Rational Inattention and Oversensitivity of Retirement to the State Pension Age*

Jamie Hentall MacCuish†

Abstract

This paper presents evidence that incorporating costly thought, modelled with rational inattention, solves two well-established puzzles in the retirement literature. The first puzzle is that, given incentives, the extent of bunching of labour market exits at legislated state pension ages (SPA) seems incompatible with rational expectations (e.g. Cribb, Emmerson, and Tetlow, 2016). Adding to the evidence for this puzzle, this paper includes an empirical analysis focusing on whether liquidity constraints can account for this bunching and find they cannot. The nature of this puzzle is clarified by exploring a life-cycle model with rational agents that does match aggregate profiles. This model succeeds in matching these aggregates only by overestimating the impact of the SPA on poorer individuals whilst underestimating its impact on wealthier people. The second puzzle is that people are often mistaken about their own pension provisions (e.g. Gustman and Steinmeier, 2001). Concerning this second puzzle, I incorporate rational inattention to the SPA into the aforementioned life-cycle model, thus allowing for mistaken beliefs. To the best of my knowledge, this paper is the first to incorporate rational inattention into a life-cycle model. Rational inattention not only improves the aggregate fit of the data but better matches the response of participation to the SPA across the wealth distribution, hence simultaneously offering a resolution to the first puzzle. This paper researches these puzzles in the context of the ongoing reform to the UK female state pension age.

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

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

*I thank Fabien Postel-Vinay and Eric French for detailed discussions and encouragement. I also thank George-Marios Angeletos, Antonio Cabrales, Andrew Chesher, Monica Costas Dias, Jonathan Cribb, Martin Cripps, Mariacristina De Nardi, Rory McGee, Robert Moffitt, Cormac O'Dea, Ran Spiegler, Morten Ravn, and participants at a number of seminars for their helpful comments.

†University College London and Institute for Fiscal Studies

Contents

List of Figures

List of Tables

1 Introduction

This paper explores the rationality of retirement decisions and whether allowing for the costly nature of thought explains observed regularities better. In particular, it focus on two puzzles for a rational expectations (RE) explanation of retirement choices. Firstly, that it is difficult to reconcile the large number of labour market exits at legislated pension entitlement ages with RE. Secondly, that individuals are frequently mistaken about their own retirement provisions. I find that acknowledging costly thought accommodates the second puzzle whilst generating a mechanism that helps explain the first.

Ageing populations have forced many governments to increase the state pension age. These reforms revealed the first puzzle: labour market exits are more sensitive to legislated state pension ages than RE can accommodate given the limited incentives to retire exactly at these ages (Behaghel, and Blau, 2012; Cribb, Emmerson, and Tetlow, 2016; Seibold, 2017; Lalive, Magesan and Staubli, 2017). This paper contributes to the evidence for this puzzle, within the context of the ongoing reform to the UK female state pension age (SPA), by studying the dependence of the response to the SPA on asset holdings. I find an indistinguishable participation response to the SPA across the wealth distribution largely ruling out liquidity constraints as a RE consistent explanation of this puzzle.

This claimed oversensitivity is in tension with studies that successfully match observed retirement decisions without abandoning RE (French, 2005; O'Dea, 2018), although these are not studies of a pension reforms. One achievement of this paper is to resolve this tension. This is achieved by investigating whether a RE model that successfully matches aggregate labour market profiles around SPA also matches the response to changes in the SPA across the wealth distribution. The model succeeds in matching the aggregate only by exaggerating the response to the SPA amongst the bottom half of the wealth distribution whilst shrinking the response amongst the top half.

The second puzzling regularity at odds with RE is that individuals are frequently mistaken about their pension provision (Gustman and Steinmeier, 2001; Rohwedder and Kleinjans, 2006; Crawford and Tetlow, 2010; Amin-Smith and Crawford, 2018). Traditionally life-cycle models treat institutional factors like the SPA as static parameters known, without cost, by everyone; an approach which precludes any explanation of why people are systematically mistaken. By acknowledging the stochasticity of government pension policy and incorporating costly thought, this paper attempts to explain this observed ignorance. Costly thought is modelled using rational inattention (RI), an approach that includes a utility cost of information acquisition.

Allowing for these incorrect beliefs explains the bunching of labour market exits with greater success than the RE benchmark model. The mechanism behind this result is as follows: rational inattention to the SPA introduces additional uncertainty implying greater precautionary saving which leads to greater labour market participation. As this uncertainty is resolved upon reaching SPA this induces bunching of labour market exit at the SPA. This mechanism is not strictly dependent on rational inattention and exist with the introduction of a stochastic SPA alone. However, for reasonable levels of stochasticity that match how the government has historically reformed the SPA, I find that the amplification of uncertainty inherent in rational inattention is required to produce a discernible difference from the RE benchmark.

The reform to the UK female state pension age (SPA) provides me with the opportunity to investigate these question. This reform has a staggered implementation which creates individual level variation in SPAs allowing the effect of the SPA on employment to be identified separately from effects of ageing. Additionally, the UK institutional context has two features advantageous to identifying motivations behind retirement decisions. Firstly, receipt of the UK state pension is not conditional on employment status and only provides an incentive to retire for liquidity constrained individuals. Secondly, forcing someone to retire purely due to age is illegal, ruling out firm mandated retirement as an explanation for the bunching of labour market exits. The dataset I will use is the English Longitudinal Study of Ageing (ELSA) which is a detailed panel survey of older individual.

The rest of the paper is structured as follows. In section 2, I review the literature and, in section 3, I outline the institutional context and the data used. In section 4, I present a reduced form analysis of the data re-evaluating the claim that bunching of labour market exits at SPA cannot be explained with RE. In particular, as liquidity constraints are the main explanation proposed within an RE approach, I investigate whether liquidity constraints can explain this bunching and find indications they cannot. To see if some RE based mechanism might explain the bunching, in section 5, I take a version of a recent rich structural model (O'Dea, 2018) that matches aggregate profile and investigates its mechanisms. I find that it matches the aggregate bunching of labour market exits at SPA only by exaggerating the response of those nearer the borrowing constraints and underestimating the response of those much further from it. In section 6, I introduce rational inattention to the SPA into the model presented in section 5 and present preliminary result from a calibrated version of the model. I find that, as well as improving the fit of the aggregate profiles, the introduction of RI better approximates the response to the SPA across the distribution while allowing for the fact people hold incorrect beliefs.

2 Literature Review

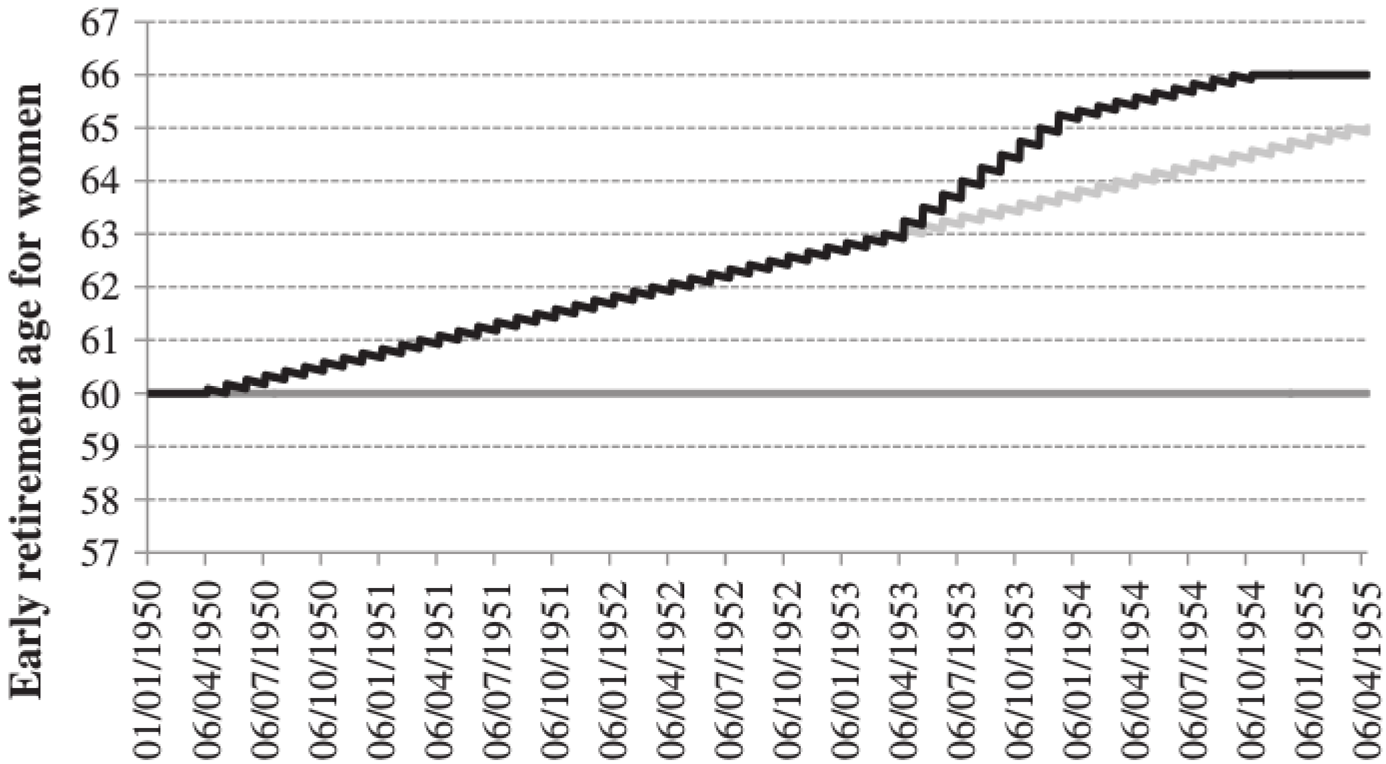
This research project will contribute to two distinct research areas: life-cycle models of retirement and RI.

