Costly Attention and Retirement *

Jamie Hentall MacCuish[†]

Abstract

Most people hold mistaken beliefs about their pension provision implying significant informational frictions. This paper introduces such frictions, a cost of attention to an uncertain pension policy, into a life-cycle model of retirement. Solving this dynamic rational inattention model with endogenous heterogeneous beliefs represents a significant methodological contribution in itself. Resulting endogenous mistaken beliefs help explain a puzzle: labour market exits concentrate at official retirement ages despite weak incentives to do so. The UK female state pension age (SPA) reform provides the context studied. I estimate the model using simulated method of moments. Costly attention significantly improves model predictions of the labour supply response to the SPA whilst accommodating the observed learning about the SPA. An extension addresses another retirement puzzle, the extremely low take-up of actuarially advantageous deferral options. Costly attention significantly increases the number of people claiming early.

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

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

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

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

1 Introduction

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

Ignoring these informational frictions limits research efforts to understand the implications of objective government policy uncertainty to individual behaviour. This is because, by ignoring them, we miss much individual uncertainty about policy: people are not only unsure about how policy may change, mistaken beliefs show often they do not know current rules. Conversely, it is easier to rationalise mistaken beliefs about government policy if we simultaneously acknowledge that government policy is objectively uncertain. Hence, there is an interplay between objective policy uncertainty and subjective mistaken beliefs. This paper investigates this interplay and its implications for retirement behaviour.

More specifically, it asks the question whether costly attention and objective policy uncertainty explain the excess employment sensitivity puzzle whilst accommodating observed mistaken beliefs. The excess employment sensitivity puzzle is that many benefit systems offer very weak incentives to retire precisely at statutory retirement ages, and yet labour market exits concentrate at them and follow them as they increase. This puzzle has been documented in multiple countries (e.g. Behaghel and Blau, 2012; Seibold, 2021). Accommodating mistaken beliefs increases the shock individuals receive upon reaching these ages because much pension policy uncertainty is resolved upon reaching eligibility. This in turn helps explain the larger-than-expected labour supply reaction.

I take advantage of the recent reform to the UK female State Pension Age (SPA) to investigate this question. This reform increased the female SPA, allowing the effect of the SPA on employment to be identified separately from the effects of ageing. Two key features UK institutions rule out common explanations for labour market exit at statutory retirement ages, making the UK an interesting context to study the excess employment sensitivity puzzle. Firstly, forcing an employee to retire purely due to age is illegal, ruling out firm-mandated retirement, and, secondly, as receipt of the state pension is not conditional on employment status it only provides an incentive to retire to the liquidity constrained.

The mode of investigation is to first document the pertinent facts, concerning mistaken beliefs and excess employment sensitivity, and then build a model with informational frictions, in the form of costly attention, that accounts for these facts. This model incorporates costly attention, modelled using rational inattention (e.g. Sims, 2003), to an uncertain pension policy into a dynamic life-cycle model of retirement (e.g. Rust and Phelan, 1997; French, 2005), thus allowing for endogenous mistaken beliefs that help explain retirement choices.

In incorporating rational inattention to an uncertain pension policy into a dynamic life-cycle model of retirement, this paper is the first, to the best of my knowledge, to solve a dynamic rational inattention model with endogenous heterogeneous beliefs. Although key to generating mistaken beliefs that help explain retirement choices, allowing for endogenous heterogeneous beliefs introduces a large state variable and so greatly complicates the solution. Weaving together recent theoretical results (Steiner et al., 2017; Caplin and Dean, 2015; Armenter et al., 2019) into a workable solution methodology for rich structural models with endogenous beliefs resulting from rational inattention is an important additional contribution of this paper.

The English Longitudinal Study of Ageing (ELSA), a panel survey of older individuals, provides the data needed to study the impact of mistaken beliefs, as both self-reported and true SPA are observable. Women subject to the reform are substantially mistaken about their SPA, most not knowing it to within a year at age 58, and these mistakes are predictive of the employment response upon reaching the SPA. So, mistaken beliefs are related to excess employment sensitivity. Moreover, the direction of the relationship indicates selection into SPA knowledge may be important hinting at a role for the endogeneity of learning.

I estimate the model on ELSA data using two-stage simulated method of moments targetting asset and labour supply profiles and find policy uncertainty and costly attention significantly improve the untargeted model predictions of the labour supply response to the SPA. I use the evidence of learning in the data to identify the cost of attention parameter, an empirical contribution to the rational inattention literature. The baseline version of the model without rational inattention matches the static aggregate profile but it fails to match the dependence of retirement on the SPA; costly attention helps rectify this shortcoming.

To investigate the excess employment sensitivity puzzle, I abstract away from the benefit claiming decision to avoid another puzzle: 87% of my sample claim the state pension immediately upon reaching it despite an actuarially advantageous adjustment of over 10.4% p.a. from deferring. An extension addresses this deferral puzzle, introducing a claiming decision, policy uncertainty over the actuarial adjustment from delayed claiming, and a cost of learning about this uncertainty actuarially adjustment. This introduction of costly attention substantially increases the proportion claiming early and so helps explain this deferral puzzle.

The rest of the paper is structured as follows. Section 2 reviews the literature. Section 3 outlines the institutional context and the data and carries out descriptive and reduced-form analysis to document the excess employment sensitivity puzzle and its relation to mistaken beliefs. Section 4 presents the model, starting with a standard model of complete information as a baseline and then building in objective uncertainty about pension policy and a cost of attention to this uncertain policy. As solving the model presents novel difficulties and finding a solution method for dynamic rational inattention models with endogenous heterogeneous beliefs represents a contribution, section 5 discusses the solution of the model. Section 6 discusses estimation and section 7 model fit. Section 8 presents an extension that increases demands on agents' attention whilst introducing a claiming decision which generates a mechanism that can help explain 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 mistaken beliefs. The two strands of literature this paper builds on most closely are dynamic life-cycle models of retirement and rational inattention, but it is also deeply connected to works documenting excess employment sensitivity and beliefs. Each of these strands is reviewed below and the contributions to each explained. I then briefly discuss the relation to the wider literature.

Dynamic lifecycle models of retirement have a history stretching back to Gustman and Steinmeier (1986) and Burtless (1986), and this paper includes the key features identified in this literature that are relevant to the UK. Computational limitations led early works to abstract away from uncertainty and borrowing constraints but more recent work finds these crucial. Rust and Phelan (1997) introduced uncertainty into a dynamic life-cycle model along with an extreme formulation of incomplete markets that ruled out all borrowing. French (2005) reintroduced borrowing while maintaining incomplete markets through a borrowing constraint, as well as introducing other innovations such as a fixed cost of work to help explain the retirement phenomena. Gustman and Steinmeier (2005) allow for time preference heterogeneity, van der Klaauw and Wolpin (2008) model medicare, and French and Jones (2011) add uncertain medical expense onto these accumulating innovations. The development of dynamic life-cycle model of retirement has been intimately connected to the excess employment sensitivity puzzle, and I return to many of these papers when discussing that literature. Much of this literature is US focused and some of its concerns are not relevant to the UK context which I study (e.g. medical insurance). Some of the key features included in this paper are uncertainty, borrowing constraints, and individual heterogeneity. The lifecycle model of retirement most similar to this one is O'Dea (2018) who estimates a structural retirement model focusing on males in the UK.

This paper models costly attention following the rational inattention literature and, while relying on recent theoretical advances from this literature, it contributes back a novel application and important quantitative techniques. Rational inattention traces its heritage back to Sims (2003). Initially, it was used to add costly attention to macroeconomic models (e.g. Luo, 2008; Maćkowiak and Wiederholt, 2009, 2015)), but recently its domain of application has expanded. To cite a handful of examples, 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. This recent flourishing makes it impossible to do justice to the literature in its entirety, and as such, I discuss just the papers most closely related to this paper. A series of papers starting with Matějka and McKay (2015) analyse general classes of models with rationally inattentive agents; Matějka and McKay (2015) solve static discrete choice models with rationally inattentive agents, and Steiner et al. (2017) extends their analytic results to dynamic discrete choice models. These analytic results from Steiner et al. (2017) are key to solving the dynamic rational inattention model with endogenous heterogeneous beliefs resulting 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 an important contribution of this paper, and I am the first, to the best of my knowledge, to solve a model with endogenous heterogeneous beliefs. Two other papers are key to bridging the gap between elegant theory and practical solution methodology. Caplin et al. (2019) show rational inattention generically implies consideration sets, implying model solution will be sparse and provide conditions for this sparsity; leveraging these conditions greatly reduces computation required. When sparsity does not provide a shortcut solution, I follow the suggestion of Armenter et al. (2019) to use sequential quadratic programming to solve the within period rational inattention problem. This paper joins recent work, Macaulay (2021) and Porcher (2020), in applying rational inattention to rich non-experimental choice data. These other works avoid the issue of endogenous heterogeneous beliefs by assuming complete information sharing between individuals.

