ONLINE APPENDIX FOR COSTLY ATTENTION AND RETIREMENT'

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

APPENDIX A: ADDITIONAL EMPIRICAL DETAILS

A.1. *Additional Institutional Details*

A.2. *Equity Acts*

The Equality Act (2006) banned mandatory retirement below age 65. Women observed reaching SPA in ELSA waves 1–7 did so after compulsory retirement at their SPAs (60-63) became illegal. The Equality Act (2010) banned all compulsory retirement ages with specific exceptions known as EJRA (Employer Justified Retirement Ages).

As EJRAs must be over 65 and all SPAs reached in the data are below this, EJRAs are not directly relevant. However, background and anecdotes may help illustrate the strictness of UK age discrimination laws on forced retirement. Seldon v Clarkson, Wright and Jakes (2012) clarified when EJRAs are justified, setting out three criteria. One, the justification must serve a public interest (e.g., intergenerational fairness), not just firm goals. Two, this objective must be consistent with the social policy aims of the state. Three, it must be a proportionate means to that end.

In Seldon v Clarkson, the plaintiff, a law firm partner, was subject to a justified EJRA. Documented EJRAs are rare; beyond law firm partners, the most debated cases involve Oxford and Cambridge. Most other UK universities have scrapped compulsory retirement. Notably, Oxford recently lost a tribunal where its EJRA was ruled unjustified. In Ewart v University of Oxford (2019), the court found Oxford's aim (intergenerational fairness) valid, but the EJRA disproportionate —its limited effectiveness didn't outweigh its clear harms. This underscores how seriously UK law treats forced retirement as age discrimination, with few, truly exceptional, exemptions.

TABLE I

EFFECT OF SPA ON HAZARD RATE: HETEROGENEITY BY VLA

	(1)	(2)	(3)	(4)
Above SPA	**0.128**	**0.120**	**0.128**	**0.140**
s.e	(0.0239)	(0.0320)	(0.0381)	(0.0237)
Above SPA×(VLA.>Med.)			**-0.007**	
s.e			(0.0496)	
Above SPA× VLA.				**-1.23×10^{-7}**
s.e				(3.30e-08)
Obs.	7,907	3,691	7,907	7,784

Note: Column (1) presents results from the specification in Equation 1 in the main text. Column (2) repeats the regression for those with above-median Very Liquid Assets (VLA) in their last interview before SPA. Column (3) tests if treatment effects differ by fully interacting the specification with having above-median VLA. Column (4) adds an interaction between wealth and being above SPA. Controls include marital status, education, self-reported health dummies, partner's age, age squared, qualifications, partner's SPA eligibility, years of education, and household assets.

A.3. *Robustness: Excess Employment Sensitivity*

A.3.1. *Restricted Asset Categorisation*

The aim of examining treatment effect heterogeneity by asset holdings is to assess the role of liquidity constraints. The main analysis uses NHNBW, but since parts of NHNBW may be illiquid, Table I repeats the analysis using a narrower category—very liquid assets, i.e., those reasonably cashable within weeks. Results are qualitatively similar to those with NHNBW and do not suggest liquidity constraints alone explain the treatment effect. The effect remains positive for those above median assets; subgroup differences are still insignificant, and the continuous interaction shows heterogeneity is too weak for liquidity constraints to fully account for the effect.

A.3.2. *Bad Control Concerns*

Bad controls are a key concern in DID, with some arguing only time-invariant controls should be used, as controls imply parallel trends conditional on them. To address this, I take a broad approach and run the model without controls, showing that the main conclusions

TABLE II

EFFECT OF SPA ON HAZARD RATE: HETEROGENEITY BY NHNBW NO CONTROLS

	(1)	(2)	(3)	(4)
Above SPA	**0.123**	**0.093**	**0.161**	**0.136**
s.e	(0.02468)	(0.03155)	(0.03716)	(0.02599)
Above SPA $\times$(NHNBW.>Med.)			**-0.068**	
s.e			(0.04868)	
Above SPA $\times$ NHNBW.				**-8.26×10^{-8}**
s.e				(2.32e-08)
Obs.	8,119	3,963	8,119	7,898

Note: Column (1) presents results from the specification in Equation 1 in the main text. 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.

remain unchanged. Table II presents these results. As shown, they differ little from those with controls.

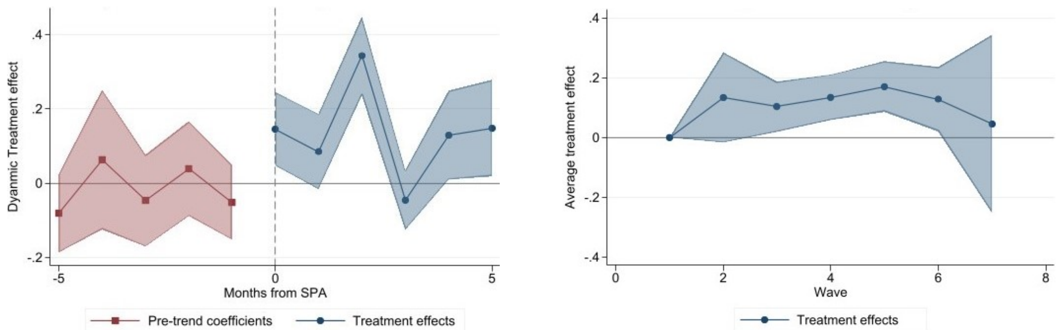
A.3.3. *Imputation Approach to DID*

Two-way fixed effects DID models assume treatment effect heterogeneity across time and units. When treatment timing drives variation—as in this paper—violating these assumptions can yield nonsensical combinations of individual-level effects. Recent literature highlights this issue and, importantly, offers solutions that relax these assumptions.

I apply the imputation method from Borusyak et al. (2024). Figure 1a shows dynamic treatment effects before and after SPA. No signs of violated parallel trends or anticipation effects appear as all pre-SPA effects are insignificant. A joint test confirms this ($p = .799$). Post-SPA, 4 of 6 effects are significant, and we reject the null of joint zero effects ($p = .000$). While the graph suggests limited variation among post-SPA effects, we cannot reject their equality ($p = .198$).

