A Life-Cycle Model with Ambiguous Survival Beliefs

Max Groneck* Alexander Ludwig† Alexander Zimper‡

December 2, 2015

Abstract

Based on a cognitive notion of neo-additive capacities reflecting likelihood insensitivity with respect to survival chances, we construct a Choquet Bayesian learning model over the life-cycle that generates a motivational notion of neo-additive survival beliefs expressing ambiguity attitudes. We embed these neo-additive survival beliefs as decision weights in a Choquet expected utility life-cycle consumption model and calibrate it with data on subjective survival beliefs from the Health and Retirement Study. Our quantitative analysis shows that agents with calibrated neo-additive survival beliefs (i) save less than originally planned, (ii) exhibit undersaving at younger ages, and (iii) hold larger amounts of assets in old age than their rational expectations counterparts who correctly assess their survival chances. Our neo-additive life-cycle model can therefore simultaneously accommodate three important empirical findings on household saving behavior.

JEL Classification: D91, D83, E21.

Keywords: Bayesian learning; likelihood insensitivity; ambiguity; Choquet expected utility; dynamic inconsistency; life-cycle hypothesis; saving puzzles

*Department of Economics, Stockholm School of Economics, P.O. Box 6501, SE-11383, Stockholm Sweden; E-mail: max.groneck@hhs.se.

†Corresponding author. SAFE, Goethe University Frankfurt, House of Finance, Theodor-W.-Adorno-Platz 3, 60629 Frankfurt am Main, Germany; E-mail: ludwig@safe.uni-frankfurt.de.

‡Department of Economics, University of Pretoria, Private Bag X20, Hatfield 0028, South Africa; E-mail: alexander.zimper@up.ac.za.

1 Introduction

Recent empirical findings on household saving behavior are puzzling for the standard rational expectations (RE) life-cycle consumption model à la Modigliani and Brumberg (1954) and Ando and Modigliani (1963). For example, Laibson et al. (1998) and Bernheim and Rangel (2007) report large gaps between self-reported behavior and self-reported plans. People save less for retirement than actually planned (Choi et al. 2006; Barsky et al. 1997; Lusardi and Mitchell 2011). They behave in a dynamically inconsistent manner. Furthermore, people hold large amounts of assets still late in life and dissave less in old age than predicted by the standard RE life-cycle model (see, e.g., Love et al. 2009; De Nardi et al. 2010; Lockwood 2014). Among many underlying assumptions the standard RE life-cycle model describes survival beliefs as objective survival probabilities. In this paper we ask whether the assumption of ambiguous rather than objective survival beliefs may jointly accommodate the aforementioned saving puzzles. As a novelty of our approach, we derive these ambiguous survival beliefs from a model of Bayesian learning with cognitive limitations.

Our approach comprises of two buildings blocks. As our first building block, we develop a model of Choquet Bayesian learning of survival beliefs which allows for *likelihood insensitivity*. Likelihood insensitivity is a well-documented cognitive limitation according to which people do not properly understand probabilities but rather over- (resp. under-) estimate small (resp. large) probabilities.[1] Our second building block combines Choquet expected utility maximization with respect to learned survival beliefs with a canonical life-cycle model. We calibrate this model with data on survival beliefs taken from the Health and Retirement Study (HRS) and asset data taken from the Survey of Consumer Finances (SCF). Our quantitative analysis investigates in how far our calibrated life-cycle model is able to accommodate the aforementioned saving puzzles.

Central to both building blocks are non-additive probability measures in the form of neo-additive capacities (Chateauneuf et al. 2007). Neo-additive capacities are empirically and theoretically very attractive because they stand for a well-interpreted (and, in some sense, minimal) deviation from the standard concept of additive probability measures to which they add two parameters only.[2] While our point of departure is a

[1] In rank dependent utility theory this cognitive limitation corresponds to an inverse-S shaped mapping from additive probabilities to probability judgments whereby "fifty-fifty" probability judgments represent an extreme form of likelihood insensitivity. For a survey of the according decision-theoretic and psychological literature see Wakker (2010).

[2] Due to their technical convenience neo-additive capacities are often used to approximate the typical inverse S-shape of (cognitive) probability judgments and (motivational) decision weights elicited for rank-dependent utility theories (e.g., Abdellaoui et al. 2011; for a survey on this literature see Wakker 2010). Moreover, Choquet decision making with respect to neo-additive capacities could be equivalently

cognitive notion of neo-additive capacities reflecting likelihood insensitivity of a representative Bayesian learner, we later employ a motivational notion of neo-additive capacities expressing ambiguity as well as ambiguity attitudes (in the form of *relative optimism*) of a representative decision maker. As a distinctive feature of our modelling approach, these motivational neo-additive capacities are not imposed ad hoc but comprehensively derived from our Choquet Bayesian learning model.

We now describe in detail the steps of our analysis. As a starting point of our first building block we extend earlier work in Ludwig and Zimper (2013) and construct a model of Choquet Bayesian learning which describes the decision maker's uncertainty about the joint distribution of the parameter and sample space of survival chances through a *cognitive* neo-additive capacity. Updating of beliefs takes place in accordance with the Generalized Bayesian update (GBU) rule (Eichberger et al. 2007). A first cognitive parameter of this neo-additive capacity reflects likelihood insensitivity. The second cognitive parameter determines in how far this likelihood insensitivity is resolved rather through over- or underestimation of additive probabilities. To combine neo-additive capacities with the GBU rule is appealing because it implies that likelihood insensitivity is increasing with age whereas over-/underestimation attitudes remain constant. Our preferred interpretation of increasing likelihood insensitivity is that older people become increasingly cognitively impaired. In contrast, we regard over-/underestimation attitudes as a personal characteristic of an individual. I.e., we think of "exaggerating" vs. "downplaying" perceived survival chances as character traits that remain constant over the life-cycle.[3]

Under simplifying assumptions, we derive a closed-form expression for the Choquet estimates of survival chances such that they only depend on the agent's age. Next, we demonstrate that these age-dependent Choquet estimates can themselves be reinterpreted as *motivational* neo-additive capacities defined on the space of the decision maker's survival events. The formal relationship between the cognitive parameters of the neo-additive capacities entering our learning model (i.e., likelihood insensitivity; over-/underestimation attitudes) and the motivational parameters of the age-dependent neo-additive survival beliefs generated by our learning model (i.e., degree of ambiguity; relative optimism) is surprisingly complex. For example, even in absence of likelihood insensitivity—i.e., even when the cognitive neo-additive capacities entering our learning model reduce to additive probability measures—the decision maker's motivational neo-additive survival beliefs only become additive in the limit of the learning process

formalized within the multiple priors framework of Ghirardato et al. (2004).

[3]Eichberger et al. (2012) provide a related argument in favor of combining (motivational rather than cognitive) neo-additive capacities with the GBU rule in terms of constant ambiguity attitudes.

where they resemble objective survival probabilities. In contrast, in presence of likelihood insensitivity, the motivational neo-additive survival beliefs that are generated by our learning model do not converge to additive probability measures.

Turning to our second building block we construct a "NEO life-cycle consumption model" according to which an ex ante representative agent maximizes her Choquet expected utility[4] over her uncertain future consumption streams. As the only deviation from the RE benchmark we assume that agents take as decision weights the motivational neo-additive survival beliefs generated by our learning model. Whenever our NEO agents do not converge to RE agents, life-cycle maximization gives rise to dynamically inconsistent behavior regardless of how much statistical information has been observed. We describe both 'naive' and 'sophisticated' NEO agents. While the former do not anticipate that their future selves deviate from ex ante optimal consumption plans, the latter are fully aware of their dynamically inconsistent behavior.

Our quantitative analysis investigates in how far our NEO life cycle model(s) could partially resolve saving puzzles. To this purpose we compare consumption and saving behavior for three different models: the naive NEO, the sophisticated NEO, and the nested RE life-cycle model, respectively. We calibrate these stochastic quantitative life-cycle models to the data. For the NEO life-cycle model(s) we use data on subjective survival beliefs from the Health and Retirement Study (HRS) for the estimation of the neo-additive survival beliefs; for the survival beliefs of the RE life-cycle model we use (projected) objective mortality rates. With the exception of the discount rate, we determine all parameters outside the life-cycle models. We estimate the discount rate through a Simulated Methods of Moments (SMM) technique by minimizing the distance of life-cycle asset holdings between the model and the data (taken from the Survey of Consumer Finance, SCF) for each model version thereby giving each model the same chance to match the data on asset holdings.

In line with the existing literature, the calibrated RE life-cycle model gives rise to the typical puzzles: The average saving rate for prime age savers of age $25-54$ is at 12.5%, compared to 9.5% in the data. Average asset holdings at ages 75, 85 and 95 relative to asset holdings at the average retirement age of 62 are 68.8%, 35.6% and 8.5%, compared to 71.9%, 53.4% and 46.7% in the data. Hence, through the lens of the RE life-cycle model, the data are puzzling: in the data the young save too little and the old decumulate assets too slowly.

In contrast, the calibrated naive NEO life-cycle model partially resolves these puzzles. The average saving rate is at 10.7% and relative asset holdings at ages 75, 85 and 95

[4]For axiomatic foundations of Choquet expected utility theory for the Anscombe-Aumann framework see Schmeidler (1989) and for the Savage framework see Gilboa (1987).

are at 78.1%, 57.9% and 36.9%, remarkably close to the data. In addition, the realized saving rate is 5.2 percentage points lower than the planned saving rate. Predictions on asset holdings for the sophisticated NEO life-cycle model are similar. Sophisticated NEO agents save a bit more than naive NEO agents and hence feature higher asset holdings in old age. The overall fit to the data is better for naive NEO agents. Our analysis therefore suggests that the naive NEO life-cycle model provides a quite accurate quantitative picture of saving behavior until about age 85.

The intuition for these quantitative findings is as follows. The calibrated survival probabilities match the HRS findings that "young" people—before age 65-70—underestimate whereas "old" people—older than 70—overestimate their survival chances.[5] Underestimation at young age is sufficiently strong so that naive NEO agents save less than their RE counterparts. In the course of the life-cycle overestimation of future survival chances lowers the speed of asset decumulation to the effect that the level of old age asset holdings is eventually higher than for RE agents. For middle aged agents, this overestimation is not too strong so that they end up saving less in each period than originally planned in the past.[6] Importantly, neither the relative strengths of these biases in survival beliefs nor relatively low young age, respectively high old age, asset holdings are direct targets in our calibration. Finally, sophisticated NEO agents correctly anticipate the more optimistic beliefs of their future selves therefore saving more than their naive counterparts.

The standard model to explain dynamic inconsistency and undersaving is the hyperbolic time-discounting model. Building on the early work by Strotz (1955) and Pollak (1968) as well as on Laibson (1997), Laibson et al. (1998) find that exponential consumers save more than hyperbolic consumers, cf. also Angeletos et al. (2001). This standard model cannot account for high old-age asset holdings because long-run discounting is identical to the rational expectations model. In contrast, over-estimating beliefs for low probabilities in our NEO life-cycle model(s) implies lower long-run effective discount rates which leads to higher old-age asset holdings. Standard explanations for insufficient old-age asset decumulation such as a bequest motive (Hurd 1989; Lockwood 2014) and precautionary savings behavior (Palumbo 1999; De Nardi et al. 2010) cannot generate undersaving at young ages. Our NEO life-cycle model therefore adds to existing explanations for saving behavior by simultaneously accommodating all three stylized findings: (i) time inconsistency, (ii) undersaving at young age and (iii) high asset holdings at old age.

[5]Similar patterns are observed in numerous other datasets, cf. Ludwig and Zimper (2013) or, for a literature survey, Nicholls and Zimper (2015).

[6]If they were to more strongly overestimate future survival rates, then they would resolve their dynamic inconsistency by actually saving more than originally planned.

The remainder of this paper is organized as follows. Section 2 constructs our model of Choquet Bayesian learning over the agent's life-cycle which is based on a cognitive notion of neo-additive capacities updated by the GBU rule. In Section 3 we use the resulting age-dependent Choquet estimators to construct, for all ages, a motivational notion of neo-additive survival beliefs. Section 4 employs these neo-additive survival beliefs as decision weights in a multi-period stochastic life-cycle model for Choquet expected utility maximizers. Calibration is outlined in Section 5. Results of the quantitative analysis are presented in Section 6. Finally, Section 7 concludes. All propositions are formally proved in Appendix A. Appendix B describes the construction of the asset data used for calibration and contains a detailed description of our SMM estimator.

2 Bayesian Learning with Likelihood Insensitivity

We assume that people can be described as Bayesian learners who update, in a Bayesian fashion, their estimators for the objective probability to survive from k to t, denoted $\psi_{k,t}$, by incorporating statistical information as they grow older. More specifically, for a fixed k and t we assume that the agent observes over her ages $h \in \{0, ..., k\}$ a non-decreasing data sample whose age-dependent sample size $e(h)$ is given by some non-decreasing experience function

$$e : \{0, ..., k\} \to \mathbb{N}.$$

This sample contains information about how many out of $e(h)$ individuals have survived from k to t whereby these individuals have the same independently and identically distributed survival chances as the representative agent. The interpretation is that this age-increasing statistical information serves as a proxy for the real-life situation that people increasingly receive news about the deaths (or critical illnesses) of acquainted people (or illnesses of themselves) or read increasingly many health studies.

We start out with the formal description of a classical Bayesian learner who is not subject to the cognitive limitation of likelihood insensitivity in that her uncertainty is captured by a unique additive probability measure. In a next step, we construct a neo-additive model of Choquet Bayesian learning that incorporates the possibility of likelihood insensitivity which is resolved by over-/underestimation attitudes.

2.1 Classical Bayesian Learning

We first consider a classical Bayesian decision maker who satisfies Savage's (1954) axioms so that her uncertainty about the joint parameter and sample space is comprehensively described by some unique subjective additive probability measure, denoted μ. The parameter space Θ is given as the Euclidean open interval $(0,1)$ with $\Sigma(\Theta)$ denoting the Borel σ-algebra on Θ. The n-dimensional sample space is given as $X^n = \times_{i=1}^{n} X_i$ with $X_i = \{0,1\}$, for all i, where 1 (resp. 0) captures the event that individual i does (resp. does not) survive from k to t. Endow each X_i with the discrete topology and denote by $\Sigma(X^n)$ the product sigma-algebra of all Borel sigma algebras $\Sigma(X_i)$, $i = 1, ..., n$. Define the infinite sample space $X^\infty = \times_{i=1}^{\infty} X_i$ with the infinite product σ-algebra $\Sigma(X^\infty)$ and denote by $\Sigma(\Theta \times X^\infty)$ the product σ-algebra of $\Sigma(\Theta)$ and $\Sigma(X^\infty)$. To model the classical Bayesian decision maker we are thus concerned with the additive probability space $(\Theta \times X^\infty, \Sigma(\Theta \times X^\infty), \mu)$.

Consider the $\Sigma(\Theta)$-measurable random variable $\tilde{\theta} : \Theta \times X^\infty \to (0,1)$ such that

$$\tilde{\theta}(\theta, x^\infty) = \theta,$$

where we interpret the value of $\tilde{\theta}$ as the true survival probability in any given state of the world. Next consider the $\Sigma\left(X_{e(h)}\right)$-measurable random variable $\tilde{I}_{e(h)}$ which counts the number of individuals $i \in \{1, ..., e(h)\}$ who survived from k to t, i.e., $\tilde{I}_{e(h)} : \Theta \times X^\infty \to \{0, ..., e(h)\}$ such that

$$\tilde{I}_{e(h)}(\theta, x^\infty) = \sum_{i=1}^{e(h)} x_i.$$

We further assume that, conditional on the true parameter value $\tilde{\theta} = \theta$, each of the $e(h)$ individuals have the same probability θ as the representative agent to survive from k to t where survival is independent across individuals. By this i.i.d. assumption of individual survivals, $\tilde{I}_{e(h)}$ is, conditional on the true survival probability $\tilde{\theta} = \theta$, binomially distributed with probabilities

$$\mu\left(\tilde{I}_{e(h)} = j \mid \theta\right) = \binom{e(h)}{j} \theta^j (1-\theta)^{e(h)-j} \text{ for } j \in \{0, ..., e(h)\}. \tag{1}$$

In the absence of any sample information the (marginal) distribution

$$\mu\left(\tilde{\theta}\right) \equiv \mu\left(\tilde{\theta} \times X^\infty\right)$$

stands for the agent's *prior* about her survival chances so that the agent's estimator for her chances to survive from k to t is defined as the (unconditional) expectation

$$\mathbb{E}\left[\tilde{\theta}, \mu\left(\tilde{\theta}\right)\right] = \int_{\theta \in (0,1)} \theta d\mu\left(\tilde{\theta}\right). \tag{2}$$

In light of random sample information $\tilde{I}_{e(h)}$, however, the agent updates her prior to the *posterior* distribution $\mu\left(\tilde{\theta} \mid \tilde{I}_{e(h)}\right)$ so that her estimator becomes the (conditional) expectation

$$\mathbb{E}\left[\tilde{\theta},\mu\left(\tilde{\theta}\mid\tilde{I}_{e(h)}\right)\right]=\int_{\theta\in(0,1)}\theta d\mu\left(\tilde{\theta}\mid\tilde{I}_{e(h)}\right). \tag{3}$$

We interpret the (random) Bayesian estimator (3) as the belief of an h-old agent to survive from age k to age h. Note that consistency results for classical Bayesian estimators establish that the posterior distributions $\mu\left(\tilde{\theta}\mid\tilde{I}_{e(h)}\right)$ concentrate almost surely at the true parameter value (i.e., the objective survival probability $\psi_{k,t}$) if $e(h)$ gets large, implying[7]

$$\lim_{e(h)\to\infty}\mathbb{E}\left[\tilde{\theta},\mu\left(\tilde{\theta}\mid\tilde{I}_{e(h)}\right)\right]=\psi_{k,t}\text{ almost surely.}$$

That is, if the classical Bayesian agent receives more and more statistical information when her age h approaches k, she will learn with certainty her true (=objective) probability to survive from k to t.