Employment being more sensitive to statutory pension ages than standard models predicts is a puzzle observed in multiple countries; this paper provides the most comprehensive evidence to date of its existence in the UK. The excess employment sensitivity puzzle was documented in the US by Lumsdaine et al. (1996) and Rust and Phelan (1997), and much of the lifecycle models of retirement literature was dedicated to explaining it. The consensus from the literature was that liquidity constraints explained the spike in labour market exits 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). These papers were unable to empirically distinguish these explanations as the US early and full retirement ages remained unchanged between 1962 and 2000. Ageing population induced the US government to increase the full retirement age, from 2004, and this reform provided the necessary variation to estimate the impact of this statutory pension age on labour supply. Much 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 the claim the puzzle was explained by medicare eligibility. [1] Ageing populations forced other governments to increase statutory pension ages, and a similar pattern was observed: 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), in Switzerland by Lalive et al. (2017), and in the UK by Cribb et al. (2016). As this paper investigates the excess employment sensitivity puzzle, I first document its existence building on the work of Cribb et al. (2016) who first document this puzzle in the UK using the

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

same female state pension age (SPA) reform I study. I build on their work, principally, by using a richer data set to rule out other potential standard complete information explanations for the bunching of labour market exits at SPA.

The use of belief data is increasing (Koşar and O'Dea, 2022), pension beliefs being an interesting case because mistakes are relatively easy to detect, and this paper contributes to the use of belief data by using mistaken pension beliefs to identify informational frictions. The earliest paper to investigate pension knowledge, such as Bernheim (1988) and Gustman and Steinmeier (2001), look at individual forecast errors about the level of their pension benefit. Forecast errors conflate misprediction of future rule changes with mistaken beliefs about current policy, and disentangling them requires gathering information on their knowledge of current social security rules. Manski (2004) documents precisely one such study, finding much individual uncertainty about their benefits is explained by a lack of understanding of current social security arrangements. Rohwedder and Kleinjans (2006) study the dynamics of these forecast errors and find they become increasingly small as individuals approach retirement, providing evidence of learning. Crawford and Tetlow (2010) look at these self-reported SPAs in the English Longitudinal Study of Ageing (ELSA) and find large errors common; Amin-Smith and Crawford (2018) document these mistakes are predictive of the labour supply response to the SPA. I find very similar patterns to Crawford and Tetlow (2010) and Amin-Smith and Crawford (2018), prevalent mistaken beliefs predictive of labour supply, and also document a similar pattern of learning to that found by Rohwedder and Kleinjans (2006); I use these patterns to identify informational frictions. This represents a relatively novel use of belief data as most papers tend to use belief data to identify an object the individuals have private information about, maintaining the assumption they are well informed.

This paper has other important connections to the broader 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 note from this literature, Luttmer and Samwick (2018) measure the welfare cost of individuals' perceived uncertainty about their social security benefits.

3 Institutional Context, Data, and Analysis

This paper studies the puzzlingly large labour supply response to the UK state pension age (SPA). A reform to the female SPA is used to separate the effects of ageing from those of the SPA and it is detailed in section in 3.1 highlighting aspects that make it particularly illuminating of this excess employment sensitivity puzzle. Section 3.2 discusses the data. Sections 3.3 - 3.4 provide descriptive and reduced form analysis, section 3.3 documenting the excess employment sensitivity puzzle, section 3.4 documenting erroneous beliefs about pension entitlement as well as their relationship to employment sensitivity to SPA. The existence of this relationship suggests mistaken beliefs about the SPA should be studied alongside the reaction to the SPA, as this paper does.

3.1 Institutional Context

The State Pension Age (SPA) is the earliest age at which retirement benefits, known as the state pension, can be claimed in the UK. In other words, it is the Early Retirement Age of the UK pension system, although, unlike in the US there is no earnings test.[2] The UK does not have a Normal or Full Retirement age, so the SPA is the sole focal age of the state pension system. Deferral of receipt does increase the generosity of the benefit; however, during the period considered this was without a cap on the deferral duration and so did not imply an implicit full retirement age. [3]

The UK State Pension came into force in 1948 with the SPA set at 65 for men and 60 for women. This remained unchanged until, the Pensions Act 1995 legislated for the female SPA to gradually rise from 60 to 65, one month every two months, over the ten years from April 2010. The Pension Act 2011 accelerated the rate of change of the female SPA from April 2016 so that it equalises with men's by November 2018. It additionally legislated an increase to both the male and female SPA to 66 years phased in between December 2018 to October 2020. Figure 1 summarises how these changes affect women in different birth cohorts.

This UK SPA reform is a convenient context to study the excess employment sensitivity puzzle, as many possible explanations for labour market exits at the early retirement age are ruled out. Firstly, firms cannot force employees to retire solely based on age: this would be classed as age discrimination under UK law[4]. So, firm mandated retirement cannot explain the sensitivity of employment to the SPA. Secondly, the state pension is not conditional on employment status. Individuals may claim the state pension and continue working and, indeed, many do [5]. Thirdly, the UK pension system does not provide major tax incentives to exit the labour market at the SPA. There is no earnings test, and although the state pension is taxable income, a component of income tax, called the National Insurance contribution, is removed upon reaching the SPA[6].

These three facts imply the state pension is essentially an anticipatable increase in non-labour income with the SPA its eligibility age. The reform increased this eligibility age without changing the benefit level and, as such, is an anticipatable decrease in non-labour income. As the reform was announced in 1995 and began in 2010, this income change was anticipatable with a horizon of at least 15 years. In a standard life-cycle model, with complete information and forward-looking agents, labour supply responses do not concentrate at an anticipatable income change unless agents are liquidity constrained. The puzzle is not that labour supply

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

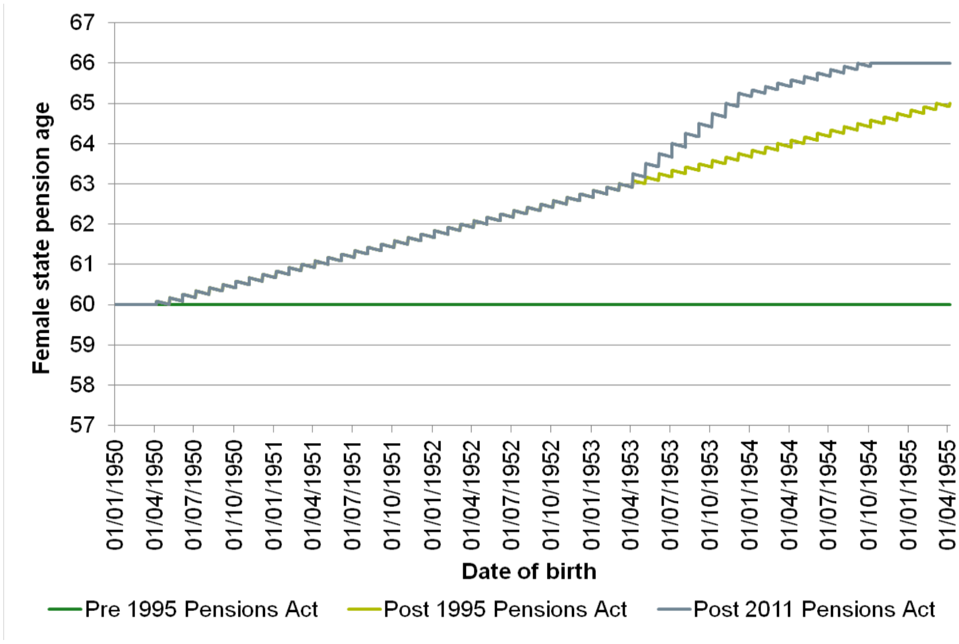
[3] Despite an extremely generous actuarial adjustment deferral was very rare, leading to a deferral puzzle, discussion of which is deferred to an extension addressing this puzzle in section 8.

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

[5] In my sample amongst women over their SPA but under 70, 24% are in work. This only falls to 22% when restricted to those also reporting non-zero state pension income.

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

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).

responds to the SPA reform, but that the response concentrates at the SPA when so much forward notice was given.

Hence, these three features remove incentives to exit the labour market at the SPA for all but the liquidity constrained[7]. Accordingly, I treat the ability of liquidity constraints to explain the sensitivity of employment to the SPA as synonymous with the ability of standard models of complete information to do so. The UK pension system's lack of other retirement ages makes ruling out liquidity constraints more difficult, and this is a major focus of section 3.3.

3.2 Data

To study the labour supply response to the State Pension Age (SPA) a dataset that samples a large number of older individuals is required. To investigate the reasons for the response rich microdata are also needed. The English Longitudinal Study of Ageing (ELSA) is the UK[8] dataset that strikes the best balance along these two aspects[9], and so it forms the principal data source for this paper.

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

[8]Technically ELSA only covers England and Wales.

[9]For example the Labour Force Survey has a larger sample of older individuals but does not contain nearly such rich data. Crucially, it does not contain sufficient information on assets or beliefs, both of which are crucial to my analysis.

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 and elicits their beliefs distribution about the level of their state pension benefit. 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. These administrative records can be used to construct average lifetime earnings, which is a useful input for predicting pension entitlements. Additionally, survey data on health, education, and family are instructive of retirement choices.

I use ELSA waves 1 (2002/03) through to 7 (2014/15) for analysis and estimation and waves 8 (2016/17) through 9 (2018/19) for model validation. As this paper is concerned with the reform to the female SPA, males are dropped from the sample. The only exception is 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 25,101 observations of 7,200 women. The implementation of the female SPA reform began in 2010 and so the first wave of ELSA after the implementation of the female SPA reform is wave 5. Having earlier waves is important to control for pre-trends and increases power when estimating inputs to the structural model. The oldest women affected by the reform were born on 6 April 1950. Having older cohorts is important as a control group and also informative when estimating exogenous processes.