Figure 1b examines whether treatment effects vary by wave. They appear fairly uniform, though we can reject equality ($p = .137$). Neither violation of homogeneity seems severe, and overall, the graphs support the baseline assumption of a homogeneous treatment effect starting at SPA, though tests show this is an approximation.

(a) Treatment Effects by Time from SPA (b) Average Treatment Effect by Wave



Note: Panel (a) Average of the individual level treatment effects estimated using the imputation approach at a given time from SPA. Panel (b) shows the within-wave average of the individual-level treatment effects estimated using the imputation approach.

These results show, allowing for arbitrary heterogeneity, something special is still happening at the SPA, which is difficult to explain in standard complete information models.

A.3.4. *Health, Wealth, Private Pensions, Joint Retirement, and Dismissals*

This section addresses alternative explanations for employment sensitivity to the SPA under a standard complete information framework. Specifically, it considers whether wealth, health, private pensions, joint retirement, or dismissals explain the labor supply response.

Wealth effects influence labor supply, and women with later SPAs are lifetime poorer, so the puzzle isn't their higher labor supply but why it drops at the SPA, despite changes being announced 15+ years in advance. In standard life-cycle models with complete information, a wealth-driven response should be spread over life, not concentrated at the SPA. In Equation 1 in the main text, lifetime wealth differences across birth cohorts (including those induced by SPA shifts) are absorbed by cohort effects. Thus, only within-cohort SPA-induced wealth differences are captured by the regressions. Additionally, the regression only captures the employment response at the SPA, so to explain the observed treatment effect via within-cohort wealth differences, the wealth effect would need to be enormous. Assuming a purely wealth-driven labor supply response implies a marginal propensity to earn (MPE) of about –0.3. This is on the high end of modern estimates (e.g. Cesarini et al., 2017), but becomes implausibly high given this captures just the final 2–3 months of a response that is spread out over 15-20 years. A wealth effect explanation also poorly explains the treatment's limited sensitivity to asset levels.

TABLE III

HETEROGENEITY BY HEALTH

	Coeff	s.e.	p=
Above SPA	0.112	0.0333	0.001
Above SPA×(V.good Health)	-0.002	0.0275	0.917
Above SPA×(Good Health)	0.353	0.0294	0.229
Above SPA×(Fair Health)	0.058	0.0457	0.208
Above SPA×(Poor Health)	0.026	0.0674	0.697

Note: Results of conditioning the treatment effect estimates from Table I in the main text on self-declared health status.

Health is a key driver of retirement decisions (e.g. De Nardi et al., 2010), but there's no reason to expect it to interact with the SPA or to explain employment's sensitivity to it. During the study period, the SPA was 60–63, while average health declines occur later. All the same, given health's importance, Table III examines heterogeneity in labor supply response by health status. Only those in the poorest health group show a significantly different response. This group represents <7% of the sample, and excluding them does not alter results qualitatively.

Private pension eligibility affects retirement decisions. However, occupational pension schemes likely didn't adjust pension ages with the female SPA, as private pensions are rarely differentiated by gender[1], and this reform only affected women. Still, checking for correlation between SPA and private pension normal pension ages (NPAs) is desirable. Checking this directly in ELSA is complicated by the fact that only self-reported NPAs are available. For the SPA, where alongside self-reports, we know an individual's true SPA, these self-reported ages are unreliable, as is documented in main text Section 4.2. However, only defined benefit pension systems have NPAs, as defined contribution schemes can be accessed from age 55. Hence, dropping those with > £2,000 in DB pensions removes any unlikely SPA–NPA correlation from explaining the results. Table V shows that, despite reduced power, the treatment effect remains significant.

[1]This is likely illegal due to anti-discrimination law. The 2012 ECJ Test-Achats ruling barred gender-based pricing in insurance.

TABLE IV

PLACEBO TESTS

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

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

TABLE V

EFFECT OF SPA ON HAZARD RATE:
LESS THAN £2,000 IN DB SCHEME

Above SPA	**0.180**
s.e	(0.0458)
Above SPA $\times$(NHNBW.>Med.)	**-0.088**
s.e	(0.0612)
Obs.	5,668

Note: Table shows results of repeating regression from Column (4) of Table I in the main text on population with above £2,000 in DB wealth.

Turning to joint retirement Table VI, repeat the analysis from the main text but only for single women and those with non-working husbands. The patterns are qualitatively similar, but we can no longer rule out the treatment effect amongst the two subgroups being different from zero due to the reduced sample size. Crucially for the argument of this paper, the treatment effect in the subgroup is not significantly different from the treatment effect in the whole population.

As mentioned in the main text age, age-based mandatory retirement is illegal, and as discussed at the start of this section, this is interpreted strictly by the courts. It is still possible that firms illegally force people to retire. To address this possibility, Table VII drops all women who self-report having been forced out of their last job. Given the small numbers who self-report having been dismissed, the results do not change significantly.

TABLE VI

EFFECT OF SPA ON HAZARD RATE: SINGLES AND NON-WORKING HUSBANDS

	(1)	(2)	(3)	(4)
Above SPA	**0.096**	**0.073**	**0.099**	**0.113**
s.e	(0.03788)	(0.04855)	(0.05523)	(0.03832)
Above SPA$\times$(VLA.>Med.)			**-0.026**	
s.e			(0.07366)	
Above SPA$\times$ VLA.				**-1.58×10^{-7}**
s.e				(4.10e-08)
Obs.	3,007	1,722	3,007	2,952

Note: Repeat first four columns of Table I from the main text on the subsample of singles and women with non-working husbands.

