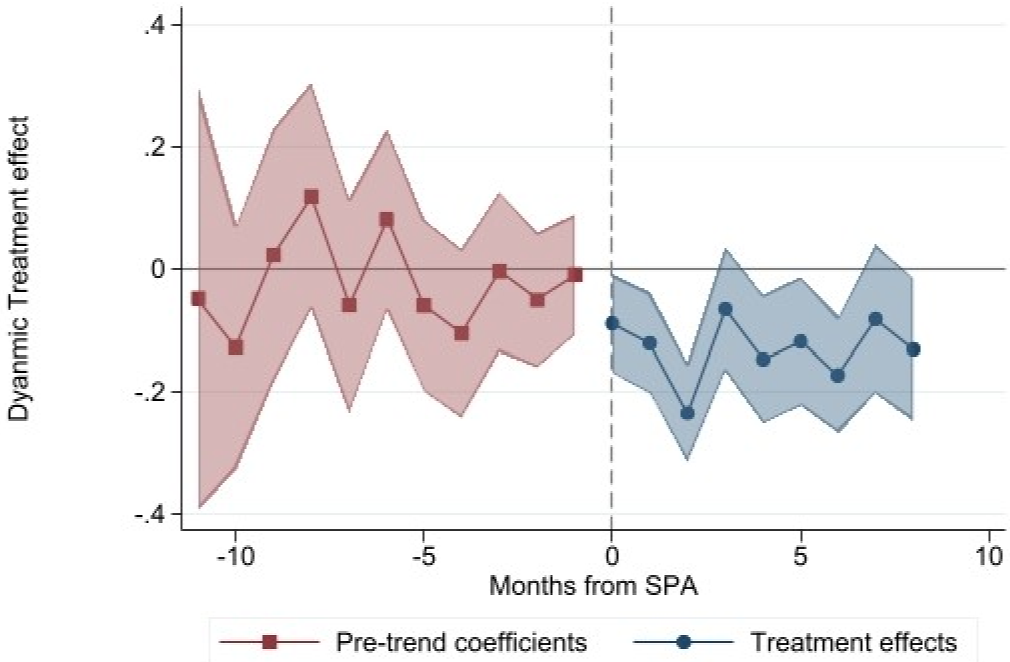
If you are more concerned about the violations of homogeneous treatment effects, then these results show that even allowing for arbitrary heterogeneity, there is something special happening at the SPA which is difficult to explain in standard complete information models.

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.

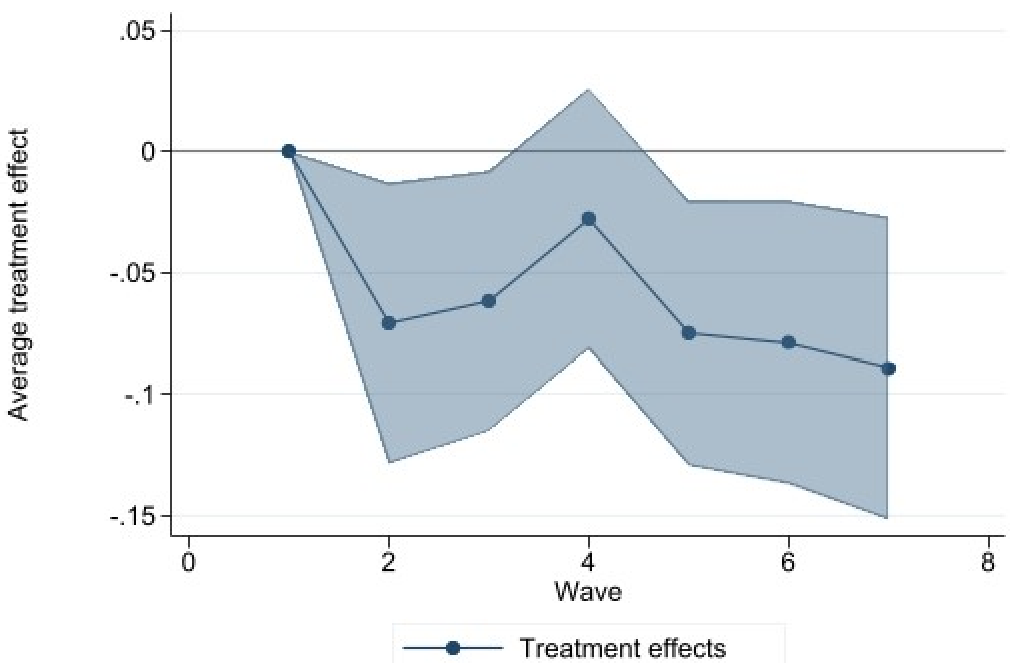
Wealth effects play an important role in determining labour supply, and women who have a later SPA are lifetime poorer. The puzzle is not that they have a higher labour supply; the puzzle is that their labour supply response should be concentrated at the SPA, the change in SPA having been announced over 15 years prior to any affected individual reaching their SPA. In standard complete information life-cycle models, the affected individuals should have a higher labour supply due to the wealth effect, but the response should be spread over their life, not concentrated at the SPA itself. In equation 1 differences in lifetime wealth, including those induced by SPA differences, between year-of-birth cohorts are absorbed