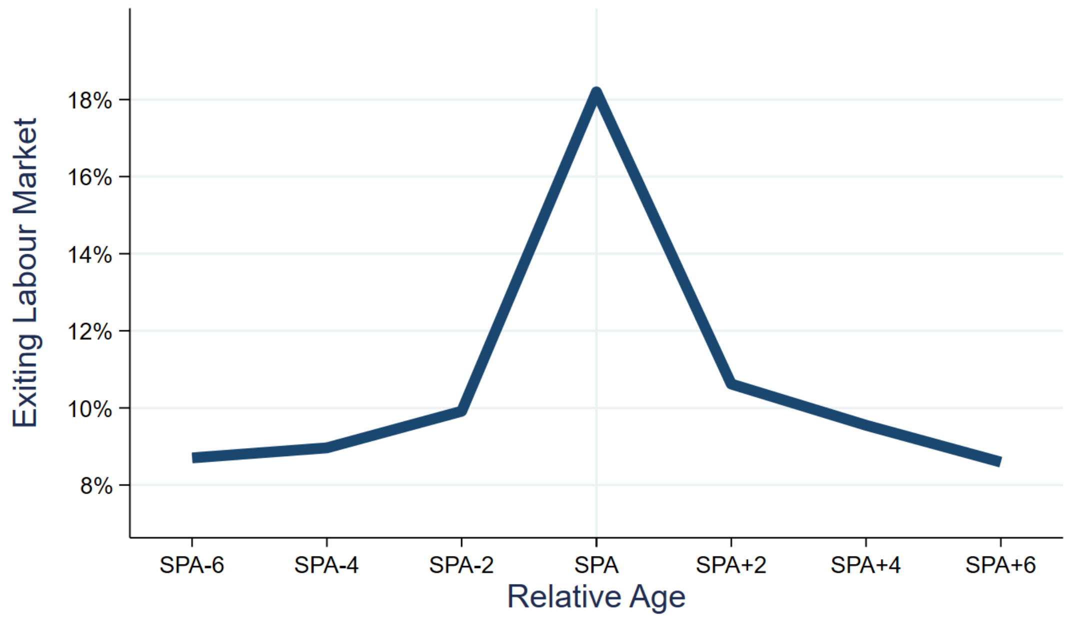
3.3 Excess Employment Sensitivity

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

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

To more separate the labour supply response to the SPA from the effects of ageing I follow the work of Cribb et al. (2016) who use this reform to identify this labour supply response and find it significant. They

Figure 2: Fraction exiting labour employment



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

argue against constraints driving their results because, whilst homeowners are less likely to be constrained than renters, the effects of the SPA on their labour market participation are indistinguishable. The focus of their paper was documenting the response to the SPA more than explaining it and homeownership is a coarse proxy for being liquidity constrained, equity in one's own home being an illiquid asset. Hence in this section, I build on the analysis of Cribb et al. (2016) using the richer data in ELSA to investigate motives more thoroughly, including ruling out liquidity constraints. In doing so I present the most detailed evidence to date of the excess employment sensitivity puzzle for the UK.

The main estimating equation used through this section builds on builds on Cribb et al. (2016) and it 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, and yearly cohort dummies; and a vector of controls[10] leading to the following specification:

$$Pr(y_{it} = 1) = \alpha \mathbb{1}[age_{it} \leq SPA_{it}] + \sum_{c=1}^{K} \gamma_c \mathbb{1}[YOB_i = c] + \sum_{a=1}^{A} \delta_a \mathbb{1}[age_{it} = a] + X_{it}\beta + \epsilon_{it} \quad (1)$$

This form assumes that there are cohort-constant age effects and age-constant cohort effects. Given these assumptions, the parameter α is a difference-in-difference estimator where the treatment is being below the

[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; 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.

Table 1: Effect of SPA on Employment: Heterogeneity by Wealth

		(1)	(2)	(3)	(4)
Below SPA		**0.112**	**0.114**	**0.126**	**0.128**
	$s.e$	(0.0272)	(0.0286)	(0.0372)	(0.0302)
	$p=$	.001	.001	.003	.000
Below SPA×(NHNBW.>Med.)				**-0.004**	
	$s.e$			(0.0311)	
	$p=$			.904	
Below SPA× NHNBW					**-7.44×10^{-8}**
	$s.e$				(2.16e-08)
	$p=$				.003
Obs.		7,947	3,126	7,947	7,947
Indv.		3,846	1,362	3,846	3,846

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. Column(4) includes an interaction between being below SPA and a continuous measure of NHNBW.

SPA[11]. Although this treatment is administered to all, variation in the duration of treatment is induced by the reform.

Like most econometric assumptions, cohort-constant age effects and age-constant cohort effects are probably only approximately true. To minimise the risk to the analysis of non-linear interaction between age and cohort I restrict the samples to those around SPA: ages 58-63. For this restricted sample I test these assumptions by interacting the fixed effects and the Wald test fails to reject the null that these interactions are zero ($p = 0.95$).

Column 1 of Table 1 presents the results of estimating equation 1. Here I Find a 0.112 increase in the probability of being in work from being below the SPA significant at the 0.1% level[12].

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)[13]. This generates a cut-off of £34,869. I impose this restriction of being in a household with above £34,869 in NHNBW in the wave before they reached their SPA, as this is when the resources to smooth labour supply affects their reaction to the SPA. The objective

[11]Being below the SPA is an interaction between cohort and age. The assumption of cohort-constant age effects is a rephrasing of the standard parallel trends assumption.

[12]I additionally test the parallel trends assumption with a placebo test where the treatment is being a year above or below the SPA. These both return null results see table 2.

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

of the median split is to restrict to a group whose retirement choices are unlikely to be affected by the liquidity constraint. Given the SPA was reformed in monthly increments, and equation 1 controls for quarterly-age and yearly-cohort fixed effects, the control for an individual is someone born in the same year and quarter, but a few months older so no longer under the SPA. Thanks to this narrow time window it is easier to argue against liquidity constraints: households having more than £34,869 in NHNBW seem unlikely to need to wait 1-3 months for the state pension to stop working.[14] The results are in column 2 of table 1. For this subpopulation, we find a treatment effect of 0.114, very similar in size to results for the whole population, and significant at the 0.1% 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.

Only considering two asset groups, above and below median assets, is an arbitrary dichotomisation and leads to a loss of information. For this reason, column 4 shows results for a specification containing an interaction between being below the SPA with the continuous variable NHNBW. As can be seen, this interaction term is highly significant but tiny, an additional £$(\frac{0.01}{7.44\times10^{-8}})$ or £134,410 of NHNBW is required to decrease the treatment effect by 1 percentage point. This indicates, unsurprisingly, that wealth does impact how important the SPA is to someone's retirement decision, but that liquidity constraints cannot completely explain the sensitivity of labour market exits to the SPA. For example, these results imply a woman from a household at the 95% percentile of the distribution, with £409,000 in NHNBW, would experience only a 3.02 percentage point decrease in her response to the SPA, and this response would still be significantly positive. 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.

Table 1 offers an embarrassment of riches, whilst it will be useful to have a single summary of the excess sensitivity puzzle. Column 4 allows for more flexible heterogeneity than column 3 which largely encapsulates columns 1 and 2. On the other hand, columns 1 and 2 more clearly epitomise excess employment sensitivity in the following two facts: one, there is a significant employment response and, two, this employment response is constant across a median split. For this reason, 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.

Other factors that are not considered here are important to labour supply amongst older individuals. Chief amongst these factors neglected for brevity in this section are health, private pension, and joint retirement. Appendix A considers whether any of these factors can explain the excess employment sensitivity puzzles and

[14] However, given the arbitrary nature of the median split, I consider other data-driven cut-offs in appendix A.

Table 2: Placebo Tests

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

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

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.

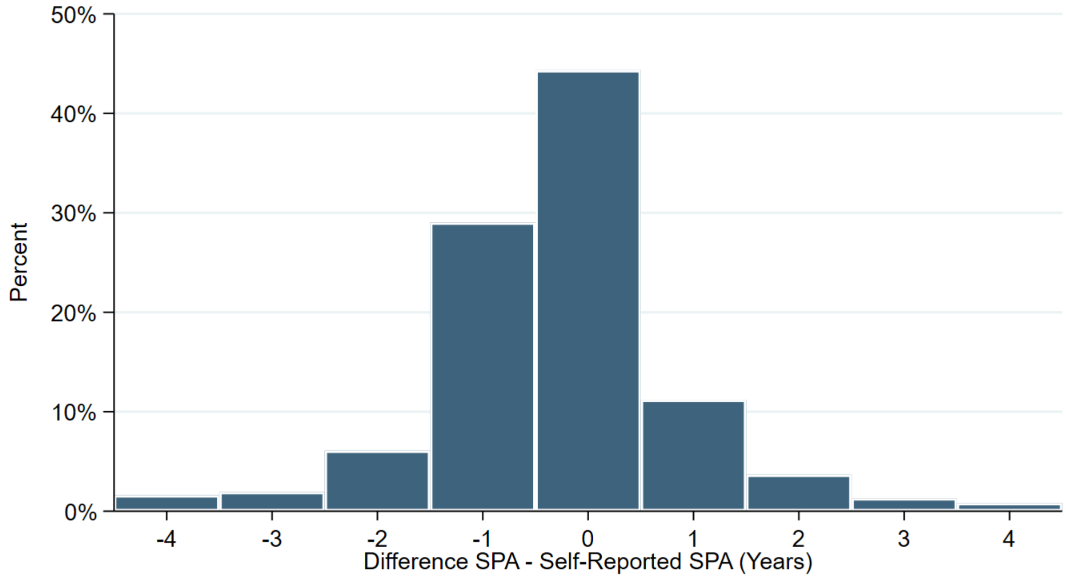
This section has gone to lengths to, firstly, present causal estimates of the labour supply response to the SPA, and, then rule out standard complete information explanations of these effects. Thus indicating these results present a puzzle. What follows does not rest on the causal nature of these estimates; I use these regressions as an untargeted auxiliary model to my structural life-cycle models. As such, what is important is the model's ability to replicate the key facts, not whether the treatment effects estimates replicated are unbiased causal estimates. Of course, what follows does depend on the reader finding these results puzzling, at least as far as standard complete information models are concerned. The results of the placebo test in table 2, add weight to the claim there is a puzzle. These show the results of including indicators of being one year over, and one year under the SPA in equation 1, and as can be seen, unlike the indicator of being below the SPA these coefficients are tiny and not significant at any reasonable level. Hence it seems that the results in this section are detecting something specific about the SPA, rather than picking up some violated assumptions like non-parallel trends.