TABLE VII

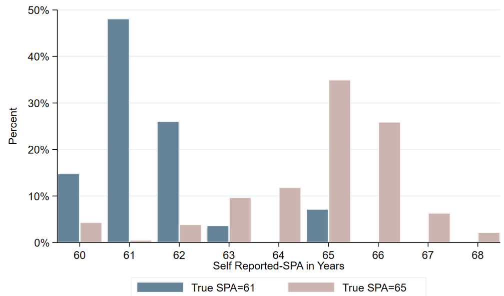
EFFECT OF SPA ON HAZARD RATE: EXCLUDING SELF-REPORTED FIRED

	(1)	(2)	(3)	(4)
Above SPA	**0.129**	**0.104**	**0.160**	**0.145**
s.e	(0.02423)	(0.03086)	(0.03750)	(0.02451)
Above SPA$\times$(VLA.>Med.)			**-0.057**	
s.e			(0.04849)	
Above SPA$\times$ VLA.				**-1.15×10^{-7}**
s.e				(2.67e-08)
Obs.	7,799	3,738	7,799	7,676

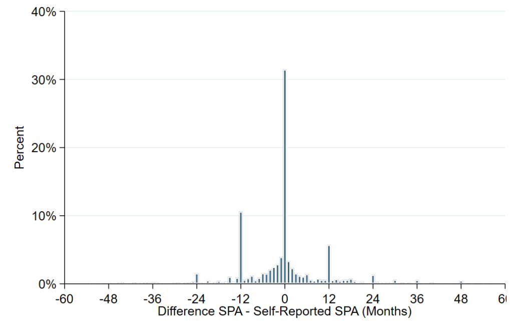
Note: Repeat first four columns of Table I from the main text excluding women that self-report being fired.

A.4. *Descriptive Analysis of SPA Beliefs*

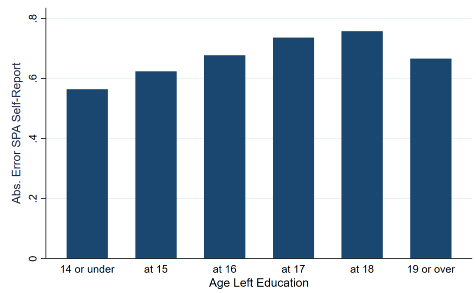
Mistaken beliefs could take on many forms. People could simply not update from the pre-reform SPA of 60 or might cling to other salient numbers like the male SPA of 65. To explore this, Figure 2a shows reported SPAs for two cohorts, with true SPAs of 61 and 65.



(a) SPA Beliefs by SPA-cohort



(b) Monthly SPA Misbeliefs Age 58



(c) Age 58 SPA Misbeliefs by Education

Note: Panel (a): self-Perceived SPA for two SPA-cohorts. One with a rounded SPA of 61 and one with a rounded SPA of 65. Panel (b): plot of error in self-reported SPA. The graph shows the frequency by which respondents gave mistaken answers about their SPA with errors at the true monthly level of SPA variation. Panel (c): SPA misbeliefs at age 58 by education.

While there's minor clustering around salient ages, reports mainly center on the true SPA, resembling a noisy signal. This is consistent with a model of costly information acquisition.

Figure 2b shows self-reported SPA errors at age 58 using monthly bins, rather than the yearly ones in the main text. Little of model relevance is gained from this, 31% report their SPA to the exact month. The main new insight is the spike in errors of 12 months.

Figure 2c shows SPA misbeliefs at age 58 by education. These rise with age left school until the 19 years or over category, suggesting more educated people (up to that point) are more mistaken. On the one hand, this is surprising as we expect more educated people to have a higher information processing capacity. On the other, the State Pension matters more for less educated individuals, giving them stronger incentives to learn. Thus, the pattern supports the modeling choices to focus on incentive heterogeneity rather than on ex-ante attention cost heterogeneity.

A.5. *Other Pension Belief and Knowledge Questions*

ELSA includes more data on beliefs about the State Pension and awareness of the reform than just SPA beliefs. I briefly discuss two question sets that seem highly relevant but are ultimately less so.

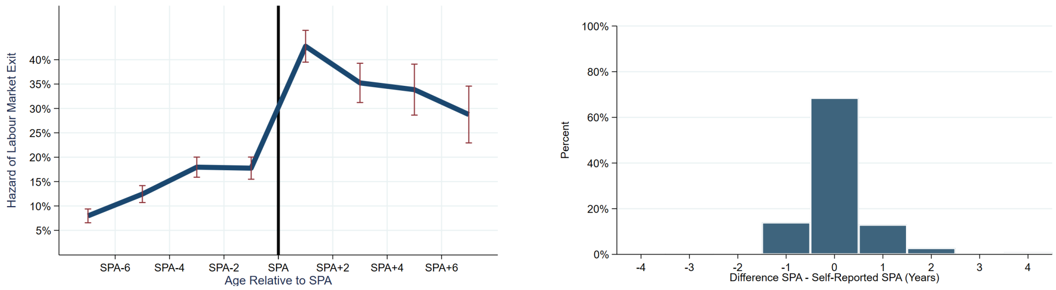
From wave 3, ELSA asked if individuals were aware of the female SPA reform. Interestingly, a total lack of awareness of the reform does not drive SPA misbeliefs, with only 6.62% reported being unaware. While the unaware were more misinformed on average (mean error at age 58 of 1.4 vs. 0.9 years), the error distributions overlap. Moreover, dropping the unaware 6.62% does not materially change the patterns. Thus, I conclude that total unawareness is less informative than a nuanced view, allowing for partial misinformation, as per the main text.

During a single wave (wave 3), ELSA collected subjective probability distributions on the level of pension benefits, but, as this was a single wave, using this data loses the panel dimension. Additionally, as those below SPA were asked these questions, the number of observations is very small.: 548 reported upper and lower bounds on expected State Pension income, and just 221 provided probabilities. Moreover, the complexity of the benefit formula makes identifying mistakes harder than with SPA beliefs. While we cannot observe mistaken beliefs directly, the narrowing range of responses as people near SPA mirrors the decline in mean squared error in SPA reporting. Average expected income range drops from £14.48 at age 55 to £6.39 at 59. Given the small sample size, the difficulty in computing true entitlements, and the computational difficulties of including two sources of pension uncertainty in the model, I focus on SPA beliefs in this project.