It has long been known that labour force participation responds strongly to the eligibility ages of social security programs. Gruber and Wise (2004) survey evidence from 11 developed countries and find labour force exits concentrated around legislated retirement ages. The response of labour market participation to reforms in the social security eligibility ages has been widely studied (Staubli and Zweimuller, 2013; Manoli and Weber, 2012; Atalay and Barrett, 2015). Often this literature has focused on estimating elasticities and fiscal impacts; however, recently a group of studies has argued that RE is unable to match the magnitude of labour market exits at legislated pension ages which is the principal puzzle this paper attempts to explain (Behaghel, and Blau, 2012; Cribb, Emmerson, and Tetlow, 2016; Seibold, 2017; Lalive, Magesan and Staubli, 2017). Lalive et al. (2017) study a Swiss reform that increased women's full retirement age (FRA), the age they can claim their full pension, and introduced an Early Retirement Age (ERA), the earliest eligibility age to a reduced pension. They find, whilst incentives encourage most women to claim early, most delay. Seibold (2017) studies retirement in Germany where there exists an ERA, a FRA, and a NRA (normal retirement ages, a purely nominal age), along with multiple pathways into retirement; this arrangement creates over 600 kinks and notches in life-time budget constraints. These non-differentiabilities can be classified into statutory retirement ages where some expectation of retirement exists (i.e. ERA, FRA and NRA) and pure financial incentives. Siebold (2017) documents that the bunching of labour market exits is higher at the non-differentiability associated with a statutory retirement age rather than those only associated with a financial incentive and interprets this as evidence for reference dependent preference. Both Seibold (2017) and Lalive et al. (2017) present structural model that attempt to explain these findings. Cribb et al. (2016) study the same increase to the UK female SPA as this paper. Since the UK state pension age is the earliest the pension can be claimed it is an ERA for international comparison. They produce reduced form estimates of the impact of reaching ERA on labour force participation and, although their paper is purely empirical, argue that their findings are difficult to explain with RE. As I borrow from and build on the methodology of Cribb et al. (2016) their methods are discussed in more detail in section 4.

The second puzzle from the empirical literature on which I focus, people mistaken beliefs about their own pension provision, goes back to at least Gustman and Steinmeier (2001). They compare reported expected benefits to objective calculations based on social security records and employer provided pension description and find misinformation the norm. Rohwedder and Kleinjans (2006) study the dynamics of these mistaken beliefs and find that their expectations become increasingly accurate on average as individuals approach retirement. Finally, Crawford and Tetlow (2010) look at women subject to the UK female SPA reform and find they hold substantially incorrect beliefs about their own SPA. An advantage of the structural model presented here over those of Seibold (2017) and Lalive et al. (2017) is that it accommodates both the fact people hold incorrect beliefs and that labour market exits bunch at SPA. I also rely on a different mechanism, rational inattention to the SPA, rather than reference dependent preferences (Seibold, 2018) or the proportion of agents being unable to choose when to retire (Lalive et al., 2017).

By embedding rational inattention (RI) in a life-cycle model this paper contributes to the RI literature. To the best of my knowledge this is the first paper to do this; moreover, this is achieved in a rich structural model. RI was developed by Sims (1998; 2003) and originally applied to macroeconomic problems although recently it has been more widely used (e.g. Matejka and McKay, 2014; Ravid, 2018). In rational inattention agents make choices about their level of attention and suffer a utility penalty for gaining information. In some

Figure 1: SPA as a Function of Date of Birth for Women in the UK



work, perhaps better termed limited attention, the limitations are exogenously set (Sims, 2003; Maćkowiak and Wiederholt, 2009), but in more recent work agents choose their level of attention trading the cost of obtaining the information against the information's value (Maćkowiak and Wiederholt, 2015; Wiederholt, 2010; Matejka and McKay, 2014). This paper extends and uses a recent result by Steiner, Stewart, and Matejka (2017) who provide a general solution to dynamic discrete choice problems with RI in order to incorporate RI to the ERA into a life-cycle model.

3 Data and Institutional Context

The context for this research project is the UK reform to the female state pension age (SPA). The Pensions Act 1995 legislated for the female SPA to rise gradually from 60 to 65 over the ten years from April 2010, rising by one month every two months, reaching age 65 by April 2020. The Pension Act 2011 accelerated the rate of change of the female SPA from April 2016 so that it equalises with men's at 65 by November 2018. It additionally legislated an increase to both the male and female SPA to 66 years phased in between December 2018 to October 2020. Figure 1 from Cribb et al. (2016) summarises how these changes affect women in different birth cohorts.

An advantage of studying the UK reform is that UK law prohibits firms compulsorily retiring people based on age, so this cannot explain the bunching of labour market exits at SPA. Another advantage of the UK context is the state pension is not conditional on employment status and does not provide major tax incentive to exit from the labour market at this age (Cribb et al., 2016). Together this removes financial incentives to retire at the SPA for all but the liquidity constrained. A disadvantage with the UK reform is the UK has a single retirement age at which pension benefits are claimed. This makes it difficult to rule out liquidity constraints driving the bunching of retirement at SPA, as the ability to borrow against future pension benefits is severely limited. Cribb et al. (2016) argue against credit constraints being the primary driver on the basis that, whilst homeowners are less likely to be liquidity constrained than renters, the effects of SPA on their labour market participation are indistinguishable. Homeownership, however, is a coarse measure of wealth and equity in one's own home is an illiquid asset. To investigate the role liquidity constraints play, in section 4, I repeat the analysis of Cribb et al. (2016) on the English Longitudinal Study of Ageing (ELSA) dataset which contains detailed asset holding information this allows me to control more precisely for assets.

ELSA is the principal dataset used it this investigation. It is a panel dataset at a biennial frequency containing a representative sample of the English private household population aged 50 and over modelled on the Health and Retirement Study (HRS) in the USA. ELSA contains detailed data on: labour market circumstances, earnings, and the amount and composition of asset holdings. The first wave of ELSA after the start of the initial implementation of the female SPA reform was wave 5 which covered 2010/11 and I use all waves of ELSA from wave 1 (2002/03) thought to wave 7 (2014/15).

4 Reduced Form Evidence

The reduced form analysis presented in this section builds on Cribb et al. (2016) study the impact of the UK female SPA reform on labour market participation. They regress probability of participation (y_{it}) on: an indicator of being below the SPA; a full set of age, and year of birth dummies; and a vector of controls leading to the following specification:

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

They interpret the parameter α as a difference-in-difference estimator of the effect of the treatment of being below the SPA and find a statistically significant increase in the probability of being employed from the treatment. They estimate this equation as both a linear probability model and as a probit model with error clustered at the level of the individual. As mentioned, they argue that liquidity constraints cannot explain the treatment effect α because a similar size effect is observed for home-owners and renters. Here I repeat their analysis using ELSA which allows me to more control carefully for assets.

I first present results of estimating equation 1 as a random effect linear probability model with errors clustered at the level of the individual. I use a random effect specification because the small sample size means controlling for both autocorrelation and heteroskedascitiy by clustering and arbitrary fixed effect leads to imprecise estimates. The random effects assumption was tested with a Durbin-Wu-Hausman test on the treatment effect and the null, of no difference between the random effect and fixed effects coefficients, was not rejected. For those uncomfortable with the random effects assumption I also repeat and present all regression as fixed effects regression with clustered standard errors. Finally, as linear probability models have many well know flaws, I repeat the analysis presented here with a random effects probit specification model with clustered standard errors. The full list of controls used is: a full set of marriage status, years of eduction, education qualifications, and self reported health dummies; partners age; partners age squared; the aggregate unemployment rate during the quarter of interview; dummies for partner eligible for SPA, and for being one and two years above and below SPA; and assets of household according to one of the two specification discussed below.

Obviously, the households whose retirement decisions are least likely to be affected by liquidity constraints are those with substantial liquidity assets. I consider two categories of assets which are the two most liquid categories in Carrol and Samwick (1996). Firstly, I look at very liquid assets (VLA) which cover any assets that could be liquidated almost immediately. This includes bank account balances, money market funds, certificates of deposit, government savings bonds, mutual funds, and publicly traded stocks. Secondly, I consider non-housing, non-business wealth (NHNBW) which consist of VLA + all other assets and liabilities not related to the primary residence or personally owned businesses; these have in common that the household could liquidate them without losing their home or primary income.

Table 1 contains the results of regressing equation 1 as a random effect linear probability model with errors clustered at the level of the individual. Row 1 of the table shows the results of running this regression for the whole population. Rows 2 show the results of running the regression on the subpopulations who have more than the median assets, taken over the whole sample, in the interview immediately before their SPA using the NHNBW asset category. Row 3 does the same for the the VLA asset categories. The treatment effect is significant at the 5% level for the whole population and for the population restricted to having above median NHNBW. For the population restricted by VLA the treatment effect is only significant at the 10% level. The size of the treatment effects also does not appear to vary much between the groups. For table 2, rows 1 and 2, I rigorously test whether there is a significant difference between the treatment effects of the two groups by adding an interaction term to equation 1 and testing its coefficients. As can be seen, there is no significant difference at any reasonable level. These results contrasts with the predictions of a liquidity constraint explanation for the bunching of labour market exits at SPA.

Table 1: Treatment Effect different subpopulations: Random Effects Model

Population	Observations	People	Treatment Effect	P>|z|	95%CI
Whole Population	5,710	2,882	0.101 (0.0312)	0.001	[0.0402,0.1623]
NHNBW > Median	2,573	1,154	0.099 (0.0441)	0.025	[0.0125,0.1855]
VLA > Median	2,563	1,155	0.084 (0.0443)	0.059	[-0.0033,0.1702]
LC Classification 1 NHNBW	4,750	2,384	0.077 (0.0347)	0.027	[0.0085,0.1446]
LC Classification 1 VLA	4,723	2,372	0.077 (0.0348)	0.027	[0.0086,0.1454]
LC Classification 2 NHNBW	2,871	1,260	0.106 (0.0422)	0.012	[0.0234,0.1890]
LC Classification 2 VLA	2,842	1,251	0.101 (0.0423)	0.017	[0.0176,0.1836]
LC Classification 3 NHNBW	2,539	1,139	0.100 (0.0443)	0.025	[0.0125,0.1865]
LC Classification 3 VLA	2,557	1,153	0.084 (0.0444)	0.057	[-0.0025,0.1714]

However, we cannot tell a priori how far up the wealth distribution liquidity constraints affect choices. For this reason, I construct three additional classifications of whether an individual is liquidity constrained each addressing this issue in distinct ways. The first classification considers an individual to be liquidity constrained if they are from a household without assets greater or equivalent to the total of their wage from their interview until their SPA. This classification, however, takes no account of exogenous risk or precautionary saving. The second classification takes account of the precautionary saving motive. It classifies an individual as liquidity constrained if their household has insufficient assets to cover their wage plus the level of asset decumulation at the bottom 25th percentile of the distribution of changes in asset between the periods covering their SPA. The idea behind this classification is that this amount of asset decumulation represent a measure of the exogenous risk a household faces at this point in their life-cycle. However, as both of these classifications select on wages when labour force participation is the dependent variable, they produce biased estimates. As people without a job before SPA are less likely to be excluded, the estimates are biased downwards. As I still find a significant treatment effect despite this bias, it is not as large an issue as it would first seem. I still, however, consider a third classification which does not select on wages: having sufficient assets to cover decumulation at the bottom 15th percentile of the distribution of changes in asset between the periods covering their SPA. Rows 4-9 of table 1 show the results of regressing equation 1 excluding liquidity constrained individuals; each row corresponds to a different combination of classifications and asset categories. As can been seen all treatment effects are positive and significant at the 5% level bar one which is significant at 10% level. The magnitudes are also little changed between subpopulation giving little indication that liquidity constraints even form part of the explanation of the observed effect of the SPA on labour force participation.