3.4 Mistaken Beliefs and Relation to Excess Employment Sensitivity

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

The SPA being such a simple aspect of the benefit system, confusion about it is both puzzling and simple to demonstrate. The SPA is a deterministic function of date of birth, recorded in ELSA, and from wave 3 women under 60 are asked what their state pension age is. Any discrepancy demonstrates less than perfect knowledge of one's SPA. Figure 3 shows this difference between the true and reported SPA of 58-year-old women subject

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



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

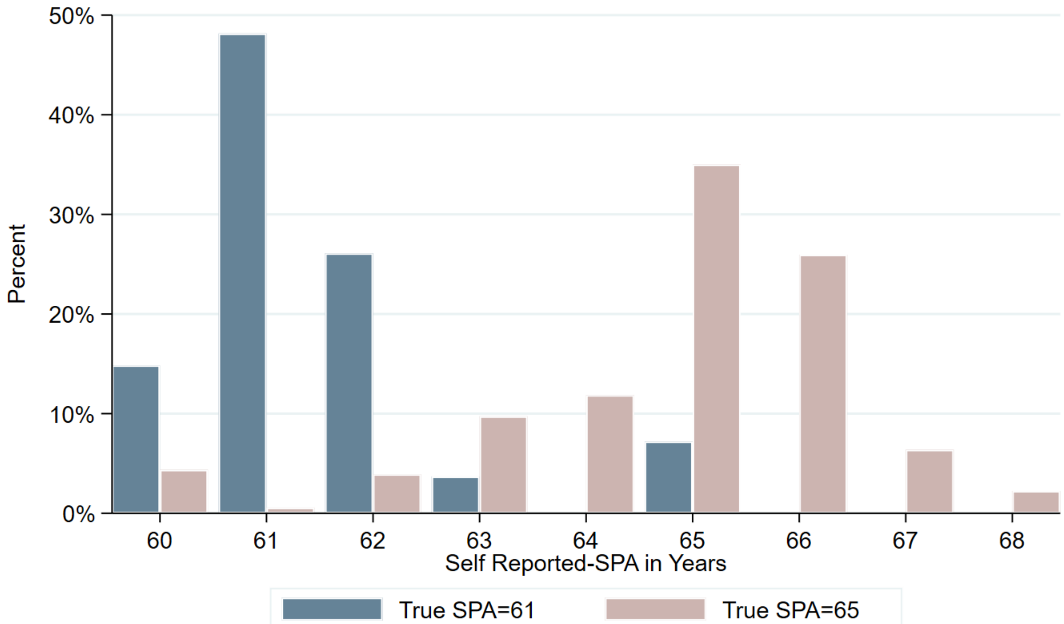
to the reform. Although the largest group are those who know their SPA to within a year, this contains those mistaken by a margin of months and still leaves over 40% who are out by a year or more. Striking evidence of the prevalence of mistaken pension beliefs in the UK.

Mistaken beliefs could, of course, 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 4 plots reported SPAs for two SPA cohorts, one with a true SPA of 61 and one with a true SPA of 65. The self-reports cluster around the true SPA for each cohort, looking very much like a noisy signal of the true SPA. Just the sort of pattern we would expect to emerge from a model of costly information acquisition.

Another prediction of costly information acquisition, supported by the data, is learning. This can be seen in figure 5 that plots, against age, the variance of errors in self-reported SPAs. A clear declining age profile can be seen with the variance of state pension age answers shrinking towards the truth as these women age towards their SPA. This declining variance of reported SPAs is the key moment used by the model of costly attention to identifying the cost of attention. Using belief data to directly estimate the cost of attention is a novel contribution to the rational inattention literature that adds empirical validity.

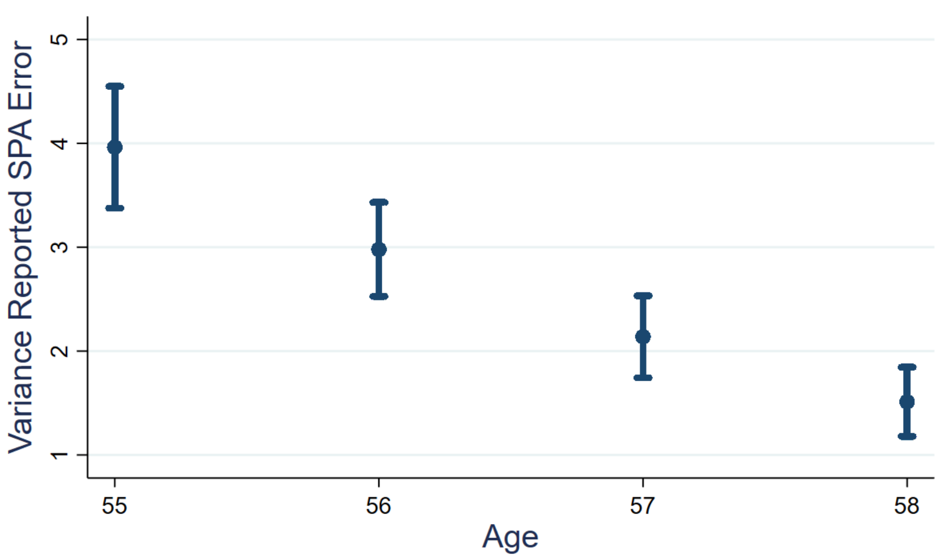
So, mistaken beliefs are a feature of reality, but if they were a feature unrelated to the excess sensitivity puzzle then models attempting to explain this puzzle could safely ignore them. This is not, however, the case. Table 3 documents the heterogeneity of the labour supply response to the SPA by the degree of mistaken belief.

Figure 4: SPA Beliefs by SPA-cohort



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

Figure 5: Variance of Error in Reported SPA



Notes: Variance in error of SPA self-reports plotted against respondents' age.

Table 3: Heterogeneity by SPA Knowledge

Below SPA		**0.179**
	$s.e$	(0.0372)
	$p=$	.001
Below SPA×(abs. Error in SPA report)		**-0.062**
	$s.e$	(0.0215)
	$p=$	.015
Error in SPA report		**0.042**
	$s.e$	(0.0177)
	$p=$	.038
Obs.		4,249
Indv.		1,870

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.

This is found by introducing into specification 1 the size of the error in self-reported SPA in the last wave before reaching 60, after which this question is no longer asked, and an interaction between this error and the indicator of being below the SPA. The interaction is significant and negative indicating that, on average, for each additional year the individual is out by in their SPA self-report the labour supply response decreases by 6.2 percentage points.

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 discovering the truth. 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 .

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

4 Model

This section presents the model: section 4.1 a baseline standard complete information model, capturing the relevant features of the UK retirement context, and section 4.2 introduces two additions: objective uncertainty about government policy and costly information acquisition about this uncertain policy. This allows the model to capture the interplay between individuals' confusion about government policy and their reaction to it.

4.1 Complete Information Baseline

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

In more detail, households are divided into four types indexed by k, based on the high or low education status of the female and the presence or absence of a partner. Households choose how much to consume c_t, how much to invest in a risk-free asset a_t with return r, and, if not involuntarily unemployed, whether the women work full-time, part-time or not 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 are treated as type-specific functions of average life time earning $AIME_t$: $S^{(k)}(.)$ the state pension and $P^{(k)}(.)$ the private pension. From age 60 the women face a probability s_t^k of surviving the period. Finally, households value bequest 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 $X_t = (a_t, w_t, AIME_t, ue_t, t)$[15] and below I detail how they impact the model.

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

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. 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.

Initial and terminal conditions The model starts with women aged 55. The reasons to start so far into the life-cycle are, firstly, the ELSA dataset only starts interviewing people at 50 and, secondly, I am interested in the period around retirement so 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 knowledge responses and variation in the true SPA; thus, allowing me to initialise the state variable from the data. 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:

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

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 $\epsilon_1 \sim N(0, \sigma^2_{\epsilon,55})$. The quadratic form of the deterministic component of wages captures the observed hump-shaped profile and is common in the literature.

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

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

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

This specification average out and abstract away from both idiosyncratic spousal income and mortality risk. In effect, the household dies when the woman dies, and the husband's mortality risk only turns up in so far as it affects average income; as if husbands were a pooled resource amongst married women. This allows me to ignore transitions between married and single which, while important to understand wider labour supply behaviours of older individuals (e.g. Casanova, 2010), is second order at best when considering labour supply responses to the SPA. Since spousal pension benefits are not modelled separately $y^{(k)}(t)$ amalgamates his labour and non-labour income into a single variable. Both female wage and spousal income are post-tax.