A.6. *Men: Misbeliefs and Employment around SPA*

Due to the lack of policy variation, the employment response to SPA cannot be causally estimated for men. Thus, the main text focuses on how misbeliefs affect women's employment response. That does not mean that similar mechanisms are not at play for men.

Figure 3a shows a similar jump in men's hazard rate at SPA. While it is not possible to separate the SPA effects from aging, it is notable that the jump also occurs for men at SPA. Figure 3b shows mistaken beliefs for men at age 58. Despite no SPA reform and the male SPA unchanged since 1948, nearly 40% didn't know their SPA within a year at age



(a) Fraction exiting labour employment - Men (b) Mistaken SPA Beliefs of Men at Age 58

Note: Panel (a): pooled average faction exiting employment market at ages relative to the SPA. Data was plotted at two yearly intervals due to the biennial frequency of ELSA waves. Panel (b): plot of error in self-reported State Pension Age (SPA). The graph shows the frequency by which respondents gave mistaken answers about their SPA, with errors binned at the yearly level.

58. Though lower than the 60% for women, it supports the idea that misbeliefs are relevant in the absence of reform. If attention is costly, the mere possibility of reform could lead to mistaken beliefs. Thus, this evidence is consistent with the paper's proposed mechanism.

APPENDIX B: ADDITIONAL MATHEMATICAL DETAILS

B.1. *Finding Unique Actions Using Second Order Conditions*

Using the Kuhn-Tucker conditions of Equation 13 from the main text Caplin et al. (2019) provide an alternative formulation of the solution of the model. If the CCP satisfy Equation 12 from the main text and for all possible actions ($\forall d = (c, l) \in \mathcal{C}$)

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

with equality if $q_t(d) > 0$, then the CCPs solve the model. This new condition from (Caplin et al., 2019) replaces the need for the unconditional choice probabilities to solve the log-sum-exp of Equation 13 from the main text.

If an action $d^\star = (c^\star, l^\star)$ satisfies Equation 14 from the main text repeated here:

$$\sum_{spa} \pi_t(spa) \frac{\exp\left(n^{(k)} \frac{((c/n^{(k)})^\nu l^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \beta \overline{V}_{t+1}^{(k)}(d, X_t, spa, \underline{\pi_t})\right)}{\exp\left(n^{(k)} \frac{((c^\star/n^{(k)})^\nu l^{\star 1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \beta \overline{V}_{t+1}^{(k)}(d^\star, X_t, spa, \underline{\pi_t})\right)} < 1, \tag{2}$$

for all $d = (c, l) \in \mathcal{C}$. That is, $d^\star$ produces such a high utility in all states that, in expectation, the exponentiated utility of any other payoff divided by its exponentiated utility is below 1.

If such a $d^\star$ exists then it automatcally statisfies 1 to have $q_t(d^\star) = 1$, because substituting $q_t(d^\star) = 1$ into 1 yeilds 14 from the main text with an non-binding constraint.

APPENDIX C: ADDITIONAL COMPUTATIONAL DETAILS

C.1. *Solving the Models without Costly Attention*

The models are solved by backward induction starting at age 101 when the household dies with certainty. The household problem is modeled as a discrete choice. When rational inattention does not complicate this within-period discrete choice optimization, it is solved by grid search, selecting the value that maximizes the household's utility. States are discretized with 30 grid points for assets (a_t), 4 for average earnings ($AIME_t$), 5 for wages (w_t), two for the unemployment shock (ue_t), and in the model with policy uncertainty the state pension age (SPA_t) has 8 gris points as it ranges from 60 to 67.

A finer grid of 500 points is offered to the household when making their saving choice. This keeps the size of the state space manageable whilst not unduly constraining households and is equivalent to having a finer grid for consumption than for assets. When evaluating continuation values of off-grid values, I use linear interpolation of the value function.

C.2. *Solving the Models with Costly Attention*

Belief Distribution Costly attention introduces a high-dimensional state variable: the belief distribution ($\underline{\pi_t}$). To discretize it, I consider all ways of reallocating fixed-size probability masses across the eight possible SPAs (60–67). Since Bayesian updating cannot shift probability from zero, I want to avoid having beliefs assigning zero weight to any SPA in the belief grid. So, each SPA gets a minimum probability of 0.01, with the movable masses allocated on top.

Specifically, I use four movable probability masses. In the absence of the minimum probability requirement, each mass would be 0.25. With the minimum probability requirement, the size of the movable masses changes as the support of SPA_t changes. For example, in periods where all eight SPAs are possible (because $t < 60$ and the women have not aged past any possible SPA), these probability masses are of the size $\frac{1-0.08}{4} = 0.23$. These four masses distributed over eight SPAs yield $\binom{7+4}{4} = 330$ grid points (because each combination can be thought of as an ordering of the four masses and the seven breaks between the eight grid points). As individuals age and fewer SPAs remain, the grid shrinks—e.g., to $\binom{1+4}{4} = 5$ at $t = 65$ when only SPAs of 66 and 67 are possible. With no natural ordering over Δ^7, I cycle through combinations lexicographically. As robustness increases, I increase the number of movable probability masses to five, finding that it does not materially change results.

High Dimensional Interpolation When the prior with which a household starts the next period is off this grid, I use k-nearest neighbor inverse distance weighting to do multi-dimensional interpolation. I use the difference in means between the distributions as an approximation to the Wasserstein, or earth mover, metric as the distance concept in the inverse distance weighting. High-dimensional interpolation is computationally intensive and prone to approximation error. To mitigate this, I start with the two nearest grid points; if the fixed point loop for the unconditional choice probabilities (q_t) fails to converge within 25 iterations, I incrementally increase the number of neighbors up to a maximum of $2^8 = 256$.

