Online Appendix for Costly Attention and Retirement

A Additional Empirical Details

A.1 Additional Institutional Details

A.2 Equity Acts

The equality Act (2006) banned mandatory retirement below age 65. Since all women observed to age past their SPAs in ELSA waves 1-7 had a SPA of 60-63, it would have been illegal for their SPAs to have coincided with a compulsory retirement age. The equality Act (2010) went further and banned all compulsory retirement ages with a handful of specific exceptions known as employer justified retirement ages (EJRA).

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

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

Table 1: 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)
p=	.000	.000	.001	.000
Above SPA×(VLA.>Med.)			**-0.007**	
s.e			(0.0496)	
p=			.882	
Above SPA× VLA.				**-1.23×10^{-7}**
s.e				(3.30e-08)
p=				.000
Obs.	7,907	3,691	7,907	7,784

Notes: Column (1) shows the results of running the two-way fixed effect specification in 1 from the main text with controls used: a full set of marriage status, years of education, education qualifications, and self-reported health dummies; partners age; partners age squared; the aggregate unemployment rate during the quarter of interview; dummies for partner eligible for SPA, and for being one and two years above and above SPA; and assets of the household. Column (2) repeats this regression on the subsample with above median Very Liquid Assets (VLA) in the last interview before their SPA. Column(3) tests whether the different treatment effects observed in columns (1) and (2) are different by introducing an interaction between being above the SPA and having above median VLA. Column(4) includes an interaction between being above SPA and a continuous measure of VLA.

A.3 Excess Employment Sensitivity

A.3.1 Restricted Asset Categorisation

The goal of investigating treatment effect heterogeneity by asset holdings is to understand the role played by liquidity constraints. The main text is restricted to NHNBW. However, parts of NHNBW can be illiquid, and so in Table 1 repeat the analysis but for a more restricted asset category, very liquid asset, which is only assets that can reasonably be liquidated in a matter of weeks. As can be seen, the results are qualitatively very similar to those using NHNBW and do not support liquidity constraints alone explaining away the treatment effect. The treatment effect for those with above median assets is still positive, the difference between the two subgroups remains insignificant, and the continuous interaction specification shows that this heterogeneity is too weak for the treatment to be completely explained by liquidity constraints.

A.3.2 Bad Control Concerns

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

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

Table 2 shows the results of this exercise of dropping controls. As can be seen, the results are very little

Table 2: 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)
p=	.000	.004	.000	.000
Above SPA×(NHNBW.>Med.)			**-0.068**	
s.e			(0.04868)	
p=			.164	
Above SPA× NHNBW.				**-8.26×10^{-8}**
s.e				(2.32e-08)
p=				.001
Obs.	8,119	3,963	8,119	7,898

Notes: Column (1) shows the results of running the two-way fixed effect specification in 1 from the main textwith no controls used. Column (2) repeats this regression on the subsample with above median NonHhousing Non-Business Wealth (NHNBW) in the last interview before their SPA. Column(3) tests whether the different treatment effects observed in columns (1) and (2) are different by introducing an interaction between being above the SPA and having above median NHNBW. Column(4) includes an interaction between being above SPA and a continuous measure of NHNBW.

changed from those with controls.

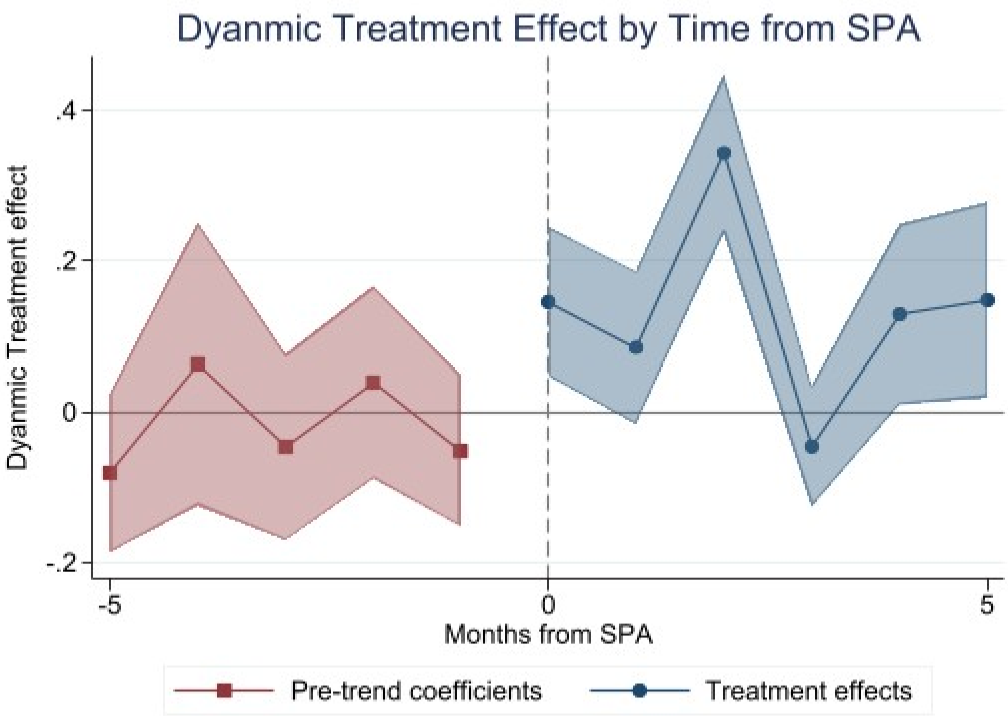
A.3.3 Imputation Approach to DID

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

Here I implement the imputation approach of Borusyak et al. (2024). Figure 1 shows the dynamic treatment effects before and after the SPA. There is no indication of violated parallel trends or anticipation effects as none of the pre-SPA treatment effects are significantly different from zero. Indeed jointly testing for violations of parallel trends fails to reject the null of parallel trends ($p = .799$). Conversely, 4 of 6 post-SPA treatment effects are individually significant, and we can easily reject the null ($p = .000$) of them being jointly zero. The graph also doesn't provide much indication that the post-SPA treatment effects differ from each other, although we can reject that hypothesis ($p = .198$).

Figure 2 looks at whether these individual treatment effects vary between waves. The treatment effects look quite uniform across waves, although, again, we can reject this hypothesis ($p = .137$). However, neither violation of homogeneity seems to SEVIERE, and generally, the graphs look supportive of the interpretation of a homogenous treatment effect that turns on at the SPA (as assumed in the baseline), although the statistical test show this is only an approximation.

Figure 1: Dynamic Treatment Effects by Time from SPA



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

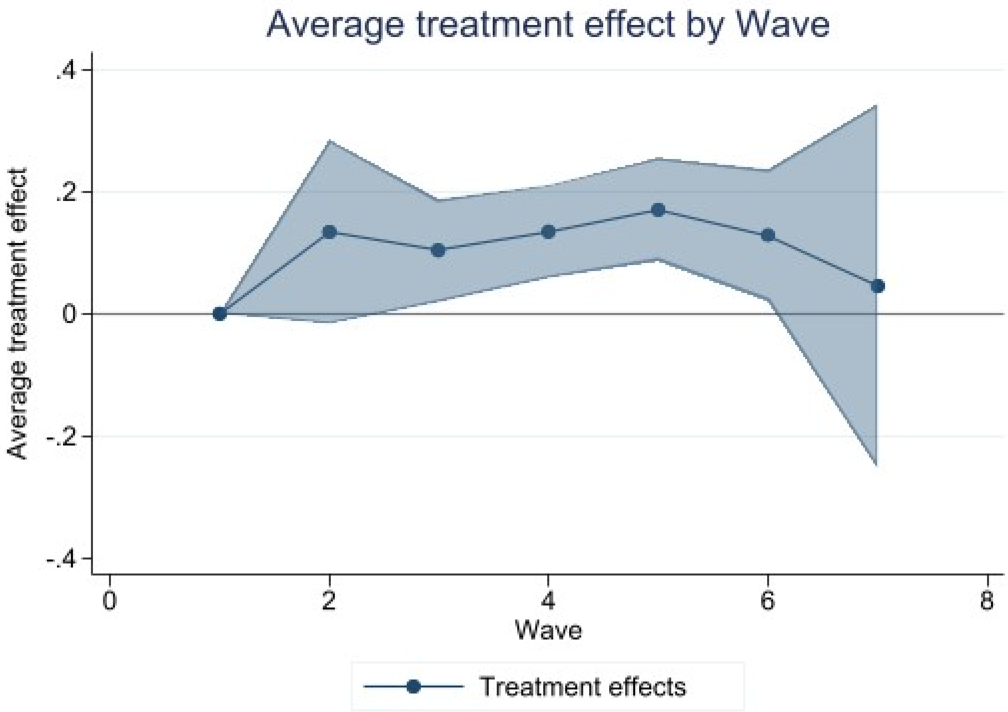
If you are more concerned about the violations of homogeneous treatment effects, then these results show that even allowing for arbitrary heterogeneity, there is something special happening at the SPA which is difficult to explain in standard complete information models.