While this limit result holds for general well-specified priors $\mu\left(\tilde{\theta}\right)$, we are foremostly interested in an analytically convenient closed-form expression that specifies (3) for any given $\tilde{I}_{e(h)}$. To this purpose we restrict attention to priors $\mu\left(\tilde{\theta}\right)$ given as some Beta distribution with parameters $\alpha,\beta>0$, implying $\mathbb{E}\left[\tilde{\theta},\mu\left(\tilde{\theta}\right)\right]=\frac{\alpha}{\alpha+\beta}$. That is, we assume that

$$\mu\left(\tilde{\theta}=\theta\right)=K_{\alpha,\beta}\theta^{\alpha-1}\left(1-\theta\right)^{\beta-1},$$

where $K_{\alpha,\beta}=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$ is a normalizing constant.[8] Given the Binomial distribution (1), we obtain by Bayes' rule the following conditional distribution of $\tilde{\theta}$

$$\begin{aligned}\mu\left(\tilde{\theta}=\theta\mid\tilde{I}_{e(h)}=j\right)&=\frac{\mu\left(\tilde{I}_{e(h)}=j\mid\theta\right)\mu\left(\theta\right)}{\int_{(0,1)}\mu\left(\tilde{I}_{e(h)}=j\mid\theta\right)\mu\left(\theta\right)d\theta}\\&=K^{\alpha+j-1}_{\alpha+j,\beta+e(h)-k}\theta^{\alpha+j-1}\left(1-\theta\right)^{\beta+e(h)-j-1}\text{ for }\theta\in(0,1).\end{aligned}$$

Note that $\mu\left(\tilde{\theta}\mid\tilde{I}_{e(h)}=j\right)$ is itself a Beta distribution with parameters $\alpha+j,\beta+e\left(h\right)-j$. The agent's subjective survival belief (3) conditional on information $\tilde{I}_{e(h)}=j$, $j\in\{0,...,e\left(h\right)\}$, is thus given as

$$\begin{aligned}\mathbb{E}\left[\tilde{\theta},\mu\left(\tilde{\theta}\mid\tilde{I}_{e(h)}=j\right)\right]&=\frac{\alpha+j}{\alpha+\beta+e\left(h\right)}\\&=\left(\frac{\alpha+\beta}{\alpha+\beta+e\left(h\right)}\right)\mathbb{E}\left[\tilde{\theta},\mu\left(\tilde{\theta}\right)\right]+\left(\frac{e\left(h\right)}{\alpha+\beta+e\left(h\right)}\right)\frac{j}{e\left(h\right)}.\end{aligned} \tag{4}$$

[7] Convergence to the true parameter value only occurs if the prior is, as in our case, well-specified, i.e., has this true value in its support (the seminal contribution is Doob 1949). For a more general convergence result–including misspecified priors–in terms of minimization of the Kullback-Leibler divergence, see Berk (1966).

[8] The gamma function is defined as $\Gamma\left(y\right)=\int\limits_{0}^{\infty}x^{y-1}e^{-x}dx$ for $y>0$.

That is, the updated estimator $\mathbb{E}\left[\tilde{\theta}, \mu\left(\tilde{\theta} \mid \tilde{I}_{e(h)}\right)\right]$ is a weighted average of the agent's prior estimator $\mathbb{E}\left[\tilde{\theta}, \mu\left(\tilde{\theta}\right)\right]$ and the observed fraction $\frac{j}{e(h)}$ of individuals who survived from k to t. From (4) the convergence behavior of classical Bayesian estimators to objective probabilities is easy to see: If the experience function $e(h)$ goes to infinity, the law of large numbers implies that the fraction $\frac{j}{e(h)}$ of individuals who have survived from k to t converges almost surely to the objective survival probability $\psi_{k,t}$ whereby this fraction receives more and more weight because $\frac{e(h)}{\alpha+\beta+e(h)}$ converges to one. The Classical Bayesian learning model therefore implies convergence of all subjective survival beliefs to objective survival probabilities as the agent gains more experience when growing older.

2.2 Choquet Bayesian Learning

As in the classical Bayesian set-up we consider the measurable space $(\Theta \times X^{\infty}, \Sigma(\Theta \times X^{\infty}))$ where $\Theta \times X^{\infty}$ denotes the joint parameter and sample space. As a generalization of the Savage decision maker, however, we describe a Choquet Bayesian learner who resolves her uncertainty through a unique neo-additive capacity that is updated in accordance with the GBU rule. The cognitive parameters of these updated neo-additive capacities govern a model of Bayesian learning over the life-cycle, which is subject to age-dependent cognitive limitations in the form of increasing likelihood insensitivity. To incorporate likelihood insensitivity into Bayesian learning over the life-cycle is, in our opinion, arguably more realistic than the assumption of a classical Bayesian learner to whom cognitive limitations do not apply.

As our point of departure, let us reformulate the classical Bayesian learning model of the previous subsection within Choquet decision theory for general conditional non-additive probabilities. Recall that a non-additive (=not necessarily additive) probability measure $\kappa : \Sigma(\Theta \times X^{\infty}) \rightarrow [0, 1]$ has to satisfy normalization and monotonicity; that is,

(i) $\kappa(\emptyset) = 0$, $\kappa(\Theta \times X^{\infty}) = 1$

(ii) $A \subset B \Rightarrow \kappa(A) \leq \kappa(B)$ for all $A, B \in \Sigma(\Theta \times X^{\infty})$.

In Choquet decision theory random variables are integrated via the Choquet integral. Formally, the Choquet integral of a bounded $\Sigma(\Theta \times X^{\infty})$-measurable function $f : \Theta \times X^{\infty} \rightarrow \mathbb{R}$ with respect to the capacity κ is defined as the following Riemann integral (cf.

Schmeidler 1986)[9]:

$$\int^{Choquet} f d\kappa \equiv \int_{-\infty}^{0} \left(\kappa\left(\left\{(\theta, x^{\infty}) \in \Theta \times X^{\infty} \mid f(\theta, x^{\infty}) \geq z\right\}\right) - 1\right) dz + \int_{0}^{+\infty} \kappa\left(\left\{(\theta, x^{\infty}) \in \Theta \times X^{\infty} \mid f(\theta, x^{\infty}) \geq z\right\}\right) dz. \quad (5)$$

In analogy to the classical Bayesian approach, we define by

$$\kappa\left(\tilde{\theta}\right) \equiv \kappa\left(\tilde{\theta} \times X^{\infty}\right)$$

the agent's (non-additive) prior about her survival chances and we define the Choquet estimator for her chances to survive from k to t as the (unconditional) Choquet expectation

$$\mathbb{E}\left[\tilde{\theta}, \kappa\left(\tilde{\theta}\right)\right] = \int_{\theta \in (0,1)}^{Choquet} \theta d\kappa\left(\tilde{\theta}\right). \quad (6)$$

We further define the Choquet estimator in light of sample information $\tilde{I}_{e(h)}$ as the (conditional) Choquet expectation

$$\mathbb{E}\left[\tilde{\theta}, \kappa\left(\tilde{\theta} \mid \tilde{I}_{e(h)}\right)\right] = \int_{\theta \in (0,1)}^{Choquet} \theta d\kappa\left(\tilde{\theta} \mid \tilde{I}_{e(h)}\right) \quad (7)$$

where $\kappa\left(\tilde{\theta} \mid \tilde{I}_{e(h)}\right)$ denotes some updated non-additive posterior in light of the sample information $\tilde{I}_{e(h)}$.

In what follows we specify the Choquet estimator (7) such that the conditional non-additive probability measure $\kappa\left(\tilde{\theta} \mid \tilde{I}_{e(h)}\right)$ is given as a cognitive neo-additive capacity that is updated in accordance with the GBU rule. In addition we impose ad hoc assumptions that greatly simplify our technical analysis.

Cognitive Neo-additive Capacities

Denote by $\mathcal{N}$ the set of *null events*, i.e., $\mathcal{N}$ collects all events that the decision maker deems impossible.

[9]For an f taking on m different values such that $A_1, ..., A_m$ is the unique partition of $\Theta \times X^{\infty}$ with $f\left((\theta, x^{\infty})_1\right) > ... > f\left((\theta, x^{\infty})_m\right)$ for $(\theta, x^{\infty})_i \in A_i$, the Choquet integral (5) becomes

$$\mathbb{E}[f, \kappa] = \sum_{i=1}^{m} f\left((\theta, x^{\infty})_i\right) \cdot \left[\kappa\left(A_1 \cup ... \cup A_i\right) - \kappa\left(A_1 \cup ... \cup A_{i-1}\right)\right],$$

which is the familiar method of integrating up some utility function f over gains in rank dependent utility theories.

Definition 1. *Fix some set of null-events* $\mathcal{N} \subset \Sigma(\Theta \times X^{\infty})$ *for the measurable space* $(\Theta \times X^{\infty}, \Sigma(\Theta \times X^{\infty}))$. *The cognitive neo-additive capacity,* ν*, is defined, for some* $\delta, \lambda \in [0,1]$ *by*

$$\nu(A) = \delta \cdot \nu_{\lambda}(A) + (1-\delta) \cdot \mu(A) \tag{8}$$

for all $A \in \Sigma(\Theta \times X^{\infty})$ *where* μ *is some additive probability measure satisfying*

$$\mu(A) = \begin{cases} 0 & \text{if } A \in \mathcal{N} \\ 1 & \text{if } \Theta \times X^{\infty} \backslash A \in \mathcal{N} \end{cases}$$

and the non-additive probability measure ν_{λ} *is defined as follows*

$$\nu_{\lambda}(A) = \begin{cases} 0 & \text{iff } A \in \mathcal{N} \\ \lambda & \text{else} \\ 1 & \text{iff } \Theta \times X^{\infty} \backslash A \in \mathcal{N}. \end{cases}$$

In this paper, we are exclusively concerned with the empty set as the only null event, i.e., $\mathcal{N} = \{\emptyset\}$. In this case, the neo-additive capacity ν in (8) simplifies to

$$\nu(A) = \delta \cdot \lambda + (1-\delta) \cdot \mu(A)$$

for all $A \neq \emptyset, \Theta \times X^{\infty}$.

The cognitive parameter $\delta \in [0,1]$ measures the empirical phenomenon of likelihood insensitivity. If there is no likelihood insensitivity ($\delta = 0$), ν reduces to the additive probability measure μ. If there is likelihood insensitivity ($\delta > 0$), the cognitive parameter $\lambda \in [0,1]$ measures in how far the agent resolves this likelihood insensitivity with respect to the additive probability μ of an event A through over- (high values of λ) versus underestimation (low values of λ) with respect to the additive probability $\mu(A)$. In short, λ measures over-/underestimation attitudes.

The following observation extends a result (Lemma 3.1) of Chateauneuf et al. (2007) for finite random variables to the more general case of random variables with a bounded range (cf. Zimper (2012) for a formal proof).

Observation 1. *Let* $f : \Theta \times X^{\infty} \to \mathbb{R}$ *be a* $\Sigma(\Theta \times X^{\infty})$*-measurable function with bounded range. The Choquet expected value (5) of* f *with respect to a neo-additive capacity (8) is then given by*

$$\mathbb{E}[f, \nu] = \delta(\lambda \sup f + (1-\lambda) \inf f) + (1-\delta)\, \mathbb{E}[f, \mu].$$

Substituting the neo-additive prior $\nu\left(\tilde{\theta}\right)$ for $\kappa\left(\tilde{\theta}\right)$ in (6) gives, by Observation 1, the following Choquet estimator in the absence of any sample information

$$\begin{aligned}\mathbb{E}\left[\tilde{\theta},\nu\left(\tilde{\theta}\right)\right] &= \delta\left(\lambda\sup\theta+(1-\lambda)\inf\theta\right)+(1-\delta)\cdot\mathbb{E}\left[\tilde{\theta},\mu\left(\tilde{\theta}\right)\right] \\ &= \delta\cdot\lambda+(1-\delta)\cdot\mathbb{E}\left[\tilde{\theta},\mu\left(\tilde{\theta}\right)\right].\end{aligned}$$

Obviously, if there is no likelihood insensitivity, i.e., $\delta=0$, this Choquet estimator reduces to the classical Bayesian estimator (2) with respect to the additive prior $\mu\left(\tilde{\theta}\right)$.

Generalized Bayesian Updating

There exist multiple perceivable Bayesian update rules for non-additive probability measures κ (cf. Gilboa and Schmeidler 1993). We suppose that our Bayesian learner applies the GBU rule which is defined[10] such that, for all non-null $A,B\in\Sigma\left(\Theta\times X^{\infty}\right)$,

$$\kappa\left(A\mid B\right)\equiv\frac{\kappa\left(A\cap B\right)}{\kappa\left(A\cap B\right)+1-\kappa\left(A\cup\neg B\right)}. \tag{9}$$

An application of the GBU rule to a neo-additive capacity results in the following characterization of a conditional neo-additive capacity.

Observation 2. *If the Generalized Bayesian update rule (9) is applied to the cognitive neo-additive capacity (8), we obtain, for all non-null* $A,B\in\Sigma\left(\Theta\times X^{\infty}\right)$,

$$\nu\left(A\mid B\right)=\delta_B\cdot\lambda+\left(1-\delta_B\right)\cdot\mu\left(A\mid B\right) \tag{10}$$

such that

$$\delta_B=\frac{\delta}{\delta+(1-\delta)\cdot\mu\left(B\right)}.$$

Henceforth, we formalize our Choquet Bayesian learning model within the cognitive neo-additive probability space

$$\left(\Theta\times X^{\infty},\Sigma\left(\Theta\times X^{\infty}\right),\nu\left(\cdot\mid\cdot\right)\right) \tag{11}$$

such that $\nu\left(\cdot\mid\cdot\right)$ satisfies (10). By combining Observations 1 and 2, the Choquet estimator (7) in light of sample information $\tilde{I}_{e(h)}$ becomes

$$\mathbb{E}\left[\tilde{\theta},\nu\left(\tilde{\theta}\mid\tilde{I}_{e(h)}\right)\right]=\delta_{\tilde{I}_{e(h)}}\cdot\lambda+\left(1-\delta_{\tilde{I}_{e(h)}}\right)\cdot\mathbb{E}\left[\tilde{\theta},\mu\left(\tilde{\theta}\mid\tilde{I}_{e(h)}\right)\right] \tag{12}$$

[10] The GB update rule has an axiomatic foundation within Choquet decision theory in the form of the plausible behavioral axioms of *Consequentialism* and *Conditional Certainty Equivalence Consistency.* For details see, e.g., Eichberger et al. (2007) and Siniscalchi (2011).

where

$$\delta_{\tilde{I}_{e(h)}} = \frac{\delta}{\delta + (1-\delta)\cdot\mu\left(\tilde{I}_{e(h)}\right)} \tag{13}$$

and $\delta_{\tilde{I}_{e(h)}} = 0$ if, and only if, there is no initial ambiguity, i.e., $\delta = 0$.

To see that the cognitive parameters in (12) and (13) behave in accordance with our preferred interpretation (see the introduction), observe that likelihood insensitivity $\delta_{\tilde{I}_{e(h)}}$ increases with age (formally because the prior probability attached to observing a specific data-sequence $\mu\left(\tilde{I}_{e(h)}\right)$ decreases in the sample-size, i.e., age h, of the observed data) which captures the notion that older people become increasingly cognitively impaired. In contrast, the over-/underestimation attitude λ in (12), interpreted by us as character trait, remains constant over the life-cycle.

Imposing Ad hoc Assumptions

By its very nature, the Choquet estimator (12) is random because it reacts to random sample information. Our aim is, however, to derive survival beliefs from Choquet estimators in a parsimonious manner. We therefore impose the following two assumptions to further simply the Choquet estimator (12) to the effect that it becomes constant for a given age h.

Assumptions. *Fix h, k, t such that $h \leq k < t$.*

A1 *The additive measure μ in (10) gives rise to a uniform distribution $\mu\left(\tilde{\theta}\right)$.*

A2 *The observed fraction of surviving individuals coincides with the objective survival probability. That is, for every given $e(h)$,*

$$j = \arg\min_{k\in\{0,\ldots,e(h)\}} \left|\frac{k}{e(h)} - \psi_{k,t}\right|, \text{ hence we set } \frac{j}{e(h)} \approx \psi_{k,t}.$$

Assumption A1 pins down the closed form of the additive estimator $\mathbb{E}\left[\tilde{\theta}, \mu\left(\tilde{\theta} \mid \tilde{I}_{e(h)}\right)\right]$ since the uniform distribution is the Beta-distribution with parameters $\alpha = \beta = 1$. This assumption implies that—prior to any sample information—the survival chances for all k, t are identically regarded as "fifty-fifty" chances. Although A1 might appear—at a first glance—as a rather strong assumption, we use it in the calibration of the model only to initialize the dynamics at biological birth (biological age of 0). That is, when agents become economically active in our model, i.e., at the biological age of 20, they have already gathered some experience according to experience function $e(h)$ which pushes the posterior beliefs away from the fifty-fifty assessment, cf. Section 5 for further details. A1 also implies that the parameter $\delta_{\tilde{I}_{e(h)}}$ (13) will be constant across all possible sample information at a given age h because for a uniform $\mu\left(\tilde{\theta}\right)$ the unconditional probability $\mu\left(\tilde{I}_{e(h)}\right)$ will be identical for every possibly observed sample information $\tilde{I}_{e(h)}$ if h is fixed.

Assumption A2 is a technical assumption which plays the role of the law of large numbers without actually requiring that $e(h)$ is already large for every age h. In particular, A2 implies that the originally random classical estimator $\mathbb{E}\left[\tilde{\theta}, \mu\left(\tilde{\theta} \mid \tilde{I}_{e(h)}\right)\right]$ embedded in (14) becomes deterministic. As one justification of A2 observe that our representative h-old agent can be considered as the average of many h-old agents who have observed their own data samples so that even with small values of the experience function $e(h)$ the average value of the fraction $\frac{j}{e(h)}$ coincides almost surely with the objective probability $\psi_{k,t}$. Also note that A2 becomes, by the law of large numbers, rather innocuous for sufficiently large values of the experience function $e(h)$.

Proposition 1. *Under the Assumptions A1-A2, the h-old agent's Choquet estimator for the chance to survive from k to t becomes*

$$\mathbb{E}\left[\tilde{\theta}, \nu\left(\tilde{\theta} \mid \tilde{I}_{e(h)}\right)\right] = \delta_{e(h)} \cdot \lambda + \left(1 - \delta_{e(h)}\right) \cdot \mathbb{E}\left[\tilde{\theta}, \mu\left(\tilde{\theta} \mid \tilde{I}_{e(h)}\right)\right] \tag{14}$$

such that

$$\mathbb{E}\left[\tilde{\theta}, \mu\left(\tilde{\theta} \mid \tilde{I}_{e(h)}\right)\right] = \left(\frac{2}{2+e(h)}\right) \cdot \frac{1}{2} + \left(\frac{e(h)}{2+e(h)}\right) \cdot \psi_{k,t} \tag{15}$$

and

$$\delta_{e(h)} = \frac{\delta + e(h)\,\delta}{1 + e(h)\,\delta}. \tag{16}$$

Note that Proposition 1 (proved in Appendix A) pins down a closed form expression of the Choquet estimator (14) which is no longer random but completely determined by the parameters δ, λ, and $e(h)$ of our Choquet Bayesian learning model as well as by the objective survival probability $\psi_{k,t}$.[11]

3 Motivational Neo-additive Survival Beliefs

So far we have been concerned with the cognitive neo-additive probability space (11) of our Choquet Bayesian learning model which captures the agent's uncertainty about the joint parameter and sample space for fixed k and t with $k < t$. By imposing several

[11] Proposition 1 gives similar long-run dynamics as the learning model developed in Ludwig and Zimper (2013). There, however, we used an ad hoc assumption on the additive prior beliefs. In contrast, Proposition 1 derives the entire dynamics in a more rigorous and entirely consistent way.

assumptions on this learning model, we have derived the closed form expression (14) for the h-old agent's Choquet estimator to survive from k to t. In this section, we show that these Choquet estimators give rise to a unique age-dependent neo-additive capacity which describes the h-old agent's ambiguous beliefs to survive from a fixed age k to any given age t. We refer to these neo-additive survival beliefs as 'motivational' because they will enter as decision weights into the Choquet utility maximization problem of our life-cycle model to be constructed in the following section.