Range of Attention Costs When rational inattention matters because $t < SPA_t$ the main equation to solve to find the household's optimal decision is:

$$\max_{\underline{q_t}} \sum_{spa} \pi_t(spa) \log\left(\sum_{d' \in \mathcal{C}} q_t(d) \exp\left(n^{(k)} \frac{((c/n^{(k)})^{\nu} l^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \beta \overline{V}^{(k)}_{t+1}(d, X_t, SPA_t, \underline{\pi_t})\right)\right) \quad (3)$$

Following the random utility literature, I normalize the payoff in this equation by dividing it by the highest payoff across SPAs. However, the presence of λ complicates exponentiation. While data—not computation—should guide $lambda$'s value, its role as a denominator in the exponent causes exponentiated payoffs to diverge as λ shrinks, but these vanishingly small $lambda$'s values cannot be ignored. Since earlier SPAs are preferred, terms tied to SPA=60 are larger, and lower attention costs amplify these differences. Still, very small exponentiated payoffs associated with high SPAs when $\lambda is small$ cannot be ignored: as $\lambda \to 0$, $\log()$ terms diverge to $-\infty$, and their gradients explode. Thus, even tiny exponentiated payoffs materially affect the objective function. Given this and the small attention costs implied by belief data, I carefully optimized code to retain very small utility values rather than dropping them—as might be acceptable with other objective functions. I use quadruple precision floating-point numbers to store the utility values (min value $\sim 10^{-4965}$), but since compilers are optimized for double precision, this greatly slows computation. So, I resort to quadruple precision only when necessary, checking first whether normalization causes underflow in double precision.

Solving the within-period problem Culling actions that are never chosen is central to the solution method. One of the two ways this is done is by dropping strictly dominated actions (for the other, see Section B.1). While identifying strictly dominated actions is an interesting problem studied in computer science (Kalyvas and Tzouramanis, 2017), the choice set here is modest (max 1,500 resulting from 3 labor and 500 asset options), so a simple Block Nested Loop algorithm is most efficient. When culling alone does not yield a solution, I solve Equation 3 using sequential quadratic programming (Schittkowski, 2014).

High-level Pseudo Code

1: Remove d from choice set $\mathcal{C}$ that are strictly dominated across all possible combinations of SPA_t and $\underline{\pi_{t+1}}$
2: **if** $|\mathcal{C}| = 1$ **then**
3: Set $\underline{q_t}$ to degenerate distribution at unique $d \in \mathcal{C}$
4: **else**
5: Set initial value of $\underline{\tilde{q}_t}$ and Error > Tolerance
6: **while** Error > Tolerance **do**
7: Solve for $\overline{V}_{t+1}$ (Equation 11 from the main text) given $\underline{\tilde{q}_t}$
8: Remove d from $\mathcal{C}$ that are strictly dominated across all possible SPA_t given $\overline{V}_{t+1}$
9: **if** $|\mathcal{C}| = 1$ **then**

```
                Set Error = 0 < Tolerance and q_t to degenerate distribution at d ∈ |C|
            else
                if there is an action d that satisfies 14 from the main text then
                    Set Error = 0 < Tolerance and q_t to degenerate distribution at d
                else
                    Solve 13 from the main text using sequential quadratic programming for q_t
                    Set Error to distance between q_t and q̃_t
                    Update q̃_t = q_t
                end if
            end if
        end while
    end if
    Substitute q_t into 12 from the main text to solve for p_t.
```

C.3. *Simulating and Estimating*

My simulated sample consists of 50,000 randomly drawn individuals aged 55. Since the simulated sample exceeds the data size, state variables initialized directly from the empirical distribution (assets and average lifetime earnings) are sampled multiple times using random Monte Carlo draws from their joint distribution. I initialize wages with draws from its estimated distribution. I simulate one SPA cohort at a time, setting SPA_t to match the cohort's SPA. I assume the SPA response reflects draws from the individual prior belief distributions, with every one of the same type starting at age 55 with identical beliefs. Thus, I initialize beliefs using the type-specific distribution of SPA point estimates.

Given these initial conditions, I simulate the choice of the individual households using the decision rule found when solving the model and the exogenous process estimate in the first stage. I aggregate the simulated data in the same way as with observed data to construct the moment conditions, detailed in Appendix D. The method of simulated moments estimates model parameters by minimizing a GMM criterion, also described in Appendix D. To minimize the objective function, I first sample the parameter space via Sobol sequencing, then apply the BOBYQA routine (Powell, 2009) at promising starting points.

APPENDIX D: ADDITIONAL ECONOMETRIC DETAILS

D.1. *Imputing AIME*

Average lifetime earnings are observed only for women present in wave 3 who consented to link their National Insurance records. To initialize the model from the joint distribution of $AIME_{55}$ and a_{55} without introducing selection into a_{55}, I impute missing values. I first regress $AIME_{55}$ on a quintic in NHNBW and a rich set of controls, including variables on health, education, location, labor market behavior, housing tenure, cohort, age, wage, and cognitive ability.

Using predicted values alone would overstate the correlation between $AIME_{55}$ and a_{55}, so I add noise to the imputations to match observed heteroscedasticity. I regress non-imputed $AIME_{55}$ on a quintic in NHNBW (excluding controls, as they're absent in the model), then regress the squared residuals on the same polynomial. Since imputed $AIME_{55}$ is homoscedastic by construction, adding noise with variance from the second regression replicates the observed heteroscedasticity.

D.2. *Type-specific Mortality*

Life expectancy heterogeneity affects older individuals' behavior (e.g. De Nardi et al., 2009), but death is often poorly recorded in surveys. To address this, I estimate type-specific mortality without relying on ELSA death records, instead combining them with ONS survival tables following French (2005). I do this using Bayes' rule:

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

Where $Pr(type = k|death_t)$ and $Pr(type = k)$ are taken from ELSA and $Pr(death_t)$ are taken from the ONS life-tables. If measurement error affects all types equally, estimates of $Pr(type = k|death_t)$ from ELSA are unbiased, unlike those of $Pr(death_t|type = k)$, and deal with the measurement error issue.

