IFS

Institute for Fiscal Studies

Jamie Hentall-MacCuish

Working paper

25/14

Costly attention and retirement

Economic and Social Research Council

Submitted to *Econometrica*

COSTLY ATTENTION AND RETIREMENT

JAMIE HENTALL-MACCUISH
Department of Economics and Decision Science, HEC Paris

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 generates 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, Retirement, Misbeliefs, Pensions, Behavioral Macro, Structural Econometrics.

1. INTRODUCTION

Understanding the cause of apparent deviations from rationality is crucial for policy design. If they represent fixed features of household behavior, our options to address them are limited, but mistaken beliefs about the policy itself can lead to similar departures from apparent rationality. In such cases, straightforward information provision might mitigate these deviations. This paper shows misbeliefs offer an alternative, or potentially comple-

---

Jamie Hentall-MacCuish: hentall-maccuish@hec.fr

I thank Fabien Postel-Vinay and Eric French for discussions and encouragement. For helpful comments, I thank Richard Blundell, Uta Bolt, Mariacristina De Nardi, Gaetano Gaballo, Tomasz Michalski, Lars Nesheim, Cormac O'Dea, Morten Ravn, Morgane Richard, Victor Rios-Rull, Arthur Seibold, Johannes Spinnewijn, and participants at the NBER SI Behavioral Macro Session, IIPF Annual Congress, RES Junior Symposium, CESifo Public Area Conference, YES, 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. Funding from Grant Inequality and the insurance value of transfers across the life cycle (ES/P001831/1), ESRC studentship (ES/P000592/1), and HEC is gratefully acknowledged.

mentary, answer to a puzzle often attributed to fixed household behavior: the excessively large drop in employment at pension eligibility age, despite weak economic incentives to stop working precisely then.[1] To do this, it develops a solution method for dynamic rational inattention models with history-dependent beliefs and uses it to estimate a model on UK data targeting both observed beliefs and behavior.

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, 59% 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. Together, policy uncertainty and costly information can generate these misbeliefs as an optimal response. Can these endogenously generated misbeliefs, in turn, help explain excess employment sensitivity to pension eligibility?

To investigate, I first document key facts on misbeliefs and excess employment sensitivity, then I separately and sequentially introduce policy uncertainty and information frictions (in the form of costly attention) into a model of retirement. Specifically, I estimate a dynamic lifecycle model of retirement (e.g. Rust and Phelan, 1997, French, 2005) with rationally inattentive households (e.g. Sims, 2003, Matějka and McKay, 2015, Caplin et al., 2019) deciding how much information about a changeable pension policy to acquire whilst incurring a disutility cost of information. The model endogenously generates observed misbeliefs, but can it generate the otherwise puzzling sharp employment drop at pension eligibility age? The drop in employment at pension eligibility age is puzzling as UK pension benefits are not tied to employment, so State Pension Age (SPA) only incentivizes retirement for liquidity-constrained individuals unable to 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

[1] This puzzle is documented in multiple countries as summarised in Gruber and Wise (2004).

[2] Documented, for example, in Gustman and Steinmeier (2005), Lusardi and Mitchell (2011), Ciani et al. (2023).

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 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. These mechanisms exist in a model with only policy uncertainty, but by introducing policy uncertainty and costly attention separately, this paper shows historically observed policy uncertainty is too low to generate meaningful changes. Hence, misbeliefs generated by costly attention are key to amplifying these positive shocks at the SPA.

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

I focus on costly attention to the SPA rather than any other burdens on people's attention for two reasons. One, pension policy uncertainty—unlike, for example, return uncertainty—resolves, or at least diminishes, upon eligibility, potentially explaining employment responses at the SPA. Two, the SPA's simplicity (relative to other sources of pension policy uncertainty like the benefit level) makes mistaken SPA beliefs easy to measure and, hence, study. The simplicity of the SPA makes the misbeliefs we observe all the more surprising.

In the data, 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 misbeliefs, 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 beliefs 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 modeling assumption for specific situations, it limits the domain of application by implying beliefs are irrelevant to choices. 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 challenges of high-dimensional states using the sparsity shown to be a property of rational inattention models by Caplin et al. (2019).

The English Longitudinal Study of Ageing (ELSA), a micro panel survey, provides data to study misbeliefs and their impact on employment. It contains 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.

I estimate the model using two-stage simulated method of moments, targeting asset and employment profiles, and, when present, identifying attention costs from changes in individual misbeliefs over time. Targeting changes in beliefs is possible thanks to my solution method, which, by retaining beliefs as a state variable, endogenously generates belief predictions that can be compared to the data. Thus, my solution method builds a bridge between the dynamic-rational-inattention literature and the subjective-belief-data literature. Policy uncertainty combined with costly attention increases the employment response to the SPA compared to a complete information baseline, explaining 43%-88% of the shortfall. The mean household is willing to pay £11.00-£83.00 to learn today's SPA, so estimated attention costs are low (consistent with other evidence, e.g., Chetty, 2012). Large changes

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

in the employment response at SPA stem from small attention costs because the concentrated response at SPA represents an intertemporal shifting of employment, compared to the frictionless benchmark.

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 generally true. Policy experiments comparing employment increases resulting from SPA changes in versions 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, could raise old-age employment by up to 15%. In most policy experiments, the benefits to households and extra tax revenue from these letters, each separately, outweigh the costs: considered jointly, information letters are always welfare-enhancing.

Related Literature. Dynamic lifecycle models of retirement began with Gustman and Steinmeier (1986) and Burtless (1986). Key features introdoced since then include uncertainty (Rust and Phelan, 1997), borrowing constraints (French, 2005), Medicare (van der Klaauw and Wolpin, 2008), and medical expenses (French and Jones, 2011). Much of this literature is US-focused, and some of its concerns, like medical insurance, are irrelevant to 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 began as a way to add costly attention to macroeconomic models (e.g., Sims, 2003, Maćkowiak and Wiederholt, 2009, 2015), but now touches most fields, e.g., industrial organization (Brown and Jeon, 2024), or labor economics (Bartoš et al., 2016). Matějka and McKay (2015) solve a general class of static discrete choice models with rationally inattentive agents, and 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 questions that can be answered. 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 utility but lacks these performance gains. A closely related static rational inattention paper Boehm (2023) estimates a lifecycle model of older individuals, focusing on the one-shot choice of annuity.

First highlighted in the US by Lumsdaine et al. (1996), a puzzlingly large drop in employment at pension eligibility ages occurs across countries. In the US, the consensus was that liquidity constraints explained the drop at age 62, and Medicare eligibility the drop at age 65 (Rust and Phelan, 1997, French, 2005, French and Jones, 2011). Testing these explanations became possible after 2004 when the full retirement age increased. Part of the age 65 spike followed the full retirement age, despite Medicare eligibility staying at 65 (Behaghel and Blau, 2012), and Mastrobuoni (2009) found larger effects than standard models predicted. Pension age increases around the world produced similar results: larger employment responses than financial incentives implied (summarised in Gruber and Wise, 2004). I document this in the UK, extending Cribb et al. (2016) by using richer data to rule out other potential explanations. Part of the literature has recently converged towards reference-dependence as the explanation of this puzzle (e.g. Seibold, 2021, Lalive et al., 2023, Gruber et al., 2022). I compare my results to this explanation in Section 8 and online Appendix E.

The use of subjective belief data in structural microeconomic models is extensive (Koşar and O'Dea, 2022). Most papers, however, do not model belief formation, limiting counterfactual analysis (e.g. de Bresser, 2023). Modeling belief formation as an optimal response to processing costs (made possible by my solution method) allows me to match model-generated beliefs to data instead of only using beliefs as input. Early studies of pensions beliefs (e.g. Bernheim, 1988, Manski, 2004) document misbeliefs about benefit levels. Caplin et al. (2022b) 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 a parameter from patterns in beliefs (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). Bairoliya and McKiernan (2023) find using misbeliefs as inputs helps explain claiming and retirement patterns in the US, supporting the external validity of this paper's mechanisms.

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

2. BACKGROUND

The UK State Pension system has changed significantly since its 1948 introduction. I discuss the 2000-2016 system, especially 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 contributing years, and a second tier, based on earnings, both calculated over 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 entitlement is frozen a year before SPA, meaning labor supply choices near SPA do not affect the pension amount.

The Basic State Pension began in 1948. By 2013, a full pension paid £107 per week ($203 in 2022 USD). Pro-rata payments apply to those with fewer than 30 contributing years needed for the full pension. Contributing years include those in the labor force (earning above a minimum threshold) and spent caring for a child or disabled person post-1978. So, the timing of and reasons for labor market inactivity 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. It provides a 37% net replacement rate for median earners, compared to 47%, 50%, and 58% in the USA, OECD, and EU, respectively. Second, it is a relatively flat-rate benefit. This is reflected in the larger drop in replacement rate between half and one-and-a-half times median earnings—35 percentage points in the UK, versus 17, 21, and 14 in the USA, OECD, and EU (OECD, 2011).

