A Life-Cycle Model with Ambiguous Survival Beliefs

Max Groneck, Alexander Ludwig, Alexander Zimper

05-2013

MEA DISCUSSION PAPERS

Alte Nummerierung: 270-13

A Life-Cycle Model with Ambiguous Survival Beliefs[*]

Max Groneck[†]   Alexander Ludwig[‡]   Alexander Zimper[§]

March 19, 2015

Abstract

On average, "young" people underestimate whereas "old" people overestimate their chances to survive into the future. We construct a Choquet Bayesian learning model of ambiguous survival beliefs which replicates these patterns. These ambiguous survival beliefs are then embedded within a Choquet expected utility model of life-cycle consumption and saving. Our analysis shows that agents with ambiguous 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 probabilities. Our ambiguity-driven model therefore simultaneously accounts for three important empirical findings on household saving behavior.

JEL Classification: D91, D83, E21.

Keywords: Bayesian learning; Ambiguity; Choquet expected utility; Dynamic inconsistency; Life-cycle hypothesis; Saving puzzles

[†]CMR, University of Cologne; Netspar; Albertus-Magnus-Platz; 50923 Köln; Germany; E-mail: groneck@wiso.uni-koeln.de

[‡]SAFE, Goethe University Frankfurt; MEA; ZEW; 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

1 Introduction

Expectations about future survival are important for numerous economic decisions. Forming such expectations is a very difficult task. In fact, substantial biases of subjective survival beliefs relative to objective data have been documented in the literature. This paper provides a theory for the emergence of such biases and investigates their implications for life-cycle saving behavior.

Our work builds on two empirical regularities. First, according to the Health and Retirement Study (HRS), on average, "younger" people strongly underestimate their (relatively high) probability to survive to some target age. At the same time, "older" people strongly overestimate their lower survival probability. Figure 1 shows aggregated data from the HRS by plotting average age-specific biases in survival beliefs—the difference between the respective average subjective belief and the average objective data—for three waves of the HRS between 2000 and 2004. We observe that younger respondents between ages 50 and 70 persistently underestimate their survival chances by about 10 to 20 percentage points on average. Older respondents around the age of 85 persistently overestimate their survival chances by 15-20 percentage points. Also notice that the overestimation is getting more pronounced with increasing age.[1]

Second, recent empirical findings on household saving behavior proved to be puzzling for the standard "workhorse"-life-cycle 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. Another well-known puzzle is that people hold large amounts of assets still late in life and dissave less in old age than predicted by the standard model (see, e.g., De Nardi et al. 2010; Hurd and Rohwedder 2010; Lockwood 2013).

This paper asks in how far a decision theoretic explanation may simultaneously account for both empirical regularities, i.e., the observed biases in survival perceptions, on the one hand, and the empirical findings on saving behavior, on the other hand.[2]

---

[1]Similar patterns have been documented in numerous other datasets, cf., e.g., Hammermesh (1985), Elder (2013), Peracchi and Perotti (2010), and Wu et al. (2013). Also see Ludwig and Zimper (2013) for a detailed discussion of the data.

[2]Obviously, the biases of subjective survival perceptions from objective life-table data shown in Figure 1 should influence a household's consumption and saving decisions. For example, Salm (2010) estimates that a 1 percent increase in the subjective probability of mortality reduces annual future consumption of non-durable goods by around 1.8 percent. Bloom et al. (2006) find that an increased subjective survival probability leads to higher wealth accumulation thereby confirming results of Hurd et al. (1998).