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

The wife receives the state pension, once she reaches the SPA and a private pension once she reaches the eligibility age $PPA^{(k)}$. This abstracts away from the benefit claiming decision for two reasons both briefly touched upon earlier. Firstly, over 85% of people claim the state pension at the SPA so, in terms of accuracy, little is lost by this simplification. Secondly, this small fraction deferring receipt of the state pension occurs despite deferral having been actuarially advantageous during most of the period considered. This behaviour presents another puzzle to standard models of complete information as they generally imply acceptance of actuarially advantageous offers. This benefit claiming puzzle is taken up in section 8, but deferring it until then gives this baseline model a fair chance of addressing the excess sensitivity puzzle.

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

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

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

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

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

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

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

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

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

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

$$a_{t+1} \geq 0 \quad (9)$$

$$ue_t(1 - l_t) = 0 \quad (10)$$

4.2 Two Additions To The Baseline: Policy Uncertainty and Costly Attention

This section introduces two additions to the model of complete information presented above. Firstly, section 4.2.1 introduces objective policy uncertainty in the form of a stochastic SPA, capturing the observed variability of SPAs over the life-cycle resulting from pension reform. Secondly, section 4.2.2 introduces costly attention to this stochastic SPA, modelled with a disutility cost for more precise information. This allows the model to capture individual uncertainty about government policy in the form of incorrect beliefs about the SPA, and the implications of these beliefs for behaviour. Since these two changes represent a novel approach section 4.2.3 rounds off with a discussion.

4.2.1 Policy Uncertainty: the Stochastic SPA

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

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

$$SPA_{t+1} = \min(SPA_t + e_t, 68) \tag{11}$$

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 68. This captures one of the key aspects 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 68.

As modelling policy uncertainty in this way represents innovation, a word about interpretation is prudent. In the model, the variable SPA_t represents the current best available information about the age the women will reach the SPA and as such the data analogue is the SPA the government is currently announcing for the women's cohort. Only one SPA cohort is modelled at a time. So there is no conflict in having a single variable SPA_t whilst in reality, at a given point in time, different birth cohorts have different government announced SPAs.

4.2.2 Costly Attention (Rational Inattention)

The second addition is the cost of information acquisition about the stochastic SPA. This allows the model to capture the fact that people are mistaken about their SPA, and that these mistaken beliefs are the results of an endogenous learning process. As such it creates a potential for the model to replicate the observed selection into being informed about your SPA and the pattern of learning.

Staes uncertain and learnable: To make the exposition of this new feature of the model, rational inattention to the SPA, as clear as possible I introduce two notational simplifications. I group decisions into a single variable $d_t = (c_t, l_t, a_{t+1})$ and all states other than the SPA into a single state variable $X_t = (a_t, w_t, AIME_t, ue_t, t)$.[16]

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

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

Wthin period timing of learning: As the household no longer directly observes SPA_t it is a hidden state. It is still a state as it is payoff relevant, but since the household does not observe it, it cannot enter the decision rule. This introduces a new state variable $\underline{\pi_t}$ the belief distribution the household holds about SPA_t. Since the household chooses how much information about the SPA to acquire, its choice can be thought of as a two-step process: first choosing a signal and then conditional on the signal draw choosing actions.[17] Provided they pay the utility cost of information, the choice of signal is completely unconstrained; the household is free to learn about SPA_t however they want. More precisely, a household with non-hidden states X_t and $\underline{\pi_t}$ is free to choose any conditional distribution function $\underline{f_t}[X_t, \underline{\pi_t}](z|SPA_t)$ for it's signal $z_t \sim Z_t$ given the value of the hidden state SPA_t.

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

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

Then the prior at the start of next period $\underline{\pi_{t+1}}$ is formed by applying the law of motion of SPA_t, equation 11, to this posterior.

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

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

subject to the same constraints 8 - 10 as the baseline model and where now the utility function includes a cost of information that is directly proportional to the mutual information (see below) between the signal and the household's current state of knowledge about the SPA $\underline{\pi_t}$.

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

One problem hidden in this bellmen equation is the formation of next period beliefs which, due to Bayesian updating, depend upon the full distribution of signals. This means that the continuation value is not known until the solution is known, this problem will be taken up in section 5.

Entropy and mutual information: Mutual information, rigorously defined below, is a concept from information theory. It is the expected reduction in uncertainty about one variable that results from learning another as measured by the entropy. Entropy is in turn a measure of uncertainty that capture the least space[18] needed to transmit or store the information contained in a random variable.

Definition 4.1 (Entropy)**.** *The entropy $H(.)$ of a random variable $X \sim P_X(x)$ is minus the expectation of logathim of its distribution: $H(X) = E_X[\log(p(X))]$.*

Definition 4.2 (Mutual Information)**.** *The mutual information between random variables $X \sim P_X(x)$ and $Y \sim P_Y(y)$ is the expected reduction in uncertainty about X resulting from learning Y measured by the entropy: $I(X,Y) = H(X) - E_Y[H(X|Y)]$.*

Utility: Hence utility incorporating information costs takes the form:

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

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

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

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

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

4.2.3 Discussion of Costly Attention to the Stochastic SPA

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

Expected Entropy Reduction Attention Cost: It is hopefully clear why a cost of information acquisition is included: to accommodate mistaken beliefs which are predictive of labour supply response to the SPA. Why the cost of attention takes the form it does may be less clear. As this utility cost of information acquisition represents a functional form assumption not widely used in the life-cycle literature, outlining the motivation for selecting this function may be instructive. Hence, I offer three motivating arguments for selecting this functional form to model costly information acquisition.

Firstly, it should be noted that 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 this literature. 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.[19] 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.

Not all reasons for using this functional form stem directly from the existence of supporting literature. The second desirable feature is that it can endogenously generate certain heuristics, or rules of thump, observed in the wild. One example, Kõszegi and Matějka (2020) show this cost of information can lead to mental budgeting and naive diversification[20], heuristics employed by individual investors. Another example comes from Caplin et al. (2019) who show it can lead to consideration sets: an observed heuristic employed in many discrete choice setting with large choice sets. This endogenous generation of observed heuristics is a very desirable feature; a good model should replicate observed behaviour. Many researchers are, however, understandably reluctant to hard code behavioural biases, for although people certainly use heuristics, no one follows the same heuristic regardless of how circumstances change. Furthermore, models with hard-coded behavioural biases suppress one of the central insights of economics: that people respond to incentives. A cost of information function that endogenously replicates observed heuristics avoids these pitfalls by allowing the household to follow heuristics when it is optimal to do so but not to be bound to them regardless of change.

[19] 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.

[20] Mental budgeting refers to consumption in a category being independent of price shocks to other categories whereas naive diversification refers to within category consumption allocations being fixed and unconsidered.

Thirdly, strong a priori reason to think that a cost of cognition should depend on entropy reduction exist. The information-theoretic concept of entropy was developed to explain how computers process information and gives a lower bound on the efficient transmission and storage of information. The computational theory of mind Mcculloch and Pitts (1943) holds that the human mind is a computer. This is controversial and well outside the scope of this paper, but even the most stringent opponents of this theory would agree the brain performs some tasks like a computer, with information processing a primary candidate. So, if the brain process information efficiently, mutual information should enter into the ideal cost of attention function. This is not to say an ideal cost of attention function would be linear in mutual information, but if it enters into the ideal then a first-order approximation along this dimension is a reasonable approximation when information processing is our focus. [21]

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

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

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

Interpreting the choice of signal: As our SPA is a number we can look up, this choice of a signal function may be difficult to conceptualise. The first thing to note is that looking up, perfectly remembering, and assimilating into one's action is not an information acquisition strategy that is excluded by the choice of a signal function conception: it corresponds to choosing a perfectly informative signal function. Carefully reading

[21] 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).

relevant regulations is not in reality the only way people learn about government policy in general or the state pension in particular. For example, people learn about how pension reforms affect them from other people and news outlets. In both examples, there is a random component, whether there is a newspaper story or other people talk about pensions, and a component that is a choice, whether you keep reading or ask follow-up questions. This is analogous to the choice of a noisy signal function in that it is partly a choice and partly stochastic and so this choice captures much about the messy real-world learning process.

5 Model Solution

By introducing a high dimensional state $\underline{\pi_t}$ and a high dimensional choice $\underline{f_t}$, rational inattention has complicated the model to the extent that solving it represents a novel contribution. To achieve this I weave together recent theoretical results into a consistent solution methodology for dynamic rational inattention models with endogenous heterogeneous beliefs, like the one presented above. Section 5.1 explains how this is done, both to communicate the methodological innovations and to give some intuition as to how the model is solved. First I provide some other details relevant to solving the model of this paper, but not relevant to solving generic dynamic models of rational inattention.