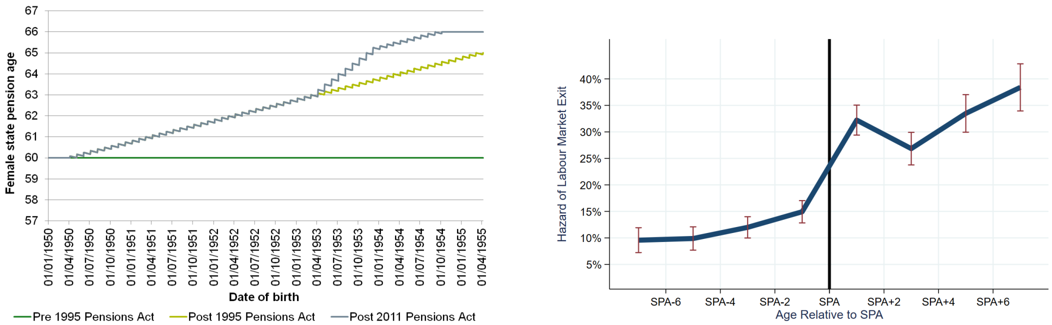
State Pension Age and its reform. The State Pension Age (SPA) is the earliest age the State Pension can be claimed, serving as the UK's early retirement age. Deferring increased benefit generosity, but without a cap on deferral duration, hence implying no effective full retirement age. [4] So, the SPA 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. Figure 1a shows how these changes affected women by birth cohort. These reforms allow estimation of the risk UK women face of SPA changes during their life, a key model input. I also use variation from the 1995 (but to avoid confounding from a benefit level change, not the 2011) reform to identify the SPA's impact on employment,

Communication and lack thereof. The government did not directly inform women affected by the reform, sending only the standard letter received by all pre-reform cohorts shortly before SPA. This lack of communication was controversial. From 2015, two campaign groups claimed the reforms discriminated against older women, with one unsuccessfully seeking to reverse the changes in the High Court. 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

[4] Despite generous actuarial adjustments, deferral was rare, presenting a puzzle. Online Appendix F offers a model extension addressing this. Elsewhere, I abstract from the deferral puzzle taking observed claiming as given.

FIGURE 1.—Pension Legislation and Employment Response to the State Pension Age



(a) SPA by Date of Birth (b) Fraction exiting labor employment

Note: Panel (a) shows State Pension Ages for women under the Pensions Act 1995, the Pensions Act 2007, and the Pensions Act 2011. Panel (b) plots the hazard of exiting employment at ages relative to SPA with data plotted at two yearly intervals to match ELSA's frequency.

this with letter-sending at SPA is beyond this paper's scope, but the absence of protests until 20 years after legislation supports the view reported misbeliefs are genuine.

Private pensions. A large private pension market supplements the State Pension. Since private pension eligibility is not tied to SPA, it has little relevance to the employment response to SPA (more evidence in online Appendix A).

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, banning it as age discrimination.[5] So, 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 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

[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.

SPA.[7] Finally, it is worth restressing that benefit entitlement is frozen the year before SPA, making it unaffected by labor supply choices near SPA.

These facts show the State Pension acts as an anticipatable increase in non-labor income, with the SPA as 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 the concentrated response at SPA despite the long notice period. In a standard life-cycle model with complete information and forward-looking agents, employment does not 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] So, liquidity constraints are the only standard explanation for employment sensitivity at the SPA.

3. DATA

Studying the employment response to the State Pension Age (SPA) requires a large sample of older individuals, and exploring its causes requires rich microdata. 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 misbeliefs. 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 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

---

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

[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.

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 inputs. The earliest affected cohort was born on 6 April 1950. Older cohorts serve as controls and also inform model inputs.

4. KEY MOTIVATING FACTS

4.1. *Excess Employment Sensitivity*

The sensitivity of employment to official retirement ages in excess of incentive is a puzzle observed in many countries (see Section 1). 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 2), I focus on whether these constraints alone can account for employment's sensitivity to the SPA.

Figure 1b 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 policy variation with which to causally estimate the SPA's effect.

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, 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:

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

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 defined if the individual was employed last period, it is 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 just

[10]Controls include marital status, education, self-reported health dummies, partner's age, age squared, qualifications, partner's SPA eligibility and education, and household assets.

rephrase the parallel trends assumption, the parameter α is a difference-in-difference estimator of the treatment of being above the SPA. The treatment is administered to all, but the reform induces variation in the duration of treatment. I test this parallel trends assumption by interacting with the fixed effects, and the Wald test fails to reject the null these interactions are zero (p = 0.5377).

Despite the well-known potential for bias of a staggered difference-in-difference, this simple difference-in-difference is preferred for the main text for ease of interpretation. Additionally, the final goal is to apply the same regression to simulated data as an auxiliary model during ex-post model validation, for which use bias is not an issue. As long as the same biased auxiliary model is used on both observations and simulated data, all that matters is the model's ability to replicate the results. However, online Appendix A addresses the potential for bias allowing for heterogeneous treatment effects with the modern imputation method of Borusyak et al. (2024). Allowing for heterogeneity does not change the conclusion about SPA sensitivity in any important way.

Column 1 of Table I 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 non-housing non-business wealth (NHNBW)[11] in the wave before reaching SPA. The resulting threshold of £28,500 targets a group unlikely to face liquidity constraints affecting retirement choices. As the SPA was reformed in monthly increments and Equation 1 controls for quarterly age and cohort effects, the control group for estimating the treatment effect consists of individuals born in the same quarter but a few months younger, thus still below 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. Column 2 of Table I show a treatment effect of 0.106 for this subgroup, similar to the full population and significant at 1%.

Column 3 of Table I 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

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

TABLE I

EFFECT OF SPA ON HAZARD OF EXITING EMPLOYMENT

	(1)	(2)	(3)	(4)	(5)	(6)
Above SPA	**0.128**	**0.106**	**0.156**	**0.145**	**0.167**	**0.189**
s.e	(0.0239)	(0.0299)	(0.0371)	(0.0242)	(0.0371)	(0.0406)
Above SPA×(NHNBW.>Med.)	—	—	**-0.050**	—	—	—
s.e			(0.0476)			
Above SPA× NHNBW	—	—	—	**-1.17×10^{-7}**	—	—
s.e				$(2.67e\times10^{-8})$		
Above SPA×(SPA $\geq$ Self-report)	—	—	—	—	**-0.078**	—
s.e					(0.0917)	
Above SPA×(abs. Error SPA)	—	—	—	—	—	**-0.049**
s.e						(0.0242)
Obs.	7,906	3,798	7,906	7,906	5,209	5,209

Note: 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. Columns (5) and (6) investigate heterogeneity by beliefs at age 58, (5) introduces an interaction with underestimating the SPA, and (6) with the absolute size of the error. Controls are a full set of marriage, years of education, and self-reported health dummies; partner's age; partner's age squared; partner's qualification and years of education; partner's SPA eligibility; and household assets.

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 extra £85,470 in NHNBW. So, while wealth matters, liquidity constraints do not fully explain the SPA's effect on employment.

Table I captures the excess sensitivity puzzle in various ways, but a simple summary to test the model against is needed. While Column 4 provides finer-grained heterogeneity than Column 3, which consolidates Columns 1 and 2, Columns 1 and 2 more clearly embody the puzzle in two key findings: one, a significant employment response, which is, two, constant across a median asset split. So, I test the model against Columns (1) and (2).

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 the 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. As mentioned, online Appendix A also relaxes the homogeneous treatment effects assumption using the modern imputation method of Borusyak et al. (2024).

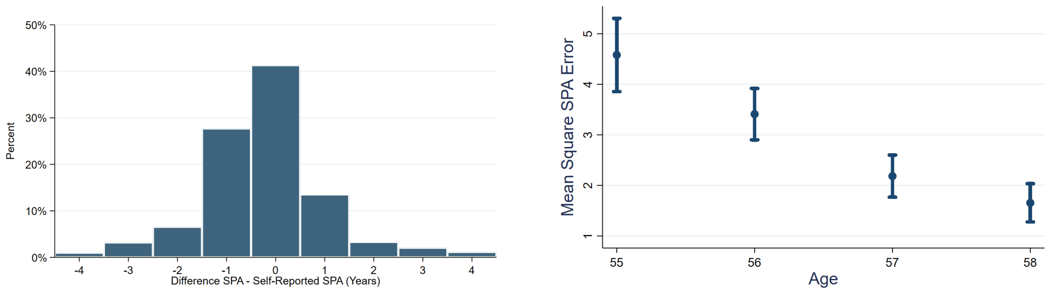
The rest of this paper does not depend on the causal nature of the estimates presented in this section but uses them as an untargeted auxiliary model for a structural model. The key is the model's ability to replicate these results, not their causal nature. However, the analysis assumes readers find these results puzzling under standard complete information models. Placebo tests, in which I drop observations over SPA and replace the treatment in Equation 1 with indicators for being one or two years below SPA, confirm with insignificant treatment effects that something specific is happening at SPA (full results in online Appendix A), This is puzzling for those with substantial liquid wealth.