3.1 The Neo-additive Probability Space of Survival Events

To construct a measurable space of survival events, define the finite state space $\Omega = \{0, 1, ..., T\}$ and denote by $\mathcal{F}$ the powerset of Ω. We interpret $D_t = \{t\}, t \in \Omega$ as the event in $\mathcal{F}$ that the agent dies at the end of period t where T stands for the maximal possible age. Define $Z_t = D_t \cup ... \cup D_T$ as the event in $\mathcal{F}$ that the agent survives (at least) until the beginning of period t.

Suppose that there exists an additive probability measure ψ on $(\Omega, \mathcal{F})$, which we interpret as the "objective" survival probability measure. Next define the conditional additive probability measure $\psi(\cdot \mid Z_k)$ on $(\Omega, \mathcal{F})$ which gives the objective probability of an agent's survival chances given that she has already survived from age 0 to age k. Recall that we already denoted by $\psi_{k,t}$ the objective probability that a k-old individual survives from k to t, implying

$$\psi_{k,t} = \psi(Z_t \mid Z_k).$$

Now observe that the Choquet estimator (14) for the chance to survive from k to t can be equivalently rewritten as

$$\mathbb{E}\left[\tilde{\theta}, \nu\left(\tilde{\theta} \mid \tilde{I}_{e(h)}\right)\right] = \delta_h \lambda_h + (1 - \delta_h)\psi_{k,t} \tag{17}$$

where[12]

$$\delta_h = \frac{2 + 3e(h)\delta + e(h)^2 \delta}{2 + 2e(h)\delta + e(h) + e(h)^2 \delta}, \tag{18}$$

$$\lambda_h = \frac{1 - \delta + 2\lambda\delta + 3\lambda e(h)\delta + \lambda e(h)^2 \delta}{2 + 3e(h)\delta + e(h)^2 \delta}. \tag{19}$$

Because $\psi(\cdot \mid Z_k)$ is an additive probability measure on $(\Omega, \mathcal{F})$, the Choquet estimators (17) for different t's can thus be interpreted as the values $\nu_k^h(Z_t)$ of a neo-additive capacity ν_k^h defined on $\left(\Omega, \mathcal{F}^h\right)$. This observation gives rise to the central definition of our paper, which translates our notion of Choquet Bayesian estimators of survival chances into a motivational neo-additive probability space for survival events.

[12] It can be shown that $0 \leq \delta_h \leq 1$ as well as $\lambda \leq \lambda_h \leq \frac{1}{2}$ if $\lambda \leq \frac{1}{2}$ and $\frac{1}{2} \leq \lambda_h \leq \lambda$ if $\lambda \geq \frac{1}{2}$.

Definition 2. *Fix some age $h = 1, ..., T$ and some $k \geq h$. Define the motivational neo-additive probability space $\left(\Omega, \mathcal{F}, \nu_k^h\right)$ such that, for all $A \in \mathcal{F}$,*

$$\nu_k^h(A) = \begin{cases} 0 & \text{if } \psi(A \mid Z_k) = 0 \\ \delta_h \lambda_h + (1 - \delta_h)\psi(A \mid Z_k) & \text{else} \\ 1 & \text{if } \psi(A \mid Z_k) = 1 \end{cases} \tag{20}$$

with parameter δ_h and parameter λ_h given by (18) and (19), respectively. For all $h \leq k < t \leq T$, we call

$$\nu_{k,t}^h \equiv \nu_k^h(Z_t) = \delta_h \lambda_h + (1 - \delta_h)\psi_{k,t} \tag{21}$$

the h-old agent's ambiguous belief to survive from k to t.

Since we use in Section 4 the neo-additive survival beliefs (21) as decision weights in a Choquet expected utility maximization problem, we interpret the age-dependent parameters in (21) as motivational parameters such that δ_h and λ_h stand for the *degree of ambiguity* and *ambiguity attitudes (relative optimism)*, respectively.[13] The following subsection investigates the formal relationship between these motivational parameters and the cognitive parameters δ (likelihood insensitivity) and λ (over-/underestimation attitude).

3.2 Discussion

As our point of departure, we have modeled Choquet Bayesian learning within the neo-additive probability space

$$(\Theta \times X^\infty, \Sigma(\Theta \times X^\infty), \nu(\cdot \mid \cdot)) \tag{22}$$

such that the conditional neo-additive capacity $\nu(\cdot \mid \cdot)$ is characterized by the cognitive parameters δ and λ combined with an application of the GBU rule. In a next step, we have constructed the survival event spaces $\left(\Omega, \mathcal{F}, \nu_k^h\right)$ such that the neo-additive capacity ν_k^h, characterized by the motivational parameters δ_h and λ_h, is defined as the h-old agent's Choquet estimator of the underlying learning model. Consequently, the motivational parameter values δ_h and λ_h are comprehensively pinned down through equations (18) and (19) by the values of the cognitive parameters δ, λ and the agent's age-dependent experience $e(h)$.

To see how the age-conditional neo-additive survival beliefs ν_k^h depend on the specification of the underlying Choquet Bayesian learning model let us consider three different

[13]Later we will interpret λ_h as a *relative optimism* parameter because it puts additional decision weight on the best possible outcome.

scenarios. First, suppose that there is no initial likelihood insensitivity about the joint distribution of the parameter- and sample space, i.e., $\delta = 0$. Even for this classical Bayesian learning model with an additive probability measure μ, the agent's survival beliefs ν_k^h will not be additive except for the limiting case in which she receives an infinite amount of statistical information.

Observation 3. *Fix the neo-additive joint parameter and sample space (11) for some k such that $\delta = 0$.*

(i) *For any value of the experience function $e(h) < \infty$, the neo-additive survival belief ν_k^h does not reduce to an additive probability measure because we have a strictly positive ambiguity parameter*

$$\delta_h = \frac{2}{2 + e(h)} > 0.$$

(ii) *As the values of the experience function $e(h)$ get large, the neo-additive survival beliefs $\nu_{k,t}^h$ converge to the objective probabilities $\psi_{k,t}$.*

Hence, the RE model is nested as a special limit case for $\delta = 0$ *and* $e(h) \to \infty$, i.e., there is no likelihood insensitivity and the agent observes an unlimited amount of data.

As a second scenario, suppose now that there is initial likelihood insensitivity in the Choquet Bayesian learning model but that there is no age-dependent learning. In this "static" scenario, the agent's neo-additive survival beliefs thus remain constant over all ages so that, for all h, $e(h) = n$ for some $n \in N$. Note that the age-independent neo-additive capacity can be interpreted as the transformation of the objective survival probability by a neo-additive probability weighting function. Bleichrodt and Eeckhoudt (2006) as well as Halevy (2008) already consider non-additive survival beliefs where some age-independent probability weighting function is applied to an additive survival probability. Since this static scenario is nested within our general notion (21) as a special case, it is straightforward to investigate the sensitivity of our results with regard to this feature of the model, cf. Section 6.2.2.

We do not believe in the plausibility of the static model because it would be in stark contrast to our everyday experience according to which people receive more and more information about survival chances. The third and, in our opinion, most plausible scenario is therefore a combination of initial likelihood insensitivity with Choquet Bayesian learning over the life-cycle so that the agent's experience function $e(h)$ strictly increases in her age h. In this scenario, the age-dependent neo-additive survival beliefs (21) do not converge through Bayesian learning to the objective survival probabilities.

Observation 4. *Fix the neo-additive joint parameter and sample space (11) for some k such that $\delta > 0$. As the values of the experience function $e(h)$ get large, the ambiguous beliefs $\nu_{k,t}^{h}$ converge to the value of the λ parameter of the Choquet Bayesian learning model, i.e.,*

$$\lim_{e(h)\to\infty} \delta_h \lambda_h + (1 - \delta_h)\, \psi_{k,t} = \lambda.$$

The data on subjective survival beliefs from the Health and Retirement Study (HRS) shows that "young" people underestimate whereas "old" people overestimate their chances to survive into the future (see, e.g., Ludwig and Zimper (2013), Nicholls and Zimper (2015) and references therein). Because neo-additive survival beliefs do, by Observation 4, not converge to objective probabilities whenever there is likelihood insensitivity in the Choquet Bayesian learning model, our calibrated learning model is able to replicate these age-dependent bias patterns. This formal fact leads to the superior empirical performance of our calibrated life-cycle model (to be constructed in the following section) over the standard RE model.

4 NEO Life-Cycle Model

This section merges our notion of ambiguous survival beliefs with a canonical life-cycle consumption model. This quantitative model features many elements that are standard in the RE literature. As the only deviation from this benchmark we take the neo-additive survival beliefs developed in the previous section as the agent's decision weights.

4.1 Demographics

We consider a large number of ex-ante identical agents (=households). Households become economically active at age (or period) 0 and live at most until age T whereby one model period corresponds to one calendar year. Households stochastically die and this survival risk is an idiosyncratic risk which washes out at the aggregate level. As we also hold constant the number of newborn households the population is stationary. We denote the number of households of age t by N_t and normalize total population to unity, i.e., $\sum_{t=0}^{T} N_t = 1$. Households work full time during periods $1, \ldots, t_r - 1$ and are retired thereafter. The working population is $\sum_{t=0}^{t_r-1} N_t$ and the retired population is $\sum_{t=t_r}^{T} N_t$.

We refer to age $h \leq t$ as the planning age of the household, i.e., the age when households make their consumption and saving plans for the future. We denote objective survival probabilities for all in-between periods k, $h \leq k < t$, by $\psi_{k,t}$ where $\psi_{k,t} \in (0, 1)$

for all $t \leq T$ and $\psi_{k,t} = 0$ for $t = T + 1$. The dynamics of the population are given by $N_{t+1} = \psi_{t,t+1} N_t$, for N_0 given.

4.2 Endowments

There are discrete shocks to labor productivity in every period $t = 0, 1, ..., , t_r - 1$ denoted by $\eta_t \in E$, E finite, which are i.i.d. across households of the same age. The reason for modeling stochastic labor productivity is to impose discipline on calibration. Our fully rational model features standard elements as used in numerous structural empirical studies on life-cycle models, cf., e.g., Laibson et al. (1998), Gourinchas and Parker (2002). By $\eta^t = (\eta_1, \ldots, \eta_t)$ we denote a history of shocks and $\eta^t \mid \eta^h$ with $h \leq t$ is the history $(\eta_1, \ldots, \eta_h, ..., \eta_t)$. Let E be the powerset of the finite set E. E^{t_r-1} are σ-algebras generated by $\mathsf{E}, \mathsf{E}, ...$ We assume that there is an objective probability space $\left(\times_{t=0}^{t_r-1} \mathsf{E}^{t_r-1}, \pi\right)$ such that $\pi_t(\eta^t \mid \eta^h)$ denotes the probability of η^t conditional on η^h.

We follow Carroll (1992), Gourinchas and Parker (2002) and others and assume that one element in E is zero (zero income). Accordingly, $\pi_t(\eta^t \mid \eta^h)$ reflects a (small) probability to receive zero income in period t. This feature gives rise to a self-imposed borrowing constraint and thereby to continuously differentiable policy functions. (Self-imposed) borrowing constraints are required to generate realistic paths of life-time consumption, saving and asset accumulation. Continuous differentiability is convenient when we model a sophisticated agent. By thereby avoiding technicalities as addressed in Harris and Laibson (2001) we keep our analysis focused. Since the zero income probability is small, results are virtually unaffected by this assumption, relative to a model with a fixed zero borrowing limit which would result in a kink in each policy function. In addition, we assume productivity to vary by age, denoted by ϕ_t, to reflect a familiar hump-shaped life-cycle earnings profile.

After retirement at age t_r households receive a lump-sum pension income, b. Retirement income is modeled in order to achieve a realistic calibration. (Without retirement income accumulated assets would be too high—ceteris paribus—which would be offset in the calibration by a higher discount rate.) Pension contributions are levied at contribution rate τ. To achieve a self-imposed borrowing constraint and continuously differentiable policy functions also during the retirement period, we assume that there is a small i.i.d. probability of default of the government on its pension obligations. Accordingly, $\eta_t \in E^r = [1, 0]$ during retirement. Correspondingly, let E^r be the powerset of the finite set E^r. ${\mathsf{E}^\mathsf{r}}^{T-t_r+1}$ are σ-algebras generated by $\mathsf{E}^\mathsf{r}, \mathsf{E}^\mathsf{r}, ...$ and $\left(\times_{t=t_r}^{T} \mathsf{E}^\mathsf{r}, \pi^r\right)$ is the objective probability space in the retirement period.

Collecting elements, income of a household of age t is given by

$$y_t = \begin{cases} \eta_t \phi_t w (1-\tau) & \text{for } t < t_r \\ \eta_t b & \text{for } t \geq t_r. \end{cases}$$

We abstract from private annuity markets.[14] The interest rate, r, is assumed to be fixed. With cash-on-hand given as $x_t \equiv a_t (1+r) + y_t$ the budget constraint writes as

$$x_{t+1} = (x_t - c_t)(1+r) + y_{t+1}. \tag{23}$$

Finally, define total income as $y_t^{tot} \equiv y_t + ra_t$, and gross savings as assets tomorrow, a_{t+1}.

4.3 Government

We assume a pure PAYG public social security system with a balanced each period:

$$\tau w \sum_{t=0}^{t_r-1} \phi_t N_t = b \sum_{t=t_r}^{T} N_t. \tag{24}$$

In addition, accidental bequests—arising because of missing annuity markets—are taxed away at a confiscatory rate of 100%. Also, in the unlikely event of default of the government on its pension obligations, the government collects the contributions to the pension system. Both these revenues are used for government consumption which is otherwise neutral.

4.4 NEO Preferences

Households face two dimensions of uncertainty, respectively risk, about period t consumption. First, agents face a risky labor income which we model in the standard objective expected utility way. Second, agents are uncertain with respect to their life expectancy which we model in terms of a Choquet expected utility maximizing agent such that decision weights are given as the motivational neo-additive survival beliefs of Definition 2. We refer to such agents as 'NEO agents'.

Given the productivity shock history η^h, denote by $\mathbf{c} \equiv (c_h, c_{h+1}, c_{h+2}...)$ a shock-contingent consumption plan such that the functions c_t, for $t = h, h+1, ...$, assign to

[14] Hence, we do not address the annuity puzzle in this paper, i.e., the observed small size of private annuity markets, see Friedman and Warshawsky (1990) for an overview. On the one hand, underestimation of survival beliefs extenuates the annuity puzzle. On the other hand, overestimation at old age reinforces the puzzle. However, on average overestimation of survival rates only sets in after the age of 70 and the average underestimation (averaged across all household in the sample) is around 27 percentage points so that the observed biases could be one explanation for the overall low demand for annuities.

every history of shocks $\eta^t|\eta^h$ some non-negative amount of period t consumption. Denote by $u(c_t)$ the agent's strictly increasing utility from consumption at age t, i.e., $u'(c_t) > 0$. We normalize $u(0) = 0$. We assume that the agent is strictly risk-averse, i.e., $u''(c_t) < 0$. Expected utility of an h-old agent from consumption in period $t > h$ contingent on the observed history of productivity shocks η^h is given as $\mathbb{E}_h[u(c_t)] \equiv \mathbb{E}\left[u(c_t), \pi\left(\eta^t|\eta^h\right)\right] = \sum_{\eta^t|\eta^h} u(c_t)\,\pi\left(\eta^t|\eta^h\right)$.

We assume additive time-separability and add a raw time discount factor $\beta = \frac{1}{1+\rho}$.[15] Fix some $s \in \{h, h+1, ..., T\}$ with the interpretation that the agent survives until period s and dies afterwards. Zero consumption in periods of death implies that $u(c_t) = 0$ for all $t > s$. Given s, the agent's von Neumann Morgenstern utility from a consumption plan $\mathbf{c}$ is defined as

$$U(\mathbf{c}(s)) = u(c_h) + \sum_{t=h+1}^{s} \beta^{t-h}\mathbb{E}_h[u(c_t)]. \tag{25}$$

We model the h-old agent as a Choquet decision maker whose survival uncertainty is expressed through the neo-additive survival beliefs of Definition 2. Thereby, we restrict attention to the neo-additive probability space $\left(\Omega, \mathcal{F}, \nu_h^h\right)$ which expresses the beliefs of an h-old agent to survive from her current age h to any age $t > h$. This NEO agent's Choquet expected utility from consumption plan $\mathbf{c}$ with respect to ν_h^h is given as (cf. Observation 1)

$$\begin{aligned}\mathbb{E}\left[U(\mathbf{c}), \nu_h^h\right] &= \delta_h\left[\lambda_h \sup_{s\in\{h,h+1,...\}} U(\mathbf{c}(s)) + (1-\lambda_h) \inf_{s\in\{h,h+1,...\}} U(\mathbf{c}(s))\right] \\ &\quad + (1-\delta_h)\cdot\sum_{s=h}^{T}\left[U(\mathbf{c}(s)), \psi(D_s \mid Z_h)\right],\end{aligned} \tag{26}$$

where $\psi(D_s \mid Z_h)$ denotes the objective probability that the h-old agent dies at the end of period s. Note that we have as best, resp. worst case, scenario for any $\mathbf{c}$ that

$$\begin{aligned}\sup_{s\in\{h,h+1,...\}} U(\mathbf{c}(s)) &= u(c_h) + \sum_{t=h+1}^{T}\beta^{t-h}\mathbb{E}_h[u(c_t)], \\ \inf_{s\in\{h,h+1,...\}} U(\mathbf{c}(s)) &= u(c_h),\end{aligned} \tag{27}$$

i.e., the least upper bound consists of the discounted sum of utilities if survival probabilities were equal to one in every period. The greatest lower bound is the utility if

[15] In line with Halevy (2008) and Andreoni and Sprenger (2012), we assume that time-preferences cannot be reduced to preferences under uncertainty. To keep the formalism as transparent as possible, we simply consider standard exponential time-discounting.

the agent does not survive to the next period. The following technically convenient characterization of (26) is derived in the appendix.

Proposition 2. *Consider a NEO agent of age h. This NEO agent's Choquet expected utility from consumption plan* $\mathbf{c}$ *is given by*

$$\mathbb{E}\left[U\left(\mathbf{c}\right), \nu^{h}\right] = u(c_h) + \sum_{t=h+1}^{T} \nu_{h,t}^{h} \cdot \beta^{t-h} \cdot \mathbb{E}_h\left[u\left(c_t\right)\right] \tag{28}$$

where the subjective probability belief to survive from age h to $t \geq h$ *is given by*

$$\nu_{h,t}^{h} = \begin{cases} \delta_h \cdot \lambda_h + (1-\delta_h) \cdot \psi_{h,t} & \text{for } t > h \\ 1 & \text{for } t = h \end{cases} \tag{29}$$

with δ_h *and* λ_h *given by (18) and (19), respectively.*

Because λ_h determines how much decision weight is (additionally) put on the best versus worst possible utility scenario in the NEO life-cycle model (26), we henceforth call λ_h age-specific measures of *relative optimism*. Note that this motivational interpretation of λ_h in terms of ambiguity attitudes for the NEO life-cycle model is different from our cognitive interpretation of λ as an "over-/underestimation" parameter in the Choquet Bayesian learning model.