D.3. *Generating Profiles*

To avoid contamination by cohort effects or macroeconomic circumstances, targetted profiles were generated with a fixed effect age regression, which included: year of birth effects, SPA-cohort specific age effects, the aggregate unemployment rate rounded to half a percentage point, and an indicator of being below the SPA. Specifically, the following regression equation was estimated:

$$y_{it} = U_t + \sum_{c\in C} \gamma_c \mathbb{1}[cohort_i = c] + \sum_{s\in S} \mathbb{1}[SPA_i = s](\sum_{a\in A} \delta_{a,s} \mathbb{1}[age_{it} = a])$$

where $cohort_i$ is the year-of-birth cohort of an individual, SPA_i is her final SPA, $age_{i,t}$ her age in years, U_t aggregate unemployment to half a percent, and the outcome variable y_{it} is either assets or employment depending on which profile is being calculated. The profiles targetted were then predicted from these regressions using average values for the pre-reform cohorts.

APPENDIX E: ADDITIONAL RESULTS

E.1. *First Stage Estimates*

Model Types A woman is classified as highly educated if she exceeds the compulsory schooling for her generation. She is considered married if married or cohabiting, as household structure matters more than legal status for the questions considered in this paper. As the model abstracts from separation, any woman ever observed as married is treated as married in all periods. This accounts for the likely receipt of alimony or a widow's pension, making 'married' the most model-consistent classification for previously married women. The resulting type shares are: 34% married/low education, 11% single/low education, 44% married/high education, and 11% single/high education.

Initial conditions Initial assets a_{55} and average earnings $AIME_{55}$ are drawn from the type-specific empirical joint distribution (summary statistics in Table VIII). As expected for this generation, married women have weaker labor market attachment, resulting in lower $AIME_{55}$ but higher household assets. Higher education raises both variables.

Labour market conditions Type-specific transition probabilities—estimated by classifying individuals as unemployed when claiming benefits—are shown in Table IX. Parameters of the stochastic wage component (persistence, innovation variance, measurement error, and initial draw) appear in Table X. The deterministic wage component generates the profiles in Figure 4a. Spousal income is shown in Figure 4b

TABLE VIII

SUMMARY STATISTICS OF INITIAL CONDITIONS (£)

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

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

TABLE IX

TYPE SPECIFIC UNEMPLOYMENT TRANSITION PROBABILITIES

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

Note: Unemployment and reemployment transition probabilities.

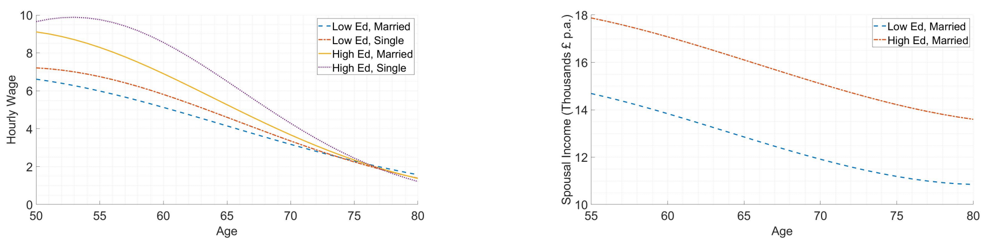
Social Insurance As noted in the main text, State Pension income differs more by marital status than by education. Among State pension claimants, high-education women re-

TABLE X

PARAMETERS OF THE STOCHASTIC COMPONENT OF THE WAGES

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

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



(a) Wage Profiles

(b) Spousal Income

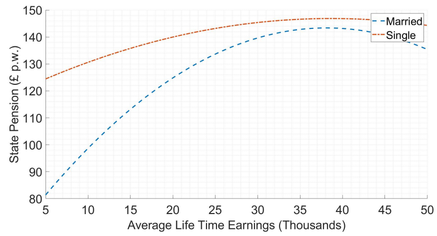
Note: Panel (a): the deterministic component of female hourly wages for the four model types plotted against female age. Panel(b): spousal income plotted against female age.

ceive £92.52 on average, low-education £87.11, while single women receive £112.50 and married women £80.89. To capture this key distinction and maximize power, I restrict State Pension heterogeneity to marital status only. The resulting functions of average lifetime earnings are shown in Figure 5a."

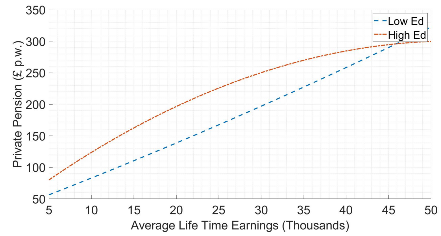
Conversely, private pension income varies more by education than by marital status. Among those with non-zero private pension income, high-education women receive £118.50 on average, low-education £61.42, while single women receive £100.78 and married women £94.24. To capture this key difference and maximize power, I restrict private pension heterogeneity to education only. The resulting functions are in Figure 5b.

E.2. *Model Fit*

As mentioned in the main text, although the different model specifications have different predictions about the labor supply response to the dynamic SPA, the static profiles are not

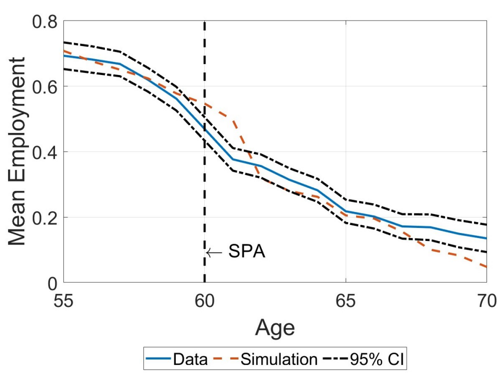


(a) State Pension Function

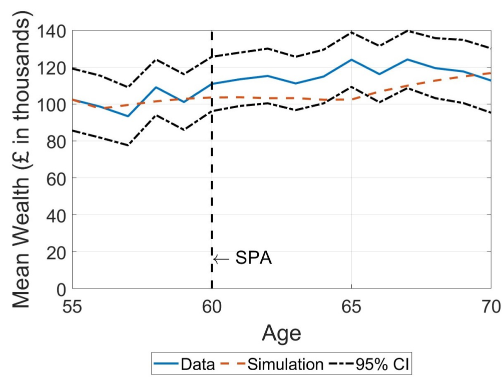


(b) Private Pension Function

Note: Panel (a): state Pension income as a function of average lifetime earnings (AIME) for married and single women. Panel(b): Private Pension income as a function of average lifetime earnings (AIME) for high and low-education women.