4.2. *Mistaken Beliefs and Employment Sensitivity*

Compared to other subjective belief data such as inflation or survival expectations, an interesting feature of pension beliefs is that a currently correct answer exists, making misbeliefs potentially observable. Pensions misbeliefs are common, though surprising, under frictionless information, as people have clear incentives to know this information. This section documents such misbeliefs about the SPA and their link to the employment response at SPA.

From wave three, ELSA asks respondents below SPA multiple questions about State Pension beliefs. This section focuses on SPA beliefs, as these are the ones I model, while online Appendix Section A.5 discusses beliefs about benefit levels, reform awareness, and how these relate to SPA beliefs. Despite ELSA's rich subjective belief data, two limitations are worth noting. First, as belief data was only collected from wave three and for those under SPA, only women under SPA in those waves are informative about beliefs, reducing the sample size. Second, ELSA only elicits point estimates for SPA beliefs, which, as De Bruin et al. (2023) notes, pose interpretation challenges. If individuals hold subjective

FIGURE 2.—SPA Beliefs



(a) Mistaken SPA Beliefs Age 58

(b) MSE in Self-reported SPA

Note: Panel (a) plots the frequency of errors in self-reported SPA at age 58 (binned to yearly accuracy). Panel (b) shows mean squared error in Self-reported SPA plotted against respondents' age.

priors, it is not clear which measure of central tendency the answer reflects or if it represents something else entirely. To operationalize the model, in Section 7, I take a stand on interpreting these point estimates, but here I remain agnostic only assuming that responses correlate with people's mean subjective SPA belief.

As the SPA is an exact function of date of birth and gender, both recorded in ELSA, SPA misbeliefs can be inferred by any discrepancy between the stated and true SPA. The fact that the SPA is such a simple facet of the benefit system makes SPA misbeliefs all the more puzzling. Figure 2a evidences the prevalence of pension belief errors in the UK showing the difference between true and reported SPA for reform-affected women at age 58, the last age when no cohort has received an SPA communication, or the closest age interviewed. Although the modal group knows their SPA to be within a year, this includes many mistakes by a margin of months, and the majority (58.7%) are off by a year or more. Online Appendix A shows self-reports cluster around each cohort's true SPA, consistent with a costly attention model.[12]

Misbeliefs 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 2b supports this,

[12]The online appendix also details self-report errors at their natural monthly frequency, and belief heterogeneity by years of education.

showing a decline in mean squared errors of self-reported SPAs as women age toward their SPA. The model uses these declining errors to identify the attention cost.

A model of endogenous SPA knowledge, like this paper's, makes two predictions about the relationship between SPA misbeliefs and the employment response to the SPA. 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 many choose not to learn it.

Column 5 of Table I shows treatment effect heterogeneity according to whether individuals over- or under-predict their SPA at 58 or the closest age observed. The point estimate goes in the predicted direction (larger amongst those who overestimate their SPA) but is not significant, potentially because of the reduced sample size. It is worth emphasizing that although the model certainly predicts a smaller response amongst those who underpredict, it does not necessarily predict no response for two reasons. Firstly, regardless of the direction of error, everyone gets a reduction in uncertainty upon reaching SPA, reducing their precautionary labor supply. Secondly, the difficulty of interpreting a point-estimated belief means people who underreport may still overestimate at the mean of their SPA distribution.

Column 6 of Table I supports the second prediction, showing Equation 1 fully interacted 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 employment response drops by 5.2 percentage points. So, those least informed about the SPA before age 60 have the smallest employment 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 a model with exogenous information acquisition, this selection mechanism would not exist. The size of the SPA error would be orthogonal to individual characteristics, leading to larger employment responses amongst the least informed 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 5.

Recent work (e.g., Seibold (2021), Lalive et al. (2023)) addresses the excess employment sensitivity puzzle by introducing reference-dependent preferences. As a complete information explanation, this does not account for the misbeliefs documented in this section or employment responses to SPA that depend on them (as shown in Table I), while the mechanism in this paper does (Section 8 and online Appendix E offers more comparisons).

I use the occurrence of the reform for identifying variation, but the mechanisms only rely on pension misbeliefs and the potential for reform. Online Appendix A documents similar employment and misbelief patterns for men, who were not subject to a reform, offering non-causal support that this misbelief channel exists in the absence of a reform.

5. MODEL

Section 5.1 presents the baseline standard complete information model. Section 5.2 introduces two additions: objective uncertainty about government pension policy and costly information acquisition about this uncertain policy.

5.1. *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, his labor supply is inelastic. The household maximizes lifetime utility from bequests, leisure, and equivalized consumption 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. Periods 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 $(1 - l_t)$ (40, 20 or 0 hours per week) 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 ($AIME_{t+1} = \frac{(1-l_{t+1})w_{t+1}+AIME_t t}{t+1}$) [13]. From age 60, the women face a probability of surviving the period (s_t^k). 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:

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

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 interviews people from 50 but the model starts with women aged 55 because this is the youngest age with significant numbers of SPA self-reports for multiple SPA-cohorts, thus allowing me to initialize state variables (a_t and $AIME_t$ but later also beliefs) from the empirical distributions for different SPA-cohorts. At age 100, 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:

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

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

The unemployment status of the woman (ue_t) evolves according to a type-specific conditional Markov process. From 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:

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

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 to employment 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 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 of solving the excess sensitivity puzzle.

Lifetime average earning ($AIME_t$) evolves until the woman reaches the age she starts to receive her private pension ($PPA^{(k)}$), 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 pension functions have the following forms

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

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

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 resulting from 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 (after the enactment of a reform). Every private pension scheme is different, but the dependence of $P^{(k)}(.)$ on $AIME_t$ reflects the dependence of most defined benefit schemes on lifetime earnings. This functional form less accurately reflects the structure of defined contribution systems, which are essentially saving accounts, but saving for retirement is captured in the model with the risk-free asset and the models starts after the statutory defined contribution eligibility age beyond which they can be accessed without penalty.

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

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

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

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

subject to the following budget, borrowing, and labor supply constraints:

$$c_t + (1+r)^{-1}a_{t+1} = a_t + w_t(1-l_t) + Y^{(k)}(t, ue_t, AIME_t), \tag{3}$$

$$a_{t+1} \geq 0, \tag{4}$$ & $$ue_t(1-l_t) = 0. \tag{5}$$

5.2. *Two Additions: Policy Uncertainty and Costly Attention*

This section adds two features to the complete information model. Section 5.2.1 introduces objective policy uncertainty via a stochastic SPA, reflecting SPA variation over the lifecycle caused by pension reform. Section 5.2.2 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 5.1, a version with policy uncertainty and informed households, and the full model with rationally inattentive households. Section 5.2.3 concludes with a discussion of these innovations.

5.2.1. *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:

$$SPA_{t+1} = \min(SPA_t + e_t, \overline{SPA}) \tag{6}$$

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 $\overline{SPA} = 67$. 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. The lowest SPA, I consider possible is the pre-reform age of 60. Hence, as the law-of-motion only allows for increases, SPA_t is bounded below by $\underline{SPA} = 60$ and above by $\overline{SPA} = 67$.

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 analog is the SPA the government is currently announcing for the woman's cohort. Only one SPA cohort is modeled 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.

5.2.2. *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 misbeliefs are the result 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 (outlined below). The other stochastic state variables, w_t and ue_t, being directly observed can be interpreted as these variables being more salient.

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 the belief distribution the household holds about SPA_t, $\underline{\pi_t} = \big(\pi(spa)\big)_{spa=\underline{SPA}}^{\overline{SPA}} \in \Delta(8) \subseteq \mathbb{R}^8$.