Figure 8: Dynamic Treatment Effects by Time from SPA



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

Figure 9: Average Treatment Effect by Wave



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

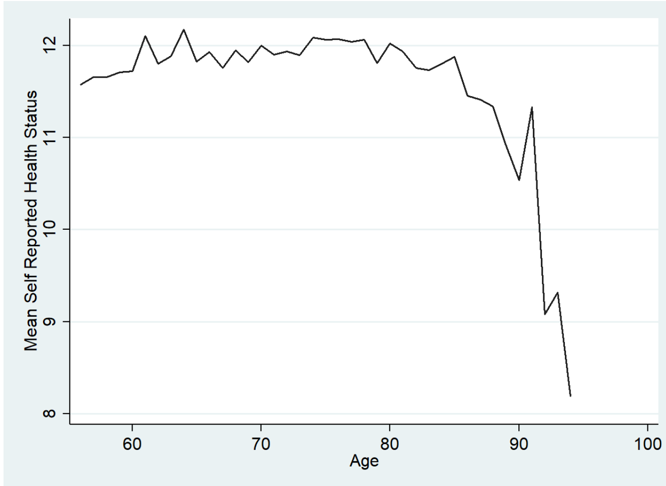


Figure 10: Self Reported Health Profile

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

Health is a major determinant of retirement behaviour (e.g. De Nardi et al., 2010). However, there is no reason to suspect it interacts with the SPA, so no reason for it to explain employment's sensitivity to the SPA. Furthermore, during the period studied, the SPA was in the range 60-63, and, at the mean, health status does not start to deteriorate until later in life. This can be seen in Figure 10, 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.

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 pensions

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

do not generally offer different eligibility ages to men and women[30], and this reform only changed the female SPA. Still, checking for a correlation between the SPA and normal pension ages (NPA) of private pension schemes would be desirable. Checking this directly in ELSA is complicated by the fact that only self-reported NPAs are available. For the SPA, where alongside self-reports, we know an individual's true SPA, these self-reported ages are unreliable, as is documented in Section 3.4. However, only defined benefit pension systems have NPAs, as defined contribution schemes can be accessed from age 55. Hence, dropping everyone with over £2,000 in a defined benefit scheme from the sample rules out an unlikely correlation between the female SPA and pension schemes NPAs from explaining the results. This is done in Table 16, and as can be seen, despite the loss of power, the treatment remains present and significant.

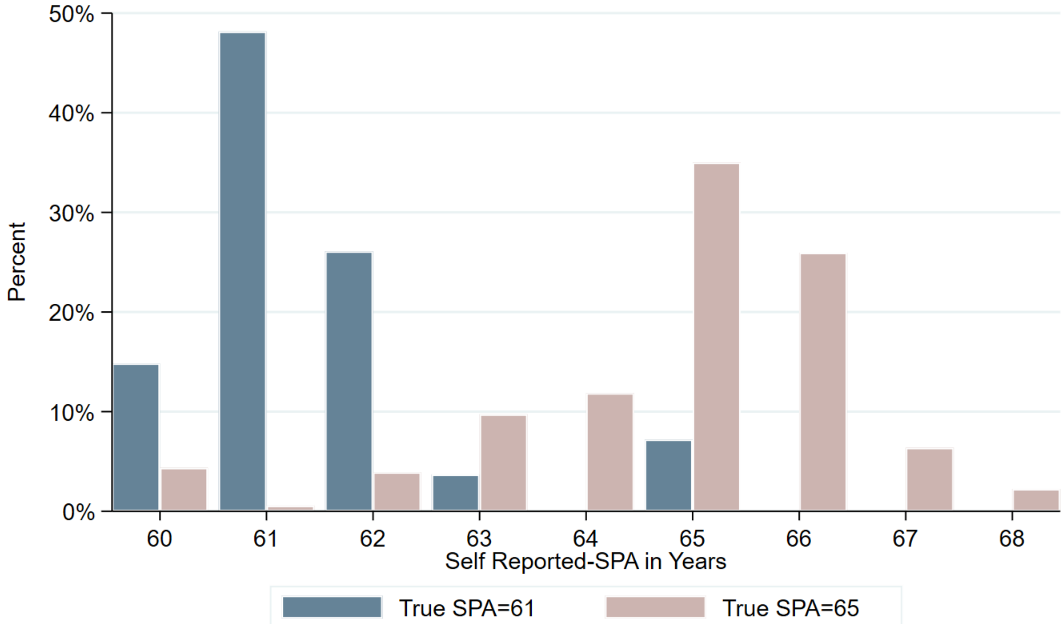
A.4 Mistaken Beliefs and Excess Employment Sensitivity

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

[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 11: 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.

acquisition.

Figure 12 shows that error in self-reported SPA at age 58 that was documented in the main text, but here at the true monthly frequency. Little that is relevant to the model is added by looking at the lower level of variation. We see that 31% know their own SPA to precisely the right month. The main thing we can glean from this graph that we cannot when the date is binned at a yearly frequency is that the spike every twelve months here show that people display an unsurprising round number bias.

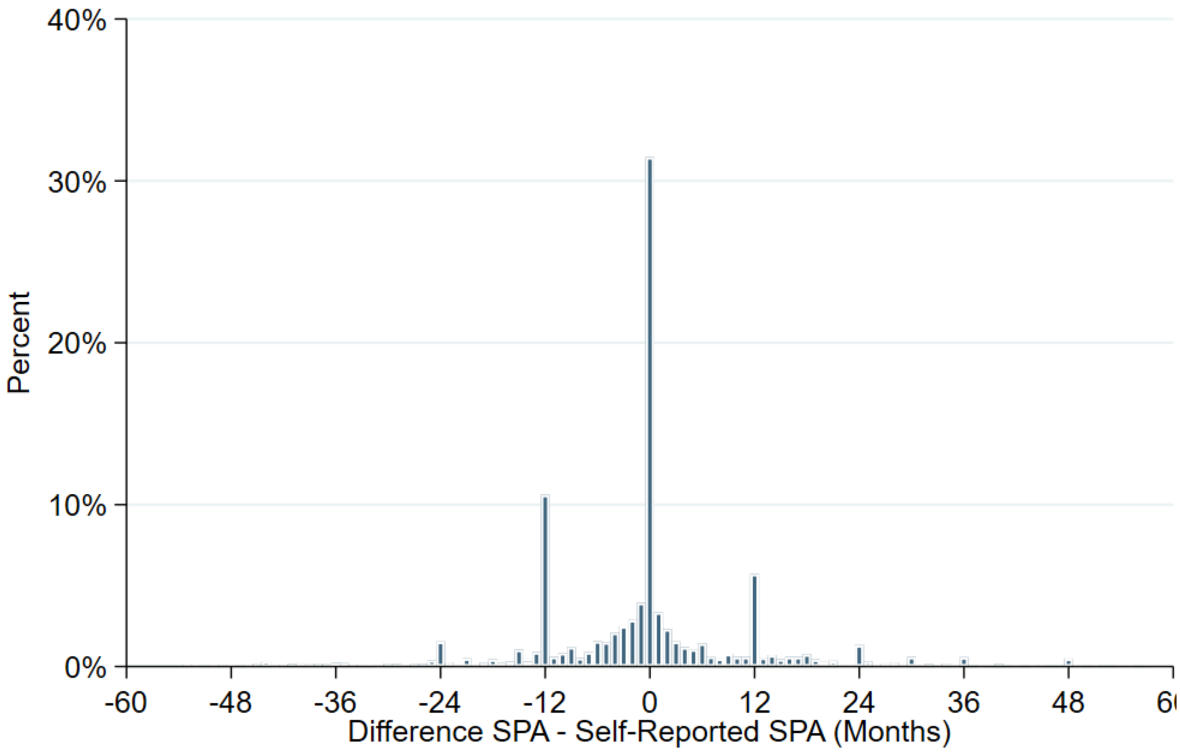
A.6 Treatment Effect Hetrogeneinety by Beliefs