In rows 3 and 4 of table 2, I test the difference between the whole population and those classed as liquidity constrained according to classification 3 above. There is no statistically significant difference for either grouping. I do not test the difference in treatment effect for the first two classification of liquidity constrained individuals because as mentioned selecting based on these two classification introduces selection bias and so the difference

Table 2: Differences in Treatment Effects: Random Effect Model

	Baseline Treatment Effect	Interaction	P>|z| (Interaction)	95%CI (Interaction)
NHNBW > Median	0.103 (0.0332)	-0.004 (0.0194)	0.852	[-0.0417,0.0344]
VLA > Median	0.109 (0.0332)	-0.013 (0.0194)	0.500	[-0.0512,0.0250]
LC Classification 3 NHNBW	0.102 (0.0331)	-0.001 (0.0194)	0.963	[-0.0389,0.0371]
LC Classification 3 VLA	0.109 (0.0332)	-0.013 (0.0194)	0.502	[-0.0511,0.0250]
Continuous Interaction NHNBW	0.110 (0.0314)	-8.05e-08 (3.18e-08)	0.011	[-1.43e-07,-1.83e-08]
Continuous Interaction VLA	0.109 (0.0314)	-8.18e-08 (3.22e-08)	0.011	[-1.45e-07,-1.86e-08]

Table 3: Treatment Effect different subpopulations: Fixed Effects Model

Population	Observations	People	Treatment Effect	P>|t|	95%CI
NHNBW > Median	2,573	1,154	0.105 (0.0514)	0.041	[0.0043,0.2060]
VLA > Median	2,563	1,155	0.082 (0.0521)	0.116	[-0.0202,0.1840]
LC Classification 1 NHNBW	4,750	2,384	0.076 (0.0412)	0.066	[-0.0049,0.1566]
LC Classification 1 VLA	4,723	2,372	0.075 (0.0415)	0.067	[-0.0055,0.1573]
LC Classification 2 NHNBW	2,871	1,260	0.108 (0.0501)	0.031	[0.0100,0.2064]
LC Classification 2 VLA	2,842	1,251	0.100 (0.0504)	0.047	[0.0012,0.1988]
LC Classification 3 NHNBW	2,539	1,139	0.107 (0.0516)	0.038	[0.0057,0.2083]
LC Classification 3 VLA	2,557	1,153	0.084 (0.0522)	0.108	[-0.0185,0.1862]

is not interpretable. The final two rows test the significance of a continuous interaction term for the two asset categorisations. For both of them the interaction term is negative and significant indicating that having more assets decrease the impact of being below SPA on the probability of being in work. However, the magnitude of the effect is tiny implying only a 50% reduction of the treatment effect to $\approx 5\%$ for someone at the 99th percentile of the wealth distribution according to NHNBW assets. With this tiny change in the treatment effect due to increased assets it is hard to argue that the treatment effect is completely explained by liquidity constraints.

Table 3 and 4 replicate tables 1 and 2 for the fixed effects specification. As can be seen in table 3, the treatment effects are now significant at the 5% level in four of the populations, at the 10% level in another two and insignificant in two populations. As the magnitudes of the point estimator are little changed, this lack of significance seems to be mostly driven by a lack of power. This is supported by table 4 where the difference between the treatment effects in the two subpopulations remains insignificant. Moreover, the impact of assets on the treatment effect is still tiny.

Table 4: Differences in Treatment Effects: Fixed Effect Model

	Baseline Treatment Effect	Interaction	P>|t| (Interaction)	95%CI (Interaction)
NHNBW > Median	0.095 (0.0395)	0.016 (0.0261)	0.534	[-0.0349,0.0673]
VLA > Median	0.103 (0.0396)	0.002 (0.0260)	0.939	[-0.0491,0.0530]
LC Classification 3 NHNBW	0.091 (0.0394)	0.022 (0.0261)	0.399	[-0.0292,0.0731]
LC Classification 3 VLA	0.103 (0.0395)	0.002 (0.0260)	0.938	[-0.0490,0.0531]
Continuous Interaction NHNBW	0.112 (0.0367)	-7.02e-08 (3.14e-08)	0.026	[-1.32e-07,-8.49e-09]
Continuous Interaction VLA	0.112 (0.0367)	-7.14e-08 (3.19e-08)	0.025	[-1.34e-07,-8.77e-09]

Table 5: Treatment Effect different subpopulations: Probit Model

Population	Observations	People	Average Marginal Effect	P>|z|	95%CI
Whole Pollination	5,706	2,881	0.091 (0.0294)	0.002	[0.0342,0.1496]
NHNBW > Median	2,570	1,152	0.083 (0.0396)	0.037	[0.0051,0.1602]
VLA > Median	2,560	1,153	0.070 (0.0395)	0.075	[-0.0071,0.1478]
LC Classification 1 NHNBW	4,745	2,382	0.069 (0.0320)	0.031	[0.0061,0.1315]
LC Classification 1 VLA	4,718	2,370	0.069 (0.0321)	0.031	[0.0064,0.1322]
LC Classification 2 NHNBW	2,868	1,258	0.092 (0.0384)	0.017	[0.0167,0.1672]
LC Classification 2 VLA	2,839	1,249	0.089 (0.0385)	0.021	[0.0133,0.1642]
LC Classification 3 NHNBW	2,536	1,137	0.085 (0.0397)	0.032	[0.0072,0.1627]
LC Classification 3 VLA	2,554	1,151	0.070 (0.0396)	0.073	[-0.0067,0.1484]

Table 6: Differences in Treatment Effects: Probit Model

	Baseline Treatment Effect	Interaction	P>|z| (Interaction)	95%CI (Interaction)
NHNBW > Median	0.859 (0.2703)	-0.159 (0.1598)	0.319	[-0.4724,0.1540]
VLA > Median	0.923 (0.2691)	-0.260 (1599)	0.105	[-0.5730,0.0539]
LC Classification 3 NHNBW	0.847 (0.2695)	-0.142 (0.1591)	0.372	[-0.4538,0.1698]
LC Classification 3 VLA	0.919 (0.2688)	-0.255 (0.1595)	0.111	[-0.5674,0.0582]
Continuous Interaction NHNBW	0.855 (0.2520)	-9.18e-07 (3.48e-07)	0.008	[-1.60e-06,-2.35e-07]
Continuous Interaction VLA	0..854 (0.2519)	-9.28e-07 (3.54e-07)	0.009	[-1.62e-06,-2.34e-07]

Table 5 and 6 replicate tables 1 and 2 for the probit specification. For interpretability, table 5 cites the average marginal effect at the median of the subpopulations wealth distribution and quotes the p values of this marginal effect; the coefficients themselves are always significant at the same level as their corresponding marginal effect. As 6 is intended to test the difference of the treatments the p values of the coefficients are quoted. The results are largely comparable to the random effect model both in terms of significance level and in terms of the magnitude of the effect.

5 Rational Expectation Model

In this section I use a model to investigate the ability of RE to match the treatment effect of being below the SPA on labour market participation across the wealth distribution observed in the preceding section. This model is based on O'Dea (2018) who develops an RE life-cycle model that closely matches asset and participation profiles of older individuals in the UK. The model here, adapted from O'Dea(2018), incorporates sufficient features to match aggregate profiles but drops some features irrelevant to the present investigation.

The model contains 4 types, $i \in \{1, 2, 3, 4\}$, differentiated by high or low education and having access or not to a direct benefit (DB) pension scheme. Agents are rational expected utility maximisers who choose how much to consume c_t, how much to invest in a risk-less asset $a_t$with return r, and whether to work dependent on not being involuntarily unemployed. The agent, conditional on not receiving a negative unemployment shock

$u_t = 1$, receives a stochastic income offer y_t each period. Unemployment status is considered verifiable so that only if $u_t = 1$ can the agent claim benefit b. If the agent is not unemployed, $u_t = 0$, she receives a stochastic income offer y_t, accepting the offer gives them an income of y_t and leisure time of $l_t = 1 - w_h$. Her partner is modelled deterministically and earns a fixed amount $spouseInc$ each period until he reaches 65 after which time he retires and earns the state pension p. The agent receives the same state pension, p, as her partner once she reaches the SPA which is a parameter that is varied to mimic the UK reform. However, unlike the partner she does not automatically retire and she receive this transfer whether she works or not. Types that have access to a defined benefit pension can claim this at age 65 conditional on leaving work. The value of the defined benefit pension is a function of average life time earning $AIME_t$. From age 60 the agent is exposed to a stochastic survival probability s_t. Finally, agents value bequest through a warm glow bequest function (De Nardi, 2004; French, 2005).

Since the impact of the state pension on retirement decisions is the focus of this working paper, it is worth interrupting outlining the rest of the model to explain why I model the state pension as I do. Modelling the state pension as a fixed transfer p upon reaching the SPA is incorrect in two regards: individuals can choose to delay receipt of the state pension and the state pension does have components that are dependent both directly and indirectly on life-time earnings. The reason to ignore the first issue is, as Crawford and Tetlow (2010) find, only ≈2% take up this option to defer receipt of the state pension, despite it being more than actuarially fair, which is difficult to account for using RE. Introducing this option would add an additional puzzle to explain. The lump sum transfer modelling of the state pension carried out here does capture the essence of the basic state pension which is a fixed transfer. However, the basic state pension is reduced by a proportional amount if individuals do not meet the minimum number of years of national insurance contributions. Additionally, the state second pension is directly based on earnings. Further complications arise as individuals now claiming the state pension have entitlements that were accrued under different systems, such as SERPS, producing complex and abstruse rules (Bozio, Crawford, and Tetlow, 2010 provide a detailed history). However, the potential to accrue additional pension entitlements is very limited within the UK pension system and I do not believe that this current oversimplification has significant implications for qualitative predictions of this model.