The household chooses what information about the SPA to acquire, and its choice can be thought of as a two-step process: first, choosing a signal distribution and then choosing actions based on the signal draw. The choice of signal is unrestricted (the household is free to learn about SPA_t however they want), but information is subject to a utility cost (outlined below). Specifically, a household with observed states (X_t and $\underline{\pi_t}$) can choose any conditional distribution function ($\underline{f_t}[X_t, \underline{\pi_t}](z|SPA_t)$) for its signal ($z_t \sim Z_t$), conditioning on the unobserved state (SPA_t). After observing the signal, they select an action ($d_t[X_t, \underline{\pi_t}](z_t)$). So, the value of information is the instrumental value of making better saving and labor supply choices, while its cost is a direct utility cost.

[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

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:

$$Pr_t(spa|z_t) = \frac{f_t(z_t|spa)\pi_t(spa)}{Pr_t(z_t)} = \frac{f_t(z_t|spa)\pi_t(spa)}{\sum_{spa'=60}^{\overline{SPA}} f_t(z_t|spa')\pi_t(spa')} \tag{7}$$

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 6, to this posterior:

$$\pi_{t+1}(spa) = (1-\rho)Pr_t(spa|z_t) + \rho Pr_t(spa-1|z_t). \tag{8}$$

Entropy and mutual information. Entropy, in the 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. The attention cost is proportional to the mutual information, which measures the expected reduction in uncertainty about one variable, quantified by entropy, after learning another variable's value.

DEFINITION—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—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. After incorporating information costs, utility takes the form:

$$u^{(k)}(d_t, \underline{f_t}, \underline{\pi_t}) = n^{(k)} \frac{((c_t/n^{(k)})^{\nu} l_t^{1-\nu})^{1-\gamma}}{1-\gamma} - \lambda I(\underline{f_t}; \underline{\pi_t}) \tag{9}$$

[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 that computers compress files using these concepts.

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

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

Revelation of uncertainty. Upon reaching SPA_t, the woman learns her true SPA_t and starts receiving the state pension. So, the household knows that if they do not receive 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 uncertainty is resolved upon reaching SPA_t can be thought of as reflecting the communication of eligibility and the general process of information disclosure triggered by claiming. At the time in the UK, eligibility was communicated by letter, and claiming involved a telephone conversation in which the implications of claiming were spelled out explicitly.

Dynamic programming problem. The full set of states for the model is:

$$(X_t, SPA_t, \underline{\pi_t}) = (a_t, w_t, AIME_t, ue_t, t, SPA_t, \underline{\pi_t}),$$

and the Bellman equation:

$$\begin{aligned} &V_t^{(k)}(X_t, SPA_t, \underline{\pi_t}) = \\ &\max_{d_t, \underline{f_t}} E\Big[u^{(k)}(d_t, \underline{f_t}, \underline{\pi_t}) + \beta\big(s_t^{(k)} V_{t+1}^{(k)}(X_{t+1}, SPA_{t+1}, \underline{\pi_{t+1}}) + (1 - s_t^{(k)}) T(a_{t+1})\big)\Big] \end{aligned} \quad (10)$$

subject to the same constraints in Equations 3 - 5 as the baseline model and where now the utility function includes a cost as per Equation 9.

A challenge buried in this Bellman equation is the formation of next-period beliefs, which, due to Bayesian updating, depend upon the full distribution of the signal. Hence, we need the solution to form the continuation value. This problem is taken up in Section 6.

5.2.3. *Discussion of Costly Attention to the Stochastic SPA*

Functional form of attention cost. The information acquisition cost is key to the model mechanisms. I assume it is 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 as 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 become 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 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]

Thirdly, reasons exist to believe that the cost of cognition depends on entropy. The information-theoretic concept of entropy sets a lower bound on efficient transmission and storage of information. Thus, if the brain processes information efficiently, mutual information should factor 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. (2022a) 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 data variation exists to identify their extra parameters. As it seems that entropy enters an ideal cost function, my cost function can be considered a first-order approximation over this dimension.

Interpreting the cost of attention. Costly information is modeled abstractly, allowing various interpretations. I propose two: one broad, and one literal.

In the broader view, learning about the SPA represents learning about the state pension system in general. The pension system is multifaceted, and people find many facets confusing. The model concentrates all costs of information acquisition on tracking the SPA, which may also capture learning and the resolution of uncertainty about these other facets. Thus,

[16]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 entropy-based cost of attention.

[17]For example, Kőszegi and Matějka (2020) show this attention cost generates mental budgeting (quantity allocated to a category being fixed and composition changing) and naive diversification (composition being fixed and quantity allocated changing) in different situations. Caplin et al. (2019) show it leads to consideration sets.

SPA learning costs can reflect broader pension policy learning. An extension in online Appendix F explores household learning about actuarial adjustment for deferred claiming.

The more literal view of the cost of attention is as the cost of learning about your SPA exclusively. While your SPA is a single number available online, looking it up does not capture the full costs of learning it. These should include information processing, storage, and recall costs, as well as straightforward hassle or time costs. For illustration, the author has paid the hassle cost of looking up his SPA but not the cognitive cost of remembering it. Hence, I would show up in survey data as having SPA misbeliefs, and I cannot use my SPA in decision-making. Thus, the minimum data- and model-consistent conceptualization includes both cognitive and hassle costs.

Interpreting the choice of signal. As it is a number we can look up, a signal function choice may seem an abstract way to model learning about the SPA. But the signal function choice encompasses (in the guise of a perfectly informative signal) the idea of looking up and remembering your SPA. Moreover, people do not learn about government policy solely from government sources; they rely on news or conversations as well. These sources involve randomness, what stories are covered or discussed, and choice, whether to keep reading or ask questions. This is analogous to the choice of a signal function in that it is partly a choice and partly stochastic. So, this modeling device reflects the messy real-world learning process.

6. 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 in this paper.

The solution method can be considered general purpose because, one, it stores the belief distribution non-parametrically, and two, it does not rely on any specifics of the data-generating process. The most 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 6.1 explains the general-purpose method, and Section 6.2 details specific to solving the model of this paper.

6.1. *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 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. Section 6.1.1 outlines key results from Steiner et al. (2017). Section 6.1.2 uses these results and presents the method.

6.1.1. *Analytic Foundations of Solution Method*

Steiner et al. (2017) show that a wide class of models have logit-like solutions. The key results needed from their paper to understand the solution method are explained below using my model. If we define the effective conditional continuation values as:

$$\overline{V}^{(k)}_{t+1}(d_t, X_t, SPA_t, \underline{\pi_t}) = \\ E\big[s^{(k)}_t V^{(k)}_{t+1}(X_{t+1}, SPA_{t+1}, \underline{\pi_{t+1}}(d_t)) + (1 - s^{(k)}_t) T(a_{t+1}) \big| d_t, X_t, SPA_t, \underline{\pi_t}\big], \tag{11}$$

where expectations are over X_{t+1} and SPA_{t+1} (Section 6.1.2 belows describes finding $\underline{\pi_{t+1}}(d_t)$), then the Bellman equation 10 becomes:

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

Steiner et al. (2017) 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 $(d_t|SPA_t \sim \underline{p_t}(.|SPA_t))$ for the signal function $(\underline{f_t})$ throughout the problem. Thus, we can combine the choice of a stochastic signal function $(\underline{f_t})$ and a deterministic decision conditional on the signal $(d_t(z_t))$ into a single choice of a stochastic decision $(d_t|SPA_t \sim \underline{p_t}(.|SPA_t))$. They show that the solution to this model has actions that are distributed with conditional choice probabilities $d_t|SPA_t \sim \underline{p_t}(.|SPA_t)$ and associated unconditional probabilities $d_t \sim \underline{q_t}(.)$ (i.e., $q_t(d) = \sum_{spa=\underline{SPA}}^{\overline{SPA}} \pi(spa)p_t(d|spa)$) that satisfy:

$$p_t(d|spa) = \frac{\exp\left(n^{(k)}\frac{\left(\left(\frac{c}{n^{(k)}}\right)^{\nu} l^{1-\nu}\right)^{1-\gamma}}{\lambda(1-\gamma)} + \log(q_t(d)) + \beta\overline{V}_{t+1}^{(k)}(d, X_t, SPA_t, \underline{\pi_t})\right)}{\sum_{d'\in\mathcal{C}} \exp\left(n^{(k)}\frac{\left(\left(\frac{c'}{n^{(k)}}\right)^{\nu} l'^{1-\nu}\right)^{1-\gamma}}{\lambda(1-\gamma)} + \log(q_t(d')) + \beta\overline{V}_{t+1}^{(k)}(d', X_t, SPA_t, \underline{\pi_t})\right)}, \quad (12)$$

$$\underline{q_t} = \arg\max_{q} \sum_{spa} \pi_t(spa) \log\left(\sum_{d\in\mathcal{C}} q(d) \exp\left(n^{(k)}\frac{((c/n^{(k)})^{\nu} l^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \beta\overline{V}_{t+1}^{(k)}(d, X_t, SPA_t, \underline{\pi_t})\right)\right). \quad (13)$$

6.1.2. *General-Purpose Solution Method*

At its core, the solution method is to solve Equation 13 for $\underline{q_t}$ and substitute the solution into 12 to get $\underline{p_t}$. This basic description corresponds to an infeasible brute-force version of my solution method and conceals two major hurdles, which I explain below, culminating in a description of the 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, because we do not know the beliefs tomorrow that will result from an action today ($\underline{\pi_{t+1}}(d_t)$), and, as a state, beliefs enter V_{t+1}. To see this, substitute the distributions of actions for the distribution of signals in the Bayesian updating formula 7 and apply the results from Equations 12 and 13 to get:

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

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 6) to this posterior as per 8. That is:

$$\pi_{t+1}(spa) = (1-\rho)Pr_t(spa|d_t) + \rho Pr_t(spa-1|d_t).$$

Thus, beliefs given choices ($\underline{\pi_{t+1}}(d_t)$) are a function of the posterior, which depends not only on the exponentiated payoff but also on $\underline{q_t}$. So, we need a solution ($\underline{q_t}$) to know $\underline{\pi_{t+1}}(d_t)$ and hence to form the effective conditional continuation values (Equation 11).

