Costly Attention and Retirement *

Jamie Hentall-MacCuish†

Abstract

In UK data, I document the prevalence of misbeliefs regarding the State Pension eligibility age (SPA) and these misbeliefs predictivity of retirement. Exploiting policy variation, I estimate a lifecycle model of retirement in which rationally inattentive households learning about uncertain pension policy endogenously generate misbeliefs. Endogenous misbeliefs explain 43%-88% of the excessive (given financial incentives) drop in employment at SPA. To achieve this, I develop a solution method for dynamic rational inattention models with history-dependent beliefs. Costly attention makes the SPA up to 15% less effective at increasing old-age employment. Information letters improve welfare and increase employment.

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

JEL CLASSIFICATION: D15, D83, D91, E21, J26, H55

*I thank Fabien Postel-Vinay and Eric French for detailed discussions and encouragement. For their helpful comments, I also thank James Banks, Richard Blundell, Uta Bolt, Jonathan Cribb, Martin Cripps, Aureo de Paula, Mariacristina De Nardi, Gaetano Gaballo, Attila Lindner, Rory McGee, Tomasz Michalski , Lars Nesheim, Cormac O'Dea, Franck Portier, Johannes Spinnewijn, Ran Spiegler, Imran Rasul, Morten Ravn, Morgane Richard, Victor Rios-Rull, and participants at the NBER Summer Institute Behavioral Macro Session, IIPF Annual Congress, RES Junior Symposium, CESifo Public Area Conference, Young Economist Symposium, NETSPAR Pensions Workshop, Econometric Society European Meeting, SED Annual Meeting, CERGE-EI, ENTER, Toulouse, Cambridge, UCL, UWO, UNSW, Tilburg, Edinburgh, Royal Holloway, HEC, and the IFS. For thoughtful discussion, I thank Andrew Caplin, Dylan Moore, Leanne Nam, Nicolas Fernandez-Arias, Jim Been, Patrizia Alexandra Massner, and Sebastian Seitz. At different times funding for this work has been received from Grant Inequality and the insurance value of transfers across the life cycle (ES/P001831/1) and ESRC studentship (ES/P000592/1), and this funding is gratefully acknowledged

†HEC Paris, correspondence to hentall-maccuish@hec.fr or 1 Rue de la Libération, 78350 Jouy-en-Josas +330139677000

1 Introduction

Understanding the cause of apparent deviations from rationality is crucial for policy design. We cannot address these deviations if they stem from fixed features of household behavior, often termed behavioral biases. However, mistaken beliefs about the policy itself can cause similar departures from apparent rationality if not accounted for. In such cases, straightforward information provision might mitigate these deviations. This paper shows that misbeliefs explain a puzzle often attributed to behavioral biases: the excessively large drop in employment at pension eligibility age, despite weak economic incentives to stop working precisely then.[1] It does so by estimating a model on UK data where observed misbeliefs are generated as an optimal response to costly attention.

Retirement is a compelling context to study the impact of misbeliefs due to their prevalence.[2] Many people are confused about pensions. In my data, 60% of women affected by pension age reform are mistaken about their pension age by over a year when within 2-4 years of eligibility. Initially, these misbeliefs seem strange since the information is financially relevant and freely available However, they become less surprising when we acknowledge that government policy is objectively uncertain (changing in unpredictable ways), and information is costly. This paper examines how the interplay between policy uncertainty and costly information acquisition (or attention) driving these misbeliefs influences retirement. Specifically, it asks whether this interplay helps explain excessive employment sensitivity to pension eligibility.

To investigate, I first document key facts on mistaken beliefs and excess employment sensitivity, then build a model with information frictions in the form of costly attention to explain them. Specifically, I estimate a dynamic lifecycle model of retirement (e.g. French, 2005) with rationally inattentive households (e.g. Sims, 2003) deciding how much information about a changeable pension policy to acquire whilst incurring a disutility cost of information. The model endogenously generates observed mistaken beliefs, which help generate the otherwise puzzling sharp employment drop at pension eligibility age. This drop is puzzling as the UK pension benefits aren't tied to employment, so the State Pension Age (SPA) only incentivizes retirement for liquidity-constrained individuals who can't substitute intertemporally. Yet, employment also falls for those with substantial liquid wealth.

Counterintuitively, unawareness of the SPA is not only consistent with high employment sensitivity to the SPA but is essential to generating it. The revelation of information upon reaching eligibility explains this. In the model, households pay a utility cost to learn their eligibility age (SPA), modeled as stochastic to capture potential government reforms. Upon reaching the SPA, its value becomes fixed and is revealed, reflecting communication of eligibility and information disclosure during claiming. Thus, reaching the SPA

[1]This puzzle appears in multiple countries with varied institutional setups, e.g., the US Behaghel and Blau (2012), Germany Seibold (2021), Switzerland Lalive et al. (2023), and Finland Gruber et al. (2022). Seibold (2021), Lalive et al. (2023), and Gruber et al. (2022) attribute it to behavioral biases.

[2]Documented, e.g., in Manski (2004), Lusardi and Mitchell (2011), Bairoliya and McKiernan (2023), Ciani et al. (2023).

is a positive information shock. It is also a positive wealth shock because as households age past earlier alternative eligibility ages without receiving benefits, they rule those ages out, making now the earliest possible eligibility age. This information shock reduces precautionary labor supply, and since leisure is a normal good, the wealth shock further reduces labor supply. Mistaken beliefs from costly information acquisition amplify these positive shocks at the SPA.

These model mechanisms rely on the potential for government changes to the SPA. Reforms in 1995 and 2011 demonstrate this potential, but the mechanism depends only on the possibility of reform, not its occurrence. However, I use the occurrence of reforms for identifying variation. Firstly, to estimate the probability of reform, and second, to causally identify the effect of the SPA on employment. Since the 1995 reform affected only the female SPA, this paper focuses on female retirement.

The English Longitudinal Study of Ageing (ELSA), a micro panel survey, provides data to study mistaken beliefs and their impact on employment, containing self-reported and true SPAs along with detailed information on assets, labor market status, and demographics. It is also linked to administrative records, particularly social security contributions, enabling the estimation of individuals' State Pension entitlements.

Observing self-reported and true SPAs reveals misbeliefs. As noted, less than four years from SPA, most reform-affected women's beliefs are out by over a year. Misbeliefs about the SPA predict employment responses to it, motivating the joint study of misbeliefs and excess sensitivity. Women more mistaken about their SPA in their late 50s show a smaller response upon reaching it in their early 60s. The model replicates this pattern because varying returns to information lead to selection into attention. Women unconcerned by the SPA neither learn nor respond to it. Misbeliefs drive excessive employment responses, but selection into SPA knowledge explains why more mistaken individuals respond less. Thus, information endogeneity and return heterogeneity are crucial for replicating the relationship between beliefs and employment.

So, the endogeneity of beliefs drives the relationship between retirement and mistaken beliefs but it complicates the model by introducing a high-dimensional state (prior beliefs) and choice (learning strategy). In static rational inattention models, prior beliefs represent ex-ante heterogeneity, but in dynamic models, today's learning affects tomorrow's beliefs, making them a state variable. Many papers sidestep this by suppressing prior beliefs as a state variable.[3] While reducing the state space is beneficial and suppressing beliefs can be a good modelling assumption for specific situations, it limits the domain of application by implying beliefs are irrelevant. It cannot capture scenarios where data shows beliefs matter and vary across individuals, like UK pension beliefs. I develop a solution method for dynamic rational inattention models that accommodates history dependence by treating beliefs as a state. The method is general-purpose in that it models beliefs non-parametrically without restricting the data-generating process. It relies on theoretical results from Steiner et al. (2017) about dynamic rational inattention models and addresses computational

[3] For example Miao and Xing (2024); Armenter et al. (2024); Turen (2023); Macaulay (2021); Porcher (2020).

challenges of high-dimensional states using the sparsity shown to be a general property of rational inattention models by Caplin et al. (2019).

I estimate the model using two-stage simulated method of moments, targeting asset and employment profiles. Policy uncertainty combined with costly attention increase the employment response to the SPA compared to a complete information baseline, explaining 43%-88% of the shortfall. As my model endogenously generates beliefs, I can match predicted beliefs to the data to identify the cost of attention, thus addressing the belief-preference identification problem (e.g. Manski, 2004). The mean household is willing to pay £11.00 to learn today's SPA, so estimated attention costs are low, consistent with other findings (e.g. Chetty, 2012). Despite small attention costs, the marginal benefit of pension policy information letters outweighs the cost. Large employment changes stem from small attention costs because people near retirement are close to their participation margin.

Pension eligibility ages are considered key to increasing old-age labor force participation, which is a common policy goal (e.g. Kolsrud et al., 2024). Since costly attention increases employment response *at* the SPA compared to full information, one might assume it makes the SPA a better tool for this purpose. The opposite is often true. Policy experiments comparing employment increases resulting from SPA changes in version of the model with and without information frictions show costly attention shifts part of the informed agent's response forward but can lower the overall response. Informed agents increase labor supply immediately, while less informed individuals, facing learning costs, respond closer to their SPA. Thus, informing individuals, for example, by sending letters, can raise old-age employment by up to 15%. In most policy experiments, the marginal benefits to households and extra tax revenue from these letters each separately outweigh their costs. Considered jointly, information letters are always welfare-enhancing.

Section 2 reviews the literature. Section 3 provides context. Section 4 presents the data and Section 5 descriptive and reduced-form analysis. Section 6 introduces the model, starting with a complete information baseline then adding pension policy uncertainty and costly attention. Section 7 explains the solution method. Section 8 covers estimation. Section 9 discusses model fit and implications. Section 10 concludes.

2 Related Literature

This paper builds primarily on two literatures: dynamic lifecycle retirement models and rational inattention. It also relates to research on excess employment sensitivity and subjective beliefs. This section situates the work within these areas and the broader literature, selectively reviewing key relevant papers.

Dynamic lifecycle models of retirement. Lifecycle retirement models start with Gustman and Steinmeier (1986) and Burtless (1986). Rust and Phelan (1997) introduce uncertainty and incomplete markets by excluding savings. French (2005) reintroduces savings with borrowing constraints and innovations such as a fixed cost of work. van der Klaauw and Wolpin (2008) incorporate Medicare, and French and Jones

(2011) add medical expenses. Much of this literature is US-focused, and some of its concerns, like medical insurance, are not relevant in the UK. My model includes uncertainty, borrowing constraints, and individual heterogeneity. The closest paper from this literature is O'Dea (2018), who models male UK retirees.

Rational inattention. Rational inattention was first applied to add costly attention to macroeconomic models (e.g., Sims, 2003; Maćkowiak and Wiederholt, 2009, 2015). Its scope now extends to, for instance, industrial organization (Brown and Jeon, 2024), and field experiments(Bartoš et al., 2016). Matějka and McKay (2015) solve a general class of static discrete choice models with rationally inattentive agents. Crucially for this paper, Steiner et al. (2017) extends these results to dynamic discrete choice models. A key contribution of this paper is turning the theoretical solutions of Steiner et al. (2017) into a solution method for quantitative dynamic rational inattention models with history-dependent beliefs. Caplin et al. (2019) show rational inattention generically implies consideration sets, meaning solutions are sparse, which I leverage to reduce computational burden. Dynamic rational inattention typically avoids these computational issues by suppressing the belief distribution as a state variable (e.g. Miao and Xing, 2024; Armenter et al., 2024; Turen, 2023; Macaulay, 2021; Porcher, 2020). While reasonable for specific cases, this approach is not fully general and limits the range of solvable problems. Afrouzi and Yang (2021) also propose a method for dynamic rational inattention that incorporates beliefs as a state variable. They use the linear-gaussian-quadratic framework popular in macro rational inattention to speed up solutions, whereas my approach handles arbitrary noise and non-linear utility but lacks these performance gains. Finally, Boem (2023) estimates a lifecycle model of older individuals, focusing on how rational inattention affects the one-shot static choice of purchasing an annuity.