4.5 Recursive Problem and Dynamic Inconsistency

At each age h, the NEO agent constructs a consumption and saving plan that maximizes her lifetime utility. The age-dependent sequence of neo-additive probability spaces $\left(\Omega, \mathcal{F}, \nu_h^h\right)$, $h = 1, ..., T$, violates dynamic consistency of the NEO agent's life-cycle utility maximization problem whenever the neo-additive survival beliefs do not reduce to the limiting case of rational expectations (where, for all h, $\nu_h^h = \psi\left(\cdot \mid Z_h\right)$).[16] To characterize actual behavior in presence of dynamic inconsistency, we analyze both naive and sophisticated NEO agents, cf. Strotz (1955) or inter alia O'Donoghue and Rabin (1999) for procrastination models.

A naive NEO agent is completely unaware of this dynamic inconsistency in that she ignores that her future selves will have strict incentives to deviate from a plan that maximizes her lifetime utility from the perspective of age h. We model naifs so that, for each age h, self h implements the first action of her optimal plan expecting future selves to

[16]We refer the interested reader to the axiomatic treatment of the relationship between violations of dynamic consistency and violations of Savage's (1954) sure-thing principle (as in Choquet expected utility theory) to Epstein and Le Breton (1993), Ghirardato (2002), and Siniscalchi (2011).

implement the remaining plan. In contrast, sophisticates fully understand the dynamic inconsistency. They incorporate the anticipated utility maximization problems of their future selves as constraints into their own maximization problem. The resulting strategic situation—in which each agent effectively plays a game against her future selves—is solved through backward induction: Conditional on any observed consumption- and saving history, the optimal consumption and saving plan of self T is incorporated into self $T-1$'s optimal plan, which are both incorporated into self $T-2$'s optimal plan and so forth to the initial self 0.

Although there exists some empirical evidence suggesting that naive rather than sophisticated decision making might be more relevant (cf. O'Donoghue and Rabin (1999) and the literature cited therein), there also exists evidence according to which several investment and contractual arrangements (e.g., investment in rather illiquid assets such as real estate financed by long-term loans) serve as commitment devices through which sophisticated agents restrain the consumption behavior of their future selves (cf., e.g., Ludwig and Zimper (2006) and references therein). In the present paper, we take the pragmatic stand to consider both types of behavior.

We further assume that income risk is first-order Markov so that $\pi(\eta^t \mid \eta^{t-1}) = \pi(\eta^t \mid \eta_{t-1})$. It is then straightforward to set up the recursive formulation of lifetime utility (28). The value function of age $t \geq h$ viewed from planning age h is given by

$$V_t^h(x_t, \eta_t) = \max_{c_t, x_{t+1}} \left\{ u(c_t) + \beta \frac{\nu_{h,t+1}^h}{\nu_{h,t}^h} \mathbb{E}_t \left[V_{t+1}^h \left(x_{t+1}, \eta_{t+1} \right) \right] \right\}.$$

Maximization of the above is subject to (23).

Naive NEO Agents

The naive NEO agent's first order condition is given by the standard Euler equations.

Proposition 3. *The Euler equation for the naive NEO agent for all $t \geq h$ is given by*

$$\frac{du}{dc_t} = \beta(1+r) \cdot \frac{\nu_{h,t+1}^h}{\nu_{h,t}^h} \cdot \mathbb{E}_t \left[\frac{du}{dc_{t+1}} \right], \tag{30}$$

where

$$\frac{\nu_{h,t+1}^h}{\nu_{h,t}^h} = \begin{cases} \nu_{h,h+1}^h = \delta_h \lambda_h + (1-\delta_h)\psi_{h,h+1} & \text{for } t = h \\ \frac{\delta_h \lambda_h + (1-\delta_h)\psi_{h,t+1}}{\delta_h \lambda_h + (1-\delta_h)\psi_{h,t}} & \text{for } t > h. \end{cases}$$

By (30), the expected growth of marginal utility from h to $h+1$ is higher than under rational expectations if the household underestimates the probability of survival to the next period, i.e., if $\nu_{h,h+1}^h < \psi_{h,h+1}$, and vice versa for overestimation. From (30) we can also verify that the NEO life-cycle maximization problem is dynamically inconsistent if and only if the neo-additive survival beliefs do not reduce to additive probabilities. To see this formally compare the optimal consumption choice of an $h+1$ old agent, first, from the perspective of an h old and, second, from her actual perspective when she turns $h+1$. By Proposition 3, the optimal consumption plan for age $h+1$ from the perspective of age h requires that

$$\frac{du}{dc_{h+1}} = \beta\,(1+r)\cdot\frac{\nu^{h}_{h,h+2}}{\nu^{h}_{h,h+1}}\cdot\mathbb{E}_{h+1}\left[\frac{du}{dc_{h+2}}\right], \tag{31}$$

whereas the optimal consumption choice at age $h+1$ from the perspective of age $h+1$ requires that

$$\frac{du}{dc_{h+1}} = \beta\,(1+r)\cdot\frac{\nu^{h+1}_{h+1,h+2}}{\nu^{h+1}_{h+1,h+1}}\cdot\mathbb{E}_{h+1}\left[\frac{du}{dc_{h+2}}\right]. \tag{32}$$

Dynamic consistency with respect to the optimal consumption choice at age $h+1$ thus holds if and only if the two first order conditions (31) and (32) coincide. Because of $\nu^{h+1}_{h+1,h+1} = 1$, this is the case if and only if

$$\frac{\nu^{h}_{h,h+2}}{\nu^{h}_{h,h+1}} = \nu^{h+1}_{h+1,h+2},$$

which holds for $\delta = 0$ and $e(h) \rightarrow \infty$ implying

$$\frac{\nu^{h}_{h,h+2}}{\nu^{h}_{h,h+1}} = \frac{\psi_{h,h+2}}{\psi_{h,h+1}} = \psi_{h+1,h+2} = \nu^{h+1}_{h+1,h+2},$$

but which is violated for $\delta > 0$ since (generically)

$$\frac{\nu^{h}_{h,h+2}}{\nu^{h}_{h,h+1}} = \frac{\delta_h\lambda_h + (1-\delta_h)\,\psi_{h,h+2}}{\delta_h\lambda_h + (1-\delta_h)\,\psi_{h,h+1}} \neq \delta_{h+1}\lambda_{h+1} + (1-\delta_{h+1})\,\psi_{h+1,h+2} = \nu^{h+1}_{h+1,h+2}.$$

As in the rank dependent utility model of Halevy (2008)—which does not consider Bayesian updating of non-additive beliefs—the life-cycle maximization problem of naive NEO agents is thus dynamically inconsistent. While dynamic inconsistency in Halevy (2008) results from a fixed non-additive probability measure, dynamic inconsistency in our model comes with a sequence of age-dependent neo-additive capacities resulting from our Choquet Bayesian learning model.

Sophisticated NEO Agents

Sophisticated NEO agents are fully aware of their dynamic inconsistency. Self h tries to influence future self's $h+1$ behavior via the choice of savings, x_{t+1}. Hence, the usual

Envelope conditions which are standard in rational expectations problems no longer apply, cf., e.g., Angeletos et al. (2001). As a result, the marginal propensities to consume out of cash-on-hand (MPC), $m_{h+1} \equiv \frac{\partial c_{h+1}}{\partial x_{h+1}}$, show up explicitly in the first-order conditions. The sophisticated NEO agent's optimal behavior is therefore characterized by a "generalized Euler equation with adjustment factor":

Proposition 4. *The generalized Euler equation with adjustment factor for the sophisticated NEO agent at age h is given by*

$$\frac{du}{dc_h} = \beta (1+r) \cdot \nu^h_{h,h+1} \cdot \mathbb{E}_h \left[\Theta_{h+1} \cdot \frac{du}{dc_{h+1}} + \Lambda_{h+1} \right] \tag{33}$$

where

$$\Theta_{h+1} \equiv m_{h+1} + \frac{\nu^h_{h,h+2}}{\nu^h_{h,h+1} \cdot \nu^{h+1}_{h+1,h+2}} (1 - m_{h+1}) \tag{34}$$

and

$$\Lambda_{h+1} \equiv \beta(1+r) \frac{\nu^h_{h,h+2}}{\nu^h_{h,h+1}} (1 - m_{h+1}) \left(\frac{\partial V^h_{h+2}}{\partial x_{h+2}} - \frac{\partial V^{h+1}_{h+2}}{\partial x_{h+2}} \right). \tag{35}$$

Proof: See Appendix A.

Relative to the naive NEO agent, the FOC of the sophisticated NEO agent (33) hence features two additional terms, Θ_{h+1} and Λ_{h+1}. To interpret this condition, first assume that $\Lambda_{h+1} = 0$. Then (33)-(34) are analogous to the "generalized Euler equation" derived in the (quasi-)hyperbolic time discounting literature, cf., e.g., Harris and Laibson (2001). The latter refer to (the analogue of) expression $\beta \nu^h_{h,h+1} \Theta_{h+1}$ as the "effective discount factor". The condition is easiest to interpret by noticing that $\Theta_{h+1} > 1$ iff $\varphi_h \equiv \frac{\nu^h_{h,h+2}}{\nu^h_{h,h+1} \cdot \nu^{h+1}_{h+1,h+2}} > 1$, which holds in our calibration of the NEO life-cycle model for all h. In this case the marginal propensity to save next period, $1 - m_{h+1}$, receives a higher value than the MPC, m_{h+1}, and self h correspondingly expresses higher patience than according to the pure short-run discount factor $\beta \nu^h_{h,h+1}$. To gain further intuition observe that, as long as $\varphi_h > 1$, the effective discount factor varies inversely with next period's MPC, just as in the hyperbolic time discounting model. If the MPC of self $h+1$ increases—i.e., self $h+1$ values consumption more—then self h compensates this over-consumption of her own future self by increasing impatience, hence by consuming more today and saving less.

Next, turn to the general case where $\Lambda_{h+1} \neq 0$. For sophisticated NEO agents the value functions of selves h and $h+1$ in periods $h+2$ are age-dependent. A positive difference $\frac{\partial V^h_{h+2}}{\partial x_{h+2}} - \frac{\partial V^{h+1}_{h+2}}{\partial x_{h+2}}$ means that self h's marginal valuation of cash-on-hand in period $h+2$ is higher than self $h+1$'s. Under such a positive difference self h accordingly values savings from $h+1$ to $h+2$ more than self $h+1$. This increases the RHS of (33) thereby increasing savings already at age h.

4.6 Aggregation

Wealth dispersion within each age bin is only driven by productivity shocks. We denote the cross-sectional measure of agents with characteristics (a_t, η_t) by $\Phi_t(a_t, \eta_t)$. Denote by $\mathcal{A} = [0, \infty]$ the set of all possible asset holdings and let $\mathcal{E}$ be the set of all possible income realizations (encompassing both, the working and the retirement period). Define by $\mathcal{P}(\mathcal{E})$ the power set of $\mathcal{E}$ and by $\mathcal{B}(\mathcal{A})$ the Borel σ-algebra of $\mathcal{A}$. Let $\mathcal{Y}$ be the Cartesian product $\mathcal{Y} = \mathcal{A} \times \mathcal{E}$ and $\mathcal{M} = (\mathcal{B}(\mathcal{A}))$. The measures $\Phi_t(\cdot)$ are elements of $\mathcal{M}$. We denote the Markov transition function—telling us how people with characteristics (t, a_t, η_t) move to period $t+1$ with characteristics $t+1, a_{t+1}, \eta_{t+1}$—by $Q_t(a_t, \eta_t)$. The cross-sectional measure evolves according to

$$\Phi_{t+1}(\mathcal{A} \times \mathcal{E}) = \int Q_t\left((a_t, \eta_t), \mathcal{A} \times \mathcal{E}\right) \cdot \Phi_t\left(da_t \times d\eta_t\right)$$

and for newborns

$$\Phi_1(\mathcal{A} \times \mathcal{E}) = N_1 \cdot \begin{cases} \Pi(\mathcal{E}) & \text{if } 0 \in \mathcal{A} \\ 0 & \text{else.} \end{cases}$$

The Markov transition function $Q_t(\cdot)$ is given by

$$Q_t\left((a_t, \eta_t), \mathcal{A} \times \mathcal{E}\right) = \begin{cases} \sum_{\eta_{t+1} \in \mathcal{E}} \pi\left(\eta_{t+1} | \eta_t\right) \cdot \psi_{t,t+1} & \text{if } a_{t+1}(a_t, \eta_t) \in \mathcal{A} \\ 0 & \text{else,} \end{cases}$$

for all $(a_t, \eta_t) \in Y$ and all $(\mathcal{A} \times \mathcal{E}) \in \mathcal{Y}$.

Aggregation gives average

consumption:	$\bar{c}_t = \frac{1}{N_t} \int c_t(a_t, \eta_t) \Phi_t(da_t \times d\eta_t)$,
assets:	$\bar{a}_t = \frac{1}{N_t} \int a_t \Phi_t(da_t \times d\eta_t)$,
income:	$\bar{y}_t = \frac{1}{N_t} \int y_t(\eta_t) \Phi_t d\eta_t$,
total income:	$\bar{y}_t^{tot} = \bar{y}_t + r\bar{a}_t$,
saving rate:	$\bar{s}_t = \frac{1}{N_t} \int s_t(a_t, \eta_t) \Phi_t(da_t \times d\eta_t)$, where $s_t(a_t, \eta_t) = 1 - \frac{c_t(a_t, \eta_t)}{y_t(\eta_t) + r \cdot a_t}$.

In the quantitative section we also study average saving plans of naive NEO agents. By dynamic inconsistency, these agents update their plans in each period. As a way to compare any gap between plans made at age h and realizations at $t \geq h$ for NEO agents we denote the planned average saving rate with superscript h for the respective planning age and compute

$$\bar{s}_t^h = \frac{1}{N_t} \int s_t^h(a_t, \eta_t) \Phi_t^h(da_t \times d\eta_t), \tag{36}$$

for all t. This gives hypothetical average profiles of the saving rate in the population if households stick to their respective period-h plans in all periods $t = h, \ldots, T$. Observe that $\Phi_t^h(\cdot)$ is an artificial distribution generated by respective plans of households. By dynamic consistency, we have for both RE and sophisticated NEO agents that

$$s_t^h(a_t, \eta_t) = s_t^1(a_t, \eta_t) \qquad \text{hence} \qquad \bar{s}_t^h = \bar{s}_t,$$

for all $h = 1, \ldots, T$. These equalities hold for naive NEO agents only for $t = h$ and, independent of current age h, for $t = T$.

5 Calibration and Estimation

We calibrate parameters of the model to minimize the distance between model simulated moments and empirical counterparts. To determine all model parameters $\xi \in \Xi \subset \mathbb{R}^s$ we follow the literature and proceed in two stages.[17] We partition ξ into two subvectors, $\vartheta \in \mathbb{R}^k$ and $\varrho \in \mathbb{R}^l$ and treat ϑ as the vector of first stage parameters that can be identified without using the NEO life-cycle model. In the second stage, the remaining parameters ϱ are determined by using a Simulated Method of Moments (SMM), taking as given the first-stage parameter estimates, $\hat{\vartheta}$. We only estimate the discount rate ρ as a second stage parameter, hence $l = 1$.

5.1 First-Stage Parameters

Households enter the model at the biological age of 20 which we normalize to model age 0. The retirement age is 62, hence $t_r = 42$, according to the average retirement age reported in the Survey of Consumer Finance (SCF).[18] We set the horizon to a maximum biological human lifespan of age 125, hence $T = 105$. This choice is motivated by estimates based on Swedish female life-table data by Weon and Je (2009).

Objective Survival Rates

For objective survival rates we use average cross-sectional survival rates for the US between 2000-2010 taken from the Human Mortality Database (HMD). Data on survival rates become unreliable for ages past 100 as age-specific sample-size is low. Bebbington et al. (2011) argue that a standard Gompertz-Makeham law, cf., e.g., Preston et al. (2001), is ill-suited for estimating human survival rates at high ages.[19] This is due to the fact

[17] For examples in the life-cycle context, see Gourinchas and Parker (2002), Cagetti (2003), De Nardi et al. (2010), and Lockwood (2014).

[18] We compute the average retirement age by pooling the SCF waves 1992-2007 and exclude respondents younger than 45.

[19] However, see Gavrilova and Gavrilov (2015) for a recent criticism of this view.

that human mortality, while first increasing exponentially with age, finally decelerates for high ages past 95. To account for this mortality deceleration we follow Bebbington et al. (2011) by applying the logistic frailty model. Accordingly, the mortality rate μ_t at age t obeys

$$\mu_t = \frac{A \exp(\alpha \cdot t)}{1 + s^2 (\exp(\alpha \cdot t) - 1) \frac{A}{\alpha}} + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0, \sigma^2), \tag{37}$$

where the term in the denominator corresponds to the standard Gompertz-Makeham law. We estimate parameters to get an out of sample prediction for ages past 100 and use predicted values as objective cohort data in the simulation. The implied average mortality rate converges to a value of 0.57 at ages around 110 ($t = 90$). This is well in line with Gampe (2010) who reports an annual mortality rate of around 0.5 for persons past age 110 using data for a series of OECD countries on mortality rates of supercentenarians.

Neo-additive Survival Beliefs

We follow Ludwig and Zimper (2013) and estimate parameters δ and λ, cf. equations (18) and (19), to match the HRS data. Subjective survival rates are obtained by pooling a sample of HRS waves $\{2000, 2002, 2004\}$. Except for heterogeneity in age, we ignore all other heterogeneity across individuals. Before proceeding with the estimation, the experience function $e(h)$ remains to be specified. A general functional form with positive integers is $e(h) \equiv \omega \cdot (20 + h)$, for $\omega \in \mathbb{N}$, which assumes that experience starts at biological birth, cf. our discussion of Assumption A1 in Section 2.2. Identifying parameters in $e(h)$ is not straightforward from our data on survival beliefs alone because of the interplay with the other model parameters δ and λ. We therefore restrict the learning speed such that $\omega = 1$.[20] With this baseline specification and parametrization we get $\hat{\delta} = 0.0163$ and $\hat{\lambda} = 0.413$ and implied values for δ_h and λ_h that lie well within reasonable ranges discussed in the literature (Wakker 2010), cf. Figure 2 in Section 6. As the confidence intervals reported in Table 1 show, these parameters are estimated with high precision. The predicted subjective survival rates resulting from our model of neo-additive survival beliefs fit their empirical counterparts, i.e., the average subjective survival beliefs for each interview age h, from the HRS quite well, cf. Figure 1 in Section 6. As a robustness check, we investigate the relevance of the parametrization of the experience function by considering a static model with constant experience ($e(h) = n$ for some n) to the effect that $\delta_h = \bar{\delta}$ and $\lambda_h = \bar{\lambda}$ for all h, cf. our discussion in Section 3.2.

[20] For higher values of ω the fit of the full life-cycle model to asset data improves slightly—the "optimal" ω is about 140—without affecting our results in any relevant way.