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

The rest of the section is concerned with addressing other potential explanations for the sensitivity of employment to the SPA in a standard complete information framework. Specifically, I consider if wealth, health, private pensions, joint retirement, or dismissals can explain the labour supply response to the SPA.

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

Figure 2: Average Treatment Effect by Wave



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

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

Health is a major determinant of retirement behaviour (e.g. De Nardi et al., 2010). However, there is no reason to suspect it interacts with the SPA, so no reason for it to explain employment's sensitivity to the SPA. Furthermore, during the period studied, the SPA was in the range 60-63, and, at the mean, health status does not start to deteriorate until later in life. This can be seen in Figure 3, which shows the age profile of health status. All the same, as it is such an important factor in retirement Table 3 looks at heterogeneity in labour supply response to the SPA by health status. As can be seen, the labour supply response is only

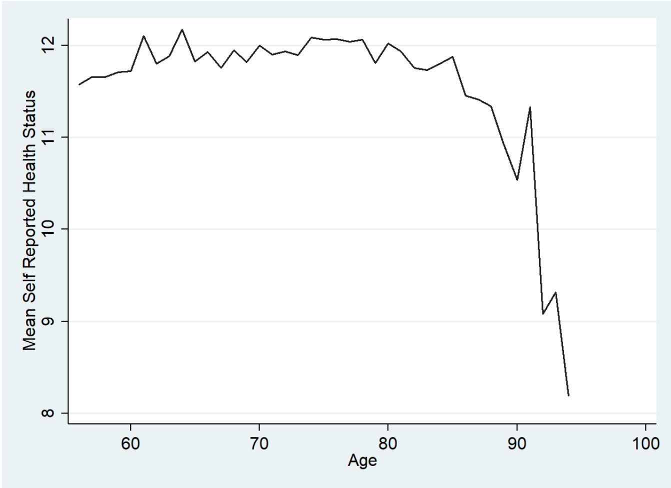


Figure 3: Self Reported Health Profile

Table 3: 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

significantly different for those with the poorest health group. This group only make up <7 % of the sample, and if dropped, do not qualitatively change the results.

The timing of private pension eligibility is important for retirement choices. However, occupational pension schemes are very unlikely to have adjusted their pension ages in line with the female SPA because private pensions do not generally offer different eligibility ages to men and women[1], and this reform only changed the female SPA. Still, checking for a correlation between the SPA and normal pension ages (NPA) of private pension schemes would be desirable. Checking this directly in ELSA is complicated by the fact that only self-reported NPAs are available. For the SPA, where alongside self-reports, we know an individual's true SPA, these self-reported ages are unreliable, as is documented in main text Section 5.2. However, only defined benefit pension systems have NPAs, as defined contribution schemes can be accessed from age 55. Hence, dropping everyone with over £2,000 in a defined benefit scheme from the sample rules out an unlikely correlation between the female SPA and pension schemes NPAs from explaining the results. This is done in Table 4, and as can be seen, despite the loss of power, the treatment remains present and significant.

Turning to joint retirement Table 5, 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

[1] Indeed it is likely to be illegal to do so on the grounds of that it would be discriminatory. For example, the 2012 European Court of Justice ruling known as Test-Achats explicitly outlawed charging men and women differently for the same insurance.

Table 4: Effect of SPA on Hazard Rate:
Less than £2,000 in DB scheme

Above SPA		**0.180**
	s.e	(0.0458)
	p=	.000
Above SPA×(NHNBW.>Med.)		**-0.088**
	s.e	(0.0612)
	p=	.151
Obs.		5,668

Table 5: Effect of SPA on Hazard Rate: Singles and Nonworking 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)
	p=	.012	.139	.075	.004
Above SPA×(VLA.>Med.)				**-0.026**	
	s.e			(0.07366)	
	p=			.723	
Above SPA× VLA.					**-1.58×10^{-7}**
	s.e				(4.10e-08)
	p=				.000
Obs.		3,007	1,722	3,007	2,952

Notes: Column (1) shows the results of running the two-way fixed effect specification in 1 in the main text with controls used: a full set of marriage status, years of education, education qualifications, and self-reported health dummies; partners age; partners age squared; the aggregate unemployment rate during the quarter of interview; dummies for partner eligible for SPA, and for being one and two years above and above SPA; and assets of the household. Column (2) repeats this regression on the subsample with above median NHNBW in the last interview before their SPA. Column(3) tests whether the different treatment effects observed in columns (1) and (2) are different by introducing an interaction between being above the SPA and having above median NHNBW. Column(4) includes an interaction between being above SPA and a continuous measure of NHNBW.

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.

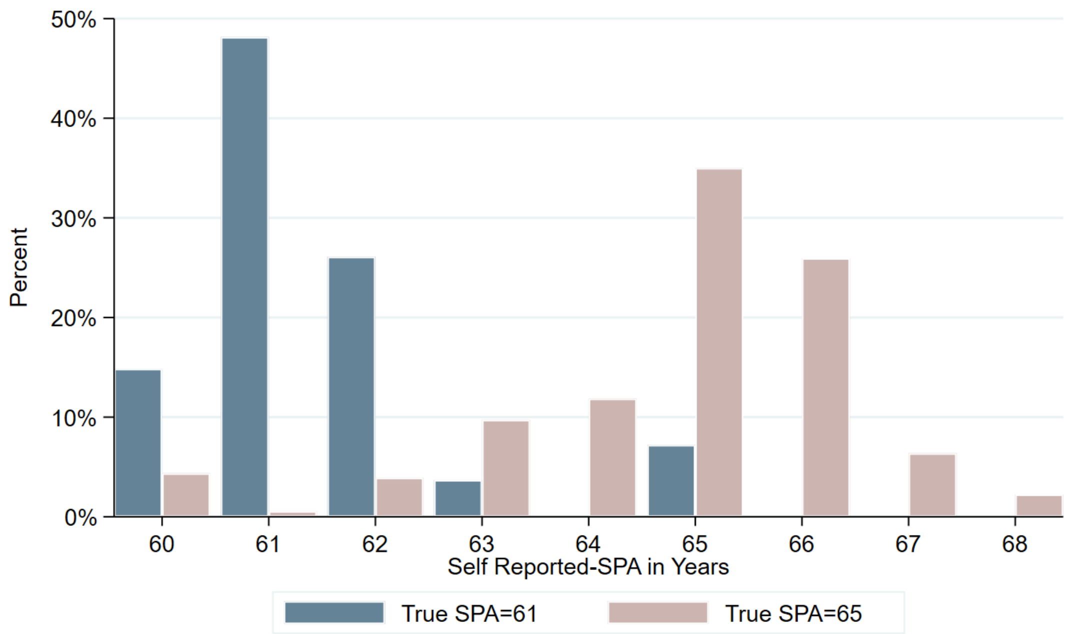
As mentioned in the main text age, age-based mandatory retirement is illegal, and as discussed at the start of this section, this is interpreted strictly by the courts. It is still possible that firms illegally force people to retire. To address this possibility, Table 6 drops all women who self-report having been forced out of their last job. Unfortunately, given the small numbers who self-report having been dismissed, the results do not change significantly.