Excess employment sensitivity. A puzzlingly large drop in employment at statutory pension ages is observed across countries. Lumsdaine et al. (1996) document this excess employment sensitivity puzzle in the US. The consensus was that liquidity constraints explained the drop at age 62, and Medicare eligibility explained the drop at age 65 (Rust and Phelan, 1997; French, 2005; French and Jones, 2011). Testing these explanations became feasible after 2004 when the full retirement age increased, allowing estimation of its employment impact. Part of the age 65 spike followed the full retirement age, even though Medicare eligibility stayed at 65 (Behaghel and Blau, 2012), and Mastrobuoni (2009) found larger effects than standard models predicted. Other governments raising pension ages saw similar outcomes: pension age increases induce larger labor supply responses than standard models predict (e.g. Seibold, 2021; Lalive et al., 2023). I document this in the UK, analyzing female state pension age reforms and extending Cribb et al. (2016) by using richer data to rule out other potential explanations for the employment response.

Subjective belief data. The use of subjective belief in just structural microeconomic models is extensive (Koşar and O'Dea, 2022). Most papers, however, do not model belief formation, limiting counterfactual analysis (e.g. Bairoliya and McKiernan, 2023; de Bresser, 2023). Modeling belief formation as an optimal response to processing costs allows me to match model-generated beliefs to data instead of treating beliefs purely as input. So, my solution method opens new avenues for using belief data with dynamic rational inattention models. Extensive research tests rational inattention on inflation beliefs (e.g. Coibion and Gorodnichenko, 2015, 2012), but even here integration into dynamic models is recent (Afrouzi et al., 2024). Early studies of pensions beliefs (e.g. Bernheim, 1988; Manski, 2004) document misbeliefs about benefit levels. Caplin et al. (2022) find substantial misbeliefs about eligibility ages in Denmark, similar to my findings in the UK. I use belief data to set initial conditions and identify parameters from patterns in beliefs and choices (patterns akin to Amin-Smith and Crawford (2018), prevalent misbeliefs predicting labor supply responses, and Rohwedder and Kleinjans (2006), errors decline as individuals age toward eligibility).

Wider literature. This paper sits in the tradition of behavioral public economics (e.g. Chetty, 2015). By examining attention to the benefit system, it complements literature on inattention to taxes (e.g. Chetty et al., 2009; Taubinsky and Rees-Jones, 2018; Farhi and Gabaix, 2020). While inattention to benefits has been less studied except for its impact on benefit take-up (e.g. Goldin et al., 2022), this paper shows its implications can be significant. The main policy experiment—sending informational letters—relates to studies on information provision in social security contexts (Mastrobuoni, 2011; Liebman and Luttmer, 2015; Dolls et al., 2018). Unlike these experimental works, the model here enables quantification of welfare and life-cycle effects rather than local effects. Luttmer and Samwick (2018) use revealed preference on hypothetical choices to estimate the welfare cost of pension policy uncertainty. My estimates are complimentary to those as they concentrate on the cost of objective uncertainty and I concentrate on the cost of misbeliefs.

3 Background

The UK State Pension system has undergone many changes since its 1948 introduction. I focuses on the system during 2000-2016, particularly post-2010, when the female SPA reform began.

State Pension benefit level. The UK State Pension comprises two parts: the Basic State Pension, based on contribution years, and a second tier, based on earnings, both calculated over the working life. Working life is defined as spanning from the tax year an individual turns 16 to the year before they reach SPA (Bozio et al., 2010). So, benefit entitlements are frozen one year before SPA meaning labor supply decisions near SPA do not affect the pension amount.

The Basic State Pension began in 1948. By 2013, a full pension paid £107.45 per week ($203 in 2022 dollars, adjusted using CPI deflators and OECD PPP exchange rates). Pro-rata payments apply to those with fewer than 30 contribution years needed for the full pension. Years in the labor force (earning above a minimum threshold) and years spent caring for a child or disabled person post-1978 count toward entitlement. So, the timing of labor participation and reasons for non-participation affect the pension amount.

The second tier of the State Pension began in 1978. Initially, it used an index-linked average of earnings between lower and upper limits over working life. Legislative changes resulted in varying accrual rates from 1978 to 2002, with a more progressive formula applied after April 2002. Thus, the timing of earnings affects second-tier entitlements. Private pension holders could opt out for reduced payroll taxes.

Even in this simple outline we see that due to protections for entitlements accrued under changing policies, the state pension benefit depends not only on total earnings and labor force participation but also on their timing and other factors (see Bozio et al., 2010, for details). Still, some general trends emerge. First, it is a relatively low benefit. OECD (2011) estimates it provides a 37.4% net replacement rate for median earners, compared to 47.3%, 50.0%, and 58.15% in the USA, OECD, and EU, respectively. Second, it is a relatively flat-rate benefit. This is evident in the sharper drop in replacement rate when moving from half to one-and-a-half times median earnings—35.2 percentage points in the UK versus 16.9, 21.2, and 14.0 in the USA, OECD, and EU (OECD, 2011). Finally, it is worth restressing that benefit entitlement is frozen the year before SPA, making it unaffected by labor supply choices near SPA.

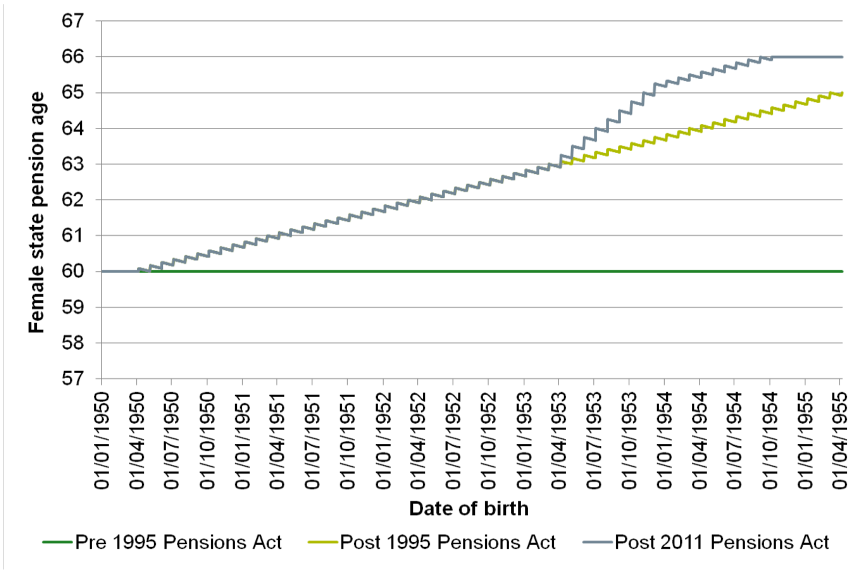
State Pension age and its reform. The UK State Pension Age (SPA) is the earliest age claim at which the State Pension, can be claimed serving as the system's early retirement age. Deferring receipt increased benefit generosity, but without a cap on deferral duration, and so did not imply an effective a full retirement age. [4] Thus, the SPA is the sole focal age of the UK state pension system.

Unlike the State Pension amount, the SPA is a simple function of birth date and gender. The SPA was 65 for men and 60 for women until the Pensions Act 1995, which raised the female SPA from 60 to 65 incrementally—one month every two months—over ten years starting April 2010. The Pensions Act 2011 accelerated this change from April 2016, equalizing SPAs by November 2018, and legislated an increase for both genders to 66, phased in from December 2018 to October 2020. Figure 1 shows how these changes affected women by birth cohort. These reforms also allow estimation of the risk UK women face of SPA changes during their life, a key input for the model. To avoid confounding from a 2016 benefit level change, I use only the variation in female SPA from the 1995 reform to identify SPA impacts on employment.

Communication and lack thereof. The government did not directly inform women affected by the reform, providing only a standard letter received by all pre-reform cohorts about four months before SPA. This lack of communication sparked controversy. In 2015, two campaign groups claimed the reforms discriminated against older women, with one unsuccessfully seeking to reverse the changes in a 2019 high court case. Their argument focused on the lack of communication. The government defended this by citing the absence of a national database in 1995, claiming direct notification was "essentially impossible". Reconciling this with their ability to send letters at SPA is beyond this paper's scope, two points stand out. First, the absence

[4] Despite generous actuarial adjustments, deferral was rare, presenting a puzzle. Online Appendix F offers a model extension addressing this, but elsewhere, I abstract from the deferral puzzle and assume observed claiming behavior.

Figure 1: SPA by Date of Birth under Different Legislation



*Notes:*State Pension Age for women under different legislation. Source: Pensions Act 1995, schedule 4; Pensions Act 2007, schedule 3 ; Pensions Act 2011, schedule 1.

of protests until 20 years after the reform supports the view that reported misbeliefs are genuine. Second, the lack of communication until SPA motivates the model assumption that SPA uncertainty resolves at SPA.

Private pensions. A large voluntary private pension market supplements the State Pension, accounting for 42% of household wealth. Since private pension eligibility is not tied to SPA, it has little relevance to the employment response to SPA (see Online Appendix A for evidence). The modest amounts in the State Pension make the strong employment response to it even more surprising.

Excess employment sensitivity and State Pension age. The UK SPA reform offers a unique opportunity to examine the excess employment sensitivity puzzle, as many common explanations for labor market exits at early retirement age are ruled out. First, UK law prohibits mandatory retirement based on age, classifying it as illegal age discrimination.[5] Thus, firm-mandated retirement cannot explain SPA employment sensitivity. Second, the state pension is not tied to employment status; individuals can claim it and continue working, and many do. Third, the UK pension system lacks major tax incentives for labor market exits at SPA. Unlike the US system, there is no earnings test,[6] and while the state pension is taxable, a component of income tax, called National Insurance contributions, is removed at SPA.[7]

These three facts show the State Pension acts as an anticipatable increase in non-labor income, with the SPA as its eligibility age. Announced in 1995 and starting in 2010, the reform provided at least 15 years of advance notice. The puzzle is not that employment responds to the reform but that the response is concentrated at the SPA despite the long notice period. In a standard life-cycle model with complete

[5]The Equality Act (2006) banned mandatory retirement below age 65, exceeding the highest SPA in this paper. The Equality Act (2010) extended the ban to all ages with exceptions in online Appendix A.

[6]Earnings tests penalize working while claiming retirement benefits, but they are *not* a feature of the UK system.

[7]Cribb et al. (2016) find changes to participation tax rates at SPA do not explain the employment response.

information and forward-looking agents, labor supply does not sharply respond to anticipatable income changes unless liquidity constraints prevent intertemporal smoothing. Liquidity-constrained individuals cannot borrow against future pension income, forcing them to wait for this income to reduce labor supply.[8] Thus, liquidity constraints are the only standard explanation for employment sensitivity to the SPA.

4 Data

Studying labor supply responses to the State Pension Age (SPA) requires a large sample of older individuals and rich microdata to explore the causes of the response. I use the English Longitudinal Study of Ageing (ELSA), as it is the UK[9] dataset best suited to these needs.