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

It is instructive to work through the transition from simple post-SPA periods to the complicated pre-SPA. We can solve the model with rational inattention by standard backward induction until we hit age 66. We can proceed as far as age 68 using standard methods because, as the state pension age process is bounded above, the woman receives her state pension with probability 1 from that point on. At age 67, because she knows the underlying data generating process just not the current value of SPA_t, if she is not in receipt of her state pension she knows her SPA is 68 with certainty. So at 67, SPA_t becomes a state variable, because whether or not the woman receives her state pension affects utility, but $\underline{\pi_t}$ is still not relevant, because she is perfectly informed. At age 66 whether or not $\underline{\pi_t}$ is a state depends on the value of SPA_t, and the same is true for all periods 60-66. If $t \geq SPA_t$ then the woman receives the state pension benefit; she knows the value of SPA_t, so $\underline{\pi_t}$ is degenerate and not a state, and she does not need to make an information acquisition choice. Hence, rational inattention is not relevant if $t \geq SPA_t$ and the period can be solved by simply searching for the optimal choice. For ages 55-59, rational inattention is always relevant as SPA_t is always above 60. Hence, because age 66 is the first period for which, when $t < SPA_t$, the true value of the SPA_t cannot simply be inferred (as it could be 67 or 68), age 66 is the first period in which the information acquisition choice is non-trivial and beliefs matter.

The solution of within period problems, when rational inattention matters, because $t < SPA_t$, is outlined immediately below in section 5.1. 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. More exhaustive computational details can be found in appendix D.

5.1 Dynamic Costly Attention Models with Endogenous Heterogeneous Beliefs

Dynamic rational inattention models with endogenous heterogeneous beliefs are complicated by the presence of a high dimensional state $\underline{\pi_t}$ and a high dimensional choice $\underline{f_t}$. This section presents my solution methodology. I use the model of retirement decision, presented earlier, to explain the methodology, but it has wider applicability: it applies to any dynamic rational inattention models with endogenous heterogeneous beliefs.

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

The rest of this section precedes as follows. Firstly, section 5.1.1 gives an outline of the proof of the main results from Steiner et al. (2017). Then section 5.1.2 will take the results from section 5.1.1 and present my solution method.

5.1.1 Analytic Foundations of Solution Methodology

Steiner et al. (2017) show that a wide class of similar models have a logit-like solution. [22] Merely citing their result would not provide any intuition. For this reason, and because an understanding of these results is needed to understand the solution methodology, in this section, I present an outline of their proof (see appendix C or the original paper for details).

Steiner et al. (2017) extend Matějka and McKay (2015)[23] to a dynamic setting and most of what is explained here applies equally to static problems. I explain what is relevant from Steiner et al. (2017) to my model using my model as a lens through which to explain their results.

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

$$\bar{V}^{(k)}_{t+1}(d_t, X_t, SPA_t, \underline{\pi_t}) = E\big[s^{(k)}_t V^{(k)}_{t+1}(X_{t+1}, SPA_{t+1}, \underline{\pi_{t+1}}) + (1 - s^{(k)}_t)T(a_{t+1})\big|d_t, X_t, SPA_t, \underline{\pi_t}\big] \tag{14}$$

where expectations are taken over the uncertainty in X_{t+1} and SPA_{t+1} and section 5.1.2 explains how to solve for $\underline{\pi_{t+1}}$. Then the Bellman equation 13 simplifies to:

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

Steiner et al. (2017) show that the solution to this model has action that are distributed with conditional choice probabilties $d_t|SPA_t \sim \underline{p_t}(d_t|SPA_t)$ and assocaited unconditional prbabitles $d_t \sim \underline{q_t}(d_t)$ that satisfy:[24]

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

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

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

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

[23] 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.

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

gives a noisy signal of the unobserved SPA_t, and which makes a decision contingent on the realisation of the signal $d(z)$.

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

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

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

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

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

and from the symmetry of mutual information: [25]

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

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

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

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

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

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

Taking $\underline{q_t}$ as given, optimality with respect to any $p_t(d|spa)$ requires the following FOC, derived from differentiating 21, be satisfied [26]

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

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

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

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

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

This derivation assumed $\underline{q_t}$ was given, but as $\underline{q_t}$ is the marginal to conditional $\underline{p_t}$ it is also chosen. The form of $\underline{q_t}$ can be found from substituting 16 into 21 and noting that the logarithm of the numerator in 16 cancels all other terms in 21 leaving only the summation from the denominator. So $\underline{q_t}$ can be found by solving:

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

5.1.2 Solution Methodology

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

The first major difficulty is that next period's beliefs given actions are not known until the full probability distribution of actions is known. This is because we do not know how strong a signal of a given SPA an action is

[26] The eagle-eyed reader may have noted that this treats the continuation value as fixed. Showing that "one can ignore the dependence of continuation values on beliefs and treat them simply as functions of histories" was a major achievement of Steiner et al. (2017) that I abstract from here to explain the intuition behind the results. I will touch again on this point briefly in section 5.1.2, but for a proper treatment please refer to the original paper.

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 $\bar{V}_{t+1}$ is not known, even when the next period's value function V_{t+1} is known, because we do not know the beliefs tomorrow that will result from an action today. Substituting the results of 16 and 17 into the Bayesian updating formula 12 gives

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

Then the prior at the start of next period $\underline{\pi_{t+1}}$ is formed by applying the law of motion of SPA_t, equation 11, to this posterior. Since the posterior depends not only on the exponentiated payoff but also on the $\underline{q_t}$ we need a solution to the model in order to know next period's beliefs given the chosen action and hence know the effective conditional continuation values:

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

Steiner et al. (2017) dodge this difficulty by removing the beliefs from the state space and replacing them with the full history of actions. They can do this because, given initial beliefs, the full history of signals, or equivalently actions, perfectly predicts the beliefs in period t. This is an inspired move for a theory paper and is a key step in extending Matějka and McKay (2015) to the dynamic case.[27] 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. [28] Redundant information in the state space is problematic because the curse of dimensionality means this is often one of the binding constraints in producing rich models. What moves this here from problematic to a non-starter is that this redundant information grows exponentially with the number of periods.

Hence, I rely on the theoretical results of Steiner et al. (2017) that used the history of action state-space representation, but in practice, I use the more compact belief state-space representation for the actual computational work. To get around the issue that I need $\underline{q_t}$ to know $\bar{V}_{t+1}$ 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 $\bar{V}_{t+1}$ I solve 17 for $\underline{q_t}$. If resulting $\underline{q_t}$ is sufficiently close to $\underline{\tilde{q}_t}$, I accept this solution otherwise I replace $\underline{\tilde{q}_t}$ with $\underline{q_t}$ and repeat. [29]

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

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

[29]Although, I have not proved this is a contraction mapping the fixed point itteration was always found to converge and generally

This solution to the first major difficulty, however, exacerbates the second, the high computational demands resulting from the high dimensional state $\underline{\pi_t}$, by increasing the computation required at each point in the state space. Here relief can be found from the results of Caplin et al. (2019), who show that generically rational inattention implies consideration sets. Hence, the solving conditional choice probabilities (CCPs) $\underline{p_t}$ are sparse. That is, various actions will never be taken. I can check for this sparsity, ex-ante, at various points in the process and remove any actions that will never be taken. This reduces the dimensionality of the optimisation in equation 17, but moreover, if after removing the actions that will never be taken we are left with a single action, then we have solved the problem without further calculation.

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

Removing actions that are strictly dominated only takes into account the ordinal characteristics of utility and not the cardinal aspect of inter-personal expected utility. Using the necessary and sufficient condition from Caplin et al. (2019), it is easily shown that if a there exists a decision $d^\star = (c^\star, l^\star)$ which satisfies

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

for all other decisions $d = (c, l)$ then it is the only action taken $q(d^\star) = 1$. [30] Unlike dropping strictly dominated alternative, which reduces the dimensionality and so makes solving equation 17 easier, checking equation 23 is only advantageous when the optimal behaviour is to take the same action in all realisations of SPA_t. As such the benefit of checking condition 23 depends on the problem faced and how frequently it shows the optimal solution without needing to solve an optimisation. For the retirement model in this paper, it was found useful.

Finally, when sparsity does not provide a shortcut to a solution I employ sequential quadratic programming to solve 17, an approach to static rational inattention problems suggested by Armenter et al. (2019). Hence bringing this together a high-level summary of the solution algorithm is:

in relatively few iterations.

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

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

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

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

else

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

while Error $>$ Tolerance **do**

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

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

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

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

else

if there is a action d that satiisfies 23 **then**

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

else

Solve 17 using sequential quadratic programming for $\underline{q_t}$

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

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

end if

end if

end while

end if

Substitute $\underline{q_t}$ into 16 to solve for $\underline{p_t}$.

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

6 Estimation

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

6.1 First Stage Estimation

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 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 starts from 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 $AIME_t$ the replicate the observed heterogeneity of $AIME_t$ with respect to assets (see appendix E for more details).

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

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

for individual j, of type k, in period t, where period t is indexed by female age and type k indicates whether high or low education and single or married. The parameters of the age-dependent deterministic component of the wage process ($\delta_{k0}, \delta_{k1}, \delta_{k2}$) are estimated by type-specific regression. The parameters of the stochastic component of the wage equation ($\rho_w, \sigma_\epsilon, \sigma_{\epsilon,55}, \sigma_\mu,$) 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 $\epsilon_t + \mu_{j,t}$.

Pension Systems: Both pension systems are type-specific functions of average lifetime earnings. These are estimated on the measure of $AIME_t$ constructed from administrative data that was described above. However, it is found the state pension is relatively insensitive to education and the private pension is relatively insensitive to marital status. Consequentially, I simplify to achieve better power and let the state pension function vary only by marital status and the private pension only by education.