Table 6: 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)
p=	.000	.001	.000	.000
Above SPA×(VLA.>Med.)			**-0.057**	
s.e			(0.04849)	
p=			.244	
Above SPA× VLA.				$\mathbf{-1.15\times10^{-7}}$
s.e				(2.67e-08)
p=				.000
Obs.	7,799	3,738	7,799	7,676

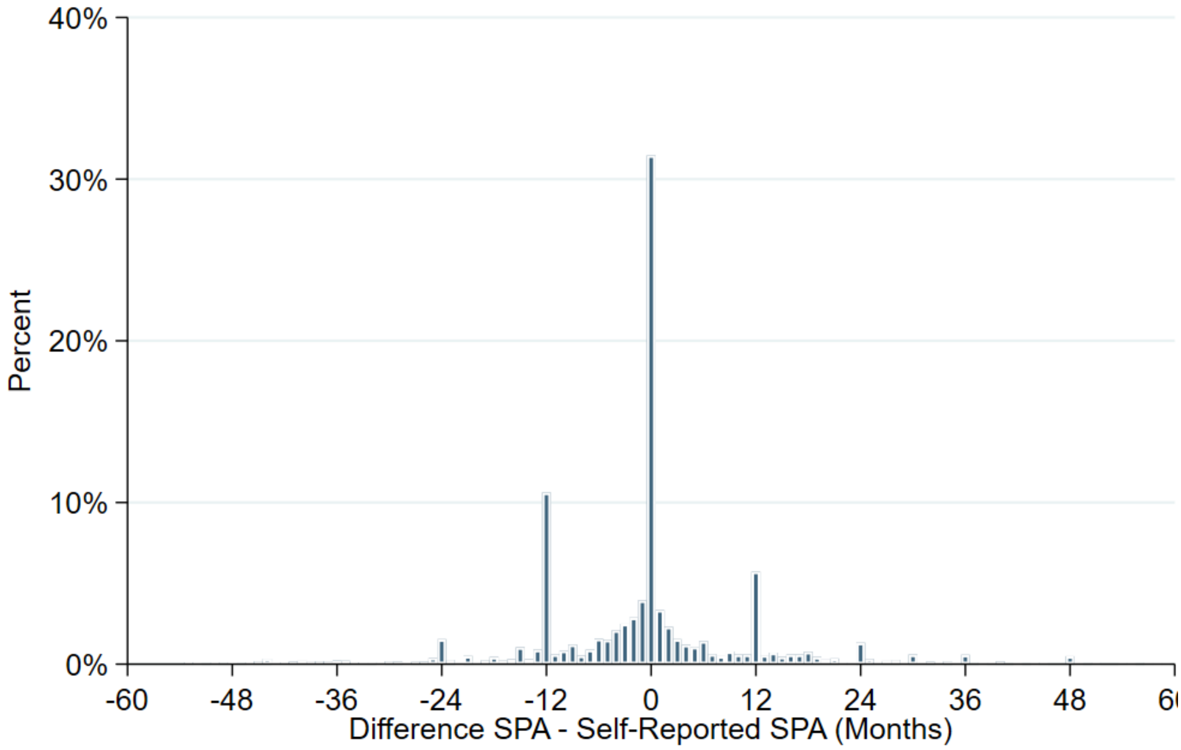
Notes: Column (1) shows the results of running the two-way fixed effect specification in 1 in the main text with controls used: a full set of marriage status, years of education, education qualifications, and self-reported health dummies; partners age; partners age squared; the aggregate unemployment rate during the quarter of interview; dummies for partner eligible for SPA, and for being one and two years above and above SPA; and assets of the household. Column (2) repeats this regression on the subsample with above median NHNBW in the last interview before their SPA. Column(3) tests whether the different treatment effects observed in columns (1) and (2) are different by introducing an interaction between being above the SPA and having above median NHNBW. Column(4) includes an interaction between being above SPA and a continuous measure of NHNBW.

Figure 4: SPA Beliefs by SPA-cohort



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

Figure 5: Mistaken SPA Beliefs of Women Subject to the Reform at Age 58 (monthly)



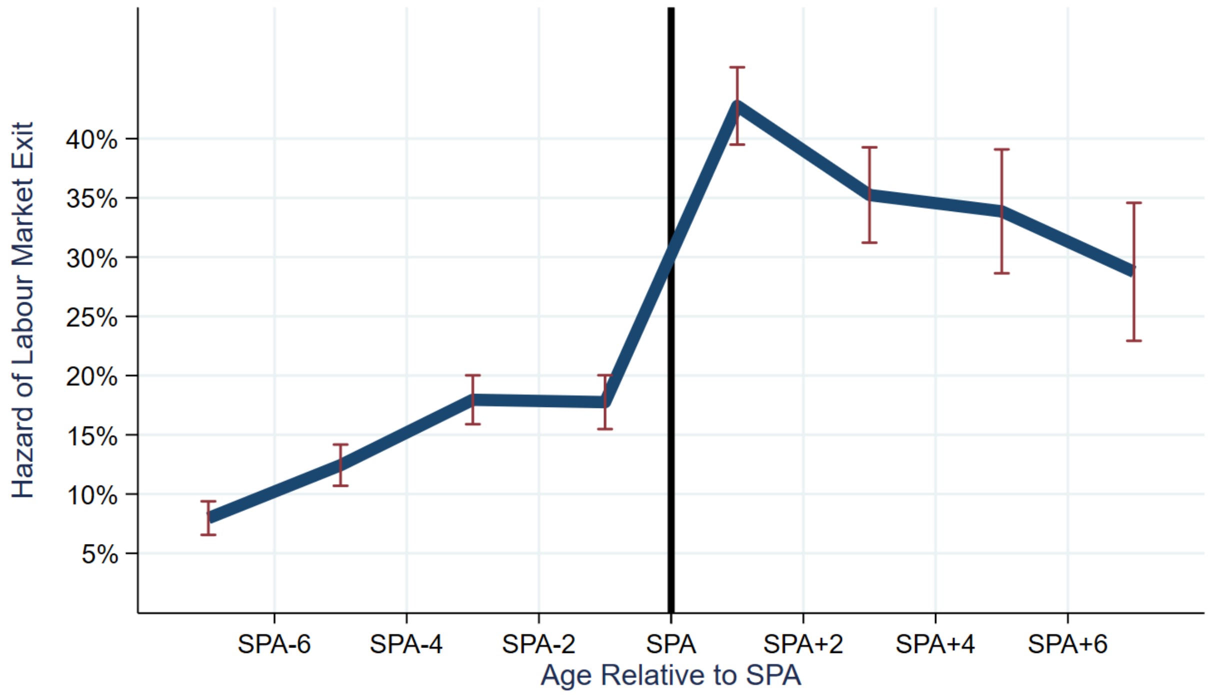
Notes: Plot of error in self-reported SPA. The graph shows the frequency by which respondents gave mistaken answers about their SPA with errors at the true monthly level of SPA variation.

A.4 Descriptive Analysis of Beliefs

Mistaken beliefs could take on many forms. People could simply not update from the pre-reform SPA of 60 or might cling to other salient numbers like the male SPA of 65. To get at these distinctions, Figure 4 plots reported SPAs for two SPA cohorts, one with a true SPA of 61 and one with a true SPA of 65. Although there is a slight increase around other salient ages, the dominant pattern is that the self-reports cluster around the true SPA for each cohort, looking very much like a noisy signal of the true SPA. Just the sort of pattern we would expect to emerge from a model of costly information acquisition.

Figure 5 shows that error in self-reported SPA at age 58 that was documented in the main text, but here at the true monthly frequency. Little that is relevant to the model is added by looking at the lower level of variation. We see that 31% know their own SPA to precisely the right month. The main thing we glean from this graph that we don't when the date is binned at a yearly frequency is that the spike every twelve months.