ELSA is a biennial panel dataset sampling the English population aged 50 and over, modeled on the US Health and Retirement Study (HRS). It provides rich microdata on labor market circumstances, earnings, and asset holdings. From wave three onward, ELSA collects data on SPA knowledge, crucial for studying erroneous beliefs. ELSA requests National Insurance numbers (equivalent to a US Social Security number) and consent to link administrative records, with 80% of respondents agreeing. These records improve pension entitlement estimates, key for modeling SPA incentives. Survey data on health, education, and family further illuminate retirement motivations.

ELSA waves 1 (2002/03) through 7 (2014/15) cover those affected by the 1995 pension age reform, forming the basis for analysis. The main sample includes women aged 55–75, capturing key SPA-related ages, with 24,968 observations of 7,165 women. Different samples are used only when estimating particular model inputs, such as the spousal income process (dropping females not males) or mortality process (including older ages). The female SPA reform began in 2010, making wave 5 the first post-reform wave. Earlier waves control for pre-trends and inform model input estimation. The oldest reform affected cohort was born on 6 April 1950. Older cohorts serve as a control group and also inform model input estimation.

5 Key Motivating Facts

5.1 Excess Employment Sensitivity

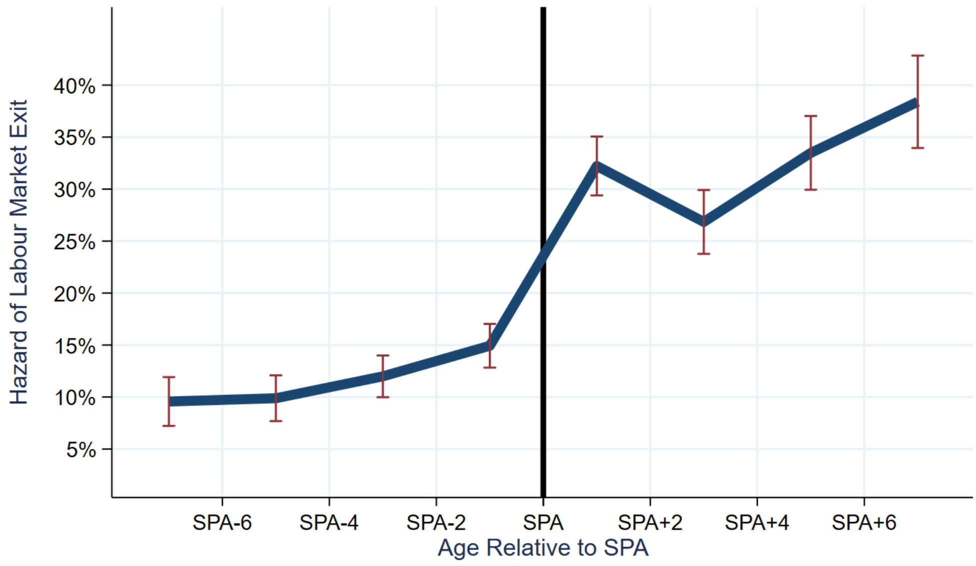
The sensitivity of employment beyond incentive to official retirement ages is a puzzle observed in many countries (see Section 2). This section examines evidence of this puzzle for the UK SPA. As liquidity constraints are the only standard complete information mechanism for explaining SPA sensitivity (see Section 3), I focus on whether these constraints alone can account for employment's sensitivity to the SPA.

Figure 2 illustrates the excess employment sensitivity puzzle, showing the mean hazard rate of exiting employment by years from SPA. A sharp rise in exits at SPA is evident. While this is a correlation, the female SPA reform provides variation for more precise identification of the SPA's effect.

[8] Loans using future pension benefits as collateral are not illegal but are not observed in practice.

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

Figure 2: Fraction exiting labor employment



Note: Hazard of exiting employment at ages relative to SPA. Data plotted at two yearly intervals to match ELSA's frequency.

To do this, I use a difference-in-difference approach, common in studies of employment responses to pension eligibility (e.g. Mastrobuoni, 2009; Staubli and Zweimüller, 2013; Behaghel and Blau, 2012; Cribb et al., 2016). The outcome variable is the hazard of exiting employment, which captures key transitions driving employment changes and accounts for shifts in overall employment levels, unlike employment drops. The main equation is:

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

This is a regression of the hazard of exiting employment (y_{it}) on an indicator of being above the SPA ($age_{it} > SPA_{it}$); a set of quarterly cohort, age, and date dummies; and a vector of controls (X_{it})[10]. The hazard (y_{it}) is an indicator that is defined if the individual was employed last period, it takes a value of one if they are no longer employed and zero otherwise.

This form assumes cohort-and-date-constant age effects, age-and-date-constant cohort effects, and cohort-and-date-constant age effects. Given these assumptions, which are just a rephrasing of the parallel trends assumption, the parameter α is a difference-in-difference estimator of the treatment of being below the SPA. I test this parallel trends assumption by interacting with the fixed effects, and the Wald test fails to reject the null that these interactions are zero (p = 0.5377). This treatment is administered to all, but the reform induces variation in the duration of treatment.

Column 1 of Table 1 presents the results of estimating equation 1. I find a 0.129 increase in the hazard of exiting work from being above the SPA significant at the 0.1% level. To investigate if liquidity constraints explain the treatment effect, I restrict the sample to women from households with above-median assets.

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

Table 1: Effect of SPA on Hazard of Exiting Employment

	(1)	(2)	(3)	(4)
Above SPA	**0.128**	**0.106**	**0.156**	**0.145**
s.e	(0.0239)	(0.0299)	(0.0371)	(0.0242)
p=	.000	.001	.000	.000
Above SPA×(NHNBW.>Med.)			**-0.050**	
s.e			(0.0476)	
p=			.299	
Above SPA× NHNBW				**-1.17×10^{-7}**
s.e				($2.67e\times10^{-8}$)
p=				.000
Obs.	7,906	3,798	7,906	7,906

Notes: Column (1) presents results from the specification in equation 1. Column (2) repeats the regression for those with above-median Non-Housing Non-Business Wealth (NHNBW) in their last interview before SPA. Column (3) tests if treatment effects differ by fully interacting the specification with having above-median NHNBW. Column (4) adds an interaction between wealth and being above SPA. Controls include marital status, education, self-reported health dummies; partner's age, age squared, qualifications; partner's SPA eligibility; years of education; and household assets. The sample is reduced to 8,119 observations as the hazard indicator is defined only for those employed in the last interview, further reduced to 7,906 due to missing wealth data.

Specifically those with non-housing non-business wealth (NHNBW) exceeding £28,500 in the wave before reaching SPA.[11] This threshold targets a group unlikely to face liquidity constraints affecting retirement choices. Given the SPA was reformed in monthly increments and with equation 1 controlling for quarterly age and cohort fixed effects, an individual's control is someone born in the same year and quarter but a few months younger, hence under SPA. This narrow window strengthens the case against liquidity constraints: women with over £28,500 in NHNBW are unlikely to need to wait 1-3 months for the state pension to stop working. Results in Table 1, column 2, show a treatment effect of 0.106 for this subgroup, similar to the full population and significant at 1%.

Column 3 of Table 1 encapsulates columns 1 and 2 by fully interacting specification (1) with an indicator for the subpopulation in specification (2). The interaction with the treatment dummy is insignificant, showing no significant difference in treatment effects between those with above- and below-median assets. Dichotomizing assets into above and below median loses information, so Column 4 includes an interaction between being below SPA and the continuous NHNBW variable. This interaction is significant but tiny: reducing the treatment effect by 1 percentage point requires an additional £($\frac{0.01}{1.17\times10^{-7}}$) or £85,470 of NHNBW. Unsurprisingly, this indicates that wealth impacts how important the SPA is to someone's retirement decision but that liquidity constraints cannot completely explain the sensitivity of labor market exits to the SPA. For instance, a woman in the 95th percentile (£409,000 NHNBW) would see only a 4.7 percentage point decrease in her SPA response, which remains significantly positive. Thus, while wealth matters,

[11] NHNBW excludes primary residence and personal business assets, per Carroll and Samwick (1996).

Table 2: Placebo Tests

	One Year Below SPA	Two Years Below SPA
Placebo Test Coefficient	**0.031**	**0.005**
s.e	(0.0256)	(0.0230)
p=	.239	.831
Obs.	4,279	4,279

Notes: Placebo test: observations over SPA dropped and treatment indicator replaced with indicator per column heading.

liquidity constraints alone cannot explain the effect of SPA on employment.

Table 1 captures the excess sensitivity puzzle in various ways, but a simple summary to test the model against is also needed. Column 4 provides finer-grained heterogeneity than Column 3, which mostly consolidates columns 1 and 2. However, columns 1 and 2 more clearly epitomize the puzzle with two key findings: one, a significant employment response which is, two, constant across a median asset split. Thus, I summarize the excess sensitivity puzzle using the results in columns (1) and (2) and use these as auxiliary models for the structural model to replicate.

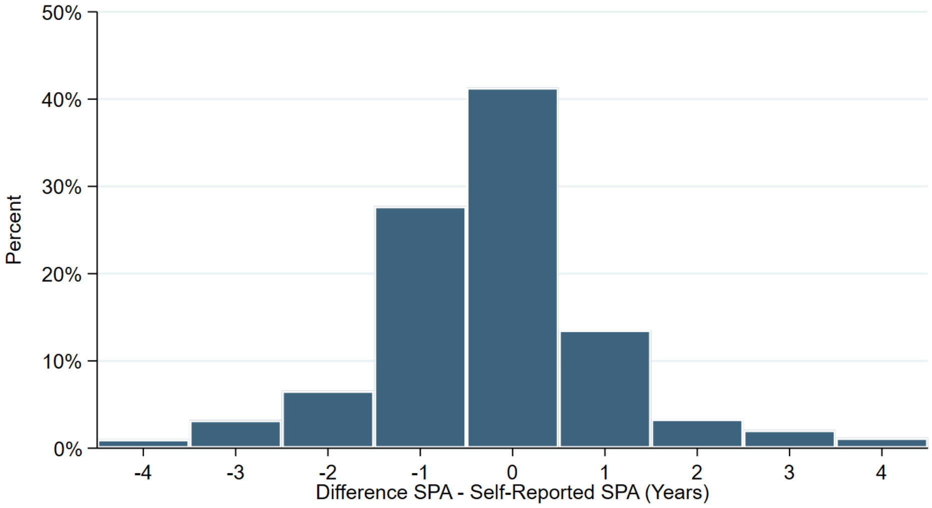
Online Appendix A provides robustness checks, including restricting to more liquid asset categories and alternative functional forms, such as dropping controls to address bad control concerns. These confirm that while assets influence the labor supply response to SPA, the effect is too weak for liquidity constraints to fully explain it. The online appendix also examines whether factors like health, private pensions, or joint retirement explain the excess sensitivity and finds they do not, as SPA does not significantly correlate with changes in these factors. Using self-declared reasons for employment termination, it also contains evidence against illegal firm-mandated retirement as a driver of the result. The traditional difference-in-difference approach here assumes homogeneous treatment effects. Online Appendix A relaxes this using the modern imputation method of Borusyak et al. (2024). Allowing for arbitrary heterogeneity produces average treatment effects estimates consistent with this sections' estimates.

The following analysis does not rely on the causal nature of these estimates but uses them as an untargeted auxiliary model for a structural model. The key is the model's ability to replicate these results, not their causality. However, the analysis assumes readers find these results puzzling under standard complete information models. Placebo test results in Table 2 reinforce the SPA puzzle. Dropping observations over SPA and replacing the treatment in equation 1 with indicators for one or two years below SPA yields negative, insignificant coefficients. Thus, this section identifies something specific to the SPA, puzzling for those with substantial liquid wealth.