Prices and Endowments

Wages are normalized to $w = 1$. We take a three-state first-order Markov chain for the income process in periods $t = 0, \ldots, t_r - 1$ with state vector $E^w = [1 + \epsilon, 1 - \epsilon, 0]$. The last entry reflects the state with zero income with probability $\zeta = 0.005$, cf. Carroll (1992). The transition matrix during the working period writes as

$$\Pi^w = \begin{bmatrix} (1-\zeta)\kappa & (1-\zeta)(1-\kappa) & \zeta \\ (1-\zeta)(1-\kappa) & (1-\zeta)\kappa & \zeta \\ 0.5 \cdot (1-\zeta) & 0.5 \cdot (1-\zeta) & \zeta \end{bmatrix}$$

for $t = 0, \ldots, t_r$. Households do not draw zero income in their first period of life, therefore the initial probability vector of the Markov chain is $\pi_0 = [0.5, 0.5, 0]'$. Values of persistence and conditional variance of the income shock process are based on the estimates of Storesletten et al. (2004) yielding $\kappa = 0.97$ and $\epsilon = 0.68$. Age specific productivity $\{\phi_t\}$ is estimated based on PSID earnings data, cf. Ludwig, Schelkle and Vogel (2012).

In retirement, for $t = t_r, \ldots, T$, we take as state vector $E^r = [1, 0]$. We assume an even smaller probability to receive zero retirement income of $\zeta^r = 0.001$ which reflects default of the government on its pension obligations. We accordingly have

$$\Pi^r = \begin{bmatrix} 1-\zeta^r & \zeta^r \\ 1-\zeta^r & \zeta^r \end{bmatrix}$$

for $t = t_r, \ldots, T$ and we take as initial probability vector $\pi_{t_r+1} = [1 - \zeta^r, \zeta^r]'$.

The interest rate is set to $r = 0.042$ based on Siegel (2002). For the social security contribution rate we take the US contribution rate of $\tau = 0.124$. Pension payments b then follow from the social security budget constraint (24).

Per-Period Utility

Recall that we normalize utility from death to zero, i.e., if the household dies at the end of period $t - 1$ we let $u(c_t) = u(0) = 0$. As to utility from survival we take a CRRA per period utility function with coefficient of relative risk aversion θ or the intertemporal elasticity of substitution (IES) of $1/\theta$. For the IES we take a conventional value chosen in the literature of 1/3, ($\theta = 3$).[21] This choice implies that a standard CRRA utility function of the form $u(c_t) = \frac{c_t^{1-\theta}}{1-\theta}$ is negative for all $c_t > 0$. This would

[21] The reason for picking a fixed value of the IES as a first-stage parameter while, at the same time, estimating the discount rate ρ endogenously as a second stage parameter, is identification: both preference parameters govern the strength of the saving motive such that joint identification is hard in a representative agent model such as ours.

violate our assumption of positive utility from survival thereby exceeding utility from death, cf. our notion of the worst, respectively best, possible outcome in Subsection 4.4. We cure this by adding an additive preference shifter to the per period utility function, denoted by $\Upsilon > 0$. With this monotone transformation we can ensure (via calibration) that utility from survival is always positive. Of course, adding a constant to all utility numbers does not affect optimal choices.[22] Collecting elements, the per-period utility function reads as

$$u(c_t) = \Upsilon + \frac{c_t^{1-\theta}}{1-\theta}$$

for some $\Upsilon > 0$.

5.2 Second Stage Estimation

In the second stage, we estimate the discount rate ρ to match empirical life-cycle asset-to-income ratios employing the Simulated Method of Moments (SMM). The SMM solves

$$\hat{\rho} = \arg\min_{\rho} J\left(\rho, \hat{\vartheta}\right). \tag{38}$$

with the loss function

$$J\left(\rho, \hat{\vartheta}\right) = m\left(\rho, \hat{\vartheta}\right)' W m\left(\rho, \hat{\vartheta}\right), \tag{39}$$

where the vector of moment conditions $m\left(\rho, \hat{\vartheta}\right)$ consists of the difference between the simulated and empirical asset-to-income ratios and W is a weighting matrix. We explain the construction of the data and the details of the implementation of the estimator in Appendix B. For our baseline results, we estimate a different subjective time discount rate ρ for each of the three models, the RE, the naive and the sophisticated NEO life-cycle model. This way, we give each model an equal chance to match the asset data.

We construct the sampling distribution of our test statistics nonparametrically, using a bootstrap procedure.[23] We follow standard approaches in the literature and ignore much of the sampling uncertainty of first-stage parameter estimates.[24] An exception is

[22]We can therefore calibrate Υ ex post. In our simulation, the lowest possible consumption realization with positive probability, $\underline{c}$, determines a lower bound to the effect that $\Upsilon > -u(\underline{c})$.

[23]Bootstrapping is used in the literature to construct sampling distributions from the data and generally yields a better finite sample approximation than asymptotic theory, cf., e.g., Horowitz (2001), Davidson and MacKinnon (2002). As we use exact aggregation to solve the model for a given set of model parameters on a relatively low dimensional state space, cf. Subsection 4.6, it is computationally feasible to construct the sampling distribution in this way, unlike in many other applications of the SMM estimator (e.g., by Gourinchas and Parker (2002) and De Nardi et al. (2010) who aggregate using computationally costly Monte-Carlo simulation methods). In addition, our model is continuously differentiable so that we can apply standard bootstrap procedures that are easy to implement.

[24]See Cagetti (2003), De Nardi et al. (2010), and Lockwood (2014) for examples. An exception is Gourinchas and Parker (2002) who consider fist-stage uncertainty.

Table 1: First-Stage Parameters

Parameter		*Source*
Technology and Prices		
$w = 1$	Gross wage	normalized
$r = 0.042$	Interest rate	Siegel (2002)
$\tau = 0.124$	Soc. sec. contr. rate	irs.gov
Income Process		
$\kappa = 0.97$	Persistence of income	Storesletten et al. (2004)
$\epsilon = 0.68$	Variance of income	Storesletten et al. (2004)
$\{\phi_t\}$	Age specific productivity	PSID
$\zeta = 0.005$	Prob. of zero lab. inc.	Carroll (1992)
$\zeta^r = 0.001$	Prob. of zero ret. inc.	
Preferences		
$\theta = 3$	Coeff. of rel. risk aversion	
Subjective Survival Beliefs		
$\delta = 0.0163$ $[0.0150, 0.0179]$	Likelihood Insensitivity	HRS
$\lambda = 0.413$ $[0.397, 0.427]$	Relative overestimation	HRS
$\omega = 1$	Learning speed	
Age Limits and Survival Data		
0	Initial model age (age 20)	
$t_r = 42$	Retirement (age 62)	SCF
$T = 105$	Maximum human lifespan	Weon and Je (2009)
$\{\psi_{k,t}\}$	Cohort survival rates	HMD
$s = 0.41$	Logistic frailty model	HMD
$\alpha = 0.13$	Logistic frailty model	HMD
$A = 2.9e - 06$	Logistic frailty model	HMD

Notes: First-stage parameters that are calibrated outside the life-cycle model. 95%-confidence intervals (CI) for subjective survival belief parameters δ, λ (reported in squared brackets in the respective rows) are bootstrapped, cf. Appendix B.

those parameters that govern subjective survival beliefs, i.e., $[\delta, \lambda] \in \vartheta$. Taking the sampling uncertainty with respect to these two parameters into account is important for our statistical tests when we compare across nested models, i.e., the two NEO alternatives to the RE model. Results on these tests are reported in Section 6.

The discount rate ρ is estimated with high precision, cf. Table 2. We observe that the estimated discount rate of naive NEO agents is lower than of the RE and sophisticated NEO agents. In Section 6.2 we investigate the implications of these differences by illustrating how results are affected when we hold the discount rate constant across model variants.

Table 2: Second-Stage Preference Parameter: The Subjective Discount Rate

	RE	naive NEO	sophisticated NEO
Discount Rate, ρ	0.0384	0.0347	0.0394
95%-Confidence Interval	$[0.0351, 0.0413]$	$[0.0320, 0.0372]$	$[0.0368, 0.0416]$

Notes: Second-stage parameter estimates for $\hat{\rho}$, cf. equation (38). Confidence intervals are bootstrapped using a percentile method, cf. Appendix B.

6 Results

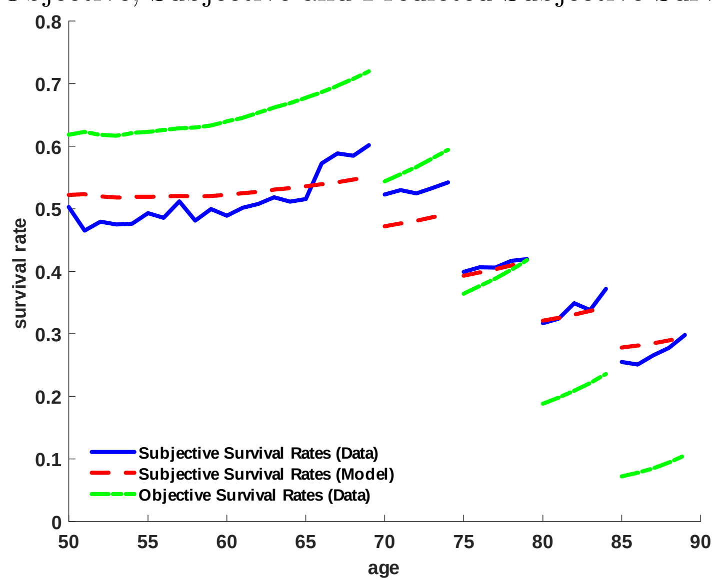
6.1 Neo-additive Survival Beliefs versus Rational Expectations

Figure 1 compares predicted subjective survival rates resulting from our model of neo-additive survival beliefs with their empirical counterparts and corresponding objective survival rates. Actual subjective survival beliefs are elicited in the HRS only for a combination of the age at interview of the individual (which is shown on the abscissa of the figure) and a corresponding target age. For example, individuals of age 69 and younger are asked about their subjective assessment to live until age 80 whereas individuals of age 70 to 74 are asked about their probability to live until age 85, and so forth, cf. Appendix B. The step function of objective probabilities in Figure 1 accordingly follows from changes in the interview age / target age assignment. E.g., the objective chance to live from 69 to 80 is much higher than the chance to live from 70 to 85. Therefore, objective survival beliefs drop discretely between interview ages 69 and 70.[25] Overall

[25] Within each interview age / target age bin, objective survival rates generally increase. For example, the chance to survive from age 60 to 80 is lower than the chance to survive from age 61 to 80. On the other hand, our cohort based prediction of objective survival rates incorporates trends in life-expectancy. In particular at relatively "young" ages it may therefore be that the objective survival rate curve is initially downward sloping within interview age / target age bins.

the data pattern of subjective beliefs mirrors findings in numerous empirical studies on subjective survival beliefs, cf. Ludwig and Zimper (2013), Nicholls and Zimper (2015) and references therein: Until the age of about 70, people underestimate whereas later in life they overestimate their chances to survive into the future. On average, these biases are relatively large.[26] For parameter estimates $\hat{\delta}$ and $\hat{\lambda}$, cf. Table 1, the figure also shows that, on average, the simulated data matches this overall pattern quite well.

Figure 1: Objective, Subjective and Predicted Subjective Survival Rates



Notes: Unconditional probabilities to survive to different target ages according to the HRS. Interview age is on the abscissa. The solid line are subjective survival beliefs, the dashed-dotted line are the corresponding objective survival rates and the dashed line are simulated subjective survival beliefs from the estimated NEO model.

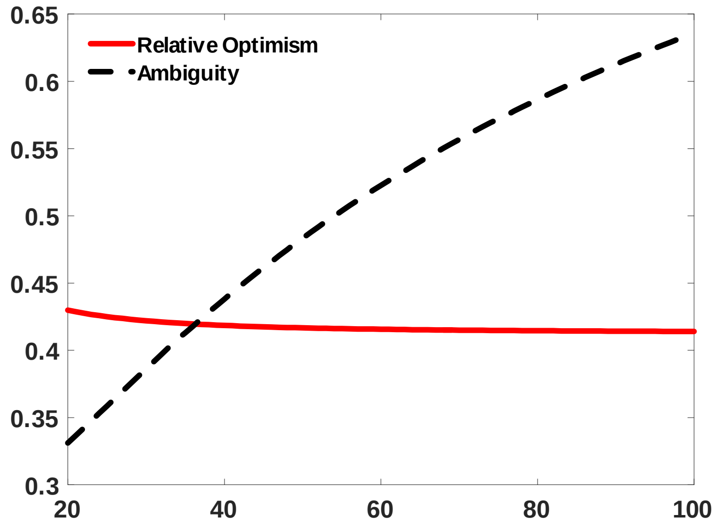
Figure 2 displays the age-specific degree of ambiguity and the degree of relative optimism, both as a function of planning age h, cf. equations (18) and (19). The degree of ambiguity, δ_h, is a monotonically increasing[27] and concave function of planning age h. Relative optimism, λ_h, is a decreasing and convex function, albeit the decrease in relative optimism in age is quantitatively small. Overall, there are two dynamics in the model:

[26] Ludwig and Zimper (2013) document that these findings are robust to focal point answers and that they are neither driven by cohort effects or differential mortality between the HRS and the population. Furthermore, results of our ongoing research where we exploit newly available data in the HRS on psychological attitudes suggest that these results are also not driven by selection on attitude: controlling for objective health measures and socioeconomic characteristics we do not find that optimists are more likely to survive.

[27] Generally, the behavior of δ_h is non-monotone. For our estimates δ_h starts to increase at the biological of age of 12.

first, ambiguity attitudes are changing over age according to the pattern in Figure 2 and second, objective survival chances decrease in age. This latter effect leads to increasingly optimistic biases of predicted subjective survival beliefs, cf. equation (29), despite the increasing relative pessimism expressed through a decreasing λ_h.

Figure 2: Degree of Ambiguity and Relative Optimism over the Life Cycle



Notes: Degree of ambiguity δ_h and relative optimism λ_h as a function of planning age h.

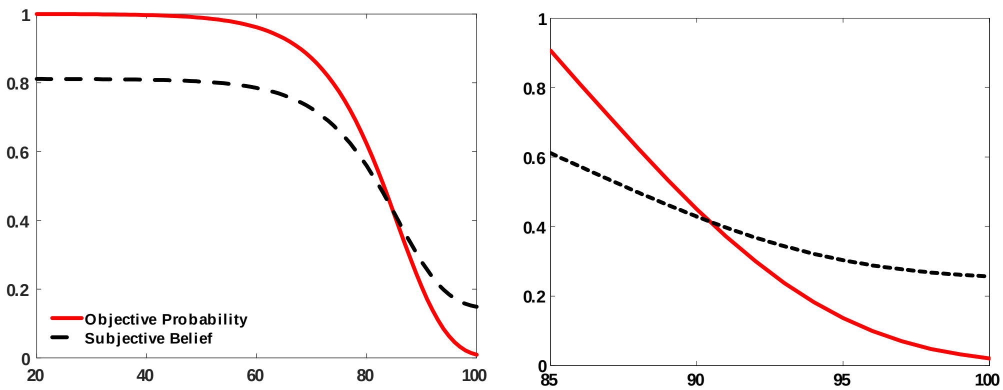
Figure 3 compares ambiguous subjective survival functions to their objective counterparts. The two panels of the figure represent two different planning ages h, planning age 20 ($h = 0$) in Panel (a) and planning age 85 ($h = 65$) in Panel (b). In each panel of the figure, (future) age $t \geq h$ is depicted on the abscissa. Within each panel, experience is unaltered and hence the ambiguity parameter δ_h and the optimism parameter λ_h is constant. Across panels, experience and hence ambiguity is increasing whereas optimism decreases slightly, according to the pattern of Figure 2. The initial point in each survival function at age $t = h$ is driven by ambiguity at that age. As planning age h increases, i.e., as we move from Panel (a) to Panel (b), the distance of this point to a survival rate of 1 increases.[28] The key observation from Figure 3 is that subjective survival functions are flatter than their objective counterparts which is in line with Hamermesh (1985), Hurd et al. (1998), Peracchi and Perotti (2010), Elder (2013) and several others. Furthermore, neo-additive survival beliefs match the stylized fact described by Wu et al. (2015): People at a specific planning (or interview) age underestimate their chances of survival to the nearer future and overestimate survival probabilities to the more distant

[28] Generally, the subjective survival functions exhibit such an initial blip which increases in h and which results from the parsimonious structure of our model but otherwise does not affect our results much, cf. Section 6.2.2 for a sensitivity analysis with respect to the size of this initial blip.

future. Also notice that the overestimation of survival probabilities becomes more pronounced as the agent gets older; that is, from the perspective of the current planning age it takes fewer years for older agents to become optimistic with respect to survival beliefs than it takes for younger agents, again compare Panel (a) and Panel (b) of Figure 3.

Figure 3: Survival Functions

(a) Planning Age 20 (b) Planning Age 85



Notes: Unconditional objective probabilities and subjective beliefs viewed from different planning ages h. Target age t is depicted on the abscissa.

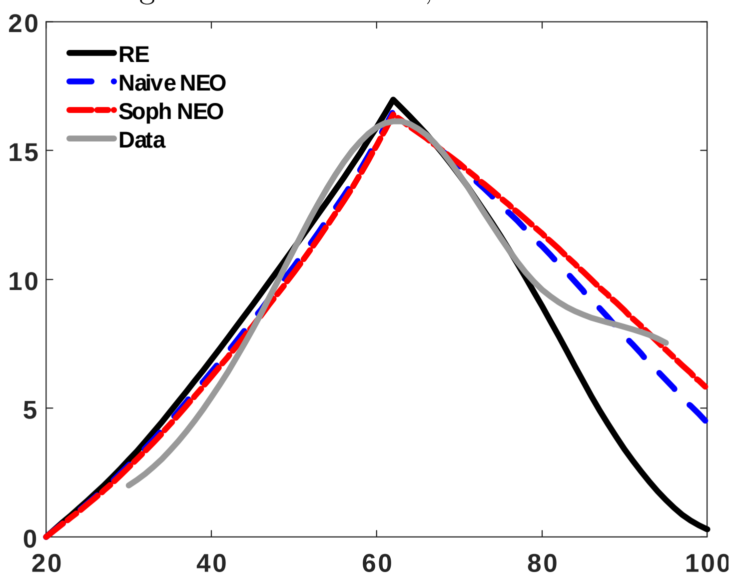
6.2 Life-Cycle Profiles with Neo-additive Survival Beliefs

6.2.1 Baseline Calibration

To highlight the effects of the patterns of subjective survival beliefs shown in the previous subsection on life-cycle savings, we conveniently compress all information by showing average asset holdings of NEO agents compared to RE agents who use objective survival data. We focus on average assets normalized by permanent income as described in the calibration section, scaling assets with the same annuity value as the one used for estimating preference parameters, cf. Section 5.2.

Figure 4 displays average (normalized) asset holdings over the life-cycle for the three types of agents—RE agents, naive and sophisticated NEO agents—and compares those to the (smoothed) data. Assets steadily increase until retirement entry and fall thereafter, i.e., saving rates are positive during working life whereas agents dissave during retirement. This life-cycle profile results from an interplay of two saving motives: Households save for life-cycle (assets are accumulated in order to finance retirement consumption) and precautionary motives. There are two forces triggering precautionary saving. One is the standard income risk, the second is the risk of drawing zero labor income. Since the latter gives rise to a self imposed borrowing constraint, asset holdings throughout the life-cycle are always positive. As agents become older, life-cycle motives for saving become more and more relevant and motives for precautionary saving become less strong, also see, e.g., Gourinchas and Parker (2002). In retirement, the only precautionary motive to save is to avoid zero resources in all income states. This motive again becomes more and more relevant as asset holdings converge towards zero when agents get older.