Figure 6: Fraction exiting labour employment - Men



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

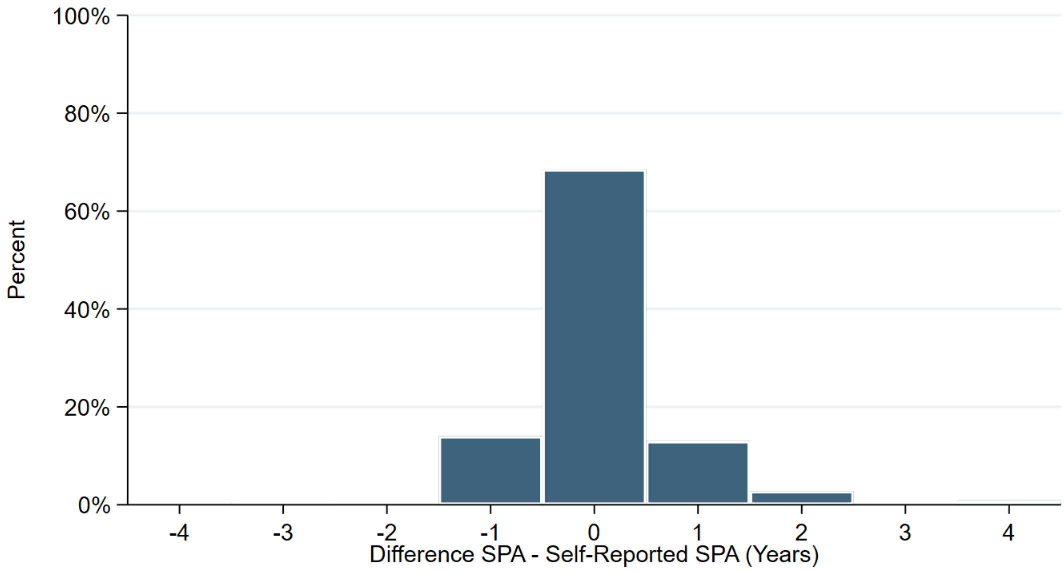
A.5 Men: Misbeliefs and Employment around SPA

Due to the lack of policy variation, it is impossible to casually estimate the employment response to the SPA for men. Hence, the main text focuses on the impact of mistaken beliefs on employment for women. However, that does not mean the similar mechanisms are not at play for men.

Figure 6 shows that men display a similar jump in the hazard rate at SPA. Although for men, it is not possible to separate the impacts of SPA from other effects like ageing, it is instructive that a jump in hazard rate also occurs at SPA.

Figure 7 plots mistaken beliefs for men at age 58. Interestingly, even though these men were not subject to SPA reform and the male SPA had remained unchanged since 1948, almost 40% of them didn't know their own SPA within a year at age 58. Although this is lower than the 60% of women who do not know their own SPA at age 58, it supports the idea that misbeliefs are relevant in the absence of reform. The potential for reform when the households is subject to costly attention would suggest that some men would be mistaken even in the absence of a realized reform. Hence, I argue that the evidence in this section is constitent with the view that the mechanism in the paper apply

Figure 7: Mistaken SPA Beliefs of Men at Age 58



Notes: 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.

B Additional Mathematical Details

B.1 Sketch proof of results from Steiner et al. (2017)

To provide some intuition, and because an understanding of these results is needed to understand the solution methodology, in this section, I present an outline of their proof using my model as a lens through which to explain their results. Steiner et al. (2017) extend Matějka and McKay (2015)[2] to a dynamic setting and so most of what is explained here applies equally to static problems.

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

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

[2] This is a more complicated step than it may sound and to show this they had to overcome various thorny issues, stemming from the information acquisition. Although I allude to some of these complexities I mostly ignore them to give the reader the intuition for the dynamic logit-like results.

conditional on the signal the decision $d(z)$ is deterministic.

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

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

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

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

and from the symmetry of mutual information: [3]

$$I(\underline{p_t};\underline{\pi_t}) = \sum_{spa} \pi_t(spa) \left(\sum_d p_t(d|spa) \log(p_t(d|spa))\right) - \sum_d q_t(d) \log(q_t(d)) \tag{2}$$

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

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

Substituting 2 into 1, rearranging, and collapsing the repeated sums gives:

$$\max_{\underline{p_t}} \sum_{spa} \pi_t(spa) \sum_{i=1}^{N} \left(n^{(k)} \frac{((c_i/n^{(k)})^\nu l_i^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \log(q_t(d_i)) - \log(p_t(d_i|spa_i)) + \beta \overline{V}_{t+1}^{(k)}(d_i, X_t, SPA_t, \underline{\pi_t})\right). \tag{3}$$

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

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

differentiating 3, be satisfied [4]

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

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

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

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

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

This derivation assumed $\underline{q_t}$ was given, but as $\underline{q_t}$ is the marginal to conditional $\underline{p_t}$ it is also chosen. The form of $\underline{q_t}$ can be found from substituting 15 from the main text into 3 and noting that the logarithm of the numerator in 15 from the main text cancels all other terms in 3 leaving only the summation from the denominator. So $\underline{q_t}$ can be found by solving:

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

B.2 Finding Unique Actions Using Second Order Conditions

Using the Kuhn-Tucker conditions of Equation 16 from the main text Caplin et al. (2019) provide an alternative formulation of the solution of the model. If the CCP satisfy equation 15 from the main text and for all possible actions ($\forall d = (c, l) \in \mathscr{C}$)

$$\sum_{spa} \pi_t(spa) \frac{\exp\left(n^{(k)} \frac{((c/n^{(k)})^{\nu} l^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \beta \overline{V}^{(k)}_{t+1}(d, X_t, spa, \underline{\pi_t})) \right)}{\sum_{d' \in \mathscr{C}} q_t(d') \exp\left(n^{(k)} \frac{((c'/n^{(k)})^{\nu} l'^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \beta \overline{V}^{(k)}_{t+1}(d', X_t, spa, \underline{\pi_t})) \right)} \leq 1, \tag{4}$$

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 16 from

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

the main text.

If an action $d^\star = (c^\star, l^\star)$ satisfies equation 17 repeated here:

$$\sum_{spa} \pi_t(spa) \frac{\exp\left(n^{(k)} \frac{((c/n^{(k)})^{\nu} l^{1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \beta \overline{V}^{(k)}_{t+1}(d, X_t, spa, \underline{\pi_t})\right)}{\exp\left(n^{(k)} \frac{((c^\star/n^{(k)})^{\nu} l^{\star 1-\nu})^{1-\gamma}}{\lambda(1-\gamma)} + \beta \overline{V}^{(k)}_{t+1}(d^\star, X_t, spa, \underline{\pi_t})\right)} < 1, \tag{5}$$

for all $d = (c, l) \in \mathscr{C}$. 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 less than 1.

If such a $d^\star$ exists then it automatcally statisfies 4 to have $q_t(d^\star) = 1$, because substituting $q_t(d^\star) = 1$ into 4 yeilds 17 from the main text with an non-binding constraint.

C Additional Computational Details

C.1 Solving the Models without Costly Attention

The models are solved by backward induction starting at age 101 when the household dies with certainty. The household problem is considered as a discrete choice problem. This within-period discrete choice optimisation problem is solved by grid search, selecting the value that maximises the household's utility. States are discretised with 30 grid points for assets (a_t), 4 for average earnings ($AIME_t$), 5 for wages (w_t), two for the unemployment shock (ue_t), and in the model with policy uncertainty the state pension age (SPA_t) has 8 gris points as it ranges from 60 to 67.

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

C.2 Solving the Models with Costly Attention

Belief Distribution Costly attention introduces a high dimensional state variable in the form of the belief distribution ($\underline{\pi_t}$). To discretise the distribution, I consider all possible combinations moving probability masses of a given size between the eight possible SPAs 60-67. As no amount of Bayesian updating can change the assignment of zero probability to an outcome, I want to avoid having beliefs that assigned zero probability to SPAs in my gridpoint of beliefs, and so I imposed a minimum probability to be assigned to each SPA of 0.01 and then had the probability masses that are moved about be in addition to this minimum amount. To make this more concrete, I broke the total probability into four masses that I moved between SPAs to form the grid over beliefs. In the absence of this minimum probability of any SPA, that would mean the probability masses being moved between SPAs was of a size of 0.25. In periods in which there are eight possible SPAs, because $t < 60$ and the women have not aged past any possible SPA, these probability masses are of the size $\frac{1-0.08}{4} = 0.23$. When $t < 60$, having these four probability masses to move between 8 possible SPAs leads to a total of $\binom{7+4}{4} = 330$ grid points because each combination can be thought of as an ordering of the four masses and the breaks between the eight grid points. As the women successively age past their SPAs, this shrinks as the number of SPAs to assign a probability mass to shrinks down to $\binom{1+4}{4} = 5$ when $t = 65$. Since there is no natural ordering over $\mathbb{R}^7$, I order these numbers in lexicographic ordering, which is convenient for constructing all possible combinations of the probability masses.