So the model can be summarised as the agent solving the following problem to find policy functions for consumption $c_t(a_t, y_t, AIME_t)$, leisure $l_t(a_t, y_t, AIME_t)$, and assets $a_{t+1}(a_t, y_t, AIME_t)$:

$$\max_{c_t, l_t, a_{t+1}} \sum_{t=start}^{End} \beta^t s_t E[u(c_t, l_t)] \tag{2}$$

$$\text{s.t } c_t + a_t = (1+r)a_{t-1} + y_t \mathbb{1}[l_t = 1 - w_h] + b\mathbb{1}[l_t = 1 \wedge u_t = 1] + p\mathbb{1}[t \geq SPA] \tag{3}$$

$$+ spouseInc\mathbb{1}[t < 65] + p\mathbb{1}[t \geq 65] + \mathbb{1}[l_t = 1]\mathbb{1}[t \geq 65]db(i, AIME_t) \tag{4}$$

$$a_t \geq 0,\ l_t|(u_t = 0) \in \{1 - w_h, 1\}, \text{ and } l_t|(u_t = 1) \in \{1\}$$

$$\text{where } u(c_t, l_t) = \frac{(c_t^{\nu} l_t^{1-\nu})^{1-\gamma}}{1-\gamma}|alive + \theta\frac{(a_t + K)^{\nu 1-\gamma}}{1-\gamma}|deceased \tag{5}$$

Average earning evolving according to:

$$AIME_t = \begin{cases} \frac{AIME_{t-1}(t-1) + \mathbb{1}[l_t = 1 - w_h]y_t}{t} & t < 65 \\ AIME_{t-1} & t \geq 65 \end{cases} \tag{6}$$

The defined benefit pension has the functional form:

$$db(AIME_t) = \begin{cases} db_1 AIME_t - db_2 AIME_t^2 & AIME_t < AI\bar{M}E \equiv \frac{db_1}{2db_2} \\ db_1 AI\bar{M}E - db_2 AI\bar{M}E^2 & \text{otherwise} \end{cases} \tag{7}$$

The log income offer, y_t, is the sum of a deterministic component, quadratic in age and specific to the agent's type, and a stochastic component:

$$log(y_t) = \delta_{i0} + \delta_{i1}t + \delta_{i2}t^2 + \epsilon_t \tag{8}$$

where ϵ_t follows an AR1 with normal error term and an initial distribution $\epsilon_1 \sim N(0, \sigma^2)$.

The income offer can be conceptualised as being equal to some underlying productivity which the agent maintains during unemployment spells. The unemployment status of the agent u_t evolves according to a conditional markov process, where the probability of unemployment is dependent on current productivity y_t and the type of the agent.

The model starts with agents aged 52. The reasons to start agents so far into the life-cylce are, firstly, the ELSA dataset only starts interviewing people over 50 and, secondly, the period I am interested in is around retirement and so modelling early life-cylce behaviour would be computationally wasteful. The reason to start at 52 rather than 50 is that this is the youngest age with interviews from a large number of people some with SPA equal to 60 and some with an SPA strictly greater than 60. The agents start life with a draw from the empirical

Table 7: Parameter Estimates

Parameter	Estimate
γ	0.288
ν	0.986
β	2.320
db_1	0.5914
db_2	-4.232E-006
θ	2.899E-002

distribution of assets at age 52. To make sure the endogeneity of the SPA to the quantity of assets chosen by age 52 does not bias the model I used a Kolmogorov–Smirnov test to test the null that assets conditional on distinct SPAs are drawn from the same distribution and found that the data do not reject this null. If the age 105 is reached the agent dies with certainty. From age 80 the agent no longer has the choice of working; this is to model some of the limitation imposed by declining health.

As described, the type of the agent introduces heterogeneity into whether the agent receives a DB pension, the earning process, and the probability of unemployment, but I do not allow for preference heterogeneity over type. The calibrations of the earning process, the unemployment probabilities, and the curvature of the warm-glow bequest are taken from O'Dea (2018). O'Dea (2018) estimates these processes for the same dataset as I use but O'Dea (2018) estimates for the principal earner in the household which would predominantly be men while I am interested in the retirement behaviour of women. There are undoubtedly differences in the male and female earning process for these generations and in future version of this working paper I will estimate the earning process for women in the ELSA dataset. For this reason, despite the fact I will match moments to estimate this model, it is more qualitative and stylistic than quantitative at present. Selection is corrected for by using the correction coefficient taken from O'Dea (2018) who implements the French correction (French, 2005). The values for the level of benefits b and the state pension p are set to the 2012 levels of job seekers allowance and the basic state pension. All prices are deflated to a common year using the RPI.

The moments used to find the parameters are the proportion of women working between 52 and 75 and the level of household non-housing, non-business wealth (NHNBW) between the same ages. These moments are used to find the preference parameters γ, ν, β, and θ as well as the parameters of the defined benefit pension function db_1, and db_2.

Figures 2 and 3 show the match of the participation and asset profile which are acceptably close given the coarseness with which the model is currently estimated (for example 30 grid points for assets). Table 7 contains the parameter estimates. Given the qualifications mentioned above, I only give to these number to show they are not wildly different from estimates in the literature and I have not calculated standard errors.

These profiles were estimated with $SPA = 60$ and against the moments of the pre-reform data. Once the

Figure 2: Participation Profile

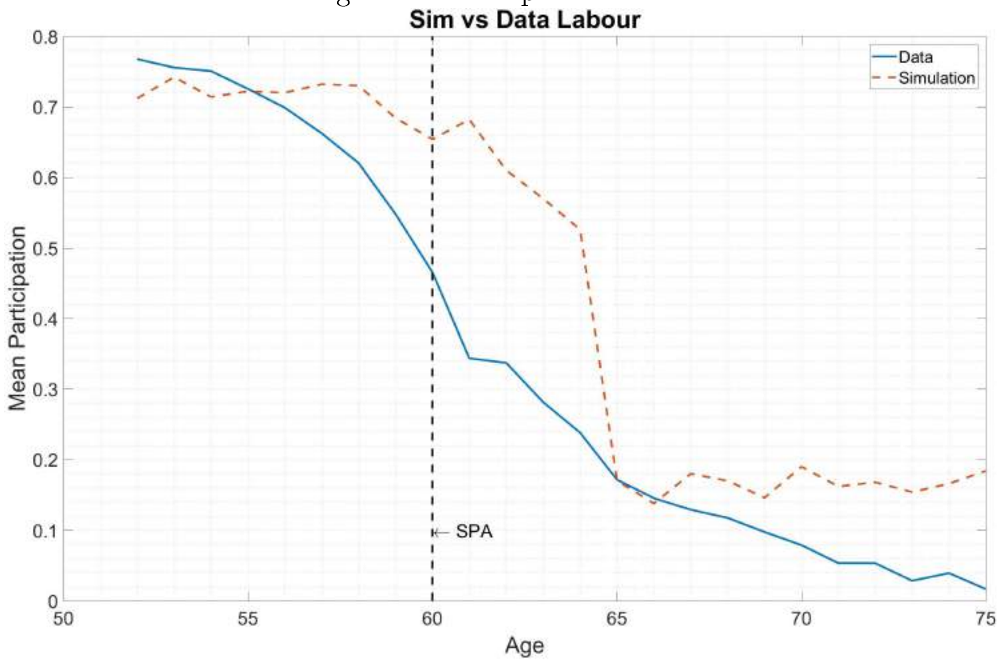


Figure 3: Asset Profile

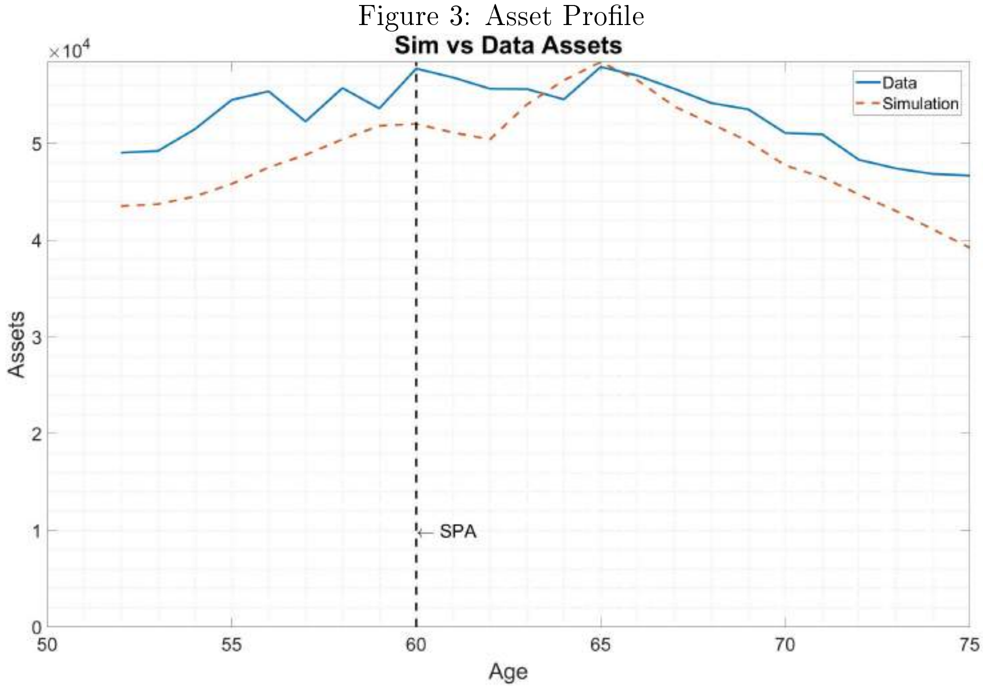


Table 8: Regression Analysis on RE Simulations Data

Population	Observations	Groups	Treatment Effect	P>|z|
1	36,000	1,500	0.0749	0.000
2	17,976	749	0.0303	0.174

parameters were estimated, I re-ran the model to generate simulated data with $SPA = 60$, $SPA = 61$, and $SPA = 62$ and re-ran the regression analysis from section 4 on this data. As can be seen in table 8 a significant treatment effect of being below the SPA is observed for the whole sample but unlike the real world data this treatment effect falls by half when we restrict to people with sufficient assets. So this rational expectation model matches the aggregate profile but fails to match the impact of the SPA on labour force participation across the distribution: exaggerating the effect on the poor and under predicting the effect on the wealthy. In the next section I attempt to ameliorate this issue by introducing rational inattention to the SPA.

6 Rational Inattentive Model

In this section I consider the implications that costly thought could have for this retirement decision. First in subsection 6.1, I describe rational inattention and how I will incorporate it into the model presented in section 5. Then in subsection 6.2, I solve and analyse the model.