Figure 4: Assets: NEO, RE and Data



Notes: Assets, normalized by permanent income, from SCF data and for NEO and RE agents. The data is smoothed and covers ages 30 through 95. Details on the data are provided in Appendix B.

With regard to differences in asset accumulation across types, first focus at the RE model. Through the lens of this model the data are puzzling: relative to the data model agents on average have higher saving and therefore stronger asset accumulation until retirement and a faster speed of asset decumulation thereafter (cf. Table 3 below for a quantification of this speed).[29] Any attempt to improve the fit of the RE life-cycle model by, e.g., decreasing the discount rate would lead to a lower speed of asset decumulation at the cost of even higher saving during the working period and vice versa.

On the contrary, the calibrated naive NEO life-cycle model gives rise to less saving during the accumulation phase and a much slower speed of asset decumulation than in

[29] According to the "mortality bias" households with low mortality risk on average have higher lifetime incomes and asset holdings, cf., e.g., Love et al. (2009) and De Nardi et al. (2010). We feed average mortality risk into an ex ante representative agent model to predict average behavior thereby facing a version of Jensen's inequality which may induce biases in our predictions of average asset holdings. Addressing this concern would only be possible using a heterogeneous agent model with differential mortalities as in De Nardi et al. (2010), also see our concluding remarks in Section 7.

the RE life-cycle model, moving it close to the data. The driving force for undersaving (relative to the RE life-cycle model) is pessimism with regard to survival prospects. The reason for high old-age asset holdings is the strong optimism with regard to surviving into the future, cf. Figure 3.

The sophisticated NEO life-cycle model generates very similar results compared to the naive NEO life-cycle model: on average saving rates during the working period are almost identical and old-age asset holdings of sophisticated agents are slightly higher: by foreseeing the optimistic biases of their own future selves, sophisticates decumulate assets at a lower speed for reasons of consumption smoothing. The close similarities between the two NEO agents only occur because we recalibrate the discount rate. It is about half a percentage point higher for the sophisticated agent, cf. Table 2. We discuss this in detail below in Section 6.2.3.

Table 3 comprises our results by reporting summary statistics for all three agent types and the data. In addition to average asset holdings already discussed above, the table reports results on average saving rates, $\bar{s}_t$. We find that the average saving rate of NEO agents during the prime saving years, ages 25-54, is at about 10.7% for naive, at about 10.2% for sophisticated CEU agents and at about 12.5% for RE agents. The corresponding average saving rate in the US is 9.5%.[30]

To further compare gaps between plans made at age h and realizations at ages $t \geq h$ for naive NEO agents we compute the planned average saving rate, $\bar{s}_t^h$. We observe that initially NEO agents plan to save substantially more and correspondingly consume less during working life which would result in higher assets: the planned average saving rate at age 20 for ages 25-54 is 4.2 percentage points higher than the realized one. Actual saving behavior deviates from plans because of time inconsistency and households save less than planned because they moderately overestimate their future survival rates. If overestimation was stronger, then they would actually save more than originally planned. These patterns are qualitatively consistent with empirical findings in the literature on undersaving.[31]

Finally, in order to evaluate the fit of the models, we compute a bootstrap analogue

[30]This is taken from Bosworth et al. (1991) who base their estimate on the Consumer Expenditure Survey (CES). The SCF does not contain quantitative questions on saving, only qualitative ones such as whether one had positive saving. Furthermore, as the SCF does not have a panel dimension, we cannot compute savings from changes in assets.

[31]Barsky et al. (1997) document that agents have a preference for constant or upward sloping consumption paths which cannot be achieved by observed saving rates. Lusardi and Mitchell (2011) present survey results showing that out of those households that made a retirement savings plan, the majority was not able to stick to their plan. Finally, Choi et al. (2006) document that two thirds of respondents in a survey have saving rates below their ideal ones.

Table 3: Summary Statistics

	RE	*NEO*		*Data*[1)]
		Naive	Soph.	
Saving rate[2)]	12.5%	10.7%	10.2%	9.5%
	[11.8, 13.3]%	[10.1, 11.5]%	[9.6, 10.9]%	–
Saving rate, plan[2)]	–	14.9%	–	–
		[14.3, 15.7]		
Assets: 75 to 62[3)]	68.8%	78.1%	80.3%	71.9%
	[67.9, 69.8]%	[76.9, 79.4]%	[79.2, 81.6]%	[64.4, 80.8]%
Assets: 85 to 62[3)]	35.6%	57.9%	62.8%	53.4%
	[34.6, 36.7]%	[56.2, 59.8]%	[61.2, 64.7]%	[44.8, 66.7]%
Assets: 95 to 62[3)]	8.5%	36.9%	44.4%	46.7%
	[8.0, 9.0]%	[35.1, 38.9]%	[42.6, 46.4]%	[18.7, 102.3]%

Notes: Bootstrapped confidence intervals are in parenthesis.

1) The data for asset decumulation is calculated from smoothed SCF data, cf. Appendix B. The saving rate is the weighted average of ages 25-54 between 1980-85 from the Consumer Expenditure Survey (CES) as reported by Bosworth et al. (1991), Table 3.

2) The average saving rate as is defined as the average of individual saving rates between ages 25 and 54. The average planned saving rate is the rate for ages 25-54 planned at age 20.

3) Average asset holdings at age 75, 85 and 95 relative to assets at retirement entry at age 62.

of the J-test for overidentifying restrictions, cf. Appendix B. As documented in the upper part of Table 4, on the basis of these tests all models are rejected at the 5%-level of statistical significance. As we only have one parameter in our model to match 66 moments this is not surprising and mimics standard results in the literature, cf., e.g., Gourinchas and Parker (2002) and De Nardi et al. (2010). More importantly, the function value is lowest for the naive NEO model (cf. the first row in Table 4) implying that the model moments come closest to their empirical counterpart which is also reflected in the highest p-value.

Our test results for model comparisons reported in the lower part of Table 4 confirm this. For those we apply a percentile method by computing the difference of the bootstrapped distributions of function values between each of the two respective NEO models and the RE model. A positive mean of these differences indicates a better data fit of the respective NEO life-cycle model. The corresponding p-value shows the fraction of bootstrapped function values where the difference is negative. This is true in only 2.4% of cases for the naive and in 12.6% of cases for the sophisticated agent model. Therefore the RE model has to be rejected in favor of the naive NEO model at conventional levels of statistical significance. The case for the sophisticated NEO model is less clear, on statistical grounds. Comparing the two NEO models, we find that the naive NEO model performs significantly better than the sophisticated NEO model with a p-value of 0.6%.

Table 4: Evaluation of Fit

	RE	naive NEO	sophisticated NEO
Overidentification Test			
Function Value $\tilde{J}\left(\hat{\rho},\hat{\vartheta}\right)$	2.5432	2.0619	2.2229
p-value	0.6%	1.4%	1.0%
Model Comparison relative to RE			
Average Difference of Function Values	–	0.5024	0.3448
p-value	–	2.4%	12.6%

Notes: The function value $\tilde{J}(\hat{\rho},\hat{\vartheta})$ is from the demeaned objective function, cf. Appendix B for details. The p-value is computed using the 95%-value of the bootstrapped distribution. The model comparison is computed as the average difference of bootstrapped function values between the respective NEO and the RE model. The p-value is the fraction of cases in our bootstrapped samples with a negative distance.

In sum, we can conclude this analysis by stating that the combination of neo-additive survival beliefs with the assumption of naivety has to be considered as the best candidate—out of our considered specifications of naive NEO, sophisticated NEO, and RE models—for accommodating the joint occurrence of time inconsistent saving behavior, low retirement savings and high old-age asset holdings.

6.2.2 The Effects of Experience

Our baseline results are based on age-dependent non-additive survival beliefs. To investigate whether a model of the Bleichrodt and Eeckhoudt (2006) or Halevy (2008) type who consider non-additive survival beliefs according to some age-independent probability weighting function, cf. our discussion in Section 3.2, fits the data better, we next analyze the importance of our assumed experience function for life-cycle asset holdings. To this end we assume constant experience by setting $e(h) = n$ for some $n \in \mathbb{N}$ which implies that $\delta_h = \bar{\delta}$ and $\lambda_h = \bar{\lambda}$ for all h. Observe that the three parameters n, δ, λ are not separately identified in this specification. We therefore directly estimate $\bar{\delta}, \bar{\lambda}$, giving $\bar{\delta} = 0.565$ and $\bar{\lambda} = 0.424$ (with respective confidence intervals of $[0.54, 0.59]$ and $[0.41, 0.44]$). This results in a counterclockwise tilting of the subjective survival belief functions relative to the baseline (shown in Figure 3) so that the initial blip at age h increases. In our baseline specification, the subjective belief of a 20-year old to survive to age 21 is 0.811, cf. Panel (a) of Figure 3, in the model variant with constant experience it is only 0.675. We also reestimate the discount rate as our second stage parameter as outlined in Section 5. For both NEO models it decreases slightly, from 3.5% to 3.2% for the naive and from 3.9% to 3.8% for the sophisticated NEO model (with respective confidence intervals of $[2.96, 3.57]\%$ and $[3.57, 4.05]\%$).

As shown in Table 5 the key quantitative implications are little affected by these changes of the experience function. Because of the increased relative pessimism (survival beliefs are initially more strongly underestimated), the average saving rate decreases slightly for both NEO types. This also leads to an increase of the difference between planned and realized saving rates for naive NEO agents. Since survival functions are flatter than in the baseline model overestimation is stronger in old age so that old-age asset holdings are also higher than in the baseline model.

Table 5: Summary Statistics: The Effects of Experience

	Naive NEO		*Sophisticated NEO*		*Data*
	baseline	$e(h) = n$	baseline	$e(h) = n$	
Saving rate	10.7%	10.6%	10.2%	10.0%	9.5%
	$[10.1, 11.5]\%$	$[10.0, 11.3]\%$	$[9.6, 10.9]\%$	$[9.4, 10.7]\%$	–
Saving rate, plan	14.9%	16.6%	–	–	–
	$[14.3, 15.7]\%$	$[16.0, 17.3]\%$			
Assets: 75 to 62	78.1%	79.6%	80.3%	81.8%	71.9%
	$[76.9, 79.4]\%$	$[78.5, 80.8]\%$	$[79.2, 81.6]\%$	$[80.8, 83.0]\%$	$[64.4, 80.8]\%$
Assets: 85 to 62	57.9%	59.9%	62.8%	65.0%	53.4%
	$[56.2, 59.8]\%$	$[58.3, 61.8]\%$	$[61.2, 64.7]\%$	$[63.4, 66.8]\%$	$[44.8, 66.7]\%$
Assets: 95 to 62	36.9%	38.4%	44.4%	46.4%	46.7%
	$[35.1, 38.9]\%$	$[36.6, 40.6]\%$	$[42.6, 46.4]\%$	$[44.6, 48.5]\%$	$[18.7, 102.3]\%$

Notes: Results for the NEO model with constant experience, $e(h) = n$. See Table 3 for a description of how the statistics are constructed.

Importantly, the p-values of the overidentification tests are lower compared to the baseline model, indicating a worse model fit. They are now at 1.2% (0.8%) for the naive (sophisticated) NEO model. This is confirmed by the tests against the RE life-cycle model which now indicates that the naive, respectively sophisticated, NEO life-cycle model performs worse in 5.6%, respectively 27.6% of cases (compared to 2.4% and 12.6% in the baseline model). When testing our baseline specification of the NEO model with changing experience against the one with constant experience then the Null hypothesis that the model with constant experience performs better than the baseline model is fully rejected for the sophisticated agent (p-value of 0.0%) and rejected for the naive agent at a 10% level of statistical significance (the p-value is 7.2%).

We can therefore conclude that an age-increasing degree of ambiguity is quite an important element for the naive NEO model's success of accommodating the joint occurrence of low retirement savings, time inconsistent saving behavior and high old-age asset holdings.

6.2.3 The Effects of Discounting

The calibrated discount rate varies across all model variants in our baseline results, cf. Table 2. These results hence do not allow us to make statements about the "pure" behavioral effects of changing the model specification. In this section, we therefore address two aspects:

1. When moving from the RE to the naive NEO model, what are the "pure" effects of changing the structure of survival belief formation?

2. When comparing the naive NEO to the sophisticated NEO model, what are the "pure" effects of sophistication?

To address the first question, we hold constant the discount rate at the value calibrated for the RE life-cycle model of 3.84% and feed this into the naive NEO life-cycle model (instead of using the parameter estimate for the discount rate of our baseline calibration of 3.47%, cf. Table 2). That way we can single out the pure effects of subjective survival belief formation according to our theory relative to using the objective survival rates. With the higher discount rate, savings of naive NEO model agents during ages 25-54 are closer to the data than in the baseline model but the speed of asset decumulation is also faster, compare columns 2 and 3 in Table 6. Overall, even with this parametrization the fit of the naive NEO model to the data is better than for the RE model: the function value is at 2.3 compared to 2.54 for the RE model (reported above in Table 4). The Null hypothesis that the two models are identical can however (just) not be rejected at the 10% level (the p-value is 12.0% instead of the 2.4% reported in Table 4 for our baseline calibration).

To address the second question, we next hold constant the value of the discount rate calibrated for the naive NEO model of 3.47% and feed this into the sophisticated NEO model (instead of using the higher baseline parameter estimate of 3.94%, cf. Table 2). The reason for this high baseline estimate is that sophisticates foresee the dynamically inconsistent saving behavior of their own future selves, in particular their own increased tendency to become optimistic. All else equal, this leads to higher old age asset holdings relative to the naive NEO model. To (at least partially) offset this, the estimation determines such a relatively higher discount rate. Holding the discount rate instead constant hence allows us to single out the pure effects of sophistication. By comparing columns 4 and 5 in Table 6 we indeed observe that holding the discount rate constant leads sophisticates to save substantially more than in the baseline model: the saving rate increases by more than one percentage point and asset holdings in old age are 2-4 percentage points higher.

Table 6: Summary Statistics: The Effects of Discounting

	Naive NEO		*Soph. NEO*		*Data*
	baseline	ρ^{RE}	baseline	$\rho^{naive\ NEO}$	
Saving rate	10.7%	9.7%	10.2%	11.5%	9.5%
	$[10.1, 11.5]\%$	$[9.0, 10.6]\%$	$[9.6, 10.9]\%$	$[10.8, 12.2]\%$	–
Saving rate, plan	14.9%	14.0%	–	–	–
	$[14.3, 15.7]\%$	$[13.2, 14.8]\%$			
Assets: 75 to 62	78.1%	76.3%	80.3%	82.6%	71.9%
	$[76.9, 79.4]\%$	$[74.9, 78.1]\%$	$[79.2, 81, 6]\%$	$[81.4, 84.0]\%$	$[64.4, 80.8]\%$
Assets: 85 to 62	57.9%	55.4%	62.8%	66.2%	53.4%
	$[56.2, 59.8]\%$	$[53.3, 57.9]\%$	$[61.2, 64.7]\%$	$[64.4, 68.3]\%$	$[44.8, 66.7]\%$
Assets: 95 to 62	36.9%	34.2%	44.4%	48.1%	46.7%
	$[35.1, 38.9]\%$	$[32.0, 37.0]\%$	$[42.6, 46.4]\%$	$[46.1, 50.3]\%$	$[18.7, 102.3]\%$

Notes: Results for the naive NEO model with constant discount rate, $\rho = \rho^{RE}$ in column 3. Results for the sophisticated NEO model with constant discount rate, $\rho = \rho^{naive\ NEO}$ in column 5. See Table 3 for a description of how the statistics are constructed.

7 Concluding Remarks

This paper constructs and parameterizes a life-cycle model that describes consumption and saving behavior of a Choquet expected utility maximizing agent with respect to age-dependent neo-additive survival beliefs. As a novelty to the literature, these motivational survival beliefs are derived from a Choquet Bayesian learning model that incorporates the well-documented cognitive shortcoming of likelihood insensitivity.

We show that the agents of our model behave dynamically inconsistent whenever the NEO life-cycle model does not reduce to the rational expectations (RE) life-cycle model, which is nested as a special limit case of our approach. Next, we calibrate our neo-additive survival beliefs with HRS data on subjective survival beliefs. As a result, we can replicate the stylized fact that young people underestimate whereas old people overestimate their survival chances. Applied to our life-cycle model, NEO agents save less at younger ages than they actually planned to save. Due to underestimation of survival at young age, NEO agents also save less than RE agents. Despite this tendency to undersave, NEO agents eventually have higher asset holdings after retirement because of the overestimation of survival probabilities in old age. Overall, the calibrated NEO model provides an accurate quantitative picture of life-cycle asset holdings until about age 85. Furthermore, the assumption of naive NEO agents better fits the data than the assumption of sophisticated NEO agents. Our (naive) NEO life-cycle model can therefore accommodate three stylized findings from the empirical literature: (i) time inconsistency of agents, (ii) undersaving at younger ages and (iii) high old age asset holdings.

We present a very specific interpretation of the data on subjective survival beliefs, develop a theory of a representative agent who learns about her prospect of survival and updates her beliefs accordingly and combine this model with an ex ante representative agent consumption life-cycle model. Our approach naturally opens several avenues of future research. First, in our ongoing empirical research we take into account the cross-sectional heterogeneity in the data by investigating how individual specific biases in survival assessments in the HRS are affected by several covariates measuring health whereby we take newly available psychological variables as well as measures of cognitive skills into account. Preliminary findings are consistent with the theory developed in this paper. Second, we plan to investigate possible alternatives to our present calibration of neo-additive survival beliefs (which enter as decision weights into a Choquet expected utility maximization problem) through these probability judgments. Experimental evidence suggests that decision weights tend to have a more pronounced inverse S-shape than probability judgments; however, the formal details of this relationship appear to be very context specific.[32] Third, building on Epper et al. (2011) we will compare the formal relationship between quasi-hyperbolic time-discounting, on the one hand, and our NEO model, on the other hand. Finally, we would like to extend our framework to address normative questions on the optimal design of the tax and transfer system, similar to Laibson et al. (1998), Imrohoroglu et al. (2003) and, more recently, Pavoni and Yazici (2012, 2013) in the hyperbolic time discounting literature.

[32]For example, in an experimental situation with option traders Fox et al. (1996) observe that probability judgments and decision weights coincide. For the according literature see the references on p. 292 in Wakker (2010).

Acknowledgements

Financial support from the Network for Studies on Pensions, Aging and Retirement (Netspar), the Research Center SAFE, funded by the State of Hessen initiative for research LOEWE, the State North Rhine-Westphalia and Economic Research Southern Africa (ERSA) is gratefully acknowledged. We thank seminar participants at various places and Hippolyte d'Albis, Christopher Busch, Nils Grevenbrock, Ben Heijdra, Christian Hellwig, Stefan Hoderlein, Marciano Siniscalchi as well as two anonymous referees for their helpful comments. We also thank Annegret Hoch for excellent research assistance.