5.2 Mistaken Beliefs and Employment Sensitivity

Mistaken beliefs about pensions are common, though surprising under frictionless information, for surely this is a topic the individual is incentivized to know about. This section documents such beliefs about the

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



Notes: Plot of errors in self-reported SPA, showing the frequency of mistaken answers binned yearly.

SPA and their link to the excess employment sensitivity.

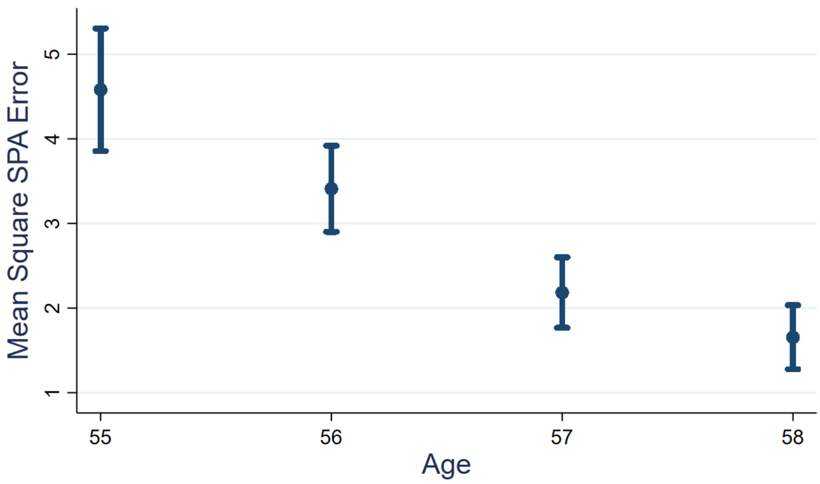
The SPA being such a simple facet of the benefit system, is both puzzling and easy to demonstrate. It is an exact function of date of birth, recorded in ELSA. From wave 3, respondents under SPA are asked their SPA, and discrepancies reveal imperfect knowledge. Figure 3 shows the difference between true and reported SPA for women subject to the reform at age 58 or the closest possible age if not interviewed then. Ideally, we would observe these mistaken beliefs upon receiving their SPA information letter, but its receipt date is unknown, only expected within four months before SPA. Thus, I focus on age 58, when no one has yet received SPA communication. The largest group knows their SPA within a year, though this contains many mistakes by a margin of months. However, 58.7% are off by a year or more, highlighting the prevalence of pension belief errors in the UK. Online Appendix A shows self-reports cluster around each cohort's true SPA, consistent with a costly attention model.[12]

Mistaken beliefs are not only prevalent but also show traits consistent with costly information, such as learning. Learning over time is likely with costly information acquisition as knowledge is retained, and the value of knowing your SPA rises with age. Figure 4 supports this, showing a decline in mean squared errors of self-reported SPAs as women age toward their SPA. The model uses this declining error as a moment to identify attention costs.

The relationship of mistaken beliefs to the employment response to the SPA remains to be seen. A model of endogenous pension knowledge, like this paper's, makes two predictions about this relationship. First, overestimating the SPA causes a larger positive wealth shock upon learning its true value, leading to a larger employment response compared to underestimators. Second, as SPA knowledge is endogenous, selection into knowing your SPA implies those most mistaken show the smallest employment response, as

[12]The online appendix also details self-report errors at their natural monthly frequency.

Figure 4: Mean Squared Error in Self-reported SPA



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

Table 3: Treatment Effect of SPA on Hazard: Heterogeneity by Learning

Weakly Overestimates SPA	**0.168**
s.e	(0.0357)
p=	.000
Underestimates SPA	**0.123**
s.e	(0.0795)
p=	.127
Obs.	5,155

Notes: Results of running specification 1 with added interaction between an indicator of over or under estimating SPA and the treatment of being above SPA. The smaller sample size, compared to Table 1, arises because the SPA knowledge question began in wave 3 and was asked only to those under 60.

many choose not to learn it.

Table 3 offers supportive evidence of the first predictions. It shows treatment effect heterogeneity according to whether individuals over- or under-predict their SPA at 58 or the closest age observed. Beliefs data was only collected from wave three onwards and beliefs questions only asked to those under SPA. Therefore, only women who were under SPA in one of those waves can be used to investigate the heterogeneity of response with respect to mistaken beliefs. This substantially reduces the sample size available. Despite these data limitations, the point estimate of the response to the SPA is larger amongst those who overestimate their SPA and is insignificant amongst those who underestimate their SPA.

Table 4 supports the second prediction. This is shown by fully interacting equation 1 with the absolute error in self-reported SPA at age 58 or the nearest age observed. The significant negative interaction suggests that for each additional year of error in SPA self-reporting, the labor supply response drops by 5.2 percentage points. So, those least informed about the SPA before age 60 have the smallest labor supply response upon reaching SPA after 60. This aligns with a model of endogenous costly information acquisition: individuals who care less about the SPA acquire less information and show smaller responses. In

Table 4: Heterogeneity by SPA Knowledge

Above SPA		**0.202**
	s.e	(0.0385)
	p=	.000
Above SPA×(abs. Error in SPA report)		**-0.052**
	s.e	(0.0246)
	p=	.039
Obs.		5,155

Notes: Results of running specification 1 with added interaction between absolute SPA self-report error and the indicator for being above SPA. The smaller sample size, compared to Table 1, arises because the SPA knowledge question began in wave 3 and was asked only to those under 60.

a model with exogenous information acquisition, this selection mechanism wouldn't exist; those least informed would be so by bad luck. Misbeliefs from bad luck, unlike from choice, typically lead to larger labor supply responses as they receive a larger shock when SPA policy uncertainty resolves. This negative relationship highlights the importance of endogenous learning in the model in Section 6.

The excess employment sensitivity puzzle challenges standard complete information models but can be explained by deviating from their assumptions. Recent work (e.g., Seibold (2021), Lalive et al. (2023)) addresses this puzzle by introducing behavioral biases. However, as complete information models, these fail to account for misbeliefs or their correlation with labor supply responses to SPA, shown in Table 4.

The mechanism in this paper accounts for these misbeliefs, and although the occurrence of the reform is used for identifying variation, the mechanism only relies on the potential for a reform to exist. I cannot causally identify the employment response to the SPA for men but the similar employment and misbelief patterns documented in online Appendix A, despite not having experienced a reform, provide some evidence for the external validity of the mechanism.

6 Model

This section presents the model: Section 6.1 a baseline standard complete information model, capturing the relevant features of the UK retirement context, and Section 6.2 introduces two additions: objective uncertainty about government pension policy and costly information acquisition about this uncertain policy.

6.1 Complete Information Baseline

Key features are summarized before diving into details. The model's decision-making unit is a household containing a couple or a single woman, but when a husband is present, their labor supply is inelastic. The household maximizes life-time utility from consumption, leisure, and bequests by choosing consumption, labor supply, and savings. Households face risk over i) whether they get an employment offer, ii) the wage associated with any offer, and iii) mortality. The households receive non-labor 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. Period are indexed by the age of the female t. Each period, households choose how much to consume c_t, how much to invest in a risk-free asset a_t with return r, and, if not involuntarily unemployed, how much of the women's time endowment (normalized to 1) to devote to wage labor $1-l_t$ (full-time, part-time or none at all) at a wage offer w_t that evolves stochastically. Unemployment ue_t, where $ue_t = 0$ indicates employment (presence of a wage offer) and $ue_t = 1$ unemployment (the absence), also evolves stochastically. The partner's labor supply is inelastic, and so his behavior 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 pensions, $S^{(k)}(.)$ the state pension and $P^{(k)}(.)$ the private pension, are treated as type-specific functions of average lifetime earning $AIME_t$ ($AIME_{t+1} = \frac{(1-l_{t+1})w_{t+1}+AIME_t t}{t+1}$) [13] . From age 60, the women face a probability s_t^k of surviving the period. Finally, households value bequests through a warm glow bequest function (De Nardi, 2004). The full vector of model state is $X_t = (a_t, w_t, AIME_t, ue_t, t)$.

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

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

where θ determines the intensity of the bequest motive, and K determines the curvature of the bequest function and hence the extent to which bequests are luxury goods. The functional form surrounding a_t+K is the utility from consumption of a household (see below), so it approximately captures the utility a descendant gains from these assets, and hence altruism as a motive, whilst keeping parameters to a minimum.

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

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

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

Initial and terminal conditions. ELSA starts interviewing people at 50 but the model starts with women aged 55 because this is the youngest age with significant numbers of SPA self-reports and variation in the true SPA, thus allowing me to initialize the state variables from the empirical distributions for different SPA-cohorts. When age 100 is reached in the model, the woman dies with certainty.

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

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

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

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

The unemployment status of the woman ue_t evolves according to a type-specific conditional Markov process. From age 80, the woman can no longer choose to work; this is to model some of the limitations imposed by declining health. As spousal income results from the confluence of wages, mortality and pension income, it follows a flexible polynomial in age:

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

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

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

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

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

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

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

These pension functions abstract away from the details of state and private pension systems but capture some of the key incentives in a tractable form. The state pension is a complex path-dependent function dependent on past and current regulations (see Bozio et al., 2010). This functional form captures the dependence of the state pension on working history without getting into these difficulties. Being type-specific allows $S^{(k)}(.)$ to capture indirect influences of education and marital status on the state pension; for example, being a stay-at-home mum counted towards state pension entitlement but only after a reform was enacted. Every private pension scheme is different, but the dependence of $P^{(k)}(.)$ on $AIME_t$ reflects the dependence of most defined benefit schemes on lifetime earnings. This functional form less accurately reflects the structure of defined contribution systems, which are essentially saving accounts, but saving for retirement is captured in the model with the risk-free asset and the models starts after the statutory defined contribution eligibility age beyond which they can be accessed without penalty.

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

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

Household maximization problem. The Bellman equation for a household of type k is:

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

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

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

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

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

6.2 Two Additions: Policy Uncertainty and Costly Attention

This section adds two features to the complete information model. Section 6.2.1 introduces objective policy uncertainty via a stochastic SPA, reflecting SPA variation over the life-cycle caused by pension reform. Section 6.2.2 adds costly attention to the stochastic SPA, in the form of disutility for more precise information. These additions are introduced independently, resulting in three model versions: the baseline from Section 6.1, a version with policy uncertainty and informed households, and the full model with rationally inattentive households. Section 6.2.3 concludes with a discussion of these innovations.

6.2.1 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.

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

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

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

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

6.2.2 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.

Directly observed vs learnable states. To make the exposition of rational inattention to the SPA as clear as possible, I introduce two notational simplifications. I group decisions into a single variable $d_t = (c_t, l_t, a_{t+1})$ and all states other than the SPA into a single state variable $X_t = (a_t, w_t, AIME_t, ue_t, t)$.[14] The stochastic SPA SPA_t is separated because, unlike other state variables, it is not directly observed by the household. Instead, the household must pay a utility cost to receive more precise information about the SPA, as outlined below. The other stochastic state variables, w_t and ue_t, being directly observed can be interpreted as these variables being more salient. I focus on costly attention 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 and hence may help explain why people respond as they do to the SPA.

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

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

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

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

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

Entropy and mutual information. The cost of attention is directly proportional to the mutual information, defined below, between signal and SPA. Mutual information is the expected reduction in uncertainty, as measured by the entropy, about one variable resulting from learning the value of another. Entropy, in this information theoretic sense, is a measure of uncertainty that captures the least space[15] needed to transmit or store the information contained in a random variable.