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

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

Firstly, those whose beliefs imply they will receive a positive shock (i.e. they overestimate the SPA) are the ones to have the largest labour supply response to it. This can be seen in Table 15, which shows how the treatment effect varies between three groups: those who think their SPA is sooner than it is, those who correctly state their SPA, and those who think it is further away than it is. The last group is the excluded category, and we can see that for this group,

Figure 12: 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

the treatment effect is positive and significant at 0.785. Then for those who underestimate their SPA, we see that their predicted treatment effect is smaller and loses significance when the sum of coefficients is tested jointly ($p = .0999$). So those who receive a negative shock do display a labour supply response to the SPA.

Secondly, those whose knowledge of the SPA got better had the largest labour supply response to the SPA. This can be seen in Table 16 which shows how the treatment effect varies between three groups: those whose SPA self-reports get worse between the first and last time they are asked, those whose self-reports stay the same, and those whose self-reports get better. It can be seen that those whose knowledge improves have the largest labour supply response to the SPA.

B Additional Mathematical Details

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

My model does not quite fit into the framework of Steiner et al. (2017) because I have made slightly different assumptions about the information the agents receive costlessly. In this section, I first present a quick summary of their model in which I highlight the assumption that is not compatible with my model. Then I discuss mapping my model into their framework and where it fails. Next I present my alternative assumption that alows me to resurect the results of Steiner et al. (2017) and I present a proof of the key lemma starting with this different timing assumpiton about costless infomration. For comparability, in this section, I adopt much of the notation of Steiner et al. (2017), and the notation is not related to the rest of the paper.

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

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

$$\max_{f,\sigma} E[\sum_{t=0}^{T} \beta^t (u(\sigma_t(z^{t-1}, x_t), \theta^t) - I(\theta^t, x_t|z^{t-1})] \tag{22}$$

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

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

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

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

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

$$\max_p E[\sum_{t=0}^{T} \beta^t (u(d^t, \theta^t) - I(\theta^t, d_t|z^{t-1})] \tag{23}$$

where we redefine $z^{t-1} = (d^{t-1}, y^t)$ the expectation is with respect to the distribution over sequences (θ_t, z_t) induced by p, the prior π, and the distributions g. Conversely, any choice rule p solving 23 induces a strategy solving the dynamic RI problem.

Proof. We proceed in steps.

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

$$E[\sum_{t=1}^{\infty} \beta^t I(\theta^t, \zeta_t|z^{t-1})] = E[\sum_{t=?}^{\infty} \beta^t (H(\theta^t|\zeta^{t-1}, y^t) - H(\theta^t|\zeta^t, y^t))] \tag{24}$$

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

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

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

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

So 24 becomes

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

Step 2: Given strategy s and the choice rule generated by it p by construction they generate the same gross utilities. Hence by step 1, 23-22 is:

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

But then $|B|\leq|X|<\infty\Rightarrow b^t$ is measurable wrt x^t and hence $E[H(\theta^t|b^t,y^{t+1})]\geq E[H(\theta^t|x^t,y^{t+1})]$ and $E[H(y_{t+1}|b^t,y^t)]\geq E[H(y_{t+1}|x^t,y^t)]$ and therefore 23≥22 .

Step 3: As $B\subset X$ if p is a probability choice rule then $f_t(x_t|w^t)=p_t(b_t|\omega^t)$ and $x_t=\sigma_t(z^{t-1},x_t)$ is a viable solution to 22. For this strategy generated by this mapping, the probability choice rule makes equation23=equation22

Step 4: If s solves 22 the corresponding PCR p must solve 23, as by step 2 the value from p in 23≥s in 22, so if p doesn't solve 23 ∃ PCR producing greater net lifetime utiltiy than s in 22. But by step 3 this produces a viable solution to 22 with greater net lifetime utility contradicting s being a solution to 22.

Step 5: If p solve 23 then by step 3 it produces a viable solution to 22 but then 23≥22 so this strategy must be the optimal solution to 22 □

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

B.2 Finding Unique Actions Using Second Order Conditions

Using the SOC of the rationally inattentive agent's problem Caplin et al. (2019) provide an alternative formulation of the solution of the model. If the CCP statisfy equation 14 and for all possble action ($\forall d = (c,l) \in \mathscr{C}$)

$$\sum_{spa} \pi_t(spa) \frac{\exp\left(n^{(k)} \frac{((c/n^{(k)})^{\nu} l^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \beta \overline{V}^{(k)}_{t+1}(d, X_t, spa, \underline{\pi_t})\right)\Big)}{\sum_{d' \in \mathscr{C}} q_t(d') \exp\left(n^{(k)} \frac{((c'/n^{(k)})^{\nu} l'^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \beta \overline{V}^{(k)}_{t+1}(d', X_t, spa, \underline{\pi_t})\right)\Big)} \leq 1, \tag{25}$$

with equality if $q_t d > 0$. This new condition from (Caplin et al., 2019) replaces the need for the unconditional choice probabilties to solve the log-sum-exp of equation 15.