References

Abdellaoui, M., A. Baillon, L. Placido, and P. P. Wakker (2011). The Rich Domain of Uncertainty: Source Functions and their Experimental Implementation. American Economic Review 101(2), 695–723.

Altonji, J. G., and Segal, L. M. (1996). Small Sample Bias in GMM Estimation of Covariance Structures. Journal of Business and Economic Statistics 14(3), 353–366.

Ando, A. and F. Modigliani (1963). The 'Life-Cycle' Hypothesis of Saving: Aggregate Implications and Tests. American Economic Review 53(1), 55–84.

Andreoni, J. and C. Sprenger (2012). Risk Preferences Are Not Time Preferences. American Economic Review 102(7), 3357–3376.

Angeletos, G.-M., D. Laibson, A. Repetto, J. Tobacman, and S. Weinberg (2001). The Hyperbolic Consumption Model: Calibration, Simulation, and Empirical Evaluation. Journal of Economic Perspectives 15(3), 47–68.

Barsky, R. B., Juster, F. T., Kimball, M. S. and Shapiro, M. D. (1997). Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study. Quarterly Journal of Economics 112(2): 537–579.

Bebbington, M., C.-D. Lai and R. Zitikis (2011). Modelling Deceleration in Senescent Mortality, Mathematical Population Studies: An International Journal of Mathematical Demography, 18(1), 18–37.

Berk, R. H. (1966). Limiting Behaviour of Posterior Distributions when the Model is Incorrect. Annals of Mathematical Statistics 31, 51–58.

Bernheim, D. and A. Rangel (2007). Behavioral Public Economics: Welfare and Policy Analysis with Nonstandard Decision-Makers. In P. Diamond and H. Vartianen (Eds.), Behavioral Economics and Its Applications. Princeton University Press.

Bleichrodt, H. and L. Eeckhoudt (2006). Survival Risks, Intertemporal Consumption, and Insurance: The Case of Distorted Probabilities. Insurance: Mathematics and Economics 38(2), 335–346.

Bosworth, B., G. Burtless and J. Sabelhaus (1991). The Decline in Saving: Evidence from Household Surveys. Brookings Papers on Economic Activity 1991(1), 183–256.

Cagetti, M. (2003), Wealth Accumulation Over the Life Cycle and Precautionary Savings, Journal of Business and Economic Statistics, 21(3), 339–353.

Carroll, C. D. (1992). The Buffer-Stock Theory of Saving: Some Macroeconomic Evidence. Brookings Papers on Economic Activity 23(2), 61–156.

Chateauneuf, A., J. Eichberger, and S. Grant (2007). Choice under Uncertainty with the Best and Worst in Mind: Neo-Additive Capacities. Journal of Economic Theory 137(1), 538–567.

Choi, J. J., D. Laibson, B. C. Madrian, and A. Metrick (2006). Saving for Retirement on the Path of Least Resistance. In E. J. McCaffrey and J. Slemrod (Eds.), Behavioral Public Finance: Toward a New Agenda, pp. 304–351. New York: Russell Sage Foundation.

Davidson, R. and J. G. MacKinnon (2002), Bootstrap J tests of Nonnested Linear Regression Models. Journal of Econometrics, 109(1), 167–193.

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.

Doob, J. L. (1949). Application of the Theory of Martingales. Le Calcul des Probabilités et ses Applications, Colloques Internationaux du Centre National de la Recherche Scientifique 13, 23–27.

Eichberger, J., S. Grant, and D. Kelsey (2007). Updating Choquet Beliefs. Journal of Mathematical Economics 43(7), 888–899.

Eichberger, J., S. Grant, and D. Kelsey (2012). When Is Ambiguity–Attitude Constant? Journal of Risk and Uncertainty 45(3), 239–263.

Elder, T. E. (2013). The Predictive Validity of Subjective Mortality Expectations: Evidence from the Health and Retirement Study. Demography 50(2), 569–589.

Epper, T., H. Fehr-Duda, and A. Bruhin (2011). Viewing the Future through a Warped Lens: Why Uncertainty Generates Hyperbolic Discounting. Journal of Risk and Uncertainty 43, 169–203.

Epstein, L. G. and M. Le Breton (1993). Dynamically Consistent Beliefs must be Bayesian. Journal of Economic Theory 61(1), 1–22.

Fox, C.R., Rogers, B.A., and A. Tversky (1996). Options Traders Exhibit Subadditive Decision Weights. Journal of Risk and Uncertainty 13(1), 5-17.

Friedman, B. and M.J. Warshawsky (1990). The Cost of Annuities: Implications for Savings Behavior and Bequests. Quarterly Journal of Economics 104(2), 135–154.

Gampe, J. (2010). Human Mortality Beyond Age 110. In: H. Maier et al. (eds.), Supercentenarians, Demographic Research Monographs, pp. 219–230. Springer 2010.

Gavrilova, N. and L. Gavrilov (2015), Biodemography of Old-Age Mortality in Humans and Rodents, Journals of Gerontology Series A: Biological Sciences and Medical Sciences, Vol. 70(1), pp.1–9.

Ghirardato, P. (2002). Revisiting Savage in a Conditional World. Economic Theory 20(1), 83–92.

Ghirardato, P., F. Maccheroni, and M. Marinacci (2004). Differentiating Ambiguity and Ambiguity Attitude. Journal of Economic Theory 118, 133–173.

Gilboa, I. (1987). Expected Utility with Purely Subjective Non-Additive Probabilities. Journal of Mathematical Economics 16(1), 65–88.

Gilboa, I. and D. Schmeidler (1993). Updating Ambiguous Beliefs. Journal of Economic Theory 59(1), 33–49.

Gourinchas, P.-O. and J. Parker (2002). Consumption over the Lifecycle. Econometrica 70(1), 47–91.

Halevy, Y. (2008). Strotz meets Allais: Diminishing Impatience and the Certainty Effect. American Economic Review 98(3), 1145–62.

Hall, P. and J. L. Horowitz (1996). Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators. Econometrica, 64(4), 891–916.

Hamermesh, D. S. (1985). Expectations, Life Expectancy, and Economic Behavior. Quarterly Journal of Economics 100(2), 389–408.

Harris, C. and D. Laibson (2001). Dynamic Choices of Hyperbolic Consumers. Econometrica 69(4), 935–957.

Horowitz, J. L. (2001): The Bootstrap, in: Heckman, J. J. and E. Leamer (eds.). Handbook of Econometrics and Statistics, Volume 5, Elsevier Sciences.

Hurd, M. (1989). Mortality Risk and Bequests. Econometrica 57(4), 173–209.

Imrohoroglu, A., S. Imrohoroglu, and D. H. Joines (2003). Time Inconsistent Preferences and Social Security. Quarterly Journal of Economics 118 (2), 745–784.

Laibson, D. I. (1997). Golden Eggs and Hyperbolic Discounting. Quarterly Journal of Economics 112 (2), 443–477.

Laibson, D. I., A. Repetto, and Tobacman (1998). Self-Control and Saving for Retirement. Brookings Papers on Economic Activity 1998(1), 91–196.

Love, D., Palumbo, M. G. and P. A. Smith (2009). The Trajectory of Wealth in Retirement. Journal of Public Economics 93, 191–208.

Lockwood, L. M. (2014). Incidental Bequests: Bequest Motives and the Choice to Self-Insure Late-Life Risks. Working Paper.

Ludwig, A., Schelkle T. and E. Vogel (2012). Demographic Change, Human Capital and Welfare, Review of Economic Dynamics, 15(1), 94–107.

Ludwig, A. and A. Zimper (2006). Investment Behavior under Ambiguity: The Case of Pessimistic Decision Makers. Mathematical Social Sciences 52, 111–130.

Ludwig, A. and A. Zimper (2013). A Parsimonious Model of Subjective Life Expectancy. Theory and Decision 75, 519–542.

Lusardi, A. and O. S. Mitchell (2011). Financial Literacy and Planning: Implications for Retirement Wellbeing. In: A. Lusardi and O. S. Mitchell (eds), Financial Literacy: Implications for Retirement Security and the Financial Marketplace, pp.17–39. Oxford University Press.

Modigliani, F. and R. Brumberg (1954). Utility Analysis and the Consumption Function: An Interpretation of Cross-Section Data. In K. K. Kurihara (ed), Post-Keynesian Economics. New Brunswick, NJ.: Rutgers University Press.

Nicholls, N. and A. Zimper (2015). Subjective Life Expectancy. In: S. Krauss Whitbourne (ed), Wiley Blackwell Encyclopedia of Adulthood and Aging, Wiley-Blackwell, forthcoming.

O'Donoghue, T. and M. Rabin (1999). Doing It Now or Later. American Economic Review 89(1), 103–124.

Palumbo, M. G. (1999). Uncertain Medical Expenses and Precautionary Saving Near the End of the Life Cycle. Review of Economic Studies 66(2), 395–421.

Pavoni, N. and H. Yazici (2012). Optimal Life-cycle Capital Taxation under Self-Control Problems, forthcoming: Economic Journal.

Pavoni, N. and H. Yazici (2013). Present-Biased Preferences and Optimal Taxation of Parental Transfers, Working Paper.

Peracchi, F. and V. Perotti (2010). Subjective Survival Probabilities and Life Tables: Evidence from Europe. EIEF Working Paper 16/10.

Pischke, J.-S. (1995). Measurement Error and Earnings Dynamics: Some Estimates From the PSID Validation Study. Journal of Business & Economic Statistics, 13, 305-314.

Pollak, R. A. (1968). Consistent Planning. Review of Economic Studies 35(2), 201–208.

Preston, S. H., P. Heuveline, and M. Guillot (2001). Demography - Measuring and Modeling Population Processes. Population and Development Review 27, 365.

Savage, L. J. (1954). The Foundations of Statistics. New York, London, Sydney: John Wiley and Sons, Inc.

Schmeidler, D. (1986). Integral Representation Without Additivity. Proceedings of the American Mathematical Society 97, 255–261.

Schmeidler, D. (1989). Subjective Probability and Expected Utility Without Additivity. Econometrica 57(3), 571–587.

Siegel, J. J. (2002). Stocks for the Long Run : The Definitive Guide to Financial Market Returns and Long-Term Investment Strategies. New York: McGraw-Hill.

Siniscalchi, M. (2011). Dynamic Choice under Ambiguity. Theoretical Economics 6(3), 379–421.

Storesletten, K., C. Telmer, and A. Yaron (2004). Cyclical Dynamics in Idiosyncratic Labor Market Risk. Journal of Political Economy 112(3), 695–717.

Strotz, R. H. (1955). Myopia and Inconsistency in Dynamic Utility Maximization. Review of Economic Studies 23(3), 165–180.

Wakker, P. P. (2010). Prospect Theory: For Risk and Ambiguity. Cambridge, UK: Cambridge University Press.

Weon, B. M. and J. H. Je (2009). Theoretical Estimation of Maximum Human Lifespan. Biogerontology 10, 65–71.

Wu, S., R. Stevens, and S. Thorp (2015), Cohort and Target Age Effects on Subjective Survival Probabilities: Implications for Models of the Retirement Phase, Journal of Economic Dynamics and Control, 55, 39–56

Zimper, A. (2012). Asset Pricing in a Lucas Fruit-Tree Economy with the Best and Worst in Mind. Journal of Economic Dynamics and Control, 36(4), 610–628.

A Appendix: Proof of Propositions

A.1 Proof of Proposition 1

First, note that for arbitrary α and β,

$$\begin{aligned}\mu\left(\tilde{I}_{e(h)}=j\right) &= \binom{e\left(h\right)}{j}\frac{\left(\alpha+j-1\right)\cdot...\cdot\alpha\cdot\left(\beta+e\left(h\right)-j-1\right)\cdot...\cdot\beta}{\left(\alpha+\beta+e\left(h\right)-1\right)\cdot...\cdot\left(\alpha+\beta\right)}, \\ \text{for } j &\in \left\{0,...,e\left(h\right)\right\}.\end{aligned} \tag{40}$$

The uniform distribution is characterized by $\alpha=\beta=1$, implying for (40) that

$$\begin{aligned}\mu\left(\tilde{I}_{e(h)}=j\right) &= \binom{e\left(h\right)}{k}\frac{k!\left(e\left(h\right)-k\right)!}{\left(e\left(h\right)+1\right)\cdot e\left(h\right)!} \\ &= \frac{1}{1+e\left(h\right)}.\end{aligned}$$

That is, for any number of possible survivors $j\in\{0,...,e(h)\}$ the ex ante probability to actually observe this number for a sample of size $e(h)$ is, by A1, identically given as $\frac{1}{1+e(h)}$. Substituting this probability back into (13) gives (16).

Next, substitute $\alpha=\beta=1$ and $\frac{j}{e(h)}=\psi_{k,t}$ in (4) to obtain (15). Finally, collect terms and substitute into (12).□

A.2 Proof of Proposition 2

Fix age h and consider the neo-additive probability space $\left(\Omega,\mathcal{F},\nu_h^h\right)$ of Definition 2. By straightforward transformations, we obtain that

$$\begin{aligned}&\sum_{t=h+1}^{T}\psi\left(D_t\mid Z_h\right)\sum_{s=h+1}^{t}\beta^{s-h}\mathbb{E}\left[u\left(c_s\right),\pi\left(\eta_s|\eta_h\right)\right] \\ =&\sum_{t=h+1}^{T}\beta^{t-h}\mathbb{E}\left[u\left(c_t\right),\pi\left(\eta_t|\eta_h\right)\right]\cdot\psi\left(D_t\cup...\cup D_T\mid Z_h\right) \\ =&\sum_{t=h+1}^{T}\beta^{t-h}\mathbb{E}\left[u\left(c_t\right),\pi\left(\eta_t|\eta_h\right)\right]\cdot\psi_{h,t}.\end{aligned}$$

Consequently, (26) can be equivalently rewritten as

$$
\begin{aligned}
\mathbb{E}\left[U\left(\mathbf{c}\right),\nu_h^h\right] = \delta_h \Bigg(\lambda_h \Bigg(u(c_h) + \sum_{t=h+1}^{T} \beta^{t-h}\mathbb{E}\left[u\left(c_t\right),\pi\left(\eta_t|\eta_h\right)\right]\Bigg) + (1-\lambda_h)u(c_h)\Bigg) \\
+ (1-\delta_h)\Bigg(u(c_h) + \sum_{t=h+1}^{T}\psi\left(D_t \mid Z_h\right)\sum_{s=h+1}^{t}\beta^{s-h}\mathbb{E}\left[u\left(c_s\right),\pi\left(\eta_s|\eta_h\right)\right]\Bigg) \\
= u(c_h) + \delta_h\lambda_h \sum_{t=h+1}^{T}\beta^{t-h}\mathbb{E}\left[u\left(c_t\right),\pi\left(\eta_t|\eta_h\right)\right] \\
+ (1-\delta_h)\sum_{t=h+1}^{T}\psi_{h,t}\cdot\beta^{t-h}\mathbb{E}\left[u\left(c_t\right),\pi\left(\eta_t|\eta_h\right)\right] \\
= u(c_h) + \sum_{t=h+1}^{T}\nu_{h,t}^{h}\cdot\beta^{t-h}\mathbb{E}\left[u\left(c_t\right),\pi\left(\eta_t|\eta_h\right)\right],
\end{aligned}
$$

which proves the proposition.□

A.3 Proof of Proposition 4

The value functions of self h in periods h and $h+1$ are given by

$$
\begin{aligned}
V_h^h\left(x_h,\eta_h\right) &= \max_{c_h,x_{h+1}}\left\{u\left(c_h\right) + \beta\nu_{h,h+1}^{h}\mathbb{E}_h\left[V_{h+1}^{h}\left(x_{h+1},\eta_{h+1}\right)\right]\right\} \\
V_{h+1}^h\left(x_{h+1},\eta_{h+1}\right) &= \max_{c_{h+1},x_{h+2}}\left\{u\left(c_{h+1}\right) + \beta\frac{\nu_{h,h+2}^{h}}{\nu_{h,h+1}^{h}}\mathbb{E}_{h+1}\left[V_{h+2}^{h}\left(x_{h+2},\eta_{h+2}\right)\right]\right\}.
\end{aligned}
$$

For self $h+1$ we accordingly have

$$
V_{h+1}^{h+1}\left(x_{h+1},\eta_{h+1}\right) = \max_{c_{h+1},x_{h+2}}\left\{u\left(c_{h+1}\right) + \beta\nu_{h+1,h+2}^{h+1}\mathbb{E}_{h+1}\left[V_{h+2}^{h+1}\left(x_{h+2},\eta_{h+2}\right)\right]\right\}.
$$

The first-order conditions with respect to consumption for selves h and $h+1$ are:

$$
\frac{du}{dc_h} = \beta R\nu_{h,h+1}^{h}\mathbb{E}_h\left[\frac{\partial V_{h+1}^{h}(\cdot)}{\partial x_{h+1}}\right] \tag{41a}
$$

$$
\frac{du}{dc_{h+1}} = \beta R\nu_{h+1,h+2}^{h+1}\mathbb{E}_{h+1}\left[\frac{\partial V_{h+2}^{h+1}(\cdot)}{\partial x_{h+2}}\right]. \tag{41b}
$$

The derivative of the value function writes as

$$\frac{\partial V_{h+1}^{h}(\cdot)}{\partial x_{h+1}} = \frac{du}{dc_{h+1}} m_{h+1} + \beta R \frac{\nu_{h,h+2}^{h}}{\nu_{h,h+1}^{h}} (1 - m_{h+1}) \mathbb{E}_{h+1}\left[\frac{\partial V_{h+2}^{h}(\cdot)}{\partial x_{h+2}}\right] \tag{42}$$

$$= m_{h+1} \underbrace{\left(\frac{du}{dc_{h+1}} - \beta R \frac{\nu_{h,h+2}^{h}}{\nu_{h,h+1}^{h}} \mathbb{E}_{h+1}\left[\frac{\partial V_{h+2}^{h}(\cdot)}{\partial x_{h+2}}\right]\right)}_{\neq 0\text{, i.e., the envelope condition does not hold.}}$$

$$+ \beta R \frac{\nu_{h,h+2}^{h}}{\nu_{h,h+1}^{h}} \mathbb{E}_{h+1}\left[\frac{\partial V_{h+2}^{h}(\cdot)}{\partial x_{h+2}}\right]$$

where $m_{h+1} \equiv \frac{\partial c_{h+1}}{\partial x_{h+1}}$.

Rewrite (41b) by adding and subtracting terms as

$$\frac{du}{dc_{h+1}} = \beta (1+r) \nu_{h+1,h+2}^{h+1} \mathbb{E}_{h+1}\left[\frac{\partial V_{h+2}^{h+1}(\cdot)}{\partial x_{h+2}}\right]$$

$$+ \beta (1+r) \nu_{h+1,h+2}^{h+1} \mathbb{E}_{h+1}\left[\frac{\partial V_{h+2}^{h}(\cdot)}{\partial x_{h+2}} - \frac{\partial V_{h+2}^{h}(\cdot)}{\partial x_{h+2}}\right]$$

to get

$$\beta R E_{h+1}\left[\frac{\partial V_{h+2}^{h}(\cdot)}{\partial x_{h+2}}\right] = \frac{du}{dc_{h+1}} \frac{1}{\nu_{h+1,h+2}^{h+1}} + \beta R E_{h+1}\left[\Delta V_{h+2}^{h,h+1}\right], \tag{43}$$

where $\Delta V_{h+2}^{h,h+1} \equiv \left[\frac{\partial V_{h+2}^{h}(\cdot)}{\partial x_{h+2}} - \frac{\partial V_{h+2}^{h+1}(\cdot)}{\partial x_{h+2}}\right]$.