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

Range of Attention Costs As explained in Section 7.1 from the main text, when rational inattention matters because $t < SPA_t$ the central equation that needs to be solved to find the households optimal decision is the following:

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

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

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

High-level Pseudo Code

Remove d from choice set $\mathscr{C}$ that are strictly dominated across all possible combinations of SPA_t and $\underline{\pi_{t+1}}$
if $|\mathscr{C}| = 1$ **then**
    Set $\underline{q_t}$ to degenerate distribution at unique $d \in \mathscr{C}$
else
    Set initial value of $\underline{\tilde{q}_t}$ and Error > Tolerance
    while Error > Tolerance **do**
        Solve for $\overline{V}_{t+1}$ (equation 14 from the main text) given $\underline{\tilde{q}_t}$
        Remove d from $\mathscr{C}$ that are strictly dominated across all possible SPA_t given $\overline{V}_{t+1}$
        if $|\mathscr{C}| = 1$ **then**
            Set Error = 0 < Tolerance and $\underline{q_t}$ to degenerate distribution at $d \in |\mathscr{C}|$
        else
            if there is an action d that satisfies 17 from the main text **then**
                Set Error = 0 < Tolerance and $\underline{q_t}$ to degenerate distribution at d
            else
                Solve 16 from the main text using sequential quadratic programming for $\underline{q_t}$
                Set Error to distance between $\underline{q_t}$ and $\underline{\tilde{q}_t}$

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

end if

end if

end while

end if

Substitute $\underline{q_t}$ into 15 from the main text to solve for $\underline{p_t}$.

C.3 Simulating and Estimating

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

Given these initial conditions, I simulate the choice of the individual households using the decision rule found when solving the model and the exogenous process estimate in the first stage. I then aggregate the simulated data in the same way we aggregate the observed data and construct moment conditions. I describe these moments in greater detail in appendix D. The method of simulated moments procedure delivers the model parameters that minimise a GMM criterion function, which we also describe in Appendix D. To find the minimum of the resulting objective function, I first sample the parameter space using Sobol sequencing and then search for a minimum using the BOBYQA (Powell, 2009) routine at promising initial conditions.

D Additional Econometric Details

D.1 Imputing AIME

Average lifetime earnings are only observed for some of the women in my sample. In order to be able to initialise the model from the joint distribution of $AIME_{55}$ and a_{55} I impute the missing observations. First, I regress $AIME_{55}$ on a quintic in NHNBW plus a very rich set of additional controls that include variables on health, education, location, labour market behaviour, housing tenure, cohort, age, wage, and measure of cognitive ability. This includes as much information as possible to impute $AIME_{55}$.

However, merely using these predictions for imputation will likely overstate the correlation between $AIME_{55}$ and a_{55}; for this reason, I add noise to the imputed variable to replicate observed heteroscedasticity.

To do this, I run regressions of the non-imputed $AIME_{55}$ values on a quintic of NHNBW without the controls (because the model does not contain the other variables) and then regress the squared residuals on the same polynomial of NHNBW. Since the imputed $AIME_{55}$ are by construction homoscedastic, adding a noise term with variance given by this last regression replicates the heteroscedasticity seen in the regression of $AIME_{55}$ on the quintic of NHNBW.

D.2 Type-specific Mortality

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

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

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

D.3 Generating Profiles

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

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

where $cohort_i$ is the year-of-birth cohort of an individual, SPA_i is her final SPA, $age_{i,t}$ her age in years, U_t aggregate unemployment to half a per cent, and the outcome variable y_{it} is either assets or employment depending on which profile is being calculated.

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

Table 7: 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

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

E Additional Results

E.1 First Stage Estimates

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

Initial conditions Initial assets a_{55} and average earning $AIME_{55}$ are set from the type-specific empirical joint distribution, some summary statistics of which are presented in Table 7. Understandable for women of this generation married women have weaker labour market attachment and so lower $AIME_{55}$ but higher household assets a_{55}. Higher education increases both variables.

Labour market conditions The type-specific transition probabilities, estimated with individuals classified as unemployed when they claim unemployment benefits, are shown in Table 8.

The parameters of the stochastic component of the wage process (persistence and the variance of innovation, measurement error, and initial draw) are shown in Table 9

Table 8: Type Specific Unemployment Transition Probabilities

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

Notes: Unemployment and reemployment transition probabilities.

Table 9: Parameters of the stochastic component of the wages

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

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

Table 10: Regression Analysis of the Determinants of Learning

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

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

The deterministic component of wages generates the wage profiles in Figure 8. Spousal Income is shown in Figure 9.

Social Insurance As mentioned in the main text, much larger differences in State Pension income are observed between married and single women than between high and low education. Amongst State pension claimers, high education women have mean state pension income of £92.52 and low education women £87.11, whereas single women have a State Pension income of £112.50 and married women £80.89. Hence to maximise power whilst capturing the key difference, I restrict heterogeneity in the State Pension process to be between married and single women only. The resulting functions of average lifetime earnings are shown in Figure 10.

Conversely, differences in private pension income are smaller between married and single women than between high and low education. Amongst those reporting non-zero private pension income, high education women have mean private pension income of £118.50 and low education women £61.42, whereas single women have State Pension income of £100.78 and married women £94.24. Hence to maximise power whilst capturing the key difference, I restrict heterogeneity in the private pension process to be between high and low-education women only. The resulting functions of average lifetime earnings are shown in Figure 11.

E.2 Model Fit

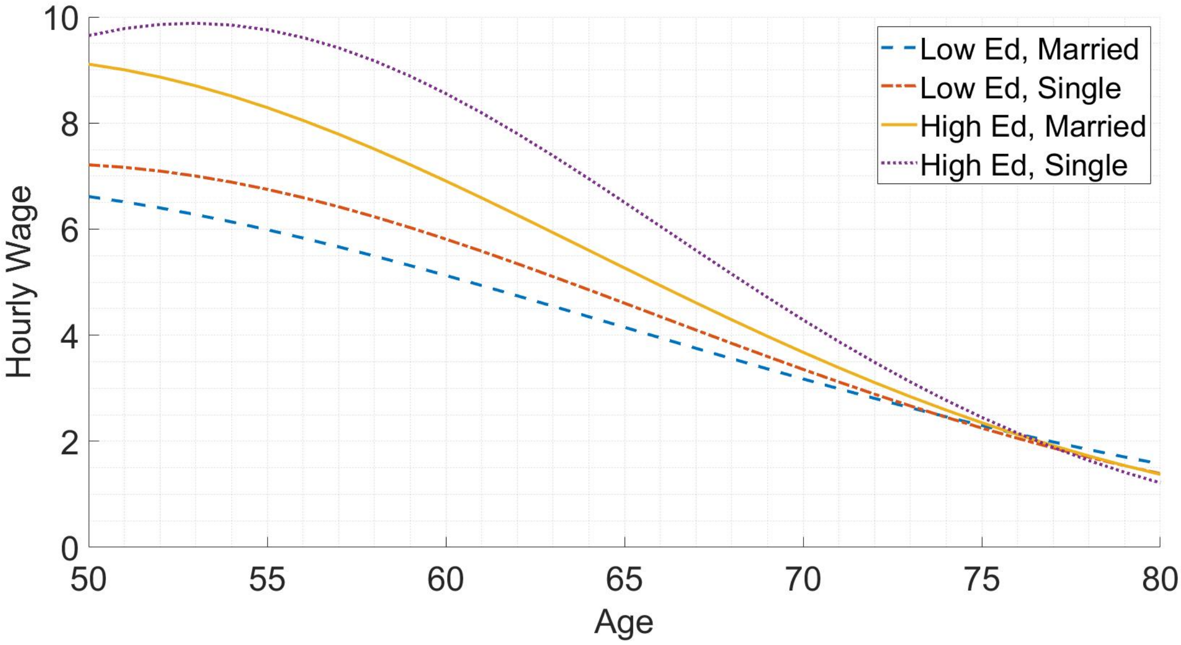
As mentioned in the main text, although the different model specifications have different predictions about the labour supply response to the dynamic SPA, the static profiles are not very sensitive to model specifications. All versions are able to match the static profiles. Figures 12-15 show the employment and asset profiles for the baseline version and the version with rational inattention with the parameter estimates of Table 5 from the main text.