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 structural modeling, it is often a non-starter as it involves introducing redundant information into the state space. If two action histories lead to the same beliefs, they do not truly represent different states.[18] Redundant information in the state space is problematic, as the curse of dimensionality often makes this the binding constraint to producing richer models. That the redundant information grows exponentially in the number of periods moves this from problematic to a non-starter for many applications.

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 13 for $\underline{q_t}$. If the 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.[19]

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).[20] Al-

---

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

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

[20] Sometimes this is justified as explicit information sharing assumption in the model. Often, it is justified by noting that local posterior invariance (Caplin et al., 2022a) extends to global posterior invariance if all actions

though potentially reasonable in specific applications, suppressing beliefs prevents dynamic rational inattention from modeling 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.

My solution method keeps the belief distribution as a state whilst leveraging results of Caplin et al. (2019) to lighten the computational burden. They show that often 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 ex-ante identify actions that will be taken with zero probability without solving the optimization problem. I then remove these from the decision problem. This filtering step always reduces the dimensionality of the optimization in Equation 13. Moreover, if a single action remains after filtering, we have solved the problem without further calculation. For my model, filtering leaves a single action in over 50% of cases.

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 for $\underline{q_t}$. However, it imposes a much stricter condition, dominant across all joint realizations of SPA_t and $\underline{\pi_{t+1}}$, than needed to drop an action, dominant across all realizations of SPA_t. Therefore, having made an initial guess for $\underline{\tilde{q}_t}$, and so having prediction for next period beliefs given any action ($\underline{\pi_{t+1}}(d_t)$) 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. Abstracting from borrowing constraints,

---

are taken with positive probability. However, Caplin et al. (2019) show that solutions are rarely strictly interior as rational inattention often implies consideration sets. Hence, the extension of local posterior invariance to a global property is restrictive.

the household faces 1,500 options, 500 saving levels, and 3 labor supply choices. A household will never assign positive probability to more actions than the random variable they are learning about (SPA_t) has points of support. 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 13.

Removing strictly dominated actions only uses ordinal information. The second criterion used to filter 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.1) that if there exists a decision $d^\star = (c^\star, l^\star)$ which satisfies:

$$\sum_{spa} \pi_t(spa) \frac{\exp\left(n^{(k)} \frac{((c/n^{(k)})^\nu l^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \beta \overline{V}^{(k)}_{t+1}(d, X_t, spa, \underline{\pi_t})\right)}{\exp\left(n^{(k)} \frac{((c^\star/n^{(k)})^\nu l^{\star 1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \beta \overline{V}^{(k)}_{t+1}(d^\star, X_t, spa, \underline{\pi_t})\right)} < 1, \tag{14}$$

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 13 easier, checking Equation 14 is only beneficial when the optimal behavior is to take the same action in all realizations of SPA_t. So, the benefits of checking condition 14 depend on how frequently, in the problem faced, it reveals the optimal choice without needing to solve an optimization. When filtering does not leave a single action, I employ sequential quadratic programming to solve Equation 13, an algorithmic choice suggested by Armenter et al. (2024). High-level pseudo code summarizing the algorithm is in online Appendix C.

Online Appendix C details 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, Equation 13 can lead to underflow problems.

6.2. *Computational 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. Beliefs ($\underline{\pi_t}$) and learning ($\underline{f_t}$) alter the nature of the within-period problem in the version 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 are solved, like periods in the other two versions, by simple search techniques to find the optimal choice amongst the discrete set of assets and labor supply choices.

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 until age $t = 66$. We can proceed back as far as age $t = 67$ because SPA_t is bounded above by 67, so 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, as she knows the data-generating process. In this period, $\underline{\pi_t}$ is not a state variable, but SPA_t is, as receipt of the state pension affects available resources.

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 backward, 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, 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 SPAs (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 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 SPA_t and the support of $\underline{\pi_t}$ is fixed.

7. 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 (a small number of parameters are also 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.

7.1. *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, and in the version with rationally inattentive households $\underline{\pi_t}$. Initial wages w_t are drawn from the estimated initial wage distribution (see below), and all agents start as employed ($ue_t = 1$). Beliefs ($\underline{\pi_t}$) are initialized from the type- and SPA-cohort specific empirical distribution, and assets (a_t) and average earnings ($AIME_t$) from their joint type- and SPA-cohort specific empirical distribution. The empirical counterpart used for assets is household non-housing non-business wealth. Using the full work histories in the administrative data linked to wave 5 of ELSA, I construct a measure of $AIME_t$. As this is only possible for a subsample, to estimate the joint distribution of $AIME_t$ and a_t, I impute missing $AIME_t$ values with a quintic in wealth and a rich set of observed characteristics (details in online Appendix D). To initialize beliefs from the point-estimate belief data, I assume that responses represent a draw from an individual's subjective beliefs distribution.[21]

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

[21] This assumption is consistent with evidence from psychology that averaging multiple responses elicited from an individual improves accuracy (Vul and Pashler, 2008). It also enables construction of an individual's subjective belief distribution from point estimates.

process:

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

for women j, of type k, and at age t. The parameters of the age-dependent deterministic component of the wage process ($\delta_{k0}, \delta_{k1}, \delta_{k2}$) are estimated by type-specific regression. The parameters of the stochastic component of the wage equation ($\rho_w, \sigma_\epsilon, \sigma_{\epsilon,55}, \sigma_\mu,$) are found minimizing the distance between the empirical covariance matrix of estimated residuals and the theoretical variance-covariance matrix of $\epsilon_t + \mu_{j,t}$ (similar to Low et al., 2010).

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. As the state pension is relatively insensitive to education and the private pension relatively insensitive to marital status, I simplify the state pension to be marital-status-specific and the private pension education-specific. I estimate the private pension claiming age ($PPA^{(k)}$) 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 unemployment.

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 6, 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.

7.2. *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 32 pre-reform moments of mean labor market participation and asset holdings from ages 55 to 70 were used to estimate $(\gamma, \nu, \beta, \theta)$. To avoid cohort effects or macroeconomic influences, a fixed-effect age regression was estimated, including birth-year effects, SPA-cohort-specific age effects, aggregate unemployment (to half a percentage point), and an indicator for being below the SPA. Target profiles were then generated using these regressions with average pre-reform cohort values (details in online Appendix D).

In the model version with rationally inattentive households, λ is identified from the reduction in self-reported SPA mean squared error between 55 and 58. The estimation of λ is done separately from the other parameters, with their values held constant at those estimated for the version with only policy uncertainty. This has three advantages: one, it reduces computation; two, it uses the variation most directly affected by costly attention to identify λ; and three, it separates the effects of costly attention from effects of changing parameter values. The trade-off is not using all available information to identify λ.