If an action $d^\star = (c^\star, l^\star)$ statifies equation **??** repeated here:

$$\sum_{spa} \pi_t(spa) \frac{\exp\left(n^{(k)} \frac{((c^\star/n^{(k)})^{\nu} l^{\star 1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \beta \overline{V}^{(k)}_{t+1}(d^\star, X_t, spa, \underline{\pi_t})\right)\Big)}{\exp\left(n^{(k)} \frac{((c/n^{(k)})^{\nu} l^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \beta \overline{V}^{(k)}_{t+1}(d, X_t, spa, \underline{\pi_t})\right)\Big)} \geq 1, \tag{26}$$

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

If such a $d^\star$ exist the only way to satisfy 25 for $d^\star$ is to have $q_t(d^\star) = 1$.

C Additional Computational Details

C.1 Solving the Models without Costly Attention

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

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 probability masses of a given size between the eight possible SPAs 60-67. As no amount of Bayesian updating can change the assignment of zero probability to an outcome, I want to avoid having beliefs that assigned zero probability to SPAs in my gridpoint of beliefs, and so I imposed a minimum probability to be assigned to each SPA of 0.01 and then had the probability masses that are moved about be in addition to this minimum amount. To make this more concrete, I broke the total probability into four masses that I moved between SPAs to form the grid over beliefs. In the absence of this minimum probability of any SPA, that would mean the probability masses being moved between SPAs was of a size of 0.25. In periods in which there are eight possible SPAs, because $t < 60$ and the women have not aged past any possible SPA, these probability masses are of the size $\frac{1-0.08}{4} = 0.23$. When $t < 60$, having these four probability masses to move between 8 possible SPAs leads to a total of $\binom{7+4}{4} = 330$ grid points because each combination can be thought of as an ordering of the four masses and the breaks between the eight grid points. As the women successively age past their SPAs, this 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 lexicographic ordering, which is convenient for constructing all possible combinations of the probability masses.

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

Range of Attention Costs As explained in Section 5.2, during periods in which and at states at which rational inattention matters because $t < SPA_t$ the central equation that needs to be solved to find the households optimal decision is the following:

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

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

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

C.3 Simulating and Estimating

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

Given these initial conditions, I simulate the choice of the individual households using the decision rule found when solving the model and the exogenous process estimate in the first stage. I then aggregate the simulated data in the same way we aggregate the observed data and construct moment conditions. I describe these moments in greater detail in appendix D. The method of simulated moments procedure delivers the model parameters that minimise 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 (Powell, 2009) 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 quintic in NHNBW plus a very rich set of additional controls that include variables on health, education, location, labour market behaviour, housing tenure, cohort, age, wage, and measure of cognitive ability. This includes as much information as possible to impute $AIME_{55}$.

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

D.2 Type-specific Mortality

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

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

Where $Pr(type=k|death_t)$ and $Pr(type=k)$ are taken from ELSA and $Pr(death_t)$ are taken from the 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 the year-of-birth cohort of an individual, SPA_i is her final SPA, $age_{i,t}$ her age in years, U_t aggregate unemployment to half a per cent, and the outcome variable y_{it} is either assets or employment depending on which profile is being calculated.

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

E Additional Results

E.1 First Stage Estimates

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

Initial conditions Initial assets a_{55} and average earning $AIME_{55}$ are set from the type-specific empirical joint distribution, some summary statistics of which are presented in Table 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}. Higher education increase both variables.

Labour market conditions The type-specific transition probabilities, estimated with individuals classified as unemployed when they claim unemployment benefits, are shown in Table 18.

The parameters of the stochastic component of the wage process (persistence and the variance of innovation, measurement error, and initial draw) are shown in Table 19

The deterministic component of wages generates the wage profiles in Figure 13. Spousal Income is shown in Figure 14.

Table 17: Sumarry Statistics of Initial Conditions (£)

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

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

Table 18: Type Specific Unemployment Transition Probabilities

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

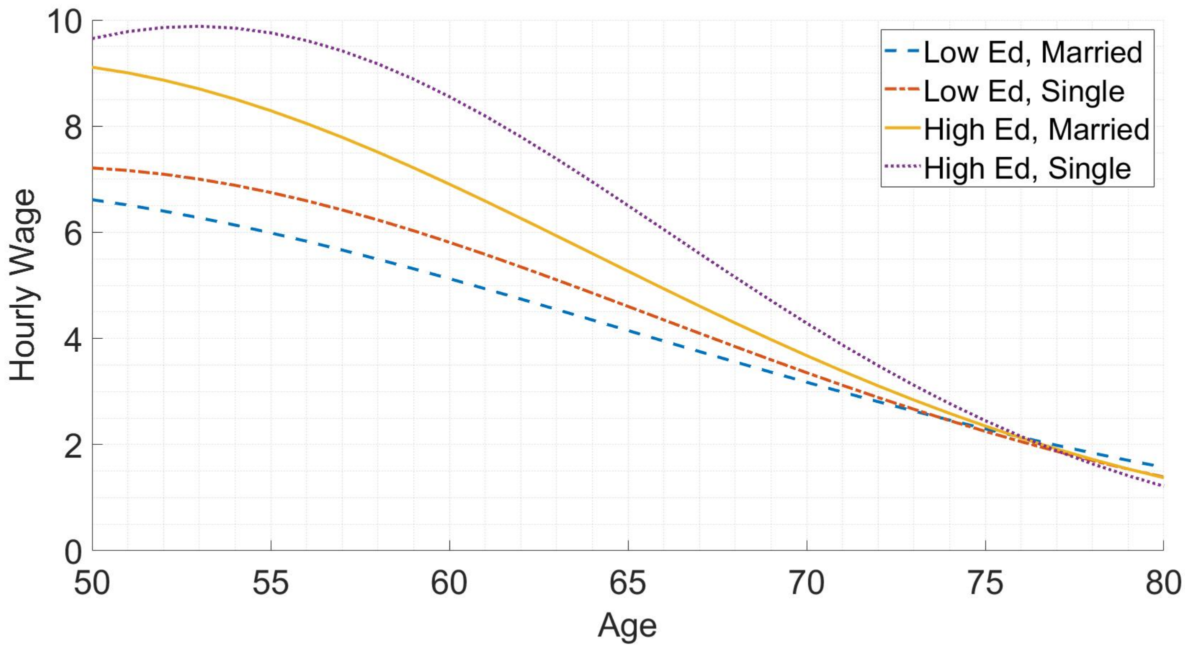
Notes: Unemployment and reemployment transition probabilities.