E.3 Results Tables

E.4 Robustness (Targeting Other Moments)

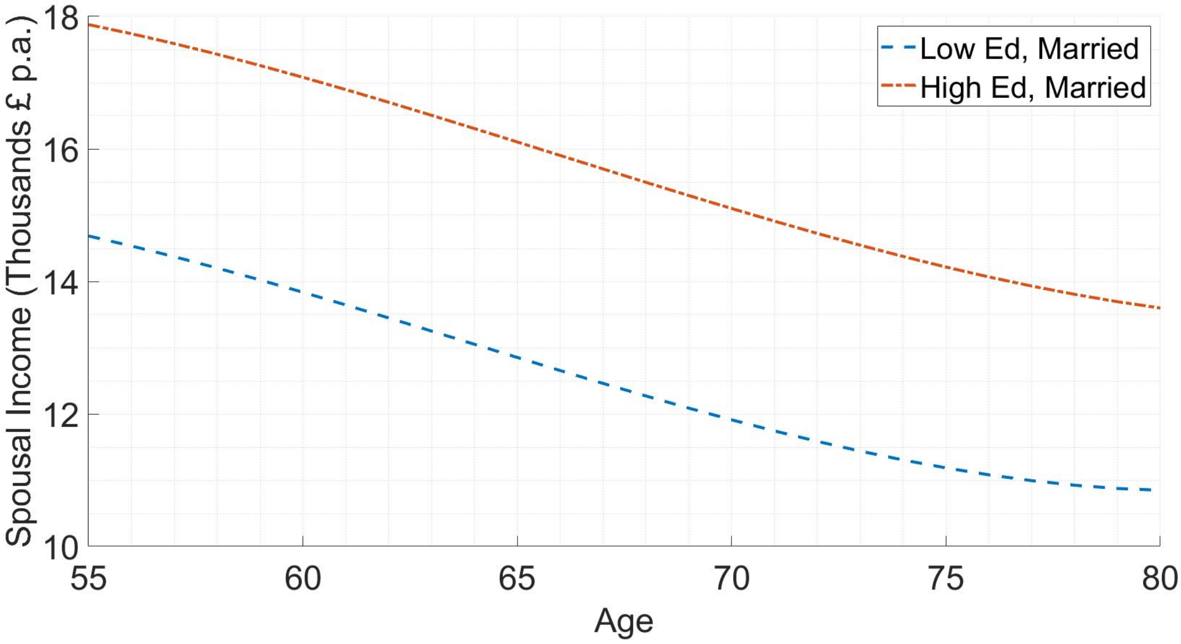
By design, the estimation procedure used in this paper does not directly target the excess employment sensitivity puzzle. This allows the model to match savings and labour supply across a range of ages and then to investigate how well a model that matches these profiles can explain excess employment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

Figure 8: Wage Profiles



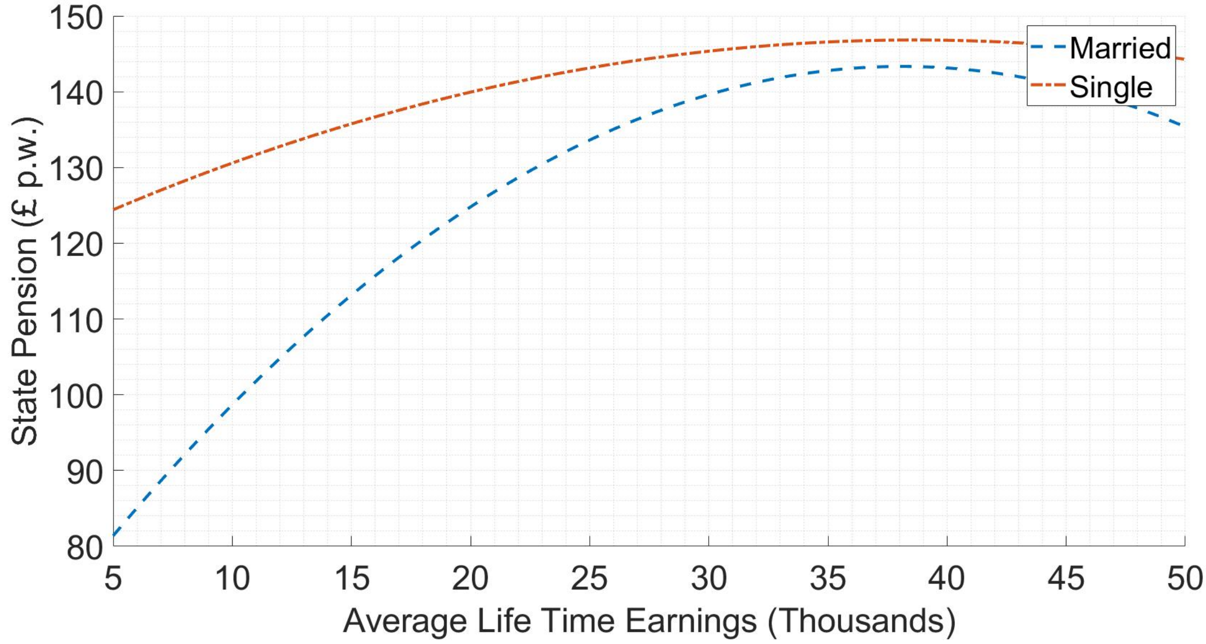
Notes: The deterministic component of female hourly wages for the four model types plotted against female age.

Figure 9: Spousal Income



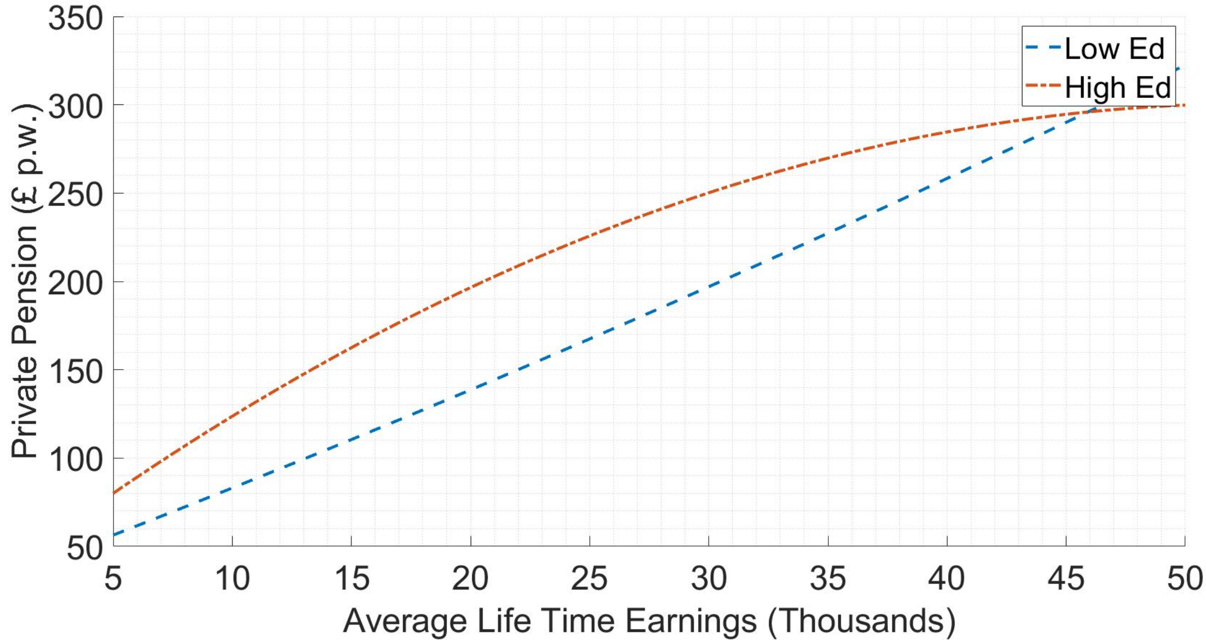
Notes: Spousal income plotted against female age.