Definition 6.1 (Entropy/conditional entropy)**.** *The entropy $H(.)$ of $X \sim P_X(x)$ is minus the expectation of the logarithm of $P_X(x)$, $H(X) = E_X[-\log(P_X(x))]$. Conditional entropy is $H(X|Y) = E_Y[H(X|Y = y)]$.*

Definition 6.2 (Mutual Information)**.** *The mutual information between $X \sim P_X(x)$ and $Y \sim P_Y(y)$ is the expected reduction in uncertainty, as measured by entropy, about X from learning Y (equally about Y from learning X) : $I(X,Y) = H(X) - H(X|Y)$.*

Utility. Incorporating information costs, utility takes the form:

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

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

[15]Taking the logarithm base 2 measures entropy in bits, but the base only affects the unit of measure. One application, that may help intuition, is by using these concepts; a computer is able to compress a file.

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

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

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

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

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

One problem hidden in this Bellman equation is the formation of next-period beliefs, which, due to Bayesian updating, depends upon the full distribution of signals. This means that the continuation value is not known until the solution is known; this problem will be taken up in Section 7.

6.2.3 Discussion of Costly Attention to the Stochastic SPA

This section discusses a modelling choice regarding and interpreations of two new features: the attention cost and the signal function choice.

Functional form of attention cost. The information acquisition cost generates the mistaken beliefs found to predict employment responses to the SPA. I choose its form to be proportional to the expected entropy reduction for three reasons.

Firstly, a cost of information acquisition that is directly proportional to mutual information is among the most common in the costly information literature, leading to two important advantages. It is tractable because many useful results are available for this functional form[16], and it follows a convention. Tractability is important in models of costly information which can be too complex to solve, and following a convention has merit because it restricts the degrees of freedom available to fit the data.

Secondly, as argued by Mackowiak et al. (2018), this functional form offers a disciplined behavioral

[16] Indeed, until Miao and Xing (2024) extended results from Steiner et al. (2017) to universally posterior separable function, we only knew how to solve the dynamic rational inattention model with this entropy-based cost of attention.

model by replicating numerous types of empirically supported departures from classical models. It endogenously generates behaviors that look like heuristics, or rules-of-thumb, observed sufficiently often to be christened as biases in the behavioral literature.[17] It would be simpler to treat these simplifying rules-of-thumb, or heuristics, as pre-ordained behavioral rules people blindly follow, but this has two major disadvantages. One, it does not explain why that particular rule of thumb and, two, it ignores the fact that people change the rule of thumb as circumstances change. Hard-coded behavioral biases suppress a central insight of economics: people respond to incentives. Endogenizing observed heuristics with this cost of attention avoids these pitfalls.

Thirdly, a priori reasons to think that a cost of cognition should depend on entropy exist. The information-theoretic concept of entropy was developed to explain how machines process information, and it gives a lower bound on the efficient transmission and storage of information. The computational theory of mind Putnam et al. (1967) holds the human mind is a computer. This is controversial and well outside the scope of this paper, but even its most stringent opponent would agree that the brain performs some tasks like a computer, with information processing as a primary candidate. So, if the brain processes information efficiently, mutual information should enter into the ideal cost of attention function. This is not to say an ideal cost of attention function would be linear in mutual information, and recent works such as Caplin et al. (2022) generalize the traditional entropy penalty in multiple ways. Laboratory evidence (e.g. Dean and Neligh, 2023) indicates that the entropy-based cost of attention omits features of human attention, such as perceptual distance, that other cost functions better capture. Outside of such a controlled setting, however, it is not always clear which departures from the entropy-based costs are most relevant or whether sufficient variation exists to identify their extra parameters. So, as it seems that entropy should be an input into the ideal cost function, my cost function can be considered a first-order approximation over this dimension.

Interpreting the cost of attention. Costly information is modelled abstractly and so open to various interpretations. I suggest two, the first broad and the second more literal.

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

[17] Two examples come from Kõszegi and Matějka (2020) who show this cost of attention generates both mental budgeting (quantity allocated to a category being fixed and composition changing) and naive diversification (composition being fixed and quantity allocated changing) depending on the circumstance. Caplin et al. (2019) show it leads to consideration sets (ignoring many options to focus on a subset).

household also learns about an uncertain actuarial adjustment to deferred claiming.

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

Interpreting the choice of signal. The choice of a signal function to learn about the SPA may be difficult to conceptualize. The SPA is a number we can just look up which seems simpler than choosing a signal function. However, looking it up is a learning strategy encompassed by the choice of a signal function conception, corresponding to choosing a perfectly informative signal function.[18] In reality, carefully reading relevant regulations is not the main way people learn about government policy in general or the state pension in particular. People often learn from other people or news outlets. In both examples, there is a random component, what stories newspapers run and what other people talk about, and a choice component, whether you keep reading or ask follow-up questions. This is analogous to the choice of a signal function in that it is partly a choice and partly stochastic, and so it captures much about the messy real-world learning process.

7 Model Solution

By introducing a high dimensional state $\underline{\pi_t}$ (beliefs) and a high dimensional choice $\underline{f_t}$ (signal), rational inattention has complicated the model to the extent that solving it represents a contribution. To achieve this, I combine theoretical results into a general-purpose solution method for dynamic rational inattention models with history-dependent beliefs, such as the one presented above.

The solution method can be considered general purposes because, one, it stores the belief distribution non-parametrically, and two, it does not rely on any specifics of the data-generating process. The only substantive restriction it imposes on the class of dynamic rational inattention model with an entropy-based cost of attention is that the problems must be discrete choice. Since any computational method requires some degree of discretization, discretizing a problem can be seen as a computational approximation. Due to this restriction, I discretize the assets and labor supply choices. Section 7.1 explains the general-purpose method. Section 7.2 explain details specific to solving the model of this paper.

7.1 Solving Dynamic Costly Attention Models with History-dependent Beliefs

Dynamic rational inattention models with history-dependent beliefs are complicated by the presence of a high dimensional state $\underline{\pi_t}$ (beliefs distribution) and a high dimensional choice $\underline{f_t}$ (signal distribution). This

[18]More precisely, a perfectly informative signal includes looking up, remembering, and assimilating information into choices.

section presents a solution method. I use the model of retirement decision from this paper to explain the method, but it applies to any dynamic rational inattention models with history-dependent beliefs.

To solve the household's problem in periods in which rational inattention is relevant, I leverage results from two theoretical papers. Most centrally, I rely on Steiner et al. (2017) who extend the static logit-like results for $\underline{f_t}$ from Matějka and McKay (2015) to a dynamic setting, showing dynamic rational inattention problems reduced to a collection of static problems. As such, it gives me analytic results that greatly simplify solving for the high dimensional choice $\underline{f_t}$. Brute force solution of the optimality conditions from Steiner et al. (2017) would provide a solution, but the high dimensional state $\underline{\pi_t}$ means doing so for all points in the state space is computationally infeasible. Combining results from Caplin et al. (2019) makes finding a solution feasible. They show rational inattention generically implies consideration sets. That is, there are many actions that the household will ignore and never take. The key to my solution method is to note that there are simple conditions indicating which action will not be in the consideration set. These conditions can be used to filter choices before attempting to solve the within-period problem. This filtering step always reduces the computational burden by reducing the dimensionality of the problem. Since households never choose more points of support for their actions than the random variable they are learning about has, this dimensionality reduction is largest when, as is the case here, there are many more actions than possible states of learnable uncertainty. Moreover, at some points, filtering leaves only one candidate action, meaning no additional computation is required to find the distribution of actions taken. At these points, the household will choose the only action in their consideration set with probability one, and the within-period optimization problem will have been effectively solved without further calculation. In my application a single action is left in over 50% of case. This reduction in the dimensionality of the within-period problems and the frequency with which they must be solved makes finding the solution feasible with finite computational resources. Hence, rather than suppressing the belief distribution as a state, as has been done by other papers, this solution method makes having beliefs as a state feasible by reducing the computational burden at each point in the state space.

The rest of this section precedes as follows. Section 7.1.1 outlines key results from Steiner et al. (2017). Then, Section 7.1.2 takes these results and presents the solution method.

7.1.1 Analytic Foundations of Solution Method

Steiner et al. (2017) show that a wide class of similar models have a logit-like solution. The key results needed from their paper for the solution method, as they apply to my model, are the following:

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

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

where expectations are over X_{t+1} and SPA_{t+1} and Section 7.1.2 describes finding $\underline{\pi_{t+1}}$, the Bellman equation 13 becomes:

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

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

$$p_t(d|spa) = \frac{\exp\left(n^{(k)}\frac{((c/n^{(k)})^{\nu}l^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \log(q_t(d)) + \beta\overline{V}_{t+1}^{(k)}(d, X_t, SPA_t, \underline{\pi_t}))\right)}{\sum_{d'\in\mathscr{C}}\exp\left(n^{(k)}\frac{((c'/n^{(k)})^{\nu}l'^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \log(q_t(d')) + \beta\overline{V}_{t+1}^{(k)}(d', X_t, SPA_t, \underline{\pi_t}))\right)}, \tag{15}$$

$$\max_{\underline{q_t}} \sum_{spa} \pi_t(spa) \log\left(\sum_{d\in\mathscr{C}} q_t(d) \exp\left(n^{(k)}\frac{((c/n^{(k)})^{\nu}l^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \beta\overline{V}_{t+1}^{(k)}(d, X_t, SPA_t, \underline{\pi_t})\right)\right). \tag{16}$$

Moreover, they show the optimal information acquisition strategy is to receive an action recommendation, which results in a one-to-one mapping from signals to actions. Using this mapping, we can substitute actions for signals and the conditional choice probabilities $\underline{p_t}(d_t|SPA_t)$ for the signal function $\underline{f_t}$ throughout the problem. Thus, nothing is lost 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 $d_t \sim \underline{p_t}(d_t|SPA_t)$.

7.1.2 Solution Method

At its core, the solution method is to solve equation 16 for $\underline{q_t}$ and substitute the solution into 15 to get $\underline{p_t}$. This basic description corresponds to a brute-force version of my solution method and conceals two major hurdles, which I explain below, culminating in a description of the actual algorithm.

The first hurdle is that knowing which belief next period will result from an action this period requires knowing the full probability distribution of actions. This follows because we do not know how strong a signal an action is of a given SPA unless we know how likely households were to take that action given other possible SPAs. It follows that the conditional effective continuation value $\overline{V}_{t+1}$ is not known, even though next period's value function V_{t+1} is known by backwards induction, because we do not know the beliefs tomorrow that will result from an action today. To see this, substitute the distributions of actions for the distribution of signals in the Bayesian updating formula 12 and apply the results from equations 15 and 16 to get:

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

Then the prior at the start of next period $\underline{\pi_{t+1}}$ is formed by applying the law of motion of SPA_t, equation 11, to this posterior. Since the posterior depends not only on the exponentiated payoff but also on the $\underline{q_t}$, we need a solution ($\underline{q_t}$) to know next period's beliefs given choices and hence to know the effective conditional continuation values (equation 14).

Steiner et al. (2017) evade this difficulty by removing the beliefs from the state space and replacing them with the full history of actions. They can do this because, given initial beliefs, the full history of signals, or equivalently actions, perfectly predicts the beliefs in period t. This is an inspired step in their proof that extends Matějka and McKay (2015) to the dynamic case as it allows them to show we can ignore the dependence of continuation values on beliefs. 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. [19] Redundant information in the state space is problematic because the curse of dimensionality means this is often the binding constraint to producing rich models. What moves this from problematic to a non-starter for many applications is that this redundant information grows exponentially with the number of periods.