Table 19: Paramteres of the stochastic component of the wages

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

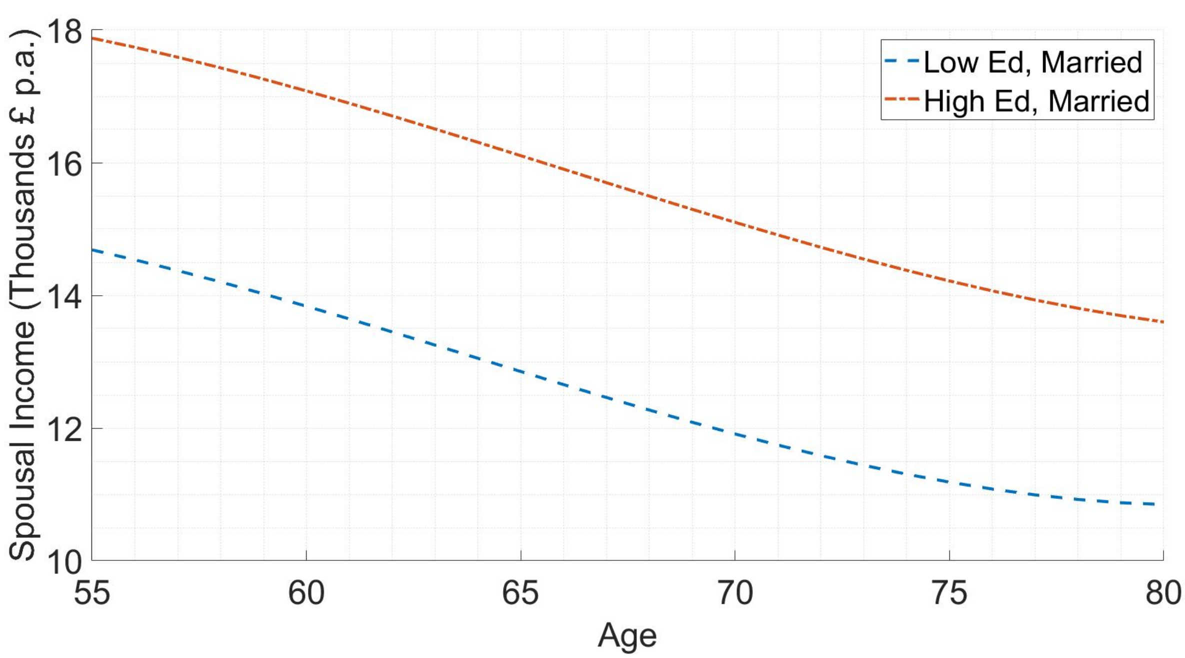
Notes: Estimates of the persistence of wages and the variance of their transitory and persistent components as well as initial distribution.

Figure 13: Wage Profiles



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

Figure 14: Spousal Income



Notes: Spousal income plotted against female age.

Table 20: Regression Analysis of the Determinants of Learning

	Constant	Assets	AIME	Wage	Age = 56	Age = 57	Age = 58	Age = 59
Coefficient	0.18	-2.898e-07	2.826e-06	9.655e-08	-6.458e-02	-7.781e-02	-9.796e-02	-8.282e-02

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

Social Insurance As mentioned in the main text, much larger differences in State Pension income are observed between married and single women than between high and low education. Amongst State pension claimers, high education women have mean State pension income of £92.52 and low education women £87.11, whereas single women have a State Pension income of £112.50 and married women £80.89. Hence to maximise power whilst capturing the key difference, I restrict heterogeneity in the State Pension process to be between married and single women only. The resulting functions of average lifetime earnings are shown in Figure 15.

Conversely, differences in private pension income are smaller between married and single women than between high and low education. Amongst those reporting non-zero private pension income, high education women have mean private pension income of £118.50 and low education women £61.42, whereas single women have State Pension income of £100.78 and married women £94.24. Hence to maximise power whilst capturing the key difference, I restrict heterogeneity in the private pension process to be between high and low-education women only. The resulting functions of average lifetime earnings are shown in Figure 16.