Figure 1: Difference of Subjective Survival Probabilities and Cohort Data

(a) Women (b) Men
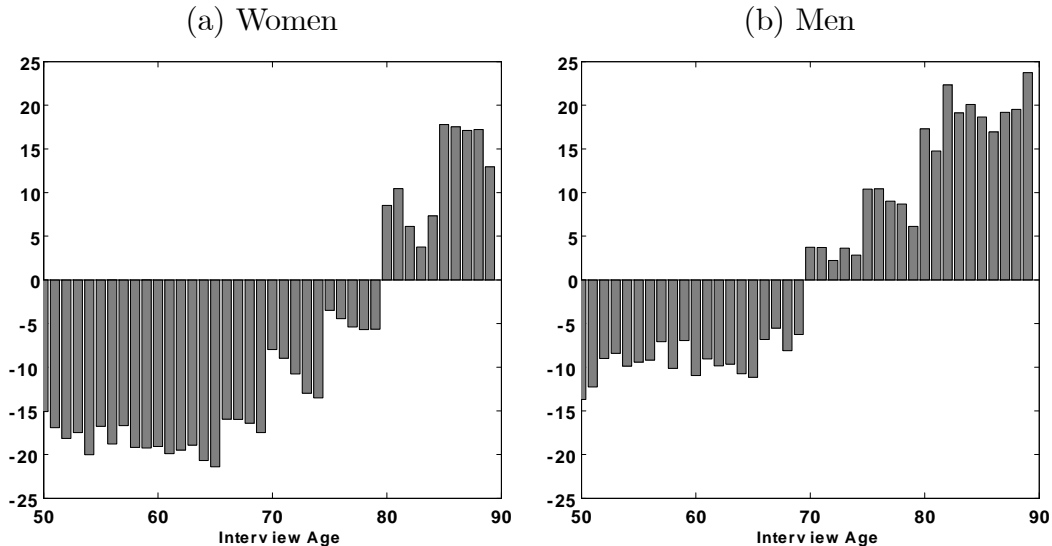
Notes: Deviations in percentage points of subjective survival probabilities from objective data. Objective survival rates are based on cohort life table data. Future objective data is predicted with the Lee-Carter procedure (Lee and Carter 1992). Each bar depicts the difference of unconditional probabilities to survive to a specific target age.

Source: Own calculations based on HRS, Human Mortality Database and Social Security Administration data.

More specifically, we consider a representative agent who expresses *ambiguity* in the sense that she does not resolve her uncertainty through a unique additive probability measure. Our approach comprises of two buildings blocks. As our first building block, we derive ambiguous survival beliefs as estimates from a model of Choquet Bayesian learning, extending earlier work in Ludwig and Zimper (2013).[3] As our second building block, we combine Choquet expected utility maximization with respect to the ambiguous survival beliefs derived from our calibrated learning model with a canonical life-cycle model. We next discuss both building blocks in turn.

Our construction of ambiguous survival beliefs provides an explanation for the biases of Figure 1 through a model of Bayesian learning. Accordingly, the decision maker is uncertain about her future survival chances (described as *parameter values*) whereby she observes with increasing age more and more data generated by the true parameter

---

[3]For axiomatic foundations of *Choquet expected utility* (CEU) theory see Schmeidler (1989) and Gilboa (1987). Because we restrict attention to gains, CEU theory in our model is equivalent to the celebrated *cumulative prospect theory* (CPT) (Tversky and Kahneman 1992; Wakker and Tversky 1993). Furthermore, note that our specific decision theoretic approach could be equivalently formalized as an "α-maxmin expected utility with multiple priors" model of Ghirardato et al. (2004, Proposition 19).

value. Clearly, the persistent biases of Figure 1 are inconsistent with classical models of Bayesian learning whose estimates converge to objective probabilities when people observe more data. In a first step, we therefore develop a model of Choquet Bayesian learning such that the decision maker expresses her ambiguity about the joint distribution of the parameter and sample space through (i) a neo-additive capacity in the sense of Chateauneuf et al. (2007) which (ii) she updates in accordance with the Generalized Bayesian update rule (Pires 2002; Eichberger et al. 2007). Under simplifying assumptions, we derive a closed-form expression for the Choquet estimates of survival chances that only depends on the decision maker's age. In a next step, we demonstrate that these age-dependent Choquet estimates can themselves be reinterpreted as neo-additive capacities defined on the space of the decision maker's survival events.