6.1 Incorporating Incorrect Beliefs with Rational Inattention

The second retirement puzzle this paper seeks to address is that people are consistently mistaken about their own retirement provision (Gustman and Steinmeier, 1999 & 2001; Rohwedder and Kleinjans, 2006). In particular, women subject to the UK SPA reform are misinformed about their SPA (Crawford and Tetlow, 2010; Amin-Smith and Crawford, 2018); for example, 41% of 58-year-old women with a SPA between 60-64 don't know their own SPA to within a year (Amin-Smith and Crawford, 2018). This is a difficult fact to accommodate in a standard RE life-cycle model for two reasons: policy parameters are not stochastic and information is acquired without cost. To address these difficulties I propose to model the SPA as stochastic and make acquisition of information about the SPA costly using a rational inattention approach.

Making a policy parameter like the SPA a stochastic variable is unusual. In fact, to the best of my knowledge, this is the first paper to do so and this I think represents a valuable contribution in itself. Hence, treating the SPA as stochastic requires some justification. Firstly the state pension does changes; a women born in 1954 would have had an SPA of 60 when she entered the workforce in the 1970s but this would have been changed to 65 in 1994 and subsequently to 66 in 2011. Assuming the SPA is a parameter known from entry into the labour force is counter factual. Secondly, although the actual cost of finding out your SPA is tiny, this model is, as are all models, a simplification from the real world where there are many more sources of uncertainty. The cost of finding out our SPA can be conceptualised as the opportunity cost of the time spent identifying the SPA or a stand in for more complicated aspects of state pension entitlement not explicitly modelled.

The stochastic process I will use to model the SPA is:

$$SPA_{t+1} = \begin{cases} SPA_t + e_t & SPA_t < 70 \\ SPA_t & SPA_t = 70 \end{cases} \tag{9}$$

$$\text{where } e_t \in \{0, 1\} \text{ and } e_t \sim Bern(p)$$

When the SPA is below 70, the process is a random walk with a skewed innovation as $e_t \in \{0, 1\}$. So the SPA either increases by one year with probability p in a given year or it stays the same. This process accommodates the idea that in recent history governments have reformed SPA upwards but generally not downward. This process is a parsimonious and analytically tractable model of pension reform. Although it does have some counter factual predictions, in particular, that the SPA can increase by at most one year per year whilst many individuals saw their SPA rise by multiple years in 1995, I believe it captures the essence of pension reform. The restriction to $SPA_t \leq 70$ was partially motivated by computational considerations but can be understood as there being some upper limit beyond which the government will not increase the SPA. It could seem unrealistic to have a model in which the small probability of the government increasing the SPA to 103 impacted on individual's decisions. Without this restriction the prior probability of SPA being greater than 70 by the time this age is reached is less than 0.02%, so this simplification should not have too large an impact whatever position is taken on the plausibility of this upper bound.

I incorporate this stochastic SPA into the model described in section 5, so the model starts with agents aged 52 but they are understood to start working life at age 20 with a SPA of 60. The agents are imperfectly informed of their SPA, which is explained in more detail below, and I make the assumption they are unable to acquire additional information before the start of the model at age 52. So they start with the posterior belief that arises from entering the workforce at age 20 believing with certainty that SPA=60 and then applying Bayesian updating in each period given the process above. Once an agent reaches their SPA their pension cannot be taken away from them. The probability p is estimated outside the model to match the actual SPAs of women born between 1950 and 1954 assuming these were generated by the process in equation 9. The probability of the SPA increasing taken from the data by this process is 6%.

A stochastic SPA alone would not explain people being mistaken about their SPA. For this I need to incorporate costly thought and this is modelled using the rational inattention approach pioneered by Sims (1998, 2003). This acknowledges the costly nature of thought whilst abstracting away from psychological details. In rational inattention agents suffer a disutility cost for receiving more information as measured by the entropy of the signal they receive about the state variables.

For ease of exposition, I will use a simplified model to explain how rational attention works in my life-cycle setting, although later I introduce rational inattention into the full model described in section 5. In this simplified model the only states are the income offer y_t, assets a_t, and the stochastic SPA SPA_t,

$$\max_{c_t, l_t, a_{t+1}} \sum_{t=0}^{T} \beta^t s_t E[u(c_t, l_t)]$$

s.t some constraints

In a fully rational model the agent solves for policy functions for consumption $c_t(a_t, y_t, SPA_t)$, leisure $l_t(a_t, y_t, SPA_t)$, and assets $a_{t+1}(a_t, y_t, SPA_t)$ to solve the problem above. In the rationally inattentive model the agent is not able to directly observe the SPA but can only perceive a noisy signal of it $Z_t \sim f_t(z_t|SPA_t, a_t, y_t)$. She can choose the distribution of the signal and make it as precise as she likes but she receives a disutility for receiving a more precise signal proportional to the mutual information between signal and SPA $I(Z_t, SPA_t)$. Her policy function can no longer depend on the SPA but only on her beliefs as to what her SPA is. As all agents start with the same prior belief at age 52 their belief in period t is completely determined by the histories of draws of Z_t they have receive up to that point, z^t. So now the agent chose $Z_t \sim f_t(z_t|SPA_t, a_t, y_t)\}$ and policy functions $c_t(a_t, y_t, z^t)$, $l_t(a_t, y_t, z^t)$, and $a_{t+1}(a_t, y_t, z^t)$ to solve:

$$\max_{c_t, l_t, a_{t+1}, f_t} \sum_{t=0}^{T} \beta^t s_t E[u(c_t, l_t) - \lambda I(SPA_t; Z_t)] \tag{10}$$

s.t some constraints

The penalty for receiving a more precise signal is proportional to the mutual information $I(Z_t, ERA_t)$ which is a concept from information theory. It measures the expected reduction in uncertainty from receiving a signal, where uncertainty is measured by entropy $H(.)$:

$$I(SPA_t; Z_t) = H(SPA_t) - E_Z[H(SPA_t|Z_t)]$$

Entropy is the central concept of information theory and is defined as $H(Z) \equiv -E_Z[\log(f_Z(Z)]$ where $Z \sim f_z$. If the base of the logarithm is taken to be 2 then the entropy is the minimum number of bytes required to communicate the information contained in a random variable; if the logarithm is to a different base then entropy represents the same quantity but measured in a different unit. As such it is an easily understandable measure of uncertainty and the most precise measure of the amount of information received by an agent.

The introduction of rational inattention greatly complicates this model for two reasons. Firstly, it introduces a very high dimensional choice variable in f_t. Since SPA_t has finite support we can restrict, without loss of generality, the support of f_t to be discreet and finite. In this case, f_t is a finite dimensional object but is still very high dimensional having a dimension of $Dim(SPA_t) - 1$. Secondly, it introduces a large and unobservable state to the agents decision problem in the form of their posterior belief or, equivalently, the history of signals they have observed z^t.

For these reason, solving rational inattention models is notoriously difficult. Most approaches either make a lot of simplifying assumptions, like quadratic utility (Maćkowiak, Matějka and Wiederholt, 2016), or use numerical methods that assume interior solutions (Maćkowiak and Wiederholt, 2015). The method I use is from Steiner, Stewart, and Matějka (2017) (henceforth SSM) who solve a general class of dynamic discrete choice models without additional simplifying assumptions. However, as SSM's result is for a discreet choice model it requires that I discretise the choice variable. As labour choice is already discreet, this only implies a need to discretise assets. Some other minor extension and adaptations of SSM's result were required and I explain these in Appendix A.

Before outlining the results from SSM which I rely on to solve this model, it is convenient to introduce some notation. Firstly, for brevity I will denote by d_t the agent's decision $d_t = (a_{t+1}, l_t)$ and then I will re-express the agent's utility function as a function of d_t and the states, X_t, by substituting out consumption via the budget constraint to give $u(d_t, X_t)$. Secondly, let W_t denote the states that the agent freely and costlessly observes, that is all the states except the SPA, $W_t = (a_t, y_t, AIME_t, u_t)$. The principal result of SSM that I use to solve this model is that the solution of a general class of dynamic discreet choice RI problems is a dynamic logit rule (Rust, 1987) with a bias in form of default rule

$$p_t(d_t|X_t) = \frac{\exp(v_t(d_t, X_t))}{\sum_{d'_t \in D} \exp(v_t(d'_t, X_t))} \tag{11}$$

where

$$v_t(d_t, X_t) = u_t(d_t, X_t) + \log q_t(d_t|W_t) + \beta E[V_{t+1}(X_{t+1})|d_t, X_t]$$

$$V_t(X_t) = \log(\sum_{d_t \in D} \exp(v_t(d_t, X_t)))$$

for the default rule $q_t(d_t|W_t) = E_{x_t}[p_t(d_t|X_t)|W_t]$.

This result differs from a standard dynamic logit only by the addition of the default rule. The default rule is the expected action at a point in the freely perceived state space W_t over the part of the state space that incurs an information cost SPA_t. Hence, the $\log q_t(d_t|W_t)$ term represents the original utility cost of information in the solution as the more information they pay to receive the more their actions will depend upon SPA_t and so the larger $\log q_t(d_t|W_t)$. Taken to the extreme we can see that if the agent takes the same action in all eventualities of SPA_t, then $q_t(d_t|W_t) = 1$ and the penalty term disappears.

This is a surprising and powerful result and it is worth emphasising a couple of points about it. Firstly, the SPA_t in the vector of states $X_t = (a_t, y_t, AIME_t, u_t, SPA_t)$ is the true SPA not the agents belief or any transformation of it. The relevant states for any agent faced with this problem would contain the posterior or history of signal received replacing the SPA_t $(a_t, y_t, AIME_t, u_t, z^t)$. SSM are able to bypass the difficulty of having this large and unobserved state by showing that the original problem is mathematically equivalent to a problem with observable states and then solve this equivalent problem. Hence, X_t is a computational state vector rather than the agent's state when faced with the problem. Secondly, the logit result is not derived by introducing preference shocks as is normally the case. Instead, the logit results arises due to a deep mathematical connection between the entropy and the logit distribution which has been known and exploited since the early information theory literature (Jaynes, 1957; Shannon, 1959).

By providing this analytic solution to rationally inattentive dynamic discreet choice models, SSM solve one of the two difficulties mentioned above: replacing the large unobserved state in the agents problem with the observed state SPA_t. However, the problem of the high dimensionality of the solution remains as $p_t(d_t|X_t)$ is a $(|Supp(d_t)| - 1)$ dimensional object and this implies a high computational cost as explained below.