Figure 10: State Pension as Function of Average Earnings



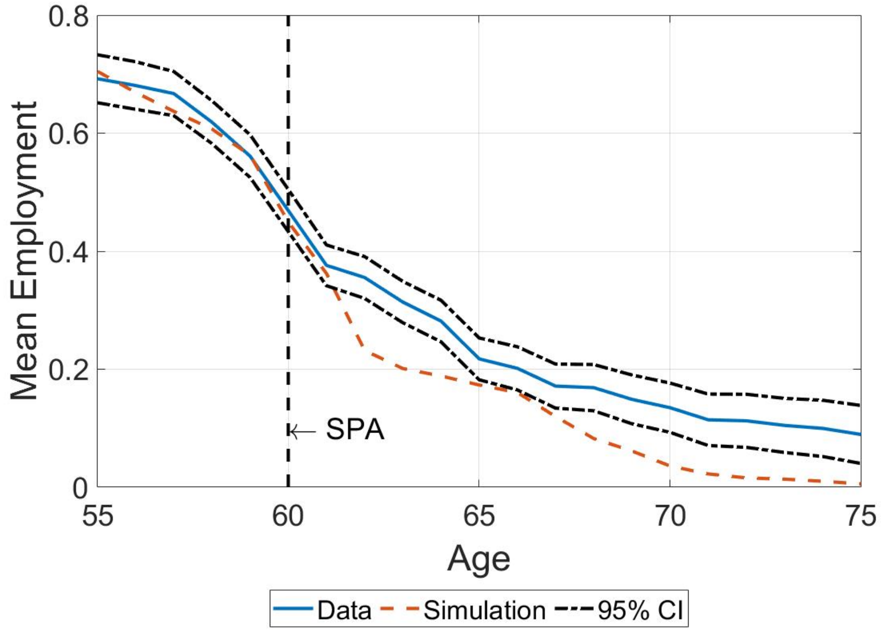
Notes: State Pension income as a function of average lifetime earnings (AIME) for married and single women.

Figure 11: Private Pension as Function of Average Earnings



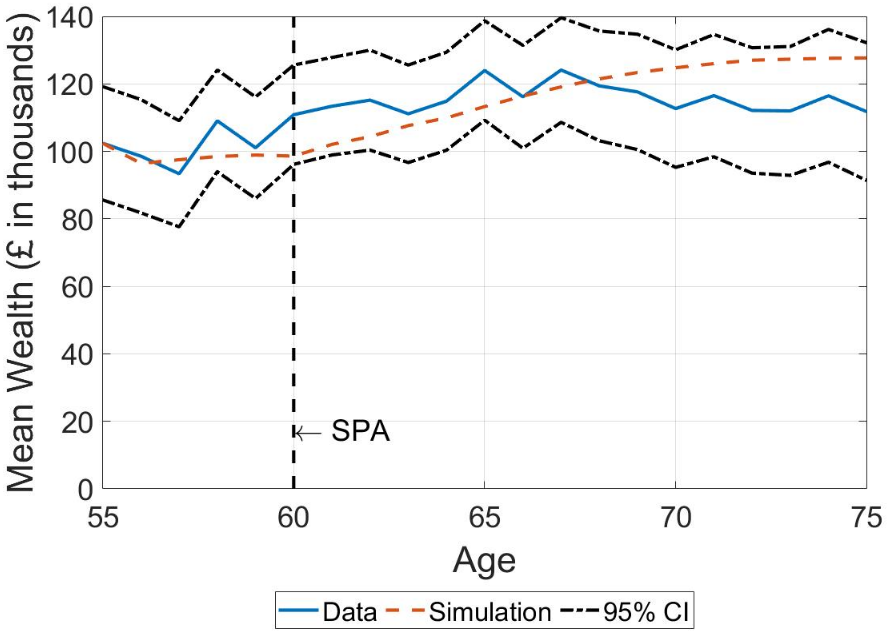
Notes: Private Pension income as a function of average lifetime earnings (AIME) for high and low education women.

Figure 12: Employment Profile Baseline



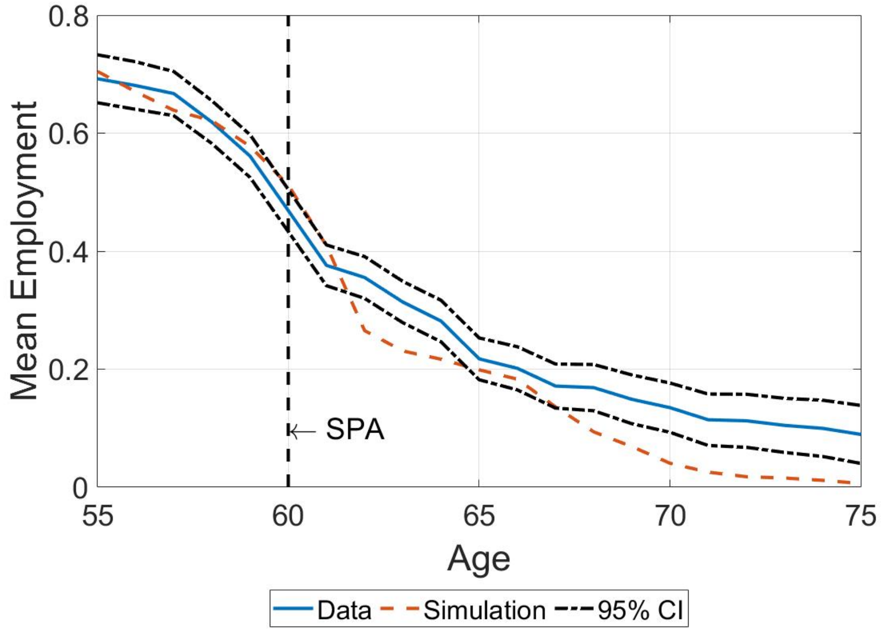
Notes: Model fit to targeted labour supply profile. The empirical profile is for the pre-reform SPA cohort with a SPA of 60. The model was simulated with an unchanging SPA of 60, mimicking the conditions faced by this cohort.

Figure 13: Asset Profile Baseline



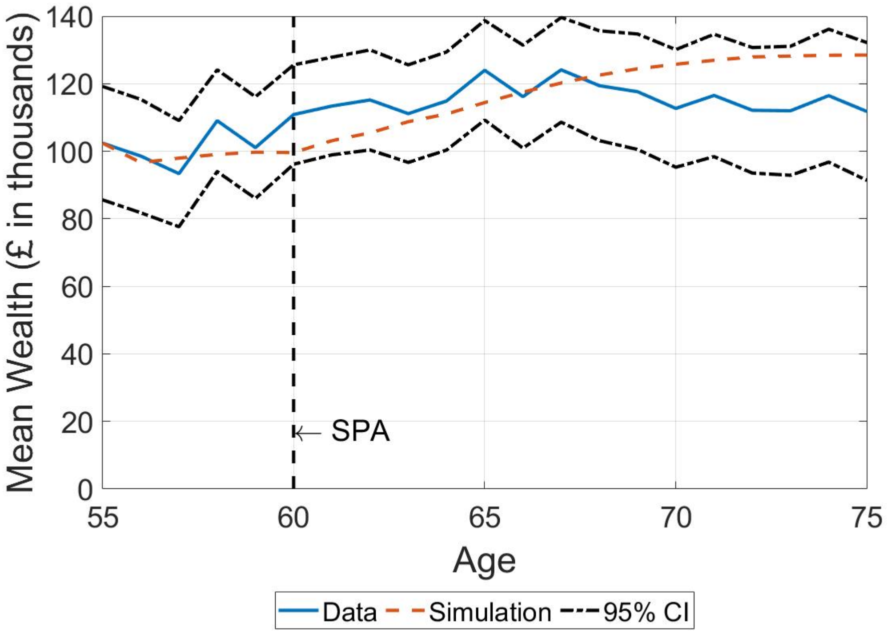
Notes: Model fit to targeted asset profile. The empirical profile is for the pre-reform SPA cohort with a SPA of 60. The model was simulated with an unchanging SPA of 60, mimicking the conditions faced by this cohort.