E.2 Model Fit

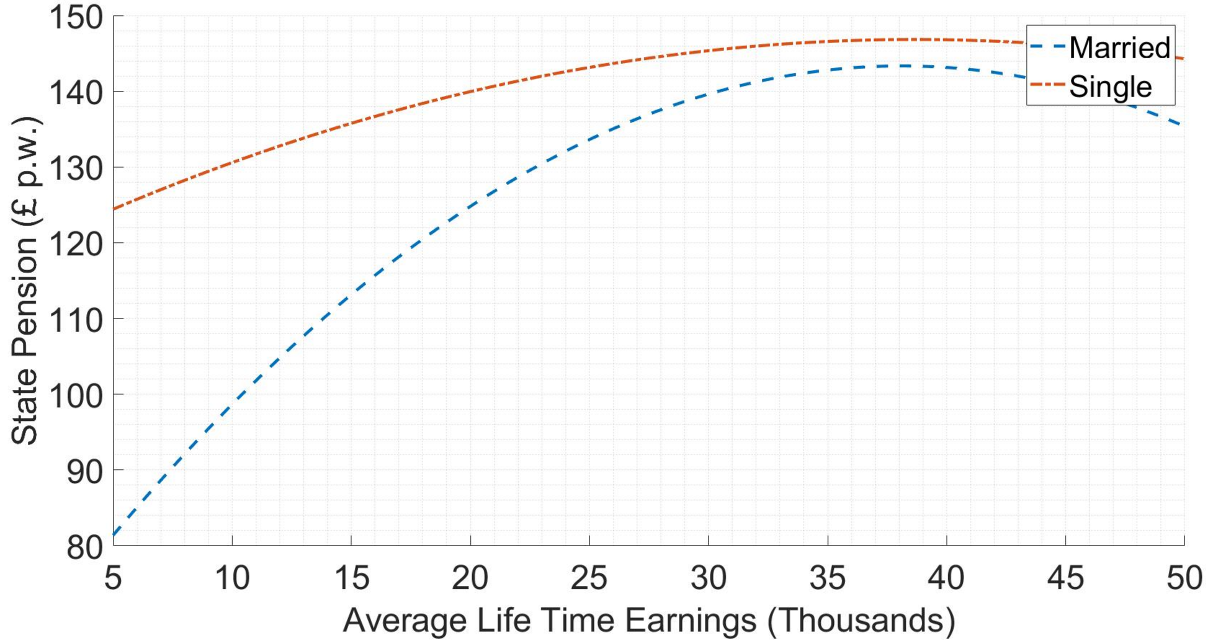
As mentioned in the main text, although the different model specifications have different predictions about the labour supply response to the dynamic SPA, the static profiles are not very sensitive to model specifications. All versions are able to match the static profiles. Figures 17-20 show the employment and asset profiles for the baseline version and the version with rational inattention with the parameter estimates of Table 4.

E.3 Results Tables

E.4 Robustness (Targetting Other Moments)

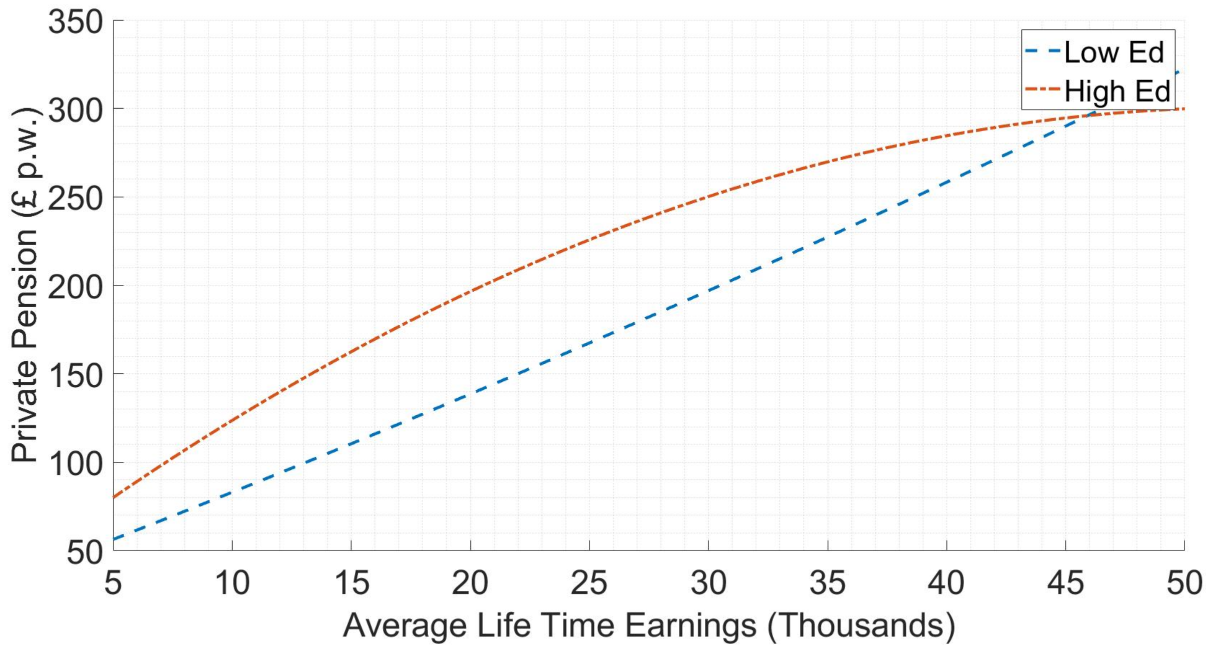
By design, the estimation procedure used in this paper does not directly target the excess employment sensitivity puzzle. This allows the model to match savings and labour supply across a range of ages and then to investigate how well a model that matches these profiles can explain excess employment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