Unlike in the traditional dynamic logit, the conditional choice probabilities now appear on both side of the equation once we substitute the definition of the default rule q_t into equation 11

$$p_t(d_t|X_t) = \frac{\exp(u_t(d_t, X_t) + \log E[p_t(d_t|X_t)|W_t] + \beta E[V_{t+1}(X_{t+1})|d_t, X_t])}{\sum_{d'_t \in D} \exp(u_t(d'_t, X_t) + \log E[p_t(d'_t|X_t)|W_t] + \beta E[V_{t+1}(X_{t+1})|d'_t, X_t])} \tag{12}$$

This equation contains conditional choice probabilities for all values of d_t in the denominator and all value of SPA_t in the penalty for a non-prescriptive default rule $\log E[p_t(d_t|X_t)|W_t]$. Hence equation 12 defines a fixed point and as p_t is a high dimensional object this is computationally costly. I use SSM's results to solve the model of section 5 incorporating both rational inattention and the stochastic SPA process described in the current section. I solve the model by backwards induction; until age 68 is reached the problem in each period is a fully rational problem and can be solved by maximising utility. At age 68 RI begins to play a part as at ages 70 and above the agent is in receipt of the state pension with certainty and at age 69 if she has not received her pension she knows the SPA must currently be 70. Once age 68 is reached I solve equation 12 as a fixed point iteration for each point in the costlessly observed state space W_t and continue to apply backward induction with this fixed point iteration replacing the UMP within the period.

6.2 Rational Inattention Results

I apply the results and methods described in the last section to solve the RI version of the model described in section 5. That is households who choose the distribution of the signal they will receive $Z_t \sim f_t$ and their actions d_t given costlessly observed states and their history of signal draws to solve:

$$\max_{d_t, f_t} \sum_{start}^{finish} \beta^t s_t E[u(d_t, X_t) - \lambda I(SPA_t, Z_t)] \tag{13}$$

subject to the constraints and exogenous process outlined in equations 3-9.

The requirement to solve a fixed point iteration at each point of the costlessly observed state space is computationally burdensome. For this reason I have yet to estimate this model. Instead I have investigated some of its dynamics for a few chosen values of λ with all other parameter values taken from the estimates of the RE model. The three values investigated are $\lambda = 1.0$, 0.01, and 0.001, the first two were too large to be interesting so only I present the result for $\lambda = 0.001$ here. This cost of attention parameter implies to be fully informed of the SPA would incur a cost per period 1.76 times larger than the life-time utility gain to the median household of a 1% increase of consumption in all contingencies.

As can be seen in figures 4 and 5, in the aggregate profiles the most noticeable change between the RE and the RI model is an increase in asset holding and participation amongst households below the SPA. Surprisingly as the value of λ was arbitrary chosen, this has improved the fit as compared to the RE model. For comparability, the RE model presented here contains a stochastic SPA, like the RI model, but, unlike the RI model, the agent is fully informed of their own SPA. This represent an additional change to the baseline presented in section 5 beyond the discretisation needed to apply SSM. This distinction, however, is unimportant as the prediction of the RE model with and without a stochastic SPA differ only in second and third order terms so as to be almost indistinguishable. This does show that RI is crucial to the results obtained here, as introducing a stochastic SPA alone produces no describable difference.

Viewed in general terms this change in the asset profile can be understood as a response to the increased precautionary saving motive induced by RI. However the distributional details are very important. The increases in mean asset holdings is driven by two groups. One group, very wealthy outliers whose small increase in savings has a disproportionately large impact on the mean. The second group, the poor, for whom precautionary saving is more important and so increase savings by a larger proportionate amount than other groups. The impact of this increase in saving amongst the poor can be detected in the summary statistics of the wealth distribution in table 9.

Neither of these groups dramatically changes their participation. The very rich do not work in either models and the poorest work in both, so the question as to what is driving the increase in participation in the RI model

Figure 4: RI vs RE Labour Market Participation Profile SPA = 60:

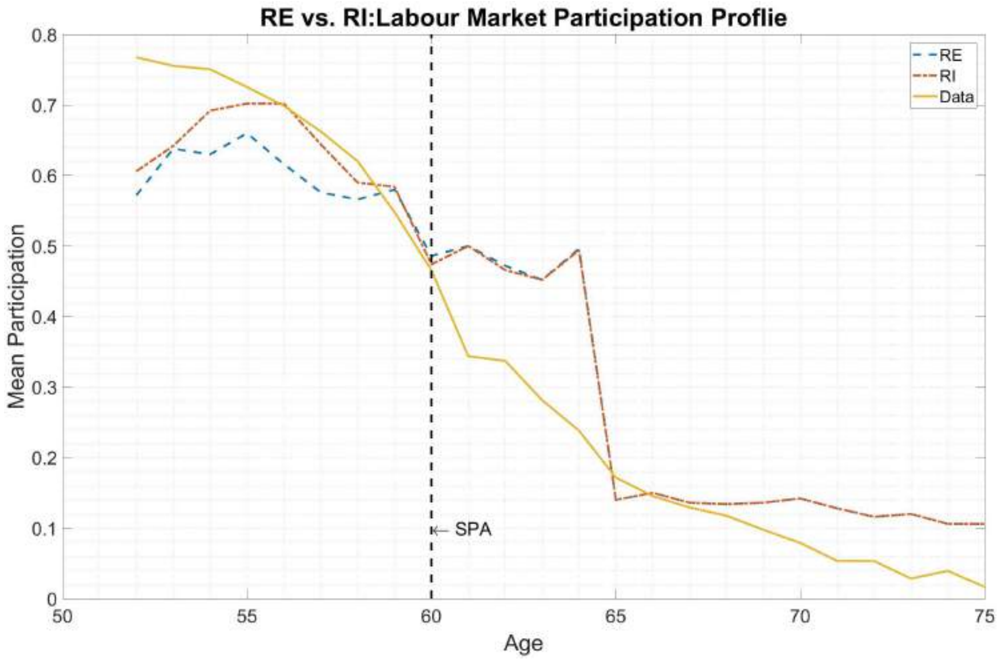


Figure 5: RI vs RE Asset Profile SPA = 60:

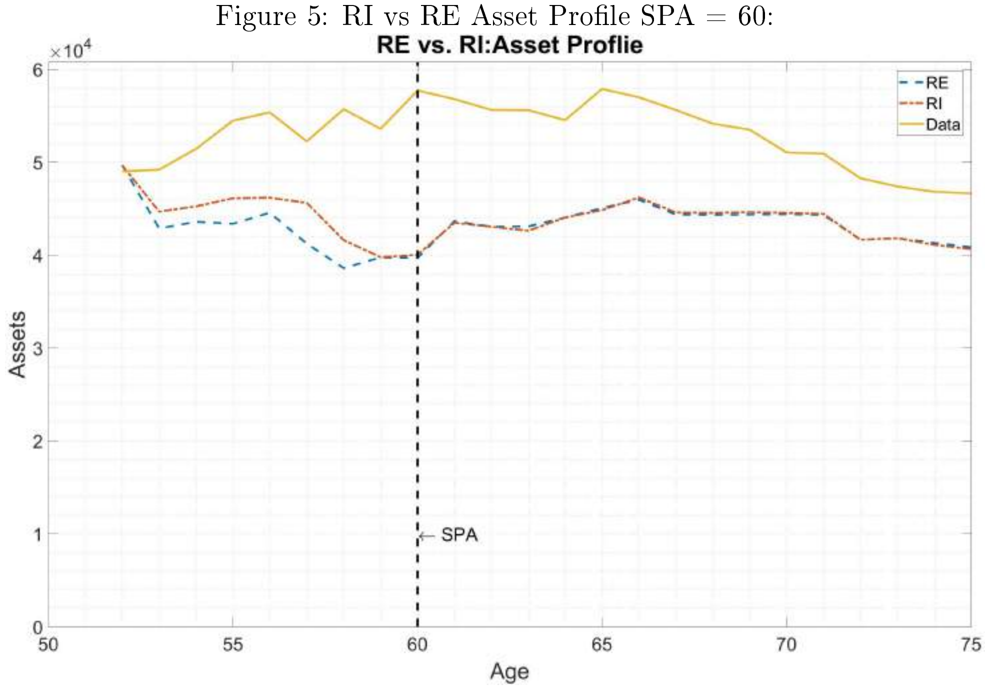


Table 9: RE vs. RI Asset Distribution Summary Statistics

Summary Statistics Assets RE					Summary Statistics Assets RI				
	Percentiles	Smallest				Percentiles	Smallest		
1.00%	0	0			1.00%	0	0		
5.00%	0	0			5.00%	0	0		
10.00%	0	0	Obs.	36,00	10.00%	0	0	Obs.	36,00
25.00%	888.54	0	Sum of Wgt.	36,00	25.00%	1545.30	0	Sum of Wgt.	36,00
50.00%	8,418.47		Mean	43315.26	50.00%	8,418.47		Mean	43645.54
		Largest	Std. Dev.	161599.6			Largest	Std. Dev.	161658
75.00%	26,057.28	2391046	Variance	2.61E+10	75.00%	26,057.28	2391046	Variance	2.61E+10
90.00%	82,276.21	2391046	Skewness	9.204104	90.00%	82,276.21	2391046	Skewness	9.189518
95.00%	154,702.70	2391046	Kurtosis	111.2387	95.00%	154,702.70	2391046	Kurtosis	111.0051
99.00%	6,20,312.00	2391046			99.00%	6,20,312.00	2391046		