8. RESULTS

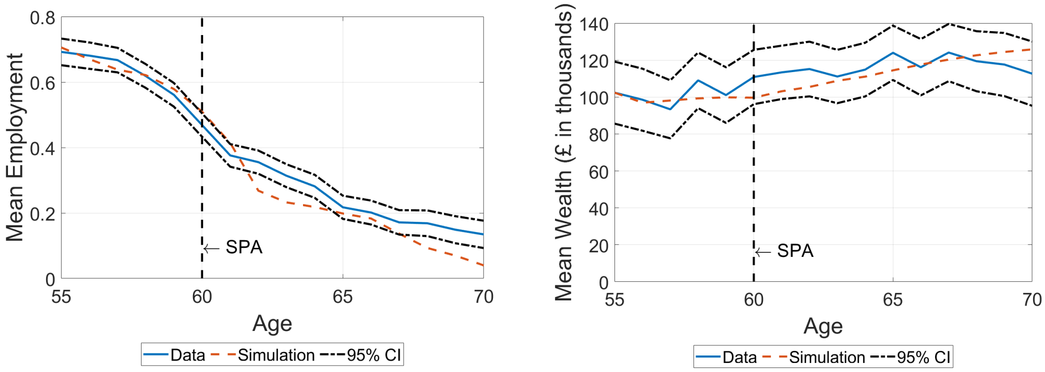
Section 8.1 evaluates model fit and ability to replicate key facts on excess employment sensitivity, misbeliefs, and their relationship. Section 8.2 explores the implications.

8.1. *Model Evaluation*

This section presents the model fit and each versions' ability to replicate the employment response to the SPA and its relation to beliefs (first stage results in online Appendix E.1).

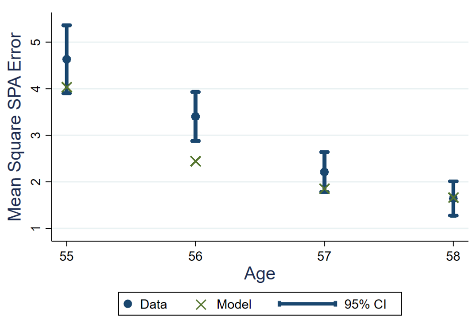
Figures 3a and 3b show the model with policy uncertainty fits pre-reform employment and asset profiles well when simulated with the pre-reform SPA of 60. Table 3d lists the estimated parameters. The baseline model and the version combining policy uncertainty with rational inattention produce similar fits to these static profiles (graphs in online Appendix E.2). However, the three versions predict very distinct responses to SPA changes.

FIGURE 3.—Model Fit and Parameter Estimates



(a) Employment Profile

(b) Asset Profile



(c) Belief Profile

TABLE

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)

(d) Parameter Estimates

Note: Panels (a)-(c) show model fit to targeted profiles, the empirical profile is for the pre-reform SPA cohort with a SPA of 60. Panel (d) shows estimated parameters (analytic standard errors in brackets calculated following Newey (1985)).

To analyze this response to the SPA, I simulate the model with the SPAs observed in ELSA waves 1-7 ($SPA = 60$, $SPA = 61$, $SPA = 62$) and repeat the regression from Section 4.1 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 period effects, which aligns with age in the model), and model counterparts to empirical controls (assets, marital status, education). As in Section 4.1, I repeat this on the subsample with above-median empirical assets (£28,500) before SPA. Results are in Table II's top panel. Column 5 repeats the empirical treatment effects from Columns 1 and 2 of Table I. The baseline model fails to match either.

TABLE II

UNTARGETTED MODEL FIT TO REGRESSION RESULTS

	Baseline	Policy Uncert.	$\hat{\lambda} = 6 \times 10^{-8}$	$\lambda = 1.0 \times 10^{-3}$	Data (95% C.I)
Treatment Effect being above SPA on employment					
Whole Population	0.019	0.014	0.041	0.095	0.128 *(0.081,0.176)*
Assets >Median(£28,500)	0.018	0.014	0.054	0.095	0.106 *(0.047,0.166)*
Treatment Effect Heterogeneity by Absolute SPA Error					
Interaction	—	—	-0.047	-0.046	-0.049 *(-0.097, -0.001)*
Treatment Effect Heterogeneity by SPA Error Positivity					
Interaction	—	—	-0.047	-0.046	-0.078 *(-0.262, 0.106)*

Note: The top panel shows employment response across the wealth distribution (Table II). The second panel shows heterogeneity in SPA labor supply response by absolute size of self-reported SPA error at 58. The second panel shows heterogeneity in SPA labor supply response by direction of self-reported SPA error at 58, and the third by absolute size of the error. Some results are identical to three decimal places but differ to four decimal places.

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

Introducing costly attention adds a parameter λ, which I identify from the reduction in mean squared error in self-reported SPAs between ages 55 and 58 for the same SPA-cohort as other targeted moments ($SPA = 60$). The mean square error of model-predicted and data beliefs are presented in Figure 3c. Beliefs at 55 are initialized from the data, so the fit in that period is mechanical (a slight undershooting results from discretizing beliefs). Beliefs at age 58 are targetted to identify λ, with beliefs at the two intervening ages (56 and 57) being untargeted moments. The value estimated is $\hat{\lambda} = 6 \times 10^{-8}$. Column 3 of Table II shows

[22] Section 4.1 highlights the ex-ante puzzling response of the wealthy, and targeting the two treatment effects directly allows the baseline to match the overall population response but not the wealthy subgroup's (results available on request). Thus, I consider the wealthy's response puzzling, though the baseline struggles most with the aggregate with the estimated parameters.

that this model version matches the employment response to the SPA significantly better than the baseline or the policy uncertainty versions but still falls short of the data. Costly attention closes 23% of the gap for the whole population and 43% for the richer subgroup, with only the richer subgroup's estimate falling within the 95% confidence interval.

The dependence on earlier misbeliefs of employment responses later in life spurred investigation into costly attention's role in the excess sensitivity puzzle. Column 6 of Table I shows individuals better informed about their SPA in their late 50s exhibit smaller labor supply responses at SPA in their 60s. Two opposing forces in the model link the accuracy of earlier SPA knowledge to labor supply responses to it. Endogenous SPA knowledge implies those least dependent on the SPA acquire less information. Conversely, households worse informed 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 middle panel of Table II shows a negative relationship, indicating the model reproduces the observed direction of this relationship. The bottom panel also shows the model replicates the (non-significant) direction of the dependence of SPA employment responses on the direction of SPA misbeliefs.

Comparison to reference point retirement A leading alternative explanation for the employment response to pension eligibility is reference-dependent preferences, which assume a shift in utility from leisure at the eligibility age. This explanation, however, does not address misbeliefs. Such studies typically introduce a parameter to directly target the employment response to the pension age (e.g. Seibold, 2021). In Column 4 of Table II, I similarly introduce a cost of attention that fits the employment response to SPA well. Costly attention now accounts for 71% of the gap for the whole population and 88% for the richer subgroup, both estimates within the 95% confidence intervals.

Nevertheless, an appeal of costly attention as an explanation is that it also accounts for misbeliefs, providing extra data to identify the parameters. When restricted by the beliefs data, costly attention only partially explains the employment response to SPA. Two potential explanations stand out. One, this paper attributes all policy learning to the SPA, whereas pension systems are complex, and individuals misunderstand many of their dimensions. This could understate learning at eligibility. Online Appendix F extends pension policy uncertainty to include learning about deferral rules, though data limitations make this work speculative. Two, misbeliefs may work alongside behavioral biases like reference dependence to shape employment responses. Intriguing evidence suggests framing effects may influence labor supply reactions to pension age changes (discussed in online Appendix E.3). Thus, online Appendix E.4 also presents results for a model with $\hat{\lambda} = 6 \times 10^{-8}$ and passive decision-makers (as in Chetty et al., 2014), who retire at SPA regardless.

Explaining misbeliefs is the key argument for costly attention as an, at least complementary, explanation for the employment response to eligibility. A potential secondary benefit is that the endogeneity of attention may explain differences in employment response across time and countries as responses to different policy environments. For instance, Deshpande et al. (2024) find smaller employment responses to the US full retirement age during reform periods. If driven by fixed preferences, such variation would not occur. With costly attention, however, misbeliefs may be lower during reform periods, especially when (as in the US) they were accompanied by major information campaigns.

8.2. *Model Implications and Predictions*

Attention cost size. λ is hard to interpret, having natural units of utils per bit. While utils are known to be non-interpretable, denominating in bits exaggerates costs, as models contain far fewer learnable bits than reality. Most models contain only single or double-digit bits of information, less than in an average sentence. Reality holds vastly more information, making per-bit information cost a larger share of total model information. To address both issues, I calculate the compensating asset that raises household utility as much as perfect SPA knowledge, effectively their willingness to pay to learn their SPA. For $\hat{\lambda} = 6 \times 10^{-8}$, compensating assets range from £6 at the 25th percentile to £14 at the 75th, with a mean of £11. For $\lambda = 1 \times 10^{-3}$, the mean is £83 (summary of compensating assets distributions for both λ values in the online Appendix).

The employment response to pension age reforms. Rising old-age dependency ratios make increasing older individuals' employment a global policy priority, with pension 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 the SPA a more effective tool. Generally, it does the opposite.