Figure 15: State Pension as Function of Average Earnings



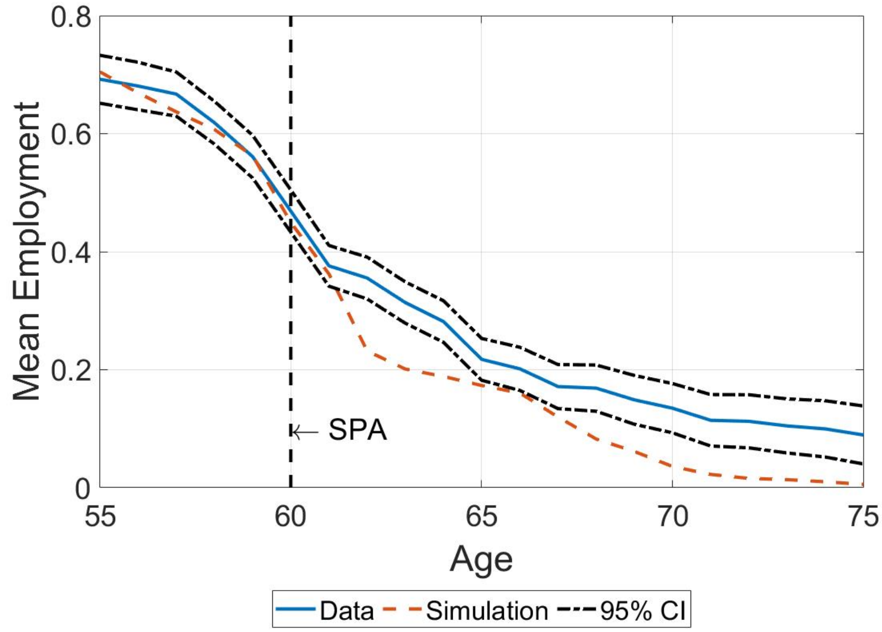
Notes: State Pension income as a function of average lifetime earnings (AIME) for married and single women.

Figure 16: Private Pension as Function of Average Earnings



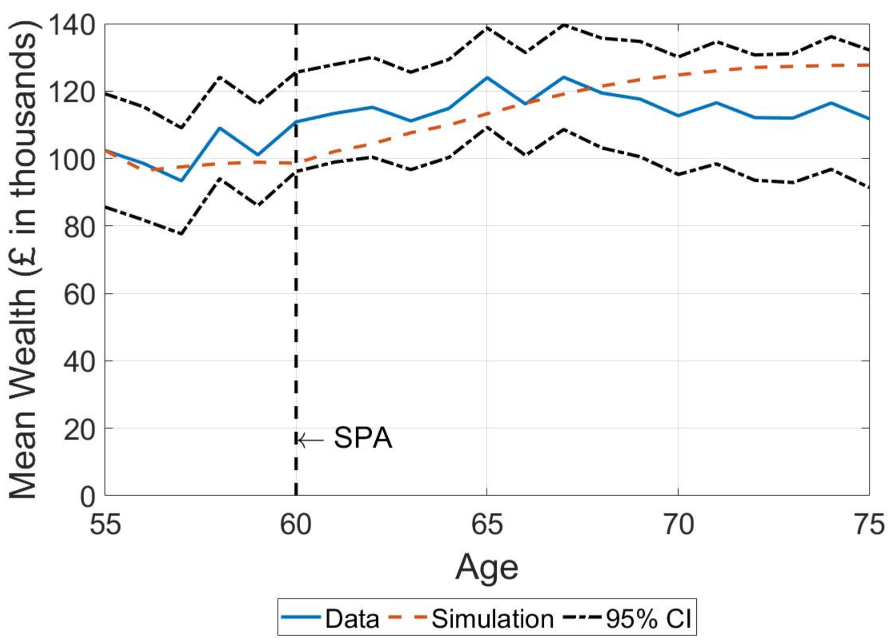
Notes: Private Pension income as a function of average lifetime earnings (AIME) for high and low education women.

Figure 17: Employment Profile Baseline



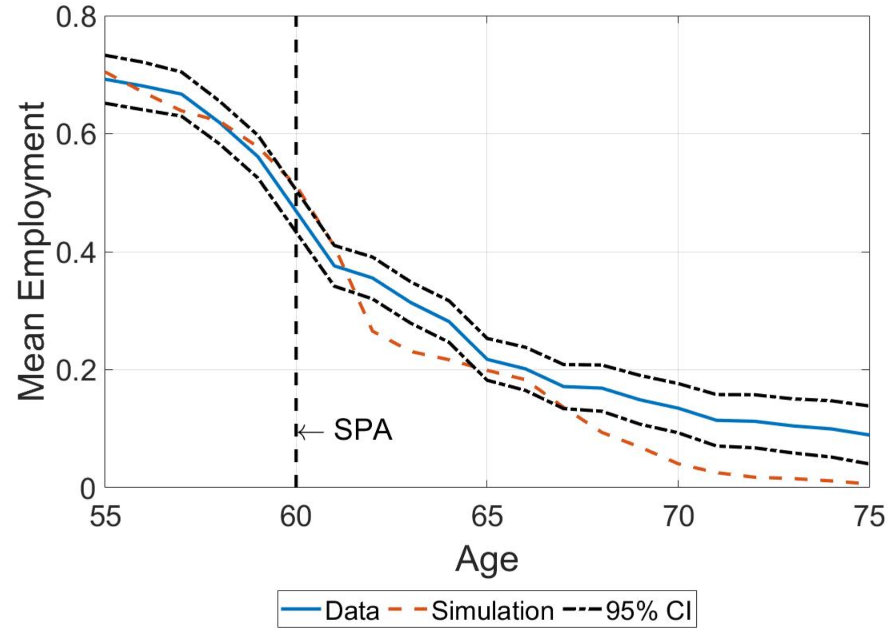
Notes: Model fit to targetted labour supply profile. The empirical profile is for the pre-reform SPA cohort with a SPA of 60. The model was simulated with an unchanging SPA of 60, mimicking the conditions faced by this cohort.