Table 10: Treatment effect RE vs RI

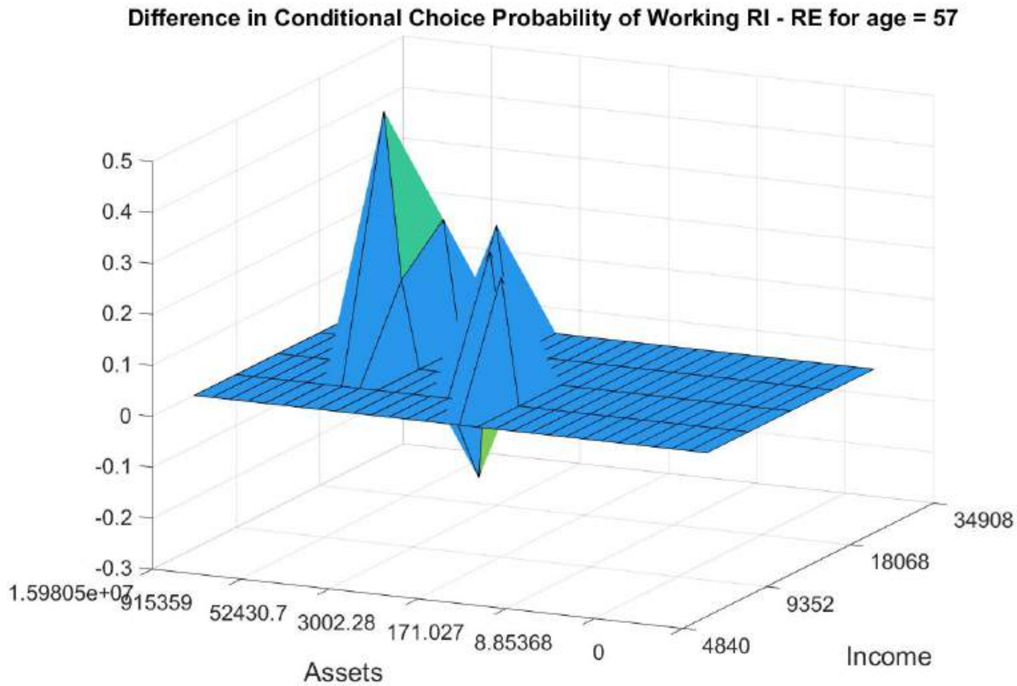
	RE		RI	
	Treatment	p	Treatment	p
Whole Population	0.092	0.000	0.093	0.000
Above Median Asset in SPA-1	0.022	0.258	0.042	0.026

remains. As can be seen from figure 6, which plots the difference in probability of participation between the RE and RI model, the increased uncertainty from RI make some agents, whose labour market attachment is marginal, switch their decision. As the participation is a discrete choice this can go both ways but the dominating effect is to increase participation. Individuals with marginal participation tend to have above median assets, hence RI increases participation amongst the top half of the asset distribution.

As the RI uncertainty is resolved upon reaching SPA this might help explain the bunching of exits at SPA and repeating the regression used in section 6.2 we see that is indeed the case. In table 10, there is a much smaller reduction in the treatment effect in the RI model when we restrict to those with above median assets in the period before SPA. Although the potential for this mechanism exists with the introduction of the stochastic SPA alone, for the empirically calibrated levels of SPA uncertainty used here, the amplification of uncertainty by RI is crucial to generate any discernible difference.

An intuition for these results can be gleaned from considering the trade-off faced by the agents. Figure 7 displays in the top panel a schematic representation of the utility function of an agent at a point in the state space if they choose to work and if they do not. Any point on the x-axis is a choice of next periods assets with the value increasing from right to left. This can be thought of roughly as consumption expressed in the standard left-to-right direction although the exact consumption bundle is different for the two functions, working and not working, as the same level of next period assets implies a higher level of consumption if you work. The reason for having the x-axis in these terms is to accord with the bottom pane which shows expected marginal

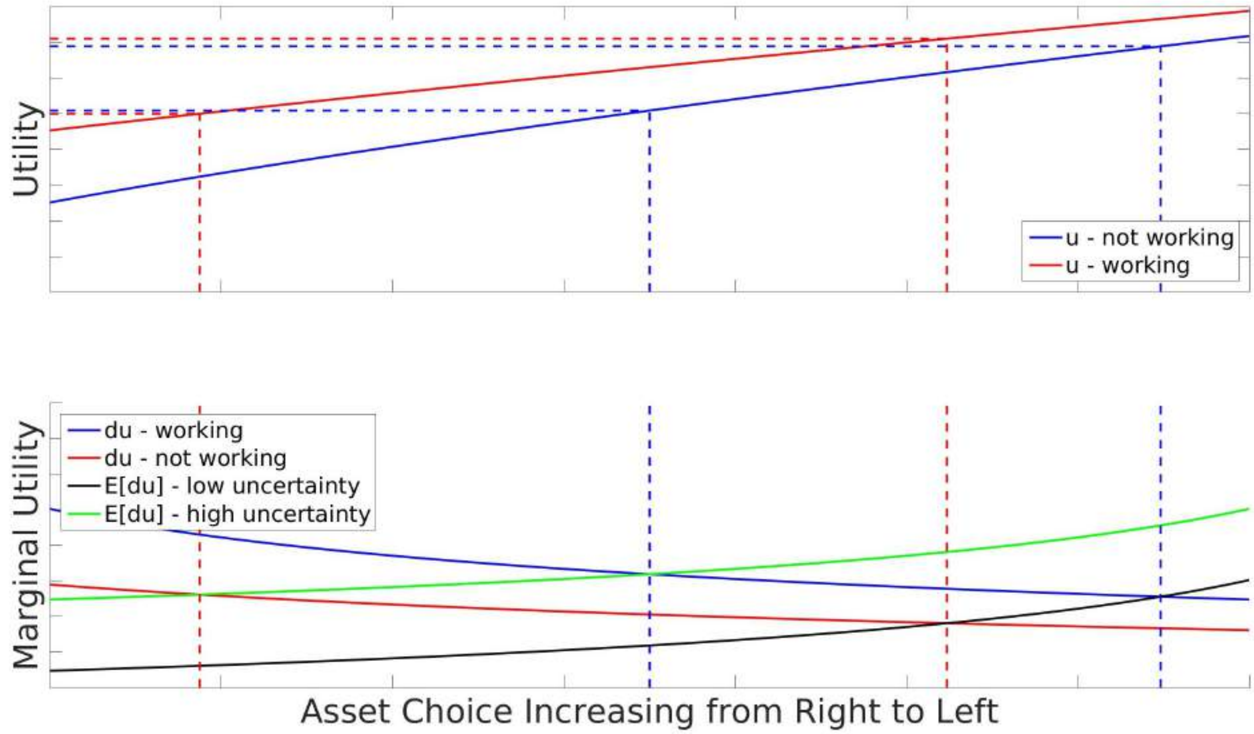
Figure 6: Difference in Conditional Probability of Working RI-RE at age 57



utility next period as a function of the asset choice. This bottom pane is used to find the asset choice level that equalise today's marginal utility with the expected marginal utility next period. As the participation decision is discrete, the agent then chooses which of these two optimal decisions, conditional on working status, produces the highest utility and selects whether to work accordingly. This can be found in the top panel by comparing which intersection point is higher. I have done this for high uncertainty, representing RI in the discussion above, and low uncertainty, representing RE. As the utility function here displays prudence, marginal utilities are convex and so, by Jensen's inequality, increasing uncertainty shifts the marginal value of assets next period upwards. In this diagram we see how this increase in uncertainty can flip the participation of an agent with marginal labour market attachment. Here I show the case where they work under low uncertainty and don't under high but I could equal draw it the other way.

RI is more than the introduction of more uncertainty. It also introduces another channel for the agent to optimise over: the precision of the signal. This channel can be understood in this diagram as the ability to shift the expected marginal value of assets tomorrow down by reducing the uncertainty but only at the cost of also shifting the utility function down. This is the central trade-off introduced by RI but my exploration of this channel is still very preliminary. Investigating this channel, and the informativeness of the signals chosen by agents at different points in the state space, has implications for why people are misinformed about their SPA. One interesting result is that, except at a handful of points, agents choose to receive very little information and these points tend to be found in the upper half of the asset distribution. This prediction agrees with a finding

Figure 7: Schematic Representation of the UMP facing a RI agent



by Rohwedder and Kleinjans (2006) that richer people are more likely informed of their social security provision.

7 Conclusion

This working paper makes contributions to two areas of research. Firstly, it adds to the evidence that retirement choices are more sensitive to legislated pension ages than rational expectations can account for. It does this in the context of the UK female state pension age reform by more carefully controlling for assets and by demonstrating that a state-of-the-art RE model fails to match the observed treatment effect for individuals across the distribution of assets: it exaggerates the impact of the SPA on poorer individuals whilst underestimating the impact on wealthier agents. Secondly, the paper contributes to the rational inattention literature by being the first, to the best of my knowledge, to incorporate RI within a life-cycle model. Doing this allows a life-cycle model to accommodate another well established finding: that people hold mistaken beliefs about their own provisions for retirement. Results from the RI model are preliminary but they offer some insight into the dynamics RI introduces and indicate that RI has the potential to illuminate bunching of labour market exits at SPA as well as people's ignorance of their pension provision. The key mechanism behind this result is that by allowing for uncertainty in the SPA, resolved upon reaching SPA, introduces additional precautionary saving; thus, inducing greater labour market participation pre-SPA. Crucially, introducing reasonable levels of uncertainty about the SPA without RI only negligible increases the sensitivity of labour force participation to the SPA. It is the differential and endogenous amplification of this uncertainty by rational inattention that allows the model to better match the sensitivity of labour market choices to the SPA across the wealth distribution.

References

[1] Amin-Smith, N. and Crawford, R. 2018. Unpublished Manuscript, 2018

[2] Atalay, K. and Barrett, G.F., 2015. The impact of age pension eligibility age on retirement and program dependence: Evidence from an Australian experiment. Review of Economics and Statistics, 97(1), pp.71-87.

[3] Behaghel, B. L., & Blau, D. M. (2012). Framing Social Security Reform : Behavioral Responses to Changes in the Full Retirement Age. American Economic Journal: Economic Policy, 4(4), 41–67.

[4] Bozio, A., Crawford, R. and Tetlow, G., 2010. The history of state pensions in the UK: 1948 to 2010.

[5] Carroll, C. D., & Samwick, A. A. (1996). The Nature of Precautionary Wealth. NBER Working Paper.

[6] Crawford, R., & Tetlow, G., 2010. Employment, retirement and pensions. Financial Circumstances, Health and Well-Being of the Older Population in England: ELSA 2008 (Wave 4), 11–75.

[7] Cribb, J., Emmerson, C. and Tetlow, G., 2013. Incentives, shocks or signals: labour supply effects of increasing the female state pension age in the UK (No. W13/03). IFS working papers.

[8] Cribb, J., Emmerson, C. and Tetlow, G., 2016. Signals matter? Large retirement responses to limited financial incentives. Labour Economics, 42, pp.203-212.

[9] De Nardi, M., 2004. Wealth inequality and intergenerational links. The Review of Economic Studies, 71(3), pp.743-768.

[10] French, E., 2005. The effects of health, wealth, and wages on labour supply and retirement behaviour. The Review of Economic Studies, 72(2), pp.395-427.

[11] Gustman, A.L. and Steinmeier, T.L., 2001. Imperfect knowledge, retirement and saving (No. w8406). National Bureau of Economic Research.

[12] Gruber, J. and Wise, D. (Eds.), 2004. Social Security Programs and Retirement around the World: Micro-Estimation. University of Chicago Press.

[13] Jaynes, Edwin T. 1957. "Information Theory and Statistical Mechanics." Physical Review 106 (4): 620–30.