TABLE III

IMPACTS OF REFORMING SPA WITH INFORMED AND UNINFORMED HOUSEHOLDS

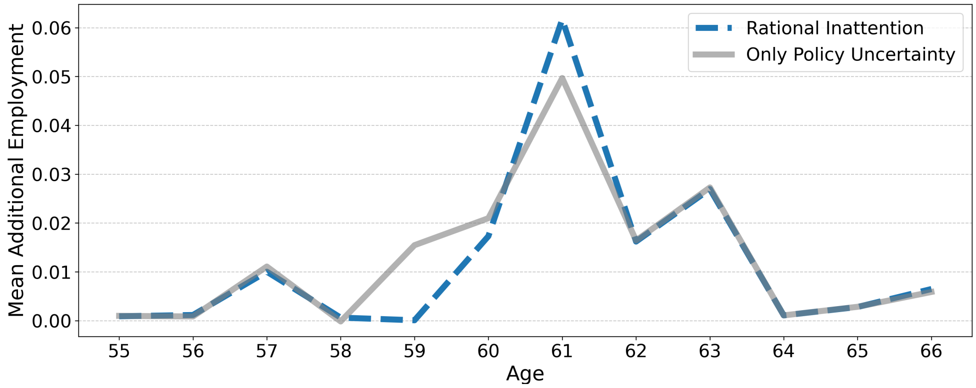
SPA increased from 60 to:	(1) - Informed Added Employment	(2) - Uninformed Added Employment	(3) MC	(4) WTP	(5) MR
61	0.07	0.06	£3.50	£4.22	£28.45
62	0.14	0.14	£4.00	£2.37	£11.78
63	0.18	0.16	£4.50	£18.34	£19.91
64	0.22	0.20	£5.00	£31.64	£4.31
65	0.31	0.27	£5.50	£44.41	£68.52

Note: Employment increases over 56-65 from raising SPA from 60 to the age in Column (1) with costly attention and in Column (2) without it. Columns 3-5 show the financial impacts of an accompanying information letter campaign that moves people from uninformed to informed. Column (3) shows the marginal cost, Column (4) the willingness to pay, and Column (5) the marginal revenue.

Column 2 of Table III shows the change in mean employment during ages 55–65 when the SPA is reformed from 60 to 61–65, based on the model with $\hat{\lambda} = 6 \times 10^{-8}$ and initializing prior beliefs and other state variables with the values of the SPA 60 cohort. Thus, this captures the response to an unanticipated SPA increase at age 55. Column 1 shows results from the model with policy uncertainty but no attention costs. Both versions show modest employment gains, with mostly larger increases under costly attention. For post-reform SPA 65, mean employment rises 0.31 years with attention costs vs. 0.27 without. So, employment rises up to 15% more under costly attention, which may seem at odds with the finding that it causes a larger employment drop at SPA. This tension resolves when noting that rationally inattentive households respond less immediately to SPA increases. Fully informed households internalize the change early, increasing work in their 50s. Inattentive households react later—often near the old SPA of 60—when they realize they must compensate for lost earnings. 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 yields smaller overall employment gains but a larger response at SPA, with much bunching driven by intertemporal shifts. Figure 4 illustrates this for a SPA rise to 62.

The impact of information on response to pension age reforms. Columns 1 and 2 of Table III show added employment from an unanticipated SPA increase at age 55 in models

FIGURE 4.—Additional Employment resulting from Increasing the SPA from 60 to 62



Note: For the two versions, employment increases resulting from a reform of the female SPA from 60 to 62.

with and without costly information. The only difference is in Column 1, households know the SPA, and in Column 2, they do not. Thus, the gap reflects the maximum potential impact of an annual information letters campaign. Columns 3–5 assess such a campaign.

Column 3 reports the marginal cost of the information letter campaign. After covering fixed costs, the only marginal cost is postage at £0.50/year (2013 prices). Column 4 shows the willingness to pay (WTP) for the information campaign under each post-reform SPA. Two forces drive WTP: higher SPAs reduce lifetime wealth (lowering WTP), but as it moves further from the pre-reform SPA of 60, the value of information rises. Initially, the first effect dominates, reducing WTP. From SPA 63 onward, the second dominates, and WTP increases. Comparing Columns 3 and 4 shows WTP for information exceeds the campaign's marginal cost for all post-reform SPAs except 62. For these reforms, the information campaign improves net welfare without accounting for added government revenue, but since the campaign also raises employment (see Columns 1 and 2), the campaign is revenue-positive as quantified in Column 5. Though modest (1950s-born women had low earnings), revenue exceeds marginal cost for all SPA reforms except 64. Combining household and government gains, Columns 3–5 show the information campaign consistently raises total welfare, with benefits exceeding costs by 3.5 to 20.5 times. Though absolute gains are modest, the experiment underscores a key point: informing individuals not only improves their welfare but also improves their responsiveness to policy.

9. CONCLUSION

Mistaken beliefs are common, but their economic impacts are still not well-understood. Using UK data, this paper shows that incorporating costly attention, which endogenously generates misbeliefs, into a retirement model explains both observed misbeliefs and the sensitivity of employment to pension eligibility ages. Costly attention accounts for 43% of the employment response gap between model and data when calibrated to observed beliefs and 88% when unconstrained. Given both pension misbeliefs and excessive employment responses are across-country regularities, these insights may be cross-nationally relevant.

Endogenous information acquisition is key to explaining retirement behavior but leads to the prior belief becoming a state variable. This high-dimensional state variable significantly increases computational demands. I propose a method for solving dynamic rational inattention models without suppressing beliefs as a state variable. From the belief data, I estimate the mean willingness to pay to learn the SPA as £11. Though small, this far exceeds the marginal cost of information letters. Policy experiments show that after most SPA reforms, households' willingness to pay for such letters exceeds their cost, but also that sending letters increases employment by up to 15%. Hence, the campaign raises additional tax revenue, which, for most SPA reforms, also 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, HASSAN AND CHOONGRYUL YANG (2021): "Dynamic rational inattention and the Phillips curve," . [6]

AMIN-SMITH, NEIL AND ROWENA 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*, 9–39. [7]

ARMENTER, ROC, MICHÈLE MÜLLER-ITTEN, AND ZACHARY R STANGEBYE (2024): "Geometric methods for finite rational inattention," *Quantitative Economics*, 15 (1), 115–144. [4, 6, 29, 31]

BAIROLIYA, NEHA AND KATHLEEN MCKIERNAN (2023): "Revisiting Retirement and Social Security Claiming Decisions," . [7]

BANKS, J., G.D. BATTY, J.J.F. BREEDVELT, K. COUGHLIN, R. CRAWFORD, M. MARMOT, J. NAZROO, Z. OLDFIELD, N. STEEL, A. STEPTOE, MARTIN. WOOD, AND P. ZANINOTTO (2021): "English Longitudinal Study of Ageing: Waves 0-9," 1998-2019 [data collection]. [10]

BARTOŠ, VOJTĚCH, MICHAL BAUER, JULIE CHYTILOVÁ, AND FILIP MATĚJKA (2016): "Attention Discrimination: Theory and Field Experiments With Monitoring Information Acquisition," *American Economic Review*, 106 (6), 1437–1475. [5]

BEHAGHEL, LUC AND DAVID BLAU (2012): "Framing Social Security Reform : Behavioral Responses to Changes in the Full Retirement Age," *American Economic Journal: Economic Policy*, 4 (4), 41–67. [6]

BERNHEIM, DOUGLAS (1988): *Social Security Benefits: An Empirical Study of Expectations and Realizations*. [6]

BOEHM, EDUARD (2023): "Intermediation, Choice Frictions, and Adverse Selection: Evidence from the Chilean Pension Market," . [6]

BORUSYAK, KIRILL, XAVIER JARAVEL, AND JANN SPIESS (2024): "Revisiting event-study designs: robust and efficient estimation," *Review of Economic Studies*, rdae007. [12, 14]

BOZIO, ANTOINE, ROWENA CRAWFORD, AND GEMMA TETLOW (2010): "The history of state pensions in the UK: 1948 to 2010," Tech. rep. [7, 8, 20]

BROWN, ZACH Y AND JIHYE JEON (2024): "Endogenous information and simplifying insurance choice," *Econometrica*, 92 (3), 881–911. [5]

BURTLESS, GARY (1986): "Social Security, Unanticipated Benefit Increases, and the Timing of Retirement," *Review of Economic Studies*, 53 (5), 781–805. [5]

CAPLIN, ANDREW, MARK DEAN, AND JOHN LEAHY (2019): "Rational Inattention, Optimal Consideration Sets, and Stochastic Choice," *Review of Economic Studies*, 86 (3), 1061–1094. [2, 4, 5, 25, 30, 31]