Figure 14: Employment Profile Model with Rational Inattention



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

Figure 15: Asset Profile Model with Rational Inattention



Notes: 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 11: Effect of SPA on Employment: Heterogeneity by Wealth

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

the whole population at the cost of greatly exaggerating liquidity constraints but cannot match the treatment effect of those with above median assets.

Table 11 shows that the model nearly perfectly replicates the treatment effect in the whole population but falls dismally short of the one seen in those with above median wealth. Figures 16-17 show that this is achieved at the cost of massively exaggerating how much households run down their assets and hence the importance of borrowing constraints. Note that this exercise is done targeting treatment 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. This slightly different treatment effect encapsulates the same excess sensitivity puzzle.

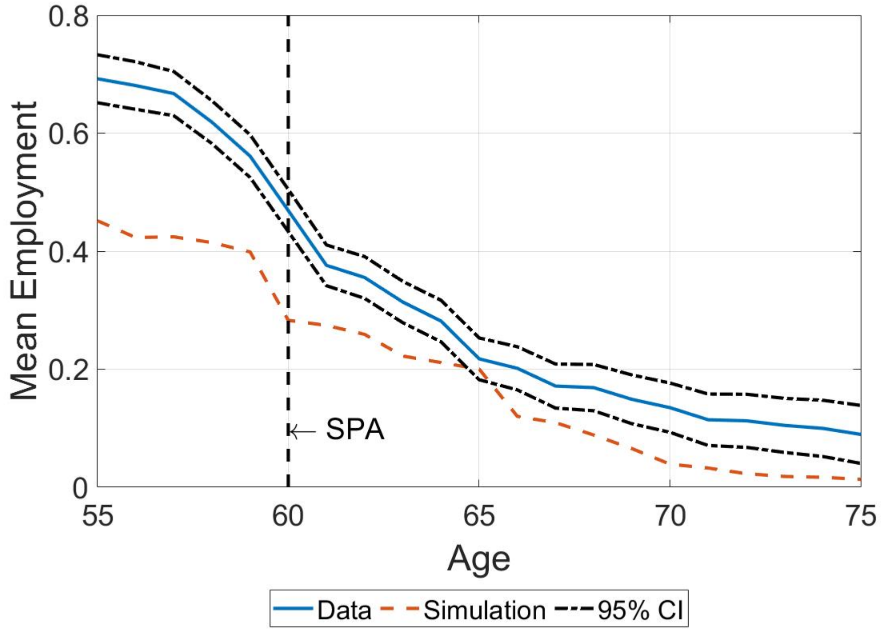
F Extension: Deferral Puzzle

Loading all policy uncertainty onto the stochastic State Pension Age (SPA) understates policy uncertainty about the State Pension. In this section, I introduce learning and uncertainty about another aspect of the state pension system: the actuarial adjustments to benefits from deferring. Combined with a claiming decision, this not only helps to align incentives by making the model more realistic but also helps explain the deferral puzzle (detailed below). Since the adjustments rate becomes irrelevant upon claiming, rational inattention to this aspect of the pension system speaks directly to this puzzle because calculations implying deferral is actuarially favourable ignore the attention benefits of claiming: claiming removes the need to pay attention to this adjustment rate. The model of Section 6.2 from the main text does not incorporate such a mechanism for two reasons. Firstly, it does not include a benefit-claiming decision. Secondly, the only source of uncertainty subject to an attention cost is the SPA, and once it is reached, the uncertainty resolves, irrespective of claiming. The simplest extension that contains this new incentive to claim is presented in the rest of this section, along with some results.

F.1 Deferral Puzzle

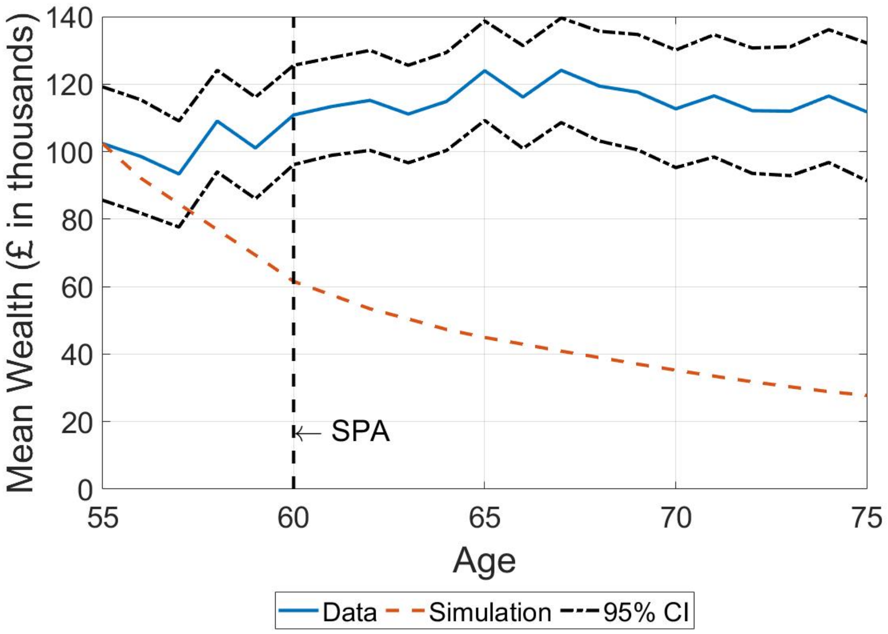
By deferral puzzle, I mean the fact deferral of state pension benefits was uncommon despite an extremely generous adjustment between April 2005 and April 2016. During this period, state pension benefits increased by 1% for every 5 weeks deferred implying an annual adjustment of 10.4%. This is an extremely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

Figure 16: Employment Profile Baseline when Targeting Treatment Effects



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

Figure 17: Asset Profile Model Baseline when Targeting Treatment Effects



Notes: 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.

had claimed by their first post-SPA interview.

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

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

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

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 claim is only an option once past the SPA, and, to keep the problem tractable, an upper limit of 70 is placed on deferral.

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

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

Table 12: Parameter Estimates - Extension

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

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

Table 13: Model Predictions - Extension with benefit claiming and uncertain deferral

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

Notes: Costly attention refers to the model with, additionally, a cost of information acquisition about the stochastic policy. The top panel shows labour supply response across the wealth distribution as per Table 6 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.

The model with policy uncertainty, to the stochastic SPA and adjustment rate, is then re-estimated to match the same pre-reform employment and assets profiles with a constant realisation of 10.5% for the deferral adjustment, which was the deferral rate these cohorts faced. Parameter estimates are in Table 12 and, for these values, only 6.2% of individuals claim the state pension before the mandatory claiming age of 70, much lower than the 99% plus claiming seen in the data.

Next, I introduce costly attention with a cost of attention corresponding to approximately £10 of consumption to the median consuming household to be fully informed. This increased the number voluntarily claiming to 22.2%, approximately a fourfold increase on the model without informational frictions, but still short of the rate observed in the data. As can be seen in Table 13, this cost of attention produced a relatively good fit along all dimensions of interest. Note that as in Section E.4 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.

References

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

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

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

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

De Nardi, M., E. French, and J. B. Jones (2009). Life expectancy and old age savings. *American Economic Review 99*(2), 110–15.

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

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

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

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

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

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

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