Next, use (43) in (42) to get

$$\frac{\partial V_{h+1}^{h}(\cdot)}{\partial x_{h+1}} = \frac{du}{dc_{h+1}} m_{h+1} + \frac{\nu_{h,h+2}^{h}}{\nu_{h,h+1}^{h}} (1 - m_{h+1}) \left(\frac{du}{dc_{h+1}} \frac{1}{\nu_{h+1,h+2}^{h+1}} + \beta R E_{h+1}\left[\Delta V_{h+2}^{h,h+1}\right]\right)$$

$$= \frac{du}{dc_{h+1}} \left(m_{h+1} + \frac{\nu_{h,h+2}^{h}}{\nu_{h,h+1}^{h} \nu_{h+1,h+2}^{h+1}} (1 - m_{h+1})\right) + \beta R \frac{\nu_{h,h+2}^{h}}{\nu_{h,h+1}^{h}} (1 - m_{h+1}) \mathbb{E}_{h+1}\left[\Delta V_{h+2}^{h,h+1}\right].$$

Using the above in (41a) and simplifying the resulting expression we finally get (33)–(35).
□

B Appendix: Details on HRS Data, SCF Data and Bootstrap Procedure

B.1 HRS Data

Interview and target ages in the HRS are assigned according to the pattern in Table 7. Otherwise see Ludwig and Zimper (2013) for a detailed description of the data.

Table 7: Interview and Target Age in the HRS

Age at Interview j	Target Age m
≤69	80
70-74	85
75-79	90
80-84	95
85-89	100

B.2 SCF Data

The Survey of Consumer Finances (SCF) is a representative triennial cross-sectional survey of U.S. families sponsored by the Federal Reserve Board in cooperation with the Department of the Treasury. We merge data from the six waves (1992, 1995, 1998, 2001, 2004 and 2007). We use households whose heads are aged 30-95.[33] Our total sample contains $N = 20.368$ respondents.

To construct the average life-cycle profile of normalized assets we define assets as net worth including housing wealth, but excluding implicit pension and social security wealth.[34] We deflate assets and income to 1992 Dollars and detrend the data. To approximate permanent income we first compute gross labor and social security income by excluding income from capital gains.[35] Using data from Cagetti (2003)—who approximates tax rates for different income percentiles—we next compute after-tax income. Due to multiple imputations, the SCF delivers five observations per respondent together with appropriate weights which we divide by 5 to get an appropriate descriptive statistic per respondent.[36] Next, we compute the weighted average of age-specific net income and net worth. Using the average we evaluate the net-present value of income and convert this to annuities using the calibrated interest rate of $r = 0.042$. This gives our permanent income approximation. Finally, we normalize assets with the computed permanent income. This leaves us with a vector of 66 data moments used for the SMM procedure.

[33] A starting age of 30 is motivated by the fact that we do not explicitly model education decisions so that our model does not match the data well at very young ages.

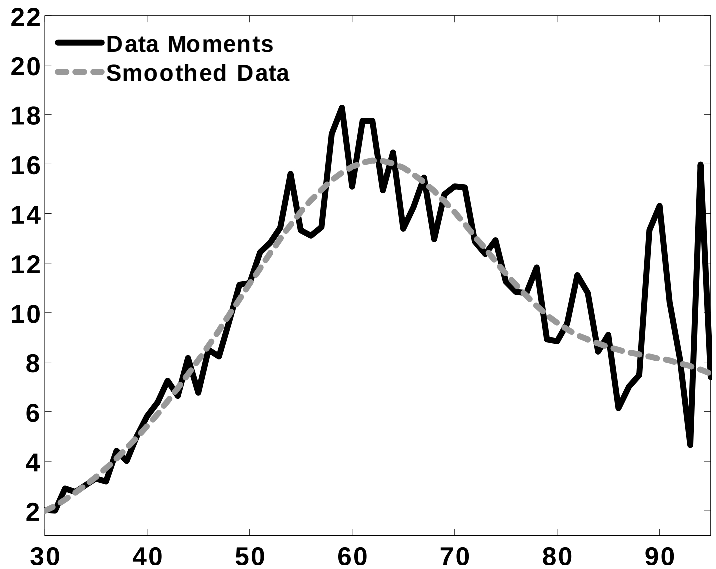
[34] To construct the data we adopt the approach described in Chris Carroll's lecture notes, cf. http://www.econ2.jhu.edu/people/ccarroll. We thank Chris Carroll for providing us the Stata code.

[35] The income measure includes 'wages and salaries', 'unemployment or worker's compensation', 'child support or alimony', 'TANF, food stamps, or other forms of welfare or assistance', 'net income from Social Security or other pensions', 'annuities, or other disability or retirement programs' and 'any other sources'. We exclude some few observations with negative income values.

[36] See the codebook for further details, e.g., http://www.federalreserve.gov/econresdata/scf/files/codebk2013.txt.

When displaying results in the main text, we compare simulated life-cycle asset profiles with smoothed data, whereby we apply a cubic spline regression. Figure 5 depicts the normalized assets from the SCF as well as the smoothed data.

Figure 5: Empirical Moments



Notes: This graph graph displays the normalized life-cycle average assets from the SCF (black line) and smoothed data with cubic spline (dashed grey line).

B.3 Simulated Method of Moments (SMM)

In the main text, we have conveniently suppressed in our notation that all model simulated variables are conditional on the values of first and second stage parameters. To formally introduce the SMM estimator we now make this contingency explicit by defining the average (or, unconditional) asset holdings at age t as

$$\bar{a}_t(\rho, \hat{\vartheta}) \equiv E\left[a_t \mid \rho, \hat{\vartheta}\right] = \frac{1}{N_t} \int a_t \Phi_t(da_t \times d\eta_t; \rho, \hat{\vartheta}),$$

where $\hat{\vartheta}$ is the vector collecting first stage parameter estimates. We normalize model simulated assets and life-cycle assets extracted from the data by permanent income. Because permanent income is not affected by δ, λ, we split $\hat{\vartheta}$ as $[\hat{\delta}, \hat{\lambda}, \hat{\vartheta}_1]$ and determine the unconditional permanent income as the constant annuity payment from the net present value of average (labor, respectively retirement) income over the life-cycle, discounted with the risk-free interest rate r. Normalized wealth simulated from the model is then $\bar{ay}_t(\rho, \hat{\vartheta}) \equiv \frac{\bar{a}_t(\rho,\hat{\vartheta})}{\bar{y}_t^p(\hat{\vartheta}_1)}$. Correspondingly, denote by $\bar{ay}_t^d$ the data analogue. We then seek to estimate the discount rate ρ from the moment conditions

$$m_t(\rho, \hat{\vartheta}) = \bar{ay}_t^d - \bar{ay}_t\left(\rho, \hat{\vartheta}\right)$$

for the age range 30 ($t_0 = 10$) to 95 ($T_0 = 75$). Collecting moment conditions in vector $m\left(\rho, \hat{\vartheta}\right)$ we solve for $\hat{\rho}$ by minimizing the loss function given in the main text:

$$J\left(\rho, \hat{\vartheta}\right) = m\left(\rho, \hat{\vartheta}\right)' W m\left(\rho, \hat{\vartheta}\right), \tag{44}$$

with weighting matrix W.

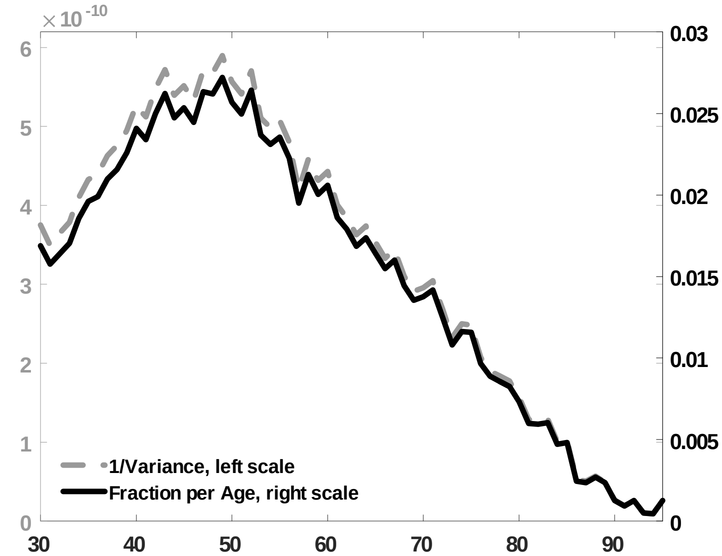
As documented in Altonji and Segal (1996), an optimal weighting matrix can be seriously biased in small samples. We instead use a 'robust' weighting matrix that does not depend on the fitted model. As in Ludwig and Zimper (2013) our weighting matrix contains the fraction of observations per age bin in the SCF data.[37] This choice of the weighting matrix ensures that the sample analogue of the moment conditions that are estimated with higher precision receive higher weight in the estimation, see Figure 6. Our approach is similar to using the inverse of the variances along the diagonal of the weighting matrix (and zeros for the off-diagonal elements) as recommended by Pischke (1995) and as is often used. Employing such an alternative weighting matrix, again see Figure 6 for the respective weights, gives almost identical results.

To conduct statistical inference, we follow Hall and Horowitz (1996) and bootstrap the distribution of the SMM estimator yielding better finite sample results than regular asymptotic theory. We apply two procedures to bootstrap our data. First, we simultaneously consider the sampling uncertainties at both stages, the one stemming from the HRS data on subjective survival beliefs at the first stage and the one from the SCF data on asset holdings at the second stage. Our results reported in the main text are based on this procedure. While this approach is appropriate when comparing across nested models, we cannot take into account the correlation between survival beliefs and savings behavior in the data because we draw from two different data sources (the subjective survival beliefs from the HRS and the asset data from the SCF). To investigate how important first stage sampling uncertainty is for statistical inference, we consider as a second approach only uncertainty with respect to normalized assets, i.e., we only bootstrap at the second stage. For both approaches, we implement the bootstrap by drawing with replacement $B = 500$ bootstrap samples $b = 1, \ldots, B$.

To construct the J test we adopt the pivotal refinement approach proposed by Hall and Horowitz (1996). The general idea is to set up a test statistic J where its distribution

[37] We adopt the same weighting scheme to derive our first-stage parameter estimates, $[\hat{\delta}, \hat{\lambda}]$, by using the fraction of observations in the HRS as weights.

Figure 6: Weights: Fraction of Observations and Inverse Variance



Notes: This graph graph displays the age-specific weights used in the estimation as the fraction of observation per age bin in the SCF data (solid line). It also displays an alternative weighting through the inverse of the variances of the data.

does not depend on unknown parameters. This is achieved by demeaning the vector of moment conditions $m(\rho, \hat{\vartheta})$.

We proceed by describing the first bootstrap procedure which takes into account first-stage sampling uncertainty. We employ the following steps:

1. Point estimates:

   (a) Based on data on subjective survival information from the HRS we estimate $[\hat{\delta}, \hat{\lambda}]$, cf. Ludwig and Zimper (2013) for details.

   (b) We collect all first-stage parameters as $\hat{\vartheta} = [\hat{\delta}, \hat{\lambda}, \hat{\vartheta}_1]$. Based on data on normalized life-cycle assets in the SCF we estimate, for given first stage parameter estimates $\hat{\vartheta}$, the parameter $\hat{\rho}$ by minimizing the loss function (44).

2. In each bootstrap iteration b:

   (a) We draw with replacement the subjective survival data in the HRS data sample to reestimate $[\hat{\delta}^b, \hat{\lambda}^b]$, cf. Ludwig and Zimper (2013) for details.

   (b) We draw with replacement all N individuals from the SCF data sample with probability $1/N$. Since each individual contains five observations in the SCF we perform a block-bootstrap such that for each draw we have five observations. For each bootstrapped sample we compute the corresponding wealth to permanent income ratio and the age-specific weights as described below. We denote the normalized wealth to permanent income ratio as $\bar{ay}_t^b$.

(c) Given the first-stage estimates $[\hat{\delta}^b, \hat{\lambda}^b]$ and the bootstrapped data $\bar{ay}_t^b$ as well as $\hat{\vartheta}$, we estimate parameter $\hat{\rho}^b$ by minimizing a re-parameterized loss function given by

$$\tilde{J}\left(\rho, \hat{\delta}^b, \hat{\lambda}^b, \hat{\vartheta}_1; \hat{\rho}, \hat{\delta}, \hat{\lambda}\right) = \tilde{m}\left(\rho, \hat{\delta}^b, \hat{\lambda}^b, \hat{\vartheta}_1; \hat{\rho}, \hat{\delta}, \hat{\lambda}\right)' W \tilde{m}\left(\rho, \hat{\delta}^b, \hat{\lambda}^b, \hat{\vartheta}_1; \hat{\rho}, \hat{\delta}, \hat{\lambda}\right)$$

where each age t entry of the vector of demeaned moment conditions $\tilde{m}(\cdot)$ is given by

$$\tilde{m}_t\left(\rho, \hat{\delta}^b, \hat{\lambda}^b, \hat{\vartheta}_1; \hat{\rho}, \hat{\delta}, \hat{\lambda}\right) = \left(ay_t^b - \bar{ay}_t\left(\rho, \hat{\delta}^b, \hat{\lambda}^b, \hat{\vartheta}_1\right)\right) - \left(\bar{ay}_t^d - \bar{ay}_t\left(\hat{\rho}, \hat{\delta}, \hat{\lambda}, \hat{\vartheta}_1\right)\right). \tag{45}$$

ay_t^d is the asset to permanent income ratio from the original data sample. Notice that term $\bar{ay}_t^d - \bar{ay}_t\left(\hat{\rho}, \hat{\delta}, \hat{\lambda}, \hat{\vartheta}_1\right)$ is a constant which depends on first and second stage parameter estimates. Minimization in bootstrap iteration b gives the bootstrap estimate

$$\hat{\rho}^b(\hat{\delta}^b, \hat{\lambda}^b, \hat{\vartheta}_1; \hat{\rho}, \hat{\delta}, \hat{\lambda}) = \arg\min_{\rho} \tilde{J}\left(\rho, \hat{\delta}^b, \hat{\lambda}^b, \hat{\vartheta}_1; \hat{\rho}, \hat{\delta}, \hat{\lambda}\right).$$

The p-value of the J-test for overidentifying restrictions is computed as

$$p\left(J\left(\hat{\rho}, \cdot\right)\right) = \frac{1}{B}\sum_{b=1}^{B} I\left(\tilde{J}(\hat{\rho}^b, \cdot) > J\left(\hat{\rho}, \cdot\right)\right),$$

where I is an indicator function equal to 1 if its argument is true and 0 otherwise, see Davidson and MacKinnon (2002). Observe that the (original) function value without demeaning, $J(\hat{\rho}, \cdot)$, enters on the right-hand-side of the inequality in the above. The model is rejected (on statistical grounds) if the p-value is low (less than 5% for a conventional level of statistical significance), i.e., if there are too few cases in which the demeaned function value exceeds the original one. The intuition is as follows: If the model is a statistically correct description of the data, then the constant term $\bar{ay}_t^d - \bar{ay}_t\left(\hat{\rho}, \hat{\delta}, \hat{\lambda}, \hat{\vartheta}_1\right)$—which is deducted from each moment condition in the process of demeaning, cf. equation (45)—is small in absolute value relative to the sampling uncertainty to the effect that the difference between function values based on demeaned moments, $\tilde{J}(\cdot)$, and original moments, $J(\cdot)$, is small. Hence there would be sufficiently many cases in which $\tilde{J}(\cdot) > J(\cdot)$.

To compute confidence intervals and to evaluate the RE model against the two NEO alternatives, i.e., to compare nested models, we employ the same steps as described above. As the only difference, we do not demean the moment conditions. The reason is that without pivotal refinement we obtain bootstrapped distributions for $J(\cdot)$ that differ between our models and that can be compared.[38] Hence, we reestimate the discount rate $\hat{\hat{\rho}}^b$ without pivotal refinement employing the steps described above but using the raw moment conditions defined as:

$$m_t\left(\hat{\hat{\rho}}^b, \hat{\delta}^b, \hat{\lambda}^b, \hat{\vartheta}_1\right) = \bar{ay}_t^b - \bar{ay}_t\left(\hat{\hat{\rho}}^b, \hat{\delta}^b, \hat{\lambda}^b, \hat{\vartheta}_1\right).$$

We compute confidence bands by using the percentile method. Similarly, we conduct inference by comparing distributions of $J(\cdot)$ across models. Accordingly, we compute the p-value as

$$p\left(\Delta J\left(\hat{\hat{\rho}}, \cdot\right)\right) = \frac{1}{B}\sum_{b=1}^{B} I\left(\Delta J(\hat{\hat{\rho}}) < 0\right)$$
$$\text{where } \Delta J(\hat{\hat{\rho}}) = J\left(\hat{\hat{\rho}}^{b,RE}, \cdot\right) - J\left(\hat{\hat{\rho}}^{b,NEO}\right),$$

and where I again is an indicator function equal to 1 if its argument is true and 0 otherwise. $\hat{\hat{\rho}}^{b,RE}$ ($\hat{\hat{\rho}}^{b,NEO}$) is the parameter estimate in bootstrap iteration b of the RE (the respective NEO) model. If the fraction $p\left(\Delta J\left(\hat{\hat{\rho}}, \cdot\right)\right)$ is small—say less than 5% for conventional levels of statistical significance—then the RE model is rejected against the respective alternative NEO model.

For our second approach we apply identical methods but do not bootstrap on the first-stage survival beliefs data. Formally, this approach is nested in the above by replacing $\hat{\delta}^b = \hat{\delta}, \hat{\lambda}^b = \hat{\lambda}$ for all $b = 1, \ldots, B$. Results reported in Table 8 are very similar because first-stage parameters are estimated with high precision, cf. Table 1.

[38] On the contrary, as noted above, the pivotal refinement ensures that the bootstrapped distribution for J does not depend on the chosen parameter and is thus (numerically) the same for all model variants.

Table 8: Sensitivity Analysis: Bootstrap Method

	naive NEO		sophisticated NEO	
	1st and 2nd	only 2nd	1st and 2nd	only 2nd
Discount Rate, $\hat{\rho}$	0.0347		0.0394	
Confidence Band	$[0.0320, 0.0372]$	$[0.0322, 0.0370]$	$[0.0368, 0.0416]$	$[0.0368, 0.0415]$
Overidentification Test				
Function Value	2.0619		2.2229	
p-value	1.4%	1.4%	1.0%	1.0%
Model Comparison relative to RE				
Average Distance	0.5024	0.5036	0.3448	0.3464
p-value	2.4%	2.2%	12.6%	13.6%

Notes: This table documents the sensitivity of our main results with respect to the bootstrap method. Our baseline results from Tables 2 and 4 are restated in column "1st and 2nd", column "only 2nd" are the corresponding results where we only consider second stage sampling uncertainty.