Hence, I rely on the theoretical results of Steiner et al. (2017) that used the history of action state-space representation, but in practice, I use the more compact belief state-space representation for the actual computational work. To get around the issue that I need $\underline{q_t}$ to know $\overline{V}_{t+1}$, I use a simple guess-and-verify fixed-point strategy. First, I guess a value $\underline{\tilde{q}_t}$ and solve the fixed point iteration for the effective conditional continuation value defined by substituting 22 into 23. Then given $\overline{V}_{t+1}$ I solve 16 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. [20]

By increasing the computation required at each state, this solution to the first hurdle, however, exacerbates the second, the high computational demands resulting from the high dimensional state $\underline{\pi_t}$. Previously, models of dynamic rational inattention have generally avoided this problem by suppressing the belief distribution as a state variable (Miao and Xing, 2024; Armenter et al., 2024; Turen, 2023; Macaulay, 2021; Porcher, 2020).[21] Although potentially reasonable in specific applications, suppressing beliefs prevents dynamic rational inattention from modelling situations in which beliefs matter and vary across individuals, as, for example, is the case for pension beliefs in the UK. Hence, suppressing beliefs as a state variable limits the domain of the applicability of rational inattention.

I developed a general-purpose solution method for dynamic rational inattention models with history-dependent beliefs. I do this using 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, households take various actions with zero probability. I propose two criteria that identify actions

[19] In Steiner et al. (2017), past actions can affect both beliefs and current utility, making two histories with the same belief different states. Hence, two histories leading to the same belief might represent different states. This is not the case here.

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

[21] Sometimes this is justified as explicit information sharing assumption in the model. More often, it is justified by noting that local posterior invariance (Caplin et al., 2022) will extend to global posterior invariance if all actions are taken with positive probability. However, Caplin et al. (2019) show that solutions are rarely strictly interior as rational inattention generically implies consideration sets. Hence, the extension of local posterior invariance to a global property is highly restrictive.

that will be taken with zero probability without solving the optimization problem. I then remove these from the decision problem. This filtering step reduces the dimensionality of the optimization in equation 16. Moreover, if we are left with a single action after removing these actions, then we have solved the problem without further calculation. I suggest three points in the procedure where these two criteria to filter out actions that will not be taken should be checked.

The first and simplest criterion for culling actions is removing strictly dominated alternatives. The agent is rationally inattentive and so will never select an action strictly dominated in all possible realizations of the SPA. Hence, all actions strictly dominated across all realizations of SPA_t can be removed.

Checking this first criterion is helpful at two points in the procedure. Firstly, before making an initial guess for $\underline{\tilde{q}_t}$, by removing any actions strictly dominated across all possible *joint* realizations of SPA_t and $\underline{\pi_{t+1}}$. Doing this before entering the loop that solves for $\overline{V}_{t+1}$ reduces unnecessary computational burden in that fixed point iteration. However, it imposes a much stricter condition, dominant across all joint realizations SPA_t and $\underline{\pi_{t+1}}$, than needed to drop an action, dominant across all realizations SPA_t. Therefore, having made an initial guess for $\underline{\tilde{q}_t}$, and so having prediction for next period beliefs $\underline{\pi_{t+1}}$ given any action and hence the conditional continuation value, I secondly remove actions strictly dominated across all realizations of SPA_t. I do this for each belief during each iteration of the loop that solves for $\overline{V}_{t+1}$.

For my model, the dimension reduction achieved from dropping strictly dominated actions is large, frequently two orders of magnitude. Before removing from a household's choice set saving levels that are unachievable due to borrowing constraints, the household faces 1,500 choices, 500 saving levels, and 3 labor supply choices. A household will never choose more actions with positive probability than the random variable they are learning about (SPA_t) has points of support, and SPA_t has two points of support at the age of 65 increasing to 8 at age 59. Filtering often reduces the initial choice set in the high hundreds to single digits or low double digits. The runtime required to perform a single filtering is negligible compared to the runtime required to solve equation 16.

Removing strictly dominated actions only uses ordinal information; the second criterion used to filter actions also uses the cardinal information encoded in expected utility. It exploits the necessary and sufficient condition from Caplin et al. (2019). using these, it is easily shown (see online Appendix B.2) that if a there exists a decision $d^\star = (c^\star, l^\star)$ which satisfies:

$$\sum_{spa} \pi_t(spa) \frac{\exp\left(n^{(k)} \frac{((c/n^{(k)})^\nu l^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \beta \overline{V}^{(k)}_{t+1}(d, X_t, spa, \underline{\pi_t}))\right)}{\exp\left(n^{(k)} \frac{((c^\star/n^{(k)})^\nu l^{\star 1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \beta \overline{V}^{(k)}_{t+1}(d^\star, X_t, spa, \underline{\pi_t}))\right)} < 1, \tag{17}$$

for all other decisions $d = (c, l)$ then it is the only action taken $q(d^\star) = 1$. Unlike dropping strictly dominated alternative, which reduces the dimensionality, making solving equation 16 easier, checking equation 17 is only advantageous when the optimal behavior is to take the same action in all realizations of SPA_t. As such, the benefit of checking condition 17 depends on the problem faced and how frequently it reveals the optimal solution without needing to solve an optimization. When filtering does not leave a single action, I employ sequential quadratic programming to solve equation 16, an algorithmic choice suggested by suggested by Armenter et al. (2024). High-level pseudo code summarizing the algorithm is in online Appendix C.

Online Appendix C detail two other computational difficulties. Firstly, the large state space also massively increases storage requirements for the solutions. With this issue, the sparsity proved by Caplin et al. (2019) is again helpful as I can use sparse matrix storage techniques. Secondly, when λ is small, the log-sum-exp of the objective function can lead to underflow problems.

7.2 Details Specific to this Model

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

In the version with rationally inattentive households, we proceed by backward induction from terminal age $t = 100$ using standard techniques for the within-period problem in the model with rationally inattentive households until age $t = 66$. We can proceed back as far as age $t = 67$ because the SPA_t is bounded above by 67, and the woman receives her state pension with certainty from this age. Standard methods can also solve the period $t = 66$ because, at this age, the household is perfectly informed. Either she has reached her SPA and policy uncertainty has been resolved, or she infers $SPA_t = 67$ with certainty because she knows the data-generating process. In this period, $\underline{\pi_t}$ is not a state variable, but SPA_t is because receipt of the state pension affects the budget constraint.

At all earlier ages t <66, if $SPA_t \leq t$, then uncertainty has been resolved, meaning the model can be solved using standard techniques. Moreover, when $SPA_t \leq t$, the exact value of SPA_t is irrelevant. All that matters to the household is they are in receipt of the benefit so that we can solve for a single representative $SPA_t \leq t$. Conversely, when the SPA is in the future $SPA_t > t$, the agent cannot infer the true value of the SPA, and so both the agent's beliefs $\underline{\pi_t}$ and the true value of the SPA SPA_t are states and the agents needs to choose a learning strategy ($\underline{f_t}$). Each year we proceed backwards, the list of future potential SPAs $SPA_t > t$ grows by one, increasing the combinations of $\underline{\pi_t}$ and SPA_t for which we need to solve a problem with uniformed learning agents that is not solvable by simple search techniques. As $\underline{\pi_t}$ is a distribution over all future SPAs ($SPA_t > t$), its points of support also grow by one with each step in the backward induction.

For example, at age $t = 65$, there are two potential future SPA (66 and 67), and if SPA_t takes on either of these values, the agent can no longer infer its true value, and so beliefs ($\underline{\pi_t}$) become a state and the choice of signal function relevant. This growth of the problem complexity along two related dimensions, rational-inattention-relevant potential future SPAs and the size of the belief distribution over them, continues until we reach $t = 59$. At this point, all SPAs 60-67 are future, and rational inattention is relevant regardless of the value of the SPA_t and the support of $\underline{\pi_t}$ is fixed.

8 Estimation

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

8.1 First Stage

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

Initial conditions. To set the initial conditions of the model I need values for $a_t, w_t, AIME_t, ue_t$. Initial wages w_t are set to a draw from the estimated initial wage distribution (see below) and all agents start as employed ($ue_t = 1$). Assets a_t and initial average earning $AIME_t$ are initialized from the type-specific empirical joint distribution. For assets, the empirical counterpart used is household non-housing non-business wealth. Wave 5 of ELSA was linked to administrative data from the UK tax authority allowing me to observe the full working histories of these individuals and so construct a measure of $AIME_t$, but, as this happened for wave five and only 80% consented, this is only true for a subsample of individuals. To avoid dropping data, and to enable the model to match initial period assets, I impute $AIME_t$ with a quintic in wealth and a rich set of observed characteristics. Imputation details can be found in **??**.

Wage equation. I assume wage data is contaminated with serially uncorrelated measurement error ($\mu_{j,t}$) leading to the following variant of equation 2 as data generation process:

$$\log(w_{j,t}) = \delta_{k0} + \delta_{k1} t + \delta_{k2} t^2 + \varepsilon_{j,t} + \mu_{j,t} \tag{18}$$

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

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

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

Stochastic State Pension age. I estimate the probability of an increase in the SPA, ρ, on the cumulative changes to the original female SPA of 60 experienced by reform-affected cohorts. That is I select the ρ to minimize the mean error in SPAs given the data generating process is equation 11, getting an estimate of $\rho = 0.102$

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

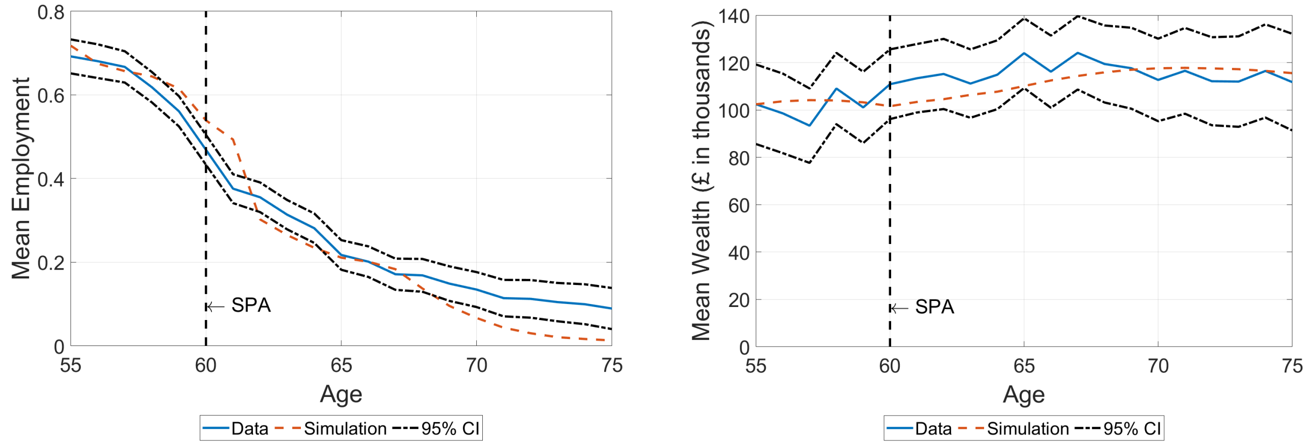
8.2 Second Stage

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

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

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

Figure 5: Fit to Targeted Profiles



(a) Employment Profile (b) Asset Profile

Notes: Model fit to targeted profiles. The empirical profile is for the pre-reform SPA cohort with a SPA of 60.