(a) Employment Profile Baseline



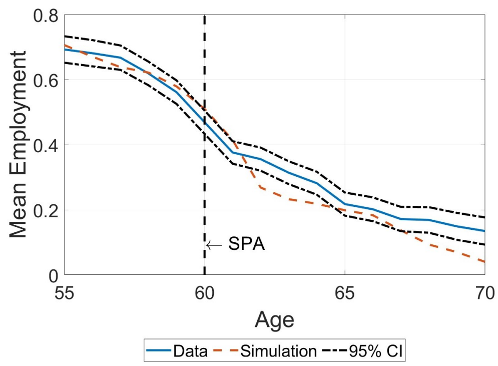
(b) Asset Profile Baseline

Note: Panel (a): model fit to targeted labor supply profile. The empirical profile is for the pre-reform SPA cohort with a SPA of 60. The model was simulated with an unchanging SPA of 60, mimicking the conditions faced by this cohort. Panel (a): model fit to targeted asset profile. The empirical profile is for the pre-reform SPA cohort with a SPA of 60. The model was simulated with an unchanging SPA of 60, mimicking the conditions faced by this cohort.

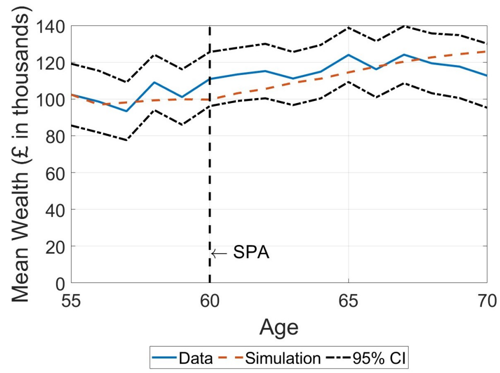
very sensitive to model specifications. All versions are able to match the static profiles. Figures 6a-7b show the employment and asset profiles for the baseline and the version with rational inattention with the parameter estimates of Table 3d from the main text.

E.3. *Reference Point Retirement Literature*

Seibold (2021) supports reference dependence through a process of elimination, ruling out alternatives. He rejects misbeliefs as an explanation, on the basis that less-educated individuals, who he argues are more prone to confusion, show a smaller employment response at eligibility. While they likely have higher processing costs, they may also be more incen-



(a) Employment Profile Model with Rational Inattention



(b) Asset Profile Model with Rational Inattention

Note: Panel (a): model fit to targeted labor supply profile. The empirical profile is for the pre-reform SPA cohort with a SPA of 60. The model was simulated with an unchanging SPA of 60, mimicking the conditions faced by this cohort. Panel (b): model fit to targeted asset profile. The empirical profile is for the pre-reform SPA cohort with a SPA of 60. The model was simulated with an unchanging SPA of 60, mimicking the conditions faced by this cohort.

TABLE XI

SUMMARY STATISTICS OF ATTENTION COST CONVERTED TO COMPENSATING ASSETS (£)

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

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

TABLE XII

IMPACTS OF REFORMING SPA WITH INFORMED AND UNINFORMED HOUSEHOLDS (FRACTION OF PASSIVE HOUSEHOLDS)

SPA increased from 60 to:	(1) - Informed Added Employment	(2) - Uninformed Added Employment
61	0.16	0.13
62	0.30	0.26
63	0.39	0.33
64	0.48	0.41
65	0.62	0.52

Note: Results of raising SPA from 60 to the age in Column (1) with costly attention and in Column (2) without it.

tivized to learn about this particular issue, dedicating more resources to it (something my model implies and belief data supports). Lalive et al. (2023) provide survey evidence in support of reference dependence, finding that eligibility is the main reason for stopping work and that many claim benefits simply because "it seems natural." Mapping survey responses to model construct is difficult, however. Eligibility could be interpreted as an implicit recommendation in the presence of costly attention, leading people to describe claiming at this age as natural. Gruber et al. (2022) presents compelling Finnish evidence: relabeling a pension age, despite minimal financial changes, caused significant employment shifts. On the one hand, this seems to strongly support reference dependence, yet many who retired due to relabeling later returned to work, which they interpret as suggesting regret. In contrast, inattention could explain both phenomena. Confusion about the relabeling prompts exit, while belief updates drive re-entry. As Gruber et al. (2022) note," since [information about optimal reitmrenet] is always attached to the statutory age itself, it is difficult to disentangle this effect empirically". I would add the caveat without gathering belief data.

E.4. *Introducing a Fraction of Passive Agents*

As a simple way of capturing a behavioral bias like reference dependence preference, I introduce a fraction of passive agents that retire at SPA but do not anticipate this fact, in the vein of (as in Chetty et al., 2014). I use this fraction to match the employment responses to the SPA of the whole population and the richer subgroup. I find that in the model with only policy uncertainty, I need 14% to be passive to match the data (treatment effect 0.119 and 0.108 for above median assets) but due to the better initial fit of the rationally inattentive version, only 10% (treatment effect 0.118 and 0.122 for above median assets)in that version of the model. Table XII shows employment responses to SPA increases in the two versions of the model with the different fraction of passive agents required to match the treatment effects. Since the difference between these two columns is not just being informed or not (because the fraction of passive agents changes, it does not make sense to analyze the impact of an information letter campaign as is done in the main text. In this table, we see that due to the mechanical effect of a fraction of passive agents, the additional employment from increasing the SPA is larger, but the relative difference between the two columns is similar to that found in the experiment without passive agents in the main text.