——— (2022a): "Rationally inattentive behavior: Characterizing and generalizing Shannon entropy," *Journal of Political Economy*, 130 (6), 1676–1715. [25, 29]

CAPLIN, ANDREW, EUNGIK LEE, SØREN LETH-PETERSEN, AND JOHAN SÆVERUD (2022b): "Communicating social security reform," Tech. rep., National Bureau of Economic Research. [6]

CARROLL, CHRISTOPHER AND ANDREW SAMWICK (1996): "The Nature of Precautionary Wealth," *NBER Working Paper*. [12]

CASANOVA, MARIA (2010): "Happy Together : A Structural Model of Couples ' Joint Retirement Choices," *Working Paper*, (5), 1–53. [19]

CHETTY, RAJ (2012): "Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply," *Econometrica*, 80 (3), 969–1018. [4]

CHETTY, RAJ, JOHN N FRIEDMAN, SØREN LETH-PETERSEN, TORBEN HEIEN NIELSEN, AND TORE 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. [39]

CIANI, EMANUELE, ADELINE DELAVANDE, BEN ETHERIDGE, AND MARCO FRANCESCONI (2023): "Policy uncertainty and information flows: Evidence from pension reform expectations," *The Economic Journal*, 133 (649), 98–129. [2]

CRIBB, JONATHAN, CARL EMMERSON, AND GEMMA TETLOW (2016): "Signals matter? Large retirement responses to limited financial incentives," *Labour Economics*, 42, 203–212. [6, 10, 11]

DE BRESSER, JOCHEM (2023): "Evaluating the Accuracy of Counterfactuals: Heterogeneous Survival Expectations in a Life Cycle Model," *Review of Economic Studies*, rdad088. [6]

DE BRUIN, WÄNDI BRUINE, ALYCIA CHIN, JEFF DOMINITZ, AND WILBERT VAN DER KLAAUW (2023): "Household surveys and probabilistic questions," in *Handbook of Economic Expectations*, Elsevier, 3–31. [14]

DE NARDI, MARIACRISTINA (2004): "Wealth inequality and intergenerational links," *The Review of Economic Studies*, 71 (3), 743–768. [18]

DE NARDI, MARIACRISTINA, ERIC FRENCH, AND JOHN BAILEY JONES (2010): "Why Do the Elderly Save? The Role of Medical Expenses," *Journal of Political Economy*, 118 (1), 39–75. [34]

DEAN, MARK AND NATHANIEL NELIGH (2023): "Experimental tests of rational inattention," *Journal of Political Economy*, 131 (12), 3415–3461. [25]

DESHPANDE, MANASI, ITZIK FADLON, AND COLIN GRAY (2024): "How Sticky Is Retirement Behavior in the United States?" *Review of Economics and Statistics*, 106 (2), 370–383. [39]

FRENCH, ERIC (2005): "The Effects of Health, Wealth, and Wages on Labour Supply and Retirement Behavior," *Review of Economic Studies*, 72 (2), 395–427. [2, 5, 6, 34]

FRENCH, ERIC AND JOHN BAILEY JONES (2011): "The Effects of Health Insurance and Self-Insurance on Retirement Behavior," *Econometrica*, 79 (3), 693–732. [5, 6]

GRUBER, JONATHAN, OHTO KANNINEN, AND TERHI RAVASKA (2022): "Relabeling, retirement and regret," *Journal of Public Economics*, 211, 104677. [6]

GRUBER, JONATHAN AND DAVID A WISE (2004): "Social Security Programs and Retirement around the World: Micro-Estimation," Tech. rep., National Bureau of Economic Research. [2, 6]

GUSTMAN, ALAN L AND THOMAS L STEINMEIER (1986): "A Structural Retirement Model," *Econometrica*, 54 (3), 555–584. [5]

——— (2005): "Imperfect knowledge of social security and pensions," *Industrial Relations: A Journal of Economy and Society*, 44 (2), 373–397. [2]

KŐSZEGI, BOTOND AND FILIP MATĚJKA (2020): "Choice Simplification: A Theory of Mental Budgeting and Naive Diversification," *The Quarterly Journal of Economics*, 135 (2), 1153–1207. [25]

KOŞAR, GIZEM AND CORMAC O'DEA (2022): " Expectations Data in Structural Microeconomic Models," *National Bureau of Economic Research*. [6]

KOLSRUD, JONAS, CAMILLE LANDAIS, DANIEL RECK, AND JOHANNES SPINNEWIJN (2024): "Retirement consumption and pension design," *American Economic Review*, 114 (1), 89–133. [5, 39]

LALIVE, RAFAEL, ARVIND MAGESAN, AND STEFAN STAUBLI (2023): "How social security reform affects retirement and pension claiming," *American Economic Journal: Economic Policy*, 15 (3), 115–150. [6, 16]

LOW, HAMISH, COSTAS MEGHIR, AND LUIGI PISTAFERRI (2010): "Wage risk and employment risk over the life cycle," *American Economic Review*, 100 (4), 1432–1467. [34]

LUMSDAINE, ROBIN L, JAMES H STOCK, AND DAVID A WISE (1996): "Why Are Retirement Rates So High at Age 65?" in *Advances in the Economics of Aging*, 61–82. [6]

LUSARDI, ANNAMARIA AND OLIVIA S MITCHELL (2011): "Financial literacy and retirement planning in the United States," *Journal of pension economics & finance*, 10 (4), 509–525. [2]

MACAULAY, ALISTAIR (2021): "Cyclical Attention to Saving *," . [4, 6, 29]

MACKOWIAK, BARTOSZ, FILIP MATEJKA, AND MIRKO WIEDERHOLT (2018): "Rational inattention: A disciplined behavioral model," *CEPR Discussion Papers*, 13243. [25]

MAĆKOWIAK, BARTOSZ AND MIRKO WIEDERHOLT (2009): "Optimal sticky prices under rational inattention," *American Economic Review*, 99 (3), 769–803. [5]

——— (2015): "Business cycle dynamics under rational inattention," *Review of Economic Studies*, 82 (4), 1502–1532. [5]

MANSKI, CHARLES (2004): "Measuring expectations," *Econometrica*, 72 (5), 1329–1376. [6]

MASTROBUONI, GIOVANNI (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. [6, 11]

MATĚJKA, FILIP AND ALISDAIR MCKAY (2015): "Rational inattention to discrete choices: A new foundation for the multinomial logit model," *American Economic Review*, 105 (1), 272–298. [2, 5, 29]

MIAO, JIANJUN AND HAO XING (2024): "Dynamic discrete choice under rational inattention," *Economic Theory*, 77 (3), 597–652. [4, 6, 25, 29]

NEWEY, WHITNEY K (1985): "Generalized method of moments specification testing," *Journal of econometrics*, 29 (3), 229–256. [36]

O'DEA, CORMAC (2018): "Insurance , Efficiency and Design of Public Pensions ," *WP*, (1), 1–63. [5, 34]

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. [8]

PORCHER, CHARLY (2020): "Migration with Costly Information," . [4, 6, 29]

ROHWEDDER, SUSANN AND KRISTIN KLEINJANS (2006): "Dynamics of Individual Information about Social Security," *RAND WP*. [7]

RUST, JOHN AND CHRISTOPHER PHELAN (1997): "How Social Security and Medicare Affect Retirement Behavior In a World of Incomplete Markets," *Econometrica*, 65 (4), 781. [2, 5, 6]

SEIBOLD, ARTHUR (2021): "Reference points for retirement behavior: Evidence from german pension discontinuities," *American Economic Review*, 11 (4), 1126–1165. [6, 16, 38]

SIMS, CHRISTOPHER (2003): "Implications of rational inattention," *Journal of Monetary Economics*, 50 (3), 665–690. [2, 5]

STAUBLI, STEFAN AND JOSEF ZWEIMÜLLER (2013): "Does raising the early retirement age increase employment of older workers?" *Journal of public economics*, 108, 17–32. [11]

STEINER, JAKUB, COLIN STEWART, AND FILIP MATĚJKA (2017): "Rational Inattention Dynamics: Inertia and Delay in Decision-Making," *Econometrica*, 85 (2), 521–553. [4, 5, 25, 27, 29]

TUREN, JAVIER (2023): "State-dependent attention and pricing decisions," *American Economic Journal: Macroeconomics*, 15 (2), 161–189. [4, 6, 29]

VAN DER KLAAUW, WILBERT AND KENNETH I. WOLPIN (2008): "Social security and the retirement and savings behavior of low-income households," *Journal of Econometrics*, 145 (1-2), 21–42. [5]

VUL, EDWARD AND HAROLD PASHLER (2008): "Measuring the crowd within: Probabilistic representations within individuals," *Psychological Science*, 19 (7), 645–647. [33]