Figure 18: Asset Profile Baseline



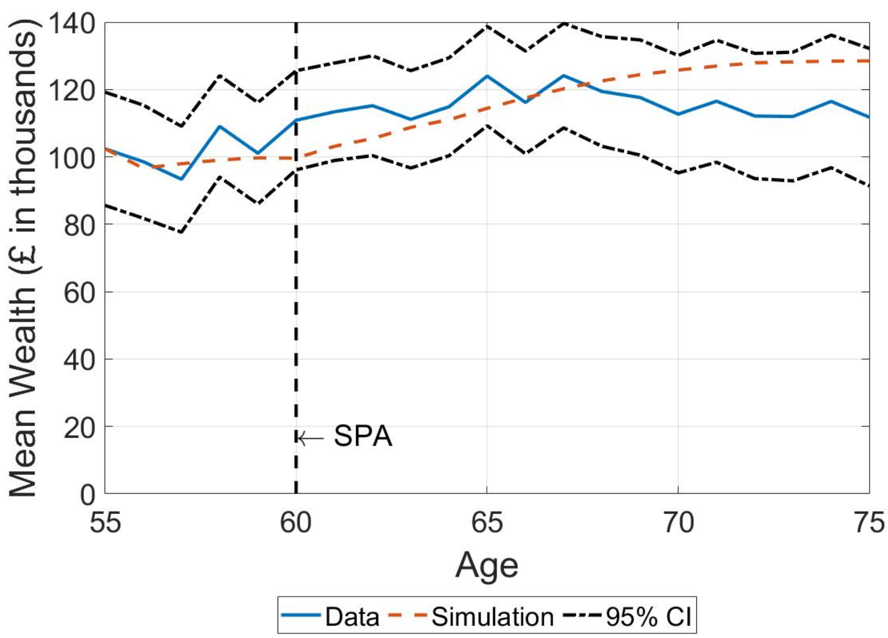
Notes: Model fit to targetted asset profile. The empirical profile is for the pre-reform SPA cohort with a SPA of 60. The model was simulated with an unchanging SPA of 60, mimicking the conditions faced by this cohort.

Figure 19: Employment Profile Model with Rational Inattention



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 20: Asset Profile Model with Rational Inattention



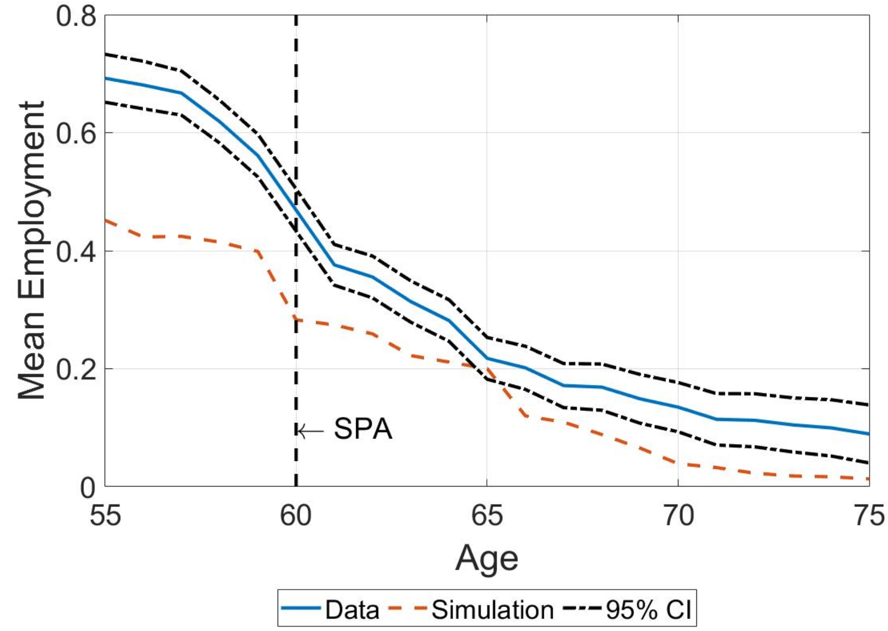
Notes: Model fit to targetted asset profile. The empirical profile is for the pre-reform SPA cohort with a SPA of 60. The model was simulated with an unchanging SPA of 60, mimicking the conditions faced by this cohort.

Table 21: Effect of SPA on Employment: Heterogeneity by Wealth

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

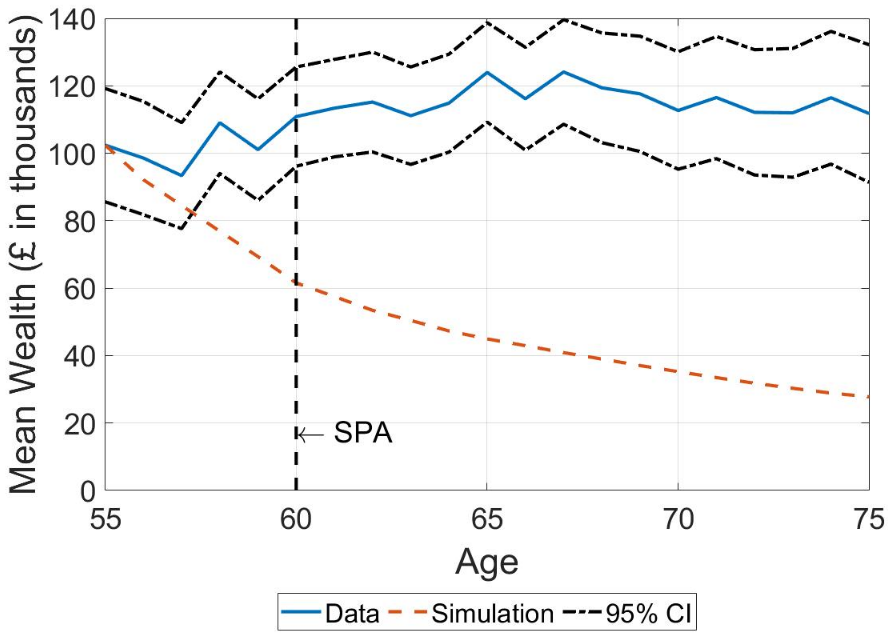
Table 21 shows that the model nearly perfectly replicates the treatment effect in the whole population but falls dismally short of the one seen in those with above median wealth. Figures 21-22 show that this is achieved at the cost of massively exaggerating how much households run down their assets and hence the importance of borrowing constraints.

Figure 21: Employment Profile Basline when Targetting Treatment Effects



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 22: Asset Profile Model Basline when Targetting Treatment Effects



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.