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

Stochastic State Pension Age: I estimate the probability of an increase in the state pension ρ based on the changes to the final state pension age of women subject to the reform. As the SPA was 60 at the start of working life for everyone and each year an individual's SPA is impacted by a Bernoulli error term SPA_t has a binomial distribution for each t. I estimate the ρ that best matches this mixture of binomial distributions, to get a final 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). All 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 about these first-stage estimates are in appendix E.

6.2 Second Stage Estimation

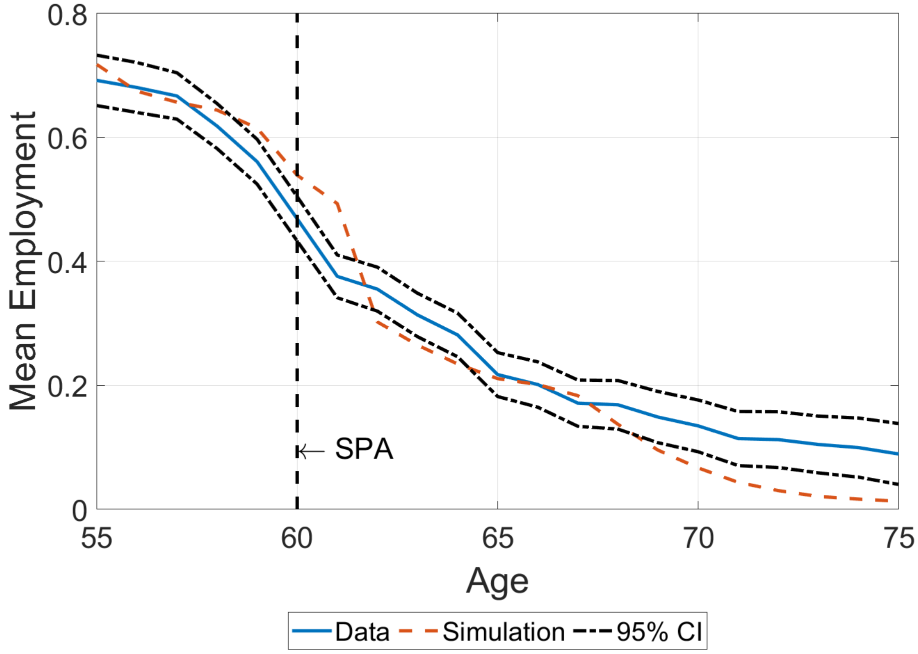
In the second step, moments are matched to estimate the preference parameters: the isoelastic curvature (γ), the consumption weight (ν), the discount factor (β), and the bequest weight (θ), as well as the cost of attention (λ) in the version with costly attention.

The moments used are the 42 pre-reform moments of mean labour market participation and asset holdings between 55 and 75. These profiles were estimated with for the $SPA = 60$ pre-reform data cohorts. To avoid contamination by cohort effects or macroeconomic circumstances a fixed effect age regression was estimated which additionally included: year of birth fixed 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.

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

Appendix E contains details on the construction of the targetted profiles.

Figure 6: Participation Profile



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

7 Results

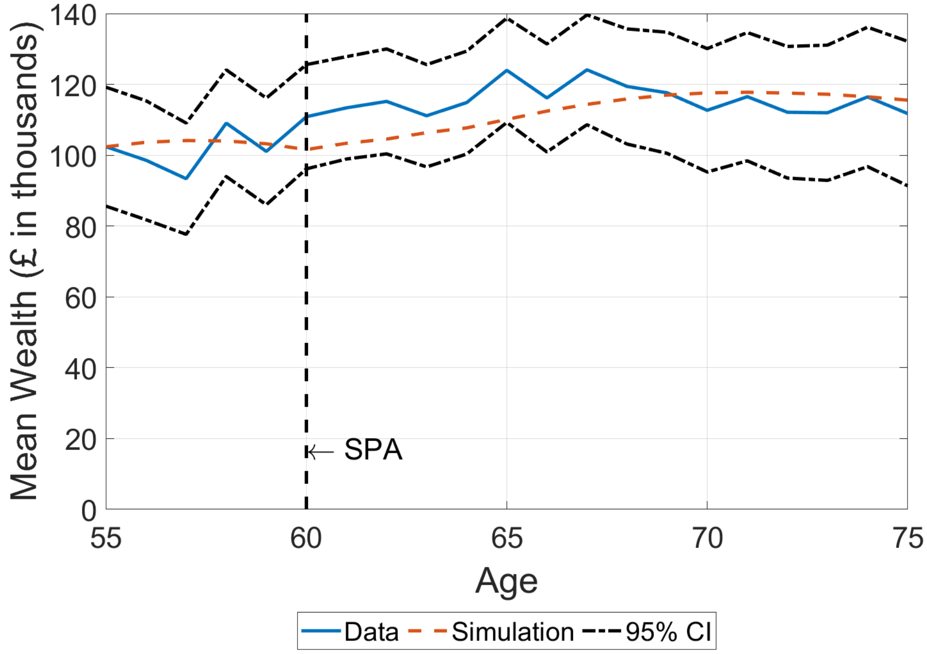
In this section, I present the model fit and given the parameter estimates investigate the model's ability to replicate the labour supply response to the SPA. Results of first stage estimation can be found in appendix F.

Figures 6 and 7 show the match of the pre-reform participation and asset profile for the baseline model with $SPA = 60$. Table 4 contains the corresponding parameter estimates. These are the goodness-of-fit for the baseline version of the model; the corresponding graphs for the versions with policy uncertainty and policy uncertainty and rational inattention can be found in appendix E but are practically indistinguishable from each other. Where the different versions of the model are clearly distinguishable is in how they replicate the dynamic reaction to the SPA as it is varied, as I show below.

With these parameters estimated, to investigate the response to the SPA I re-ran the model to generate simulated data with $SPA = 60$, $SPA = 61$, and $SPA = 62$ and re-ran the regression analysis from section 3.3 on this simulated data. The comparison between column 1 of table 5, containing the baseline model results, and column 4, containing the empirical counterparts, shows that the baseline model struggles to match both the aggregate response to SPA and the correlation of this response with wealth.

This motivates the introduction of policy uncertainty and costly attention. To see the results of each separately, I introduce them sequentially. Column 2 introduces policy uncertainty. As can be seen, policy uncertainty alone makes little to no difference. This is because the level of objective policy uncertainty is very low;

Figure 7: Asset Profile



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

Table 4: Parameter Estimates

ν: Consumption Weight	0.446
	(-)
β: Discount Factor	0.964
	(-)
γ: Relative Risk Aversion	1.873
	(-)
θ: Warm Glow bequest Weight	17,795
	(-)

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

Table 5: Regression Analysis on Simulated Data

	Baseline	Policy Uncert.	Costly Attention $\lambda \approx £490$	Costly Attention $\hat{\lambda} \approx £8$	Data
Population	Treatment Effect for being below SPA on participation				
Whole Population	0.0320	0.0342	0.0723	0.0519	0.112
Assets >Median(£34,869)	0.0525	0.0532	0.0652	0.0609	0.114

Notes: Difference-in-Difference Treatment effect estimates on simulated data of the increase in the probability of being in work from being below SPA for whole population and subpopulation with above the empirical median asset at SPA. The baseline is the model with a deterministic SPA. Policy uncertainty refers to the model with a stochastic SPA, and costly attention to the model with costly information acquisition to the stochastic SPA. The utility cost of information is converted into a monetary value by considering the consumption equivalent change for the median consuming household to equal the cost of going from completely uninformed (uniform prior) to complete information (degenerated posterior on the true SPA).

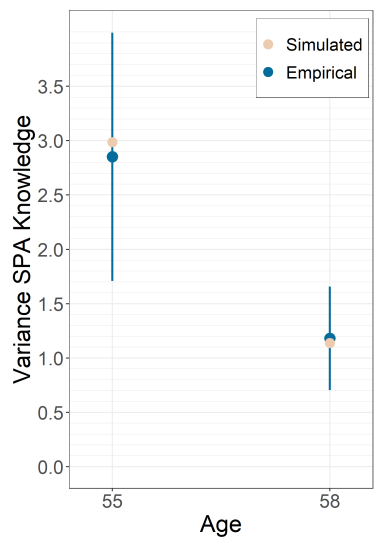
we observe changes to the SPA arrive very infrequently. Column 3 introduces costly attention to the stochastic SPA introduced in column 2. To do this a value of λ is needed, and λ is not an easily interpretable parameter having natural units of utils per bit. In column 3, I start with a relatively arbitrary value of $\lambda = 3 \times 10^{-4}$. To make this interpretable I first convert to the equivalent marginal consumption required to increase the median-consuming full-time-working household utility by this amount. This expresses λ in money units per bit, but expressing λ in this way exaggerates the cost of attention. This is because uncertainty and opportunities to learn are much more limited in this model, and all models, than in reality and so each bit represent a greater proportion of total uncertainty. Hence, I express λ as the utility cost of going from completely uninformed to perfectly informed about the current value of SPA_t for the median household; when expressed this way the cost of attention used in column 3 is £490. As can be seen, the treatment effect in both the whole population and those with above median assets move substantially in the direction of the data. It matches the treatment effect for the above-median assets subgroup but for the whole population still falls short of the empirical counterpart.

The SPA knowledge data in ELSA, however, offers the opportunity to improve on the arbitrary set value of λ, as it offers clear and direct identifying variation. Exploiting this I pin down λ by the reduction in variance in SPA answers between ages 55 and 58. Figure 8 shows the fit, which is extremely close. The estimated cost of attention λ is much lower than the one used in column 3 of table 5, implying that for the median household the utility cost of going from completely uninformed to perfectly informed of their current SPA equals the utility of an additional £8. Column 3 of 5 includes the regression analysis with this lower value of λ, and it can be seen costly attention still improves on the fit of the baseline but the improvement is much more muted.