9 Results

Section 9.1 evaluates model fit and its replication of key facts on excess employment sensitivity, mistaken beliefs, and their relationship. Section 9.2 explores the model's implications, including the distribution of information costs and the impact of information campaigns on fiscal and welfare outcomes of SPA increases

9.1 Model Evaluation

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

The model with policy uncertainty fits pre-reform employment and asset profiles well when simulated with pre-reform $SPA = 60$, as shown in Figures 5a and 5b. Table 5 lists the estimated parameters. The baseline model and the version combining policy uncertainty with rational inattention produce nearly identical fits to these profiles (graphs in online Appendix E.2). However, the three model versions predict distinct responses to SPA changes.

To analyze this response to SPA, I simulate the model with SPAs observed in ELSA waves 1-7 ($SPA = 60$, $SPA = 61$, $SPA = 62$) and repeat the regression from Section 5.1 on the simulated data. I adapt equation 1 to the model's simpler environment, estimating the treatment effect of being above SPA on the hazard of exiting employment using a two-way fixed effects difference-in-difference approach. This regression includes the treatment indicator, full age and cohort fixed effects (excluding date, which aligns with age in the model), and model counterparts to empirical controls (assets, marital status, education). As in Section 5.1, I repeat this on the subsample with above-median empirical assets (£28,500) before SPA. Results are in Table 6's top panel. Column 5 reproduces empirical treatment effects from columns 1 and 2 of Table 1. The baseline model fails to match either the aggregate SPA response or its correlation with wealth.

This baseline's failure reflects the excess employment sensitivity puzzle that prompted investigation of

Table 5: Parameter Estimates

ν: Consumption Weight	0.439
	(0.0000025)
β: Discount Factor	0.985
	(0.0000003)
γ: Relative Risk Aversion	3.291
	(0.0000116)
θ: Warm Glow bequest Weight	100
	(27.228)

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

Table 6: Model Predictions for Different Costs of Attention

	Baseline	Policy Uncert.	Costly Attention $\hat{\lambda} = 6 \times 10^{-8}$	Costly Attention $\lambda = 1.0 \times 10^{-3}$	Data
	Treatment Effect being above SPA on employment				
Whole Population	0.019	0.014	0.041	0.095	0.128
[95% C.I]					[0.081,0.176]
Assets >Median(£28,500)	0.018	0.014	0.054	0.095	0.106
[95% C.I]					[0.047,0.166]
	MSE of SPA Self-Reports at age 58				
MSE at 58	-	-	1.68	5.06	1.67
[95% C.I]					[1.29,2.02]
Coefficient	Treatment Effect Heterogeneity by SPA Error				
Interaction	-	-	-0.047	-0.046	-0.052
[95% C.I]					[-0.102,-0.002]

Notes: The top panel shows labor supply response across the wealth distribution (Table 6). The second panel shows the reduction in self-reported SPA MSE from 55 to 58. The bottom panel highlights heterogeneity in SPA labor supply response by self-reported SPA error at 58. In Columns 2 and 4, treatment effects for the two populations (rows 1 and 2) are identical to three decimal places. To four decimal places, they are 0.0947 (row 1) and 0.0953 (row 2).

policy uncertainty and costly attention. To assess their impacts separately, I introduce them sequentially. Column 2 shows policy uncertainty alone has no effect. This is because objective uncertainty is low (SPA changes are rare). Both this version and the baseline fail to match treatment effects for the whole population and those with above-median assets at SPA but are closer to the lower response of the richer subgroup.[22]

Column 4 presents results with costly attention to the stochastic SPA, for an arbitrary attention cost of $\lambda = 1 \times 10^{-3}$ that fits the treatment effects well. The treatment effects for the whole population and those with above-median assets align more closely with the data. Costly attention closes 71% of the gap for the whole population and 88% for the richer subgroup, with both estimates within the 95% confidence intervals.

Column 4 shows that misbeliefs from costly attention can generate the large employment response to SPA observed in the data. Like other explanations of this puzzle (e.g., Seibold, 2021; Gruber et al., 2022), it introduces a parameter to directly target the employment response. However, unlike competing models,

[22]Section 5.1 highlights the ex-ante puzzling response of the wealthy, and online Appendix C shows that targeting treatment effects directly allows the baseline to match the overall population response but not the wealthy subgroup's. Thus, the I consider the wealthy's response puzzling, though the baseline struggles most with the aggregate with these parameters estimates.

costly attention also explains misbeliefs, another phenomenon inconsistent with the standard models. We can use these additional predictions in combination with the subjective belief data in ELSA to identify λ. Using ELSA's belief data, I identify λ from the reduction in mean squared error in self-reported SPAs between ages 55 and 58 for the cohort with an SPA of 60, as other targeted moments are from this cohort. The middle panel of Table 6 presents these values. Column 3 shows that for a lower $\lambda = 6 \times 10^{-8}$, the model closely matches the reduction in mean squared error but fits the SPA employment response less well, though it still significantly outperforms the baseline and the policy uncertainty model. The introduction of costly attention now closes 23% of the gap for the whole population and 43% for the richer subgroup, with only the richer subgroup estimates falling within the 95% confidence interval. Introducing a norm or preference to retire at SPA could match the employment drop but fails to explain misbeliefs. Column 4 shows that for the higher $\lambda = 1 \times 10^{-3}$, SPA knowledge worsens between ages 55 and 58, suggesting that any learning is offset by drift as agents update based on known laws of motion.

The link between mistaken beliefs earlier in life and employment response later spurred investigation into endogenous information's role in the excess sensitivity puzzle. As shown in Table 4, individuals better informed about their SPA in their late 50s exhibit smaller labor supply responses at SPA in their 60s. A key question is whether the model replicates this relationship. Two opposing forces in the model link SPA knowledge to labor supply response. Endogenous SPA knowledge implies those least dependent on the SPA acquire less information, while worse-informed households, by luck rather than selection, face a larger shock upon learning their SPA, prompting a greater reaction. Which dominates determines whether the model generates a positive or negative relationship. The bottom panel of Table 6 shows a negative relationship for both λ values, indicating the model reproduces the observed direction of this relationship.

9.2 Model Implications and Predictions

Size of informational frictions. λ is hard to interpret, having as natural units of utils per bit. Gabaix (2019) suggests converting attention costs to implied price misperceptions, but this doesn't apply when the object, like SPA, isn't traded. While utils are known to be not interpretable, expressing costs per bit can exaggerate costs, as models contain far fewer learnable bits than reality. Total information in most models typically is generally counted in single or double-digit bits, less than the information in an average sentence. Total information in reality is vastly greater, so per-bit information costs represent a much larger share of total knowledge in models. To address these issues, I express λ as the compensating assets needed to raise household utility as much as perfect SPA knowledge, effectively their willingness to pay to learn their SPA.

Table 7 summarizes the distribution of compensating assets (rounded to the pound) for $\hat{\lambda} = 6 \times 10^{-8}$, estimated from belief data, and the larger $\lambda = 1 \times 10^{-3}$, also shown in Table 6. For $\hat{\lambda} = 6 \times 10^{-8}$, costs range from £6 at the 25th percentile to £14 at the 75th, with a mean of £9. The semi-elasticity of information

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

λ	Mean	Standard Deviation	Median	25th Percentile	75th Percentile	Semi-elasticity (per 10k Assets)
6×10^{-8}	£11.00	£9.00	£9.00	£6.00	£14.00	-1.82%
1×10^{-3}	£83.00	£172.00	£23.00	£10.00	£49.00	-5.26%

Notes: Distribution of compensating assets equivalent to the utility of learning your SPA today, shown for two attention costs.

Table 8: Additional Mean Employment from Increasing SPA from 60

	Only Policy Uncertainty	Policy Uncertainty + Costly Attention	Evaluting Difference as resulting from an Information Campaign (per person)		
SPA Increased to	Additional Employment (55-65)	Additional Employment (55-65)	Marginal Cost	Willingness to pay	Marginal Revenue
61	0.07	0.06	£3.50	£4.22	£28.45
62	0.14	0.14	£4.00	£2.37	£11.78
63	0.18	0.16	£4.50	£18.34	£19.91
64	0.22	0.20	£5.00	£31.64	£4.31
65	0.31	0.27	£5.50	£44.41	£68.52

Notes: Mean additional employment years (55–65) from raising SPA from 60 to the age in the first column, with and without costly attention.

valuations with respect to household wealth shows: information frictions impose the highest costs on the poorest and valuations decrease by 1.82% per £10,000 of additional wealth. For $\lambda = 1 \times 10^{-3}$, the pattern is similar, but valuations are higher.

Increasing old age participation via pension eligibility increase. Rising old-age dependency ratios have made increasing older individuals' labor force participation a global policy priority, with statutory retirement ages seen as a key tool (e.g. Kolsrud et al., 2024). This paper shows that misbeliefs from costly attention amplify employment responses at the SPA, raising the question of whether misinformation makes SPA a more effective tool. In general, the opposite is true.

Column 2 of Table 8 shows the additional mean employment during ages 55–65 when SPA is reformed to 61–65, relative to a preform SPA of 60. These results are generated by the model with $\hat{\lambda} = 6 \times 10^{-8}$, using initial prior beliefs and other state variables from the empirical belief distribution of the cohort with SPA 60. Thus, the results reflect the employment response to an unanticipated SPA increase at age 55, starting from a SPA of 60. Column 1 shows the same for the model with policy uncertainty but without costly attention. The first two columns reveal modest overall employment increases and consistently larger increases in the model with costly attention. For SPA 65, mean employment rises by 0.27 years with costly attention, compared to 0.31 years without informational frictions.

Neglecting costly attention overestimates the employment increase by up to 15%, seemingly conflicting with the claim that costly attention causes a larger employment drop at SPA. This tension resolves when considering that rationally inattentive households respond less immediately to SPA increases. Fully aware households internalize the change early, increasing labor supply in their 50s. Rationally inattentive households respond partially, only realizing near SPA they need to compensate for lost time. Arriving at the old SPA of 60 is often associated with large employment response, as they recognize the increase and work more. This compensatory effort reduces but does not eliminate the difference over 55–59 due to imperfect intertemporal substitution and lower employment at older ages. It also inflates employment just before SPA, amplifying the drop at SPA. Thus, costly attention results in a smaller overall employment increase but a larger response at SPA, as much of the bunching reflects intertemporal employment shifting. Figure 6 illustrates these dynamics for an SPA increase to 62.

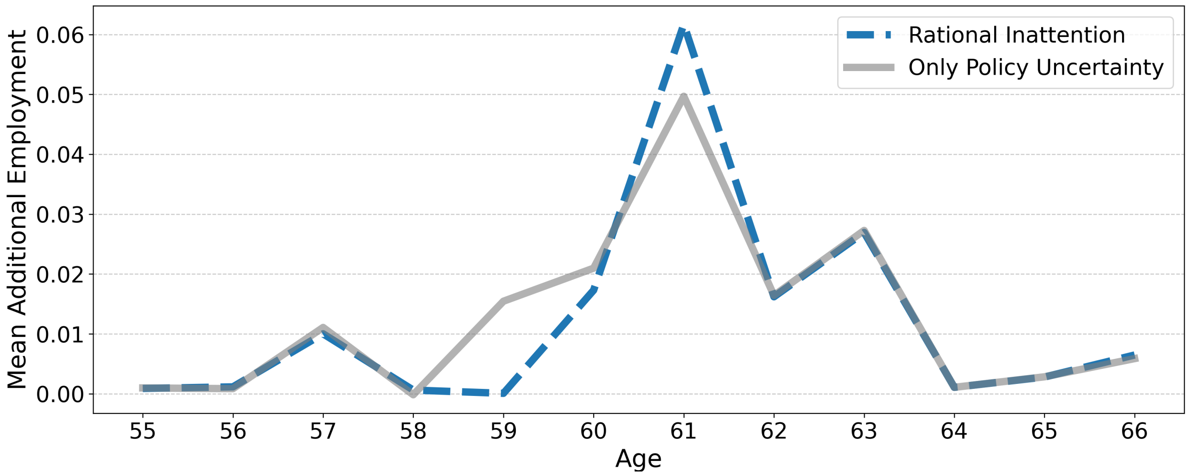
The impact of information on response to pension age reforms. Columns 1 and 2 of Table 8 show the additional employment from an unanticipated SPA increase at age 55 in models with and without costly information. The only difference is that households in column 1 know the SPA, while those in column 2 do not. Thus, the gap between the two can be interpreted as the upper limit of what an information campaign could achieve. The last three columns analyze a policy of yearly information letters that eliminate uncertainty about current pension policy.