APPENDIX F: EXTENSION: DEFERRAL PUZZLE

Attributing all policy uncertainty to the stochastic State Pension Age (SPA) understates overall pension uncertainty. This section introduces uncertainty and learning about another key feature: the actuarial adjustment from deferring benefits. Combined with a claiming decision, this addition improves realism and helps explain the deferral puzzle (discussed below). Since the adjustment rate becomes irrelevant after claiming, rational inattention speaks directly to this puzzle. While deferral may appear actuarially favorable, this overlooks the benefit of claiming due to removing the need to monitor the adjustment rate. The model in Section 5.2 omits this mechanism for two reasons: it lacks a benefit-claiming decision and assumes SPA is the only uncertainty incurring attention costs, and once SPA is reached, this uncertainty resolves, irrespective of claiming. The rest of this section presents a simple extension that introduces this new incentive and its implications.

F.1. *Deferral Puzzle*

By deferral puzzle, I refer to the rarity of deferring state pension benefits despite highly generous terms between April 2005 and April 2016. Between those dates, benefits rose by 1% for every 5 weeks deferred—an annual adjustment of 10.4%. This is an extremely generous actuarial adjustment, and yet, 86.7% of women observed past SPA in ELSA during this period had claimed by their first post-SPA interview.

What exactly constitutes actuarially fair depends on life expectancy and the interest rate, but at all plausible levels, this adjustment was generous. For women reaching SPA during this period, life expectancy ranged from 23 to 25 years. Even using a conservative estimate of 23 years, a 10.4% annual adjustment was advantageous at interest rates up to 9%. The Bank of England base rate never exceeded 5.75% and was 0.5% from March 2009 onward. Thus, the 10.4% adjustment was actuarially favorable at any realistic rate.

Even among the few women who deferred, deferral durations were short. With a conservative life expectancy of 23 years and a 5.75% interest rate, the optimal deferral is 9 years. Yet, the median deferral was 2 years, and 99.54% claimed within 8 years.

These calculations use mean life expectancy, which masks heterogeneity. However, heterogeneity alone is not a plausible explanation. It would mean 86.7% of women had significantly below mean life expectancy, implying implausible skewness.

TABLE XIII

PARAMETER ESTIMATES - EXTENSION

ν: Consumption Weight	0.5310
β: Discount Factor	0.9852
γ: Relative Risk Aversion	2.0094
θ: Warm Glow bequest Weight	20,213

Note: Estimated parameters from method of simulated moments for the model extension with a stochastic deferral rate and a benefit claiming decision.

F.2. *Model and Estimation*

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

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

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

The model with policy uncertainty, to the stochastic SPA and adjustment rate, is then re-estimated to match the same pre-reform employment and assets profiles with a constant realization of 10.5% for the deferral adjustment, which was the deferral rate these cohorts faced. Parameter estimates are in Table XIII and, for these values, only 6.2% of individuals

TABLE XIV

MODEL PREDICTIONS - EXTENSION WITH BENEFIT CLAIMING AND UNCERTAIN DEFERRAL

	Costly Attention	Data
Population	Treatment Effect for being below SPA on employment	
Whole Population	0.0416	0.080
Assets >Median(£29,000)	0.0903	0.061
Age	Variance of SPA Answers	
55	2.985	2.852
58	1.795	1.180
Coefficient	Treatment Effect Heterogeneity by SPA Error	
Treatment Effect	0.0532	0.157
Interaction	-0.0111	-0.023

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

claim the state pension before the mandatory claiming age of 70, much lower than the 99% plus claiming seen in the data.

Next, I introduce costly attention with a cost of attention corresponding to approximately £10 of consumption to the median consuming household to be fully informed. This increased the number voluntarily claiming to 22.2%, approximately a fourfold increase on the model without informational frictions, but still short of the rate observed in the data. As can be seen in Table XIV, this cost of attention produced a relatively good fit along all dimensions of interest. Note that the treatment effect displayed is the effects of being below the SPA on the probability of being in employment, rather than being above the SPA on the hazard of exiting employment used in the main text.

APPENDIX: REFERENCES

BORUSYAK, KIRILL, XAVIER JARAVEL, AND JANN SPIESS (2024): "Revisiting event-study designs: robust and efficient estimation," *Review of Economic Studies*, rdae007. [3]

BOZIO, ANTOINE, ROWENA CRAWFORD, AND GEMMA TETLOW (2010): "The history of state pensions in the UK: 1948 to 2010," Tech. rep. [23]

CAPLIN, ANDREW, MARK DEAN, AND JOHN LEAHY (2019): "Rational Inattention, Optimal Consideration Sets, and Stochastic Choice," *Review of Economic Studies*, 86 (3), 1061–1094. [10]

CESARINI, DAVID, ERIK LINDQVIST, MATTHEW J NOTOWIDIGDO, AND ROBERT ÖSTLING (2017): "The effect of wealth on individual and household labor supply: evidence from Swedish lotteries," *American Economic Review*, 107 (12), 3917–46. [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. [21]

DE NARDI, MARIACRISTINA, ERIC FRENCH, AND JOHN BAILEY JONES (2009): "Life expectancy and old age savings," *American Economic Review*, 99 (2), 110–15. [15]

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. [5]

FRENCH, ERIC (2005): "The Effects of Health, Wealth, and Wages on Labour Supply and Retirement Behavior," *Review of Economic Studies*, 72 (2), 395–427. [15]

GRUBER, JONATHAN, OHTO KANNINEN, AND TERHI RAVASKA (2022): "Relabeling, retirement and regret," *Journal of Public Economics*, 211, 104677. [21]

KALYVAS, CHRISTOS AND THEODOROS TZOURAMANIS (2017): "A survey of skyline query processing," *arXiv preprint arXiv:1704.01788*. [13]

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. [21]

POWELL, MICHAEL JD (2009): "The BOBYQA algorithm for bound constrained optimization without derivatives," *Cambridge NA Report NA2009/06, University of Cambridge, Cambridge*, 26. [14]

SCHITTKOWSKI, K (2014): "NLPQLP-Nonlinear Programming with Non-Monotone and Distributed Line Search," . [13]

SEIBOLD, ARTHUR (2021): "Reference points for retirement behavior: Evidence from german pension discontinuities," *American Economic Review*, 11 (4), 1126–1165. [19]