Expressing each cost of attention as a single monetary value is a convenient shorthand but masks the resulting heterogeneity. Table 6 shows some summary statistics of the distribution of implied attention cost by education level for the two values of λ considered so far. This gives some idea of the possible welfare gains from reducing this informational friction, for example by sending more letters.

A motivation for investigating the role of informational frictions in this excess sensitivity puzzle was that

Figure 8: Variance in SPA Knowledge



Notes: Reduction in variance in self-reported SPA between 55, the first model period, and 58. Age 58 is the last age at which we can be sure no one has received communication from the government about their SPA, as letters were sent sometime in the 6 months before reaching SPA and the youngest SPA is 60.

Table 6: Sumarry Stastics of Distribtion of Attention Cost

λ	Education	Median	5% Percentile	95% Percentile
Low	Low	7.00	4.00	14.00
Low	High	10.00	5.00	28.00
High	Low	383.00	200	843.00
High	High	599.00	294	1618.00

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

Table 7: Model Predictions for Different Costs of Attention

λ	$\approx$ £1	$\approx$ £10	$\approx$ £100	Data
Population	Treatment Effect for being below SPA on participation			
Whole Population	0.0118	0.0214	0.0386	0.109
Assets >Median(£34,869)	0.0456	0.0527	0.0827	0.089
Age	Variance of SPA Answers			
55	2.985	2.985	2.985	2.852
58	0.497	1.138	2.852	1.180
Coefficient	Treatment Effect Hetrogeneity by SPA Error			
Treatment Effect	0.0237	0.0164	0.0337	0.157
Interaction	-0.0161	0.0068	0.0043	-0.023

Notes: The columns show results from three separate costs of attention. The top panel shows labour supply response across the wealth distribution as per table 5. The second panel shows the reduction in self-reported SPA 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.

mistaken beliefs predict the labour supply response to the SPA; a natural question is whether the model can replicate this relationship. As was shown in section 3, those who are better informed of their SPA in their late 50s have a larger labour supply response upon reaching their SPA in their 60s. Two countervailing forces exist in the model linking the degree of SPA knowledge to the labour supply response to the SPA. Firstly, SPA knowledge is endogenous implying those whose actions depend least on the SPA acquire the least information about it. This mechanism pushes in the direction of the empirical finding of a negative relationship between SPA knowledge and labour supply response. Conversely, if we compare to ex-ante equivalent households where, by luck of the draw, one ended up worse informed than the other then the worse informed household will receive a larger shock upon discovering their SPA and so have a larger reaction. Hence whether the model generates a positive or negative relationship between the degree of SPA knowledge and the labour supply response to the SPA depends on which dominates. The first columns in the bottom panel of table 7 shows that for the values of λ considered we get a slightly negative relationship or no relationship.

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

8 Extension

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

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

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

8.1 Deferral Puzzle

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

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

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

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

8.2 Model and Estimation

Benefit claiming is a binary decision and having claimed is an absorbing state: once an individual claims the state pension they cannot unclaim. Benefit claim is only an option once past the SPA, and to keep the problem tractable an upper limit of 70 is placed on deferral.

To keep the state space manageable, stochastic deferral adjustment is modelled as iid with two points of support. Having only two points of support limits the growth of the state due to solving the model with different values for the adjustment rate to a factor of two. Having the uncertainty be iid means that beliefs do not enter as a state variable as yesterday's learning is not relevant to today's knowledge; the agent knows the underlying probabilities so these form their prior each period. As benefit claiming is an absorbing state an indicator of having claimed or not also expands the state space.

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

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

Next, I introduce costly attention with a cost of attention corresponding to approximately £10 of consumption to the median consuming household to be fully informed. This increased the number voluntarily claiming to

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 participation	
Whole Population	0.0416	0.109
Assets >Median(£34,869)	0.0903	0.089
Age	Variance of SPA Answers	
55	2.985	2.852
58	1.795	1.180
Coefficient	Treatment Effect Hetrogeneity by SPA Error	
Treatment Effect	0.0532	0.157
Interaction	-0.0111	-0.023

Notes: Costly attention refers to the model with, additionally, a cost of information acquisition about the stochastic policy. Top panel shows labour supply response across the wealth distribution as per table 5. The second panel shows the reduction in self-reported SPA between 55 and 58. The bottom panel shows, in the interaction term, the heterogeneity in labour supply response to the SPA by self-reported SPA error at age 58.

22.2%, approximately a fourfold increase on the model without informational frictions, but still short of the rate observed in the data. As can be seen in table 9, this cost of attention produced a relatively good fit along all dimensions of interest.

9 Conclusion

This paper shows that incorporating one empirical regularity, mistaken beliefs resulting from information frictions, into a model of retirement can help explain other puzzling empirical regularities, in particular, the excess sensitivity of employment to statutory retirement ages. I find that including rational inattention to an inherently uncertain pension policy significantly improves the model prediction of the labour supply response to the SPA.

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

By including an explicit model of belief formation this paper takes an approach to the beliefs preferences identification problem that avoids loading all explanations onto preferences by making the same sort of functional form assumptions about beliefs that are routinely made about preferences. This paper then uses beliefs data to pin down the cost of attention. Very modest costs of attention, in the range of £8-490 to become fully informed of your current SPA, rationalise the key features of the data.

Finally, I present an extension of the main model with a mechanism to explain another puzzle: that people do not take up more than actuarially advantageous deferral options. The insight offered by this extension is that the assertion that deferral is actuarially advantageous ignores the attention cost which can be avoided by claiming; hence this assertion omits an incentive not to defer.

References

Amin-Smith, N. and R. Crawford (2018). State pension age increases and the circumstances of older women. In *The Dynamics of Ageing: Evidence from the english longitudinal study of ageing*, pp. 9–39.

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

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

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

Behaghel, L. and D. Blau (2012). Framing Social Security Reform : Behavioral Responses to Changes in the Full Retirement Age. *American Economic Journal: Economic Policy 4*(4), 41–67.

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

Bozio, A., R. Crawford, and G. Tetlow (2010). The history of state pensions in the UK: 1948 to 2010. Technical report.

Braun, R. A., K. A. Kopecky, and T. Koreshkova (2017, mar). Old, Sick, Alone, and Poor: A Welfare Analysis of Old-Age Social Insurance Programmes. *The Review of Economic Studies 84*, 580–612.

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.

Coile, C. and J. Gruber (2007). Future Social Security Entitlements and the Retirement Decision. *The Review of Economics and Statistics 89* (2), 234–246.

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. Technical report.

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.

Gustman, A. and T. Steinmeier (2001). Imperfect Knowledge, Retirement amd Saving. *NBER Working Paper 3* (5).

Gustman, A. L. and T. L. Steinmeier (1986). A Structural Retirement Model. Technical Report 3.

Gustman, A. L. and T. L. Steinmeier (2005, feb). The social security early entitlement age in a structural model of retirement and wealth. *Journal of Public Economics 89* (2-3), 441–463.

Guvenen, F. (2009). An empirical investigation of labor income processes. *Review of Economic Dynamics 12* (1), 58–79.

Kõszegi, B. and F. Matějka (2020, may). Choice Simplification: A Theory of Mental Budgeting and Naive Diversification. *The Quarterly Journal of Economics 135*(2), 1153–1207.

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

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

Low, H., C. Meghir, and L. Pistaferri (2010, sep). Wage risk and employment risk over the life cycle. *American Economic Review 100*(4), 1432–1467.

Lumsdaine, R. L., J. H. Stock, and D. A. Wise (1996). Why Are Retirement Rates So High at Age 65? In *Advances in the Economics of Aging*, pp. 61–82.

Luo, Y. (2008). Consumption dynamics under information processing constraints. *Review of Economic Dynamics 11*(2), 366–385.

Luttmer, E. F. and A. Samwick (2018). The Welfare Cost of Perceived Policy Uncertainty: Evidence from Social Security. *American Economic Review 108*(2), 275–307.

Macaulay, A. (2021). Cyclical Attention to Saving *.

Maćkowiak, B. and M. Wiederholt (2009). Optimal sticky prices under rational inattention. *American Economic Review 99*(3), 769–803.

Maćkowiak, B. and M. Wiederholt (2015). Business cycle dynamics under rational inattention. *Review of Economic Studies 82*(4), 1502–1532.

Manoli, D. and A. Weber (2016, nov). Nonparametric evidence on the effects of financial incentives on retirement decisions. *American Economic Journal: Economic Policy 8*(4), 160–182.

Manski, C. (2004). Measuring expectations. *Econometrica 72*(5), 1329–1376.

Mastrobuoni, G. (2009). Labor supply effects of the recent social security benefit cuts: Empirical estimates using cohort discontinuities. *Journal of Public Economics 93*(11-12), 1224–1233.

Matějka, F. and A. McKay (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review 105*(1), 272–298.

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

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

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

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

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

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

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

Details available upon request.

B Additional Model Detail

Details available upon request.

C Additional Mathematical Details

Details available upon request.

D Additional Computational Details

Details available upon request.

E Additional Econometric Details

Details available upon request.

F First Stage Estimates

Details available upon request.

G Alternative Model Specifications

Details available upon request.