Column 3 shows the marginal costs of an information letter campaign. After covering fixed administrative and technological costs, the marginal cost is only the postage, which is £0.50 per year in 2013 prices. Column 4 examines how mean willingness to pay (WTP) to learn the SPA (from Table 7) changes with SPA increases. Two forces affect WTP: higher SPAs reduce lifetime wealth, lowering WTP, and as SPAs move into less likely regions, the instrumental value of knowing the truth increases. When the SPA is near 60, the first effect dominates, reducing WTP from the £11 in Table 7. For SPAs of 63 or higher, the second effect dominates, raising WTP above its value when SPA is 60.

Columns 4 and 5 show that WTP for information exceeds the program's marginal cost for all post-reform SPAs except 62. For these reforms, the program improves net welfare, even without considering government budget effects. Beyond direct costs, the campaign increases employment, as shown in columns 1 and 2, leading to higher revenue. Column 5 quantifies this additional revenue. Though modest, given the studied group (low-earning women born in the 1950s), it exceeds the campaign's marginal cost for all SPAs except 64, ranging from 3 to 12 times the cost. Combining the benefits to household and to government revenue, the campaign always improves net welfare.

Columns 3-5 show the information campaign always increases total welfare, with benefits outweighing costs by 3.5 to 20.5 times. While absolute welfare changes are modest, the experiment highlights an overlooked policy point: Informing individuals not only improves their welfare, it also enhances their

Figure 6: Additional Employment resulting from Increasing the SPA from 60 to 62



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

responsiveness to policy levers. For people to respond, they must know the change occurred.

Other model implications. Online Appendix E provides additional model results, examining the effects of policy uncertainty and costly attention on optimal retirement savings and examining which traits predict information acquisition. The appendix also includes results for a model version with $\hat{\lambda} = 6 \times 10^{-}8$ and a fraction of passive decision-makers, as in Chetty et al. (2014) or Lalive et al. (2023), who retire at SPA regardless. This approach incorporates a norm to stop working at SPA alongside costly attention. Evidence suggests framing effects, norms, and reference-dependent preferences may also influence labor supply responses to SPA (e.g. Seibold, 2021; Gruber et al., 2022). Although these mechanisms don't explain misbeliefs, costly attention may operate alongside them to account for observed retirement patterns. Online Appendix F enriches pension policy uncertainty to be multidimensional, but data limitations make the extension necessarily more speculative.

10 Conclusion

Despite the commonality of mistaken beliefs, their economic impacts are still being understood. This paper contributes by showing that integrating costly attention, which endogenously generates misbeliefs, into a retirement model explains both observed misbeliefs and the puzzling sensitivity of employment to pension eligibility ages, at least for the UK. Costly attention closes 43% of the employment response gap between model and data when the cost of attention is estimated to match observed belief patterns and 88% when not constrained by belief data. Given the prevalence of mistaken pension beliefs and excessive employment responses to pension ages, these results may be relevant to other countries.

Costly attention explains employment sensitivity to SPA because responses depend on both changes in circumstances and beliefs, with SPA arrival triggering a revelation of uncertainty. This amplifies the positive information shock at SPA. Hence, a key model mechanism is that, all else equal, misbelief amplifies the employment response to the SPA. The model simultaneously replicates the negative observed relationship between misbeliefs and SPA responses due to the endogeneity of information. Those who care less about the SPA neither learn nor respond to it.

So, endogenous information acquisition is key to explaining retirement behavior, but it means that this period's information choice affects next period's prior, influencing future decision-making. Thus the prior belief becomes a state variable. This high-dimensional state variable significantly increases computational demands. I propose a method for solving dynamic rational inattention models without suppressing belief as a state variable, which makes the model computationally feasible on a high-powered desktop.

I use data on mistaken beliefs to estimate attention costs, with the mean willingness to pay to learn the SPA (for households with a pre-reform SPA of 60) at £11. Though small, this is far higher than the marginal cost of information letters. Policy experiments show that after most SPA increases, the willingness to pay exceeds the letter campaign's cost. These experiments also show that sending letters boosts employment responses to SPA changes by up to 15%, as individuals must know about a change to respond. Beyond the cost, the campaign raises additional tax revenue, which for most SPA changes exceeds the cost. Considering total benefits to government and households, the campaign always improves welfare, with benefits outweighing costs by 3.5 to 20.5 times.

References

Afrouzi, H., J. P. Flynn, and C. Yang (2024). What can measured beliefs tell us about monetary non-neutrality? Technical report, National Bureau of Economic Research.

Afrouzi, H. and C. Yang (2021). Dynamic rational inattention and the phillips curve.

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 (2024). Geometric methods for finite rational inattention. *Quantitative Economics 15*(1), 115–144.

Bairoliya, N. and K. McKiernan (2023). Revisiting retirement and social security claiming decisions.

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

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

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

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

Borusyak, K., X. Jaravel, and J. Spiess (2024). Revisiting event-study designs: robust and efficient estimation. *Review of Economic Studies*, rdae007.

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

Brown, Z. Y. and J. Jeon (2024). Endogenous information and simplifying insurance choice. *Econometrica 92*(3), 881–911.

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

Caplin, A., M. Dean, and J. Leahy (2019). Rational Inattention, Optimal Consideration Sets, and Stochastic Choice. *Review of Economic Studies 86*(3), 1061–1094.

Caplin, A., M. Dean, and J. Leahy (2022). Rationally inattentive behavior: Characterizing and generalizing shannon entropy. *Journal of Political Economy 130*(6), 1676–1715.

Caplin, A., E. Lee, S. Leth-Petersen, and J. Sæverud (2022). Communicating social security reform. Technical report, National Bureau of Economic Research.

Carroll, C. and A. Samwick (1996). The Nature of Precautionary Wealth. *NBER Working Paper*.

Casanova, M. (2010). Happy Together : A Structural Model of Couples ' Joint Retirement Choices. *Working Paper* (5), 1–53.

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

Chetty, R. (2015). Behavioral economics and public policy: A pragmatic perspective. *American Economic Review 105*(5), 1–33.

Chetty, R., J. N. Friedman, S. Leth-Petersen, T. H. Nielsen, and T. Olsen (2014). Active vs. passive decisions and crowd-out in retirement savings accounts: Evidence from denmark. *The Quarterly Journal of Economics 129*(3), 1141–1219.

Chetty, R., A. Looney, and K. Kroft (2009). Salience and taxation: Theory and evidence. *American economic review 99*(4), 1145–1177.

Ciani, E., A. Delavande, B. Etheridge, and M. Francesconi (2023). Policy uncertainty and information flows: Evidence from pension reform expectations. *The Economic Journal 133*(649), 98–129.

Coibion, O. and Y. Gorodnichenko (2012). What can survey forecasts tell us about information rigidities? *Journal of Political Economy 120*(1), 116–159.

Coibion, O. and Y. Gorodnichenko (2015). Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review 105*(8), 2644–2678.

Cribb, J., C. Emmerson, and G. Tetlow (2016). Signals matter? Large retirement responses to limited financial incentives. *Labour Economics 42*, 203–212.

de Bresser, J. (2023). Evaluating the accuracy of counterfactuals: Heterogeneous survival expectations in a life cycle model. *Review of Economic Studies*, rdad088.

De Nardi, M. (2004). Wealth inequality and intergenerational links. *The Review of Economic Studies 71*(3), 743–768.

De Nardi, M., E. French, and J. B. Jones (2010). Why Do the Elderly Save? The Role of Medical Expenses. *Journal* *of Political Economy 118*(1), 39–75.

Dean, M. and N. Neligh (2023). Experimental tests of rational inattention. *Journal of Political Economy 131*(12), 3415–3461.

Dolls, M., P. Doerrenberg, A. Peichl, and H. Stichnoth (2018). Do retirement savings increase in response to information about retirement and expected pensions? *Journal of Public Economics 158*, 168–179.

Farhi, E. and X. Gabaix (2020). Optimal taxation with behavioral agents. *American Economic Review 110*(1), 298–336.

French, E. (2005). The Effects of Health, Wealth, and Wages on Labour Supply and Retirement Behavior. *Review of Economic Studies 72*(2), 395–427.

French, E. and J. B. Jones (2011). The Effects of Health Insurance and Self-Insurance on Retirement Behavior. *Econometrica 79*(3), 693–732.

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

Goldin, J., T. Homonoff, R. Javaid, and B. Schafer (2022). Tax filing and take-up: Experimental evidence on tax preparation outreach and benefit claiming. *Journal of Public Economics 206*, 104550.

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

Gustman, A. L. and T. L. Steinmeier (1986). A Structural Retirement Model. *Econometrica 54*(3), 555–584.

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*.

Kolsrud, J., C. Landais, D. Reck, and J. Spinnewijn (2024). Retirement consumption and pension design. *American Economic Review 114*(1), 89–133.

Lalive, R., A. Magesan, and S. Staubli (2023). How social security reform affects retirement and pension claiming. *American Economic Journal: Economic Policy 15*(3), 115–150.

Liebman, J. B. and E. F. Luttmer (2015). Would people behave differently if they better understood social security? evidence from a field experiment. *American Economic Journal: Economic Policy 7*(1), 275–299.

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.

Lusardi, A. and O. S. Mitchell (2011). Financial literacy and retirement planning in the united states. *Journal of pension economics & finance 10*(4), 509–525.

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 *.

Mackowiak, B., F. Matejka, and M. Wiederholt (2018). Rational inattention: A disciplined behavioral model. *CEPR Discussion Papers 13243*.

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.

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.

Mastrobuoni, G. (2011). The role of information for retirement behavior: Evidence based on the stepwise introduction of the social security statement. *Journal of Public Economics 95*(7-8), 913–925.

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.

Miao, J. and H. Xing (2024). Dynamic discrete choice under rational inattention. *Economic Theory 77*(3), 597–652.

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

OECD (2011). Net pension replacement rates: Public and private schemes. In *Pensions at a Glance 2011: Retirement-income Systems in OECD and G20 Countries*. OECD Publishing.

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

Putnam, H. et al. (1967). Psychological predicates. *Art, mind, and religion 1*, 37–48.

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.

Staubli, S. and J. Zweimüller (2013). Does raising the early retirement age increase employment of older workers? *Journal of public economics 108*, 17–32.

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

Taubinsky, D. and A. Rees-Jones (2018). Attention variation and welfare: theory and evidence from a tax salience experiment. *The Review of Economic Studies 85*(4), 2462–2496.

Turen, J. (2023). State-dependent attention and pricing decisions. *American Economic Journal: Macroeconomics 15*(2), 161–189.

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.