[14] Lalive, R., Magesan, A. and Staubli, S.,2017.Raising the Full Retirement Age: Defaults vs Incentives.WP

[15] Maćkowiak, B., & Wiederholt, M. 2009. Optimal sticky prices under rational inattention. American Economic Review, 99(3), 769–803.

[16] Maćkowiak, B., & Wiederholt, M. 2015. Real Business Cycle Dynamics under Rational Inattention. ECB Working Paper, (14089).

[17] Mackowiak, B., Matejka, F., & Wiederholt, M. (2016). The Rational Inattention Filter. CEPR Discussion Papers.

[18] Manoli, D. and Weber, A., 2012. The effects of increasing the early retirement age on social security claims and job exits. Unpublished manuscript, version of May, 11, p.2012.

[19] Matêjka, F., & McKay, A. 2015. Rational inattention to discrete choices: A new foundation for the multinomial logit model. American Economic Review, 105(1), 272–298. https://doi.org/10.1257/aer.20130047

[20] O'Dea, C., 2018. Insurance, Efficiency and Design of Public Pensions. WP

[21] Ravid, D., 2018. Bargaining with rational inattention.

[22] Rohwedder, S. and Kleinjans, K.J., 2006. Dynamics of Individual Information about Social Security. Unpublished manuscript, RAND.

[23] Rust, J. (1987). Optimal Replacement of GMC Bus Engines : An Empirical Model of Harold Zurcher. Econometrica, 55(5), 999–1033.

[24] Seibold, A. 2017. Reference Dependence in Retirement Behavior: Evidence from German Pension Discontinuities. WP

[25] Shannon, Claude E. 1959. "Coding Theorems for a Discrete Source With a Fidelity Criterion" In Institute of Radio Engineers, International Convention Record, Vol. 7, 325–50. New York: Institute of Radio Engineers.

[26] Sims, C. 1998. Stickiness. Carnegie-Rochester Conference Series on Public Policy, 49, 317–356. https://doi.org/10.1016/S0167-2231(99)00013-5

[27] Sims, C. A. 2003. Implications of rational inattention. Journal of Monetary Economics, 50(3), 665–690. https://doi.org/10.1016/S0304-3932(03)00029-1

[28] Staubli, S. and Zweimüller, J., 2013. Does raising the early retirement age increase employment of older workers?. Journal of public economics, 108, pp.17-32.

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

[30] Wiederholt, M. (2010). Rational Inattention. The New Palgrave Dictionary of Economics, (August), 1–8. https://doi.org/10.1016/j.red.2010.01.002

A Extending Steiner, Stewart, and Matejka (2017) and Mapping to Model

In this section I adopt much of the notation of SSM and notation is not related to the rest of the paper.

My variant of their setup is the following. There is a payoff relevant exogenously evolving state $\theta_t \in \Theta_t$ according to measure $\pi \in \Delta(\prod_t \theta_t)$ and agents must make a payoff relevant decision from a choice set D_t. Before making a decision, the agent first observes a costless signal $y_t \in Y_t$, $y_t \sim g_t(y_t|\theta^t, y^{t-1})$ and then can choose any costly signal about θ_t on signal space X_t. Agents get gross flow utilities $u(d^t, \theta^t)$ that can depend on the whole history of state and actions but suffer a utility cost for more precise information $\propto I(\theta^t, x_t|z^{t-1})$ where $z^t = (x^t, y^{t+1})$. It is assumed that $y_{t+1} \perp (x^t, d^t)|(\theta^t, y^t)$. The sets Θ_t, D_t,Y_t, and X_t are finite and that $|D_t| \leq |X_t|$. My setup differs from SSM's in that I adapt the timing assumption so that the costless signal is received before the action is taken each period rather than after it. This change in timing only affects the proof of lemma 1 from SSM's paper and I show below that this results still holds using a slightly different strategy to prove it.

The agent chooses information strategy $f_t(x_t|\theta^t, z^{t-1})$ and action strategies $d_t = \sigma_t(z^{t-1}, x_t)$, collectively referred to as their strategy $s_t = (f_t, \sigma_t)$ to solve

$$\max_{f,\sigma} E[\sum_{t=0}^{T} \beta^t (u(\sigma_t(z^{t-1}, x_t), \theta^t) - I(\theta^t, x_t|z^{t-1})] \tag{14}$$

where the expectation is taken with respect to the distribution over sequences (θ_t, z_t) induced by the prior π together with the strategy $s_t = (f_t, \sigma_t)$ and the distributions g_t of costless signals. The function $u(.,.)$ is assumed continuous. For notational convenience, let $\omega^t = (\theta^t, z^{t-1})$ be the current state and the agents current decision node, or information about the state, then:

Proposition 1. *(Lemma 1 in SSM) Any strategy s_t solving the dynamic RI problem generates a choice rule $p_t(d_t|\omega^t)$ solving*

$$\max_{p} E[\sum_{t=0}^{T} \beta^t (u(d^t, \theta^t) - I(\theta^t, d_t|z^{t-1})] \tag{15}$$

where we redefine $z^{t-1} = (d^{t-1}, y^t)$ the expectation is with respect to the distribution over sequences (θ_t, z_t) induced by p, the prior π, and the distributions g. Conversely, any choice rule p solving 15 induces a strategy solving the dynamic RI problem.

Proof. We precede in steps.

Step 1:First note that for random variable $\zeta_t \in \{x_t, b_t\}$

$$E[\sum_{t=1}^{\infty} \beta^t I(\theta^t, \zeta_t|z^{t-1})] = E[\sum_{t=?}^{\infty} \beta^t (H(\theta^t|\zeta^{t-1}, y^t) - H(\theta^t|\zeta^t, y^t))] \tag{16}$$

But then by the entropic chain rule and that $\theta_t \perp \zeta^{t-1}|\theta^{t-1}$

$$
\begin{aligned}
H(\theta^t|\zeta^{t-1}, y^t) &= H(\theta^{t-1}|\zeta^{t-1}, y^t) + H(\theta_t|\theta^{t-1}, \zeta^{t-1}, y^t) \\
&= H(\theta^{t-1}|\zeta^{t-1}, y^t) + H(\theta_t|\theta^{t-1}, y^t)
\end{aligned}
$$

Also $y_{t+1} \perp (x^t, b^t)|(\theta^t, y^t) \Rightarrow H(y_{t+1}|\theta^t, x^t, y^t) = H(y_{t+1}|\theta^t, y^t) = H(y_{t+1}|\theta^t, b^t, y^t)$, so by symmetry of mutual information

$$
\begin{aligned}
H(\theta^t|\zeta^t, y^t) - H(\theta^t|\zeta^t, y^{t+1}) &= I(\theta^t; y_{t+1}|\zeta^t, y^t) = I(y_{t+1}; \theta^t|\zeta^t, y^t) \\
= H(y_{t+1}|\zeta^t, y^t) - H(y_{t+1}|\theta^t, \zeta^t, y^t) &= H(y_{t+1}|\zeta^t, y^t) - H(y_{t+1}|\theta^t, y^t)
\end{aligned}
$$

So 16 becomes

$$
\begin{aligned}
&E[\sum_{t=1}^{\infty} \beta^t (H(\theta^{t-1}|\zeta^{t-1}, y^t) - H(\theta^t|\zeta^t, y^{t+1}) - H(y_t|\zeta^t, y^t) + H(y_t|\theta^t, y^{t-1}) + H(\theta_t|\theta^{t-1}, y^t))] \\
&= E[\sum_{t=1}^{\infty} (\beta^{t+1} - \beta^t) H(\theta^t|\zeta^t, y^{t+1}) - \beta^t H(y_t|\zeta^t, y^t) + \beta^t (H(y_t|\theta^t, y^{t-1}) + H(\theta_t|\theta^{t-1}, y^t))]
\end{aligned}
$$

Step 2: Given strategy s and the choice rule generated by it p by construction they generate the same gross utilities. Hence by step 1, 15-14 is:

$$
E[\sum_{t=1}^{\infty} (\beta^t - \beta^{t+1})(H(\theta^t|b^t, y^{t+1}) - H(\theta^t|x^t, y^{t+1})) + \beta^t (H(y_{t+1}|b^t, y^t) - H(y_{t+1}|x^t, y^t))]
$$

But then $|B| \leq |X| < \infty \Rightarrow b^t$ is measurable wrt x^t and hence $E[H(\theta^t|b^t, y^{t+1})] \geq E[H(\theta^t|x^t, y^{t+1})]$ and $E[H(y_{t+1}|b^t, y^t)] \geq E[H(y_{t+1}|x^t, y^t)]$ and therefore 15≥14 .

Step 3: As $B \subset X$ if p is a probability choice rule then $f_t(x_t|w^t) = p_t(b_t|\omega^t)$ and $x_t = \sigma_t(z^{t-1}, x_t)$ is a viable solution to 14. For this strategy generated by this mapping, the probability choice rule makes equation15=equation14

Step 4: If s solves 14 the corresponding PCR p must solve 15, as by step 2 the value from p in 15≥s in 14, so if p doesn't solve 15 ∃ PCR producing greater net lifetime utiltiy than s in 14. But by step 3 this produces a viable solution to 14 with greater net life-time utility contradicting s being a solution to 14.

Step 5: If p solve 15 then by step 3 it produces a viable solution to 14 but then 15≥14 so this strategy must be the optimal solution to 14 □

The remainder of the proof follow as stated in SSM for the case where the choice variables are discrete as is the case in this paper.

It is worth saying a few words about how the model in this paper maps to the class of models in this appendix based on SSM as the correspondence is not obvious. The clearest difference between the SSM setup and the model presented in section 6 is that SSM only allow for exogenous states whilst I have an endogenous state in the form of assets a_t. However, since utility can depend upon the entire history of choices and states there is a simple mapping from the endogenous states without history dependent preferences to the world of exogenous states with history dependent preferences. The state in the sense of SSM now only contains the exogenous states $\Theta_t = Supp(SPA) \times Supp(Y_t) \times Supp(AIME_t)$, $(SPA_t, y_t, u_t) = \theta_t \in \Theta_t$ but since $a_t \in d_{t-1}$ and $AIME_t = g(d^{t-1}, \theta^{t-1})$ for the function g that follows from the definition of $AIME_t$ given in section 5. Hence, we can re-express the the utility given in terms of section 6 states X_t and the current decision $u(d_t, X_t)$ in terms of the history of exogenous state θ^t and the history of decisions $u(d^t, \theta^t)$. And since the SSM agent condition their action on everything useful from $z^{t-1} = (d^{t-1}, y^t)$, they can condition on all states.