Although the ambiguous survival beliefs of our model are thus formally described as neo-additive capacities, it is important to notice that this neo-additive structure is not imposed ad hoc but is a formal implication of our Choquet Bayesian learning model. The formal relationship between the parameters of the conditional neo-additive capacities of the Choquet learning model and the parameters of the resulting age-dependent neo-additive survival beliefs has a number of interesting implications. For example, even if our Choquet Bayesian learning model reduces to classical Bayesian learning with additive probabilities (i.e., no ambiguity in the learning model), the decision maker's survival beliefs only become non-ambiguous in the limit of the learning process where they resemble objective survival probabilities. In general, however, our Choquet estimates do not converge to objective survival probabilities to the effect that the calibrated version of our learning model will be able to generate ambiguous survival beliefs which replicate the biases of Figure 1.

Having characterized survival beliefs as age-dependent neo-additive capacities, we describe in our second building block the representative agent's life-cycle utility over consumption streams as her Choquet expected utility (CEU) with respect to the neo-additive survival beliefs derived from our calibrated learning model. Our analysis then investigates whether biases in survival beliefs can partially resolve saving puzzles. To this purpose we compare consumption and saving behavior of CEU agents with the special case of rational expectations (RE) agents who are described by the limit (i.e., by an infinite amount of statistical information) of our learning model in the absence of ambiguity. Whenever CEU agents do not converge to RE agents, life-cycle maximization gives rise to dynamically inconsistent behavior. We study both 'naive' and 'sophisticated' CEU 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.

A qualitative analysis for a simplified three-period model, presented in a Supplementary Appendix[4], shows that naive as well as sophisticated CEU agents exhibit undersaving relative to their RE counterparts if they sufficiently underestimate objective survival probabilities at young ages. Furthermore, for the model to give rise to the phenomenon that households save less in the intermediate model period than originally planned, the survival beliefs of dynamically inconsistent naive CEU agents have to feature only moderate overestimation in this intermediate period (otherwise they would save more). At the same time, because of their relative optimism in the intermediate model period naive CEU agents save more out of cash on hand than the corresponding RE agent. However, whether asset holdings in the final period are higher for the CEU agent depends on the interplay between underestimation at young ages and overestimation at older ages. Whether these conditions hold and how relevant the biases in beliefs are for generating saving puzzles are quantitative questions.

To address these questions we calibrate the stochastic quantitative life-cycle model to the data. With the exception of the discount rate, we determine all parameters outside the model. We pin down the discount rate to minimize the distance of life-cycle asset holdings between the model and the data. With this strategy we do not directly target any of the aforementioned theoretical conditions required for the CEU model to (partially) resolve saving puzzles.

We indeed find that the calibrated RE model gives rise to these puzzles: The average saving rate for prime age savers of age $25-54$ is at 13.5%, 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 70.0%, 37.0% and 9.1%, compared to 72.4%, 53.0% and 47.9% in the data. Hence, through the lens of the RE model, the data are puzzling: relative to the model the young save too little and the old decumulate assets too fast in the data. The calibrated naive CEU agents model partially resolves these puzzles. The average saving rate is at 9.4% and relative asset holdings at ages 75, 85 and 95 are at 77.7%, 56.8% and 34.8%. These statistics are remarkably close to the data. In addition, the realized saving rate is 5.5 percentage points lower than the planned saving rate. Predictions on asset holdings for the sophisticated agent CEU model are similar. They save a bit more than naive CEU agents and hence feature slightly higher asset holdings in old age. Overall, the fit to the data is better for naive than for sophisticated agents. Our analysis therefore suggests that our notion of ambiguous survival beliefs combined with naivety provides an accurate quantitative picture of saving behavior until about age 85. Importantly, we also document that this success of the naive agents' CEU model does

---

[4]The Supplementary Appendix is available online at http://www.wiwi.uni-frankfurt.de/fileadmin/user_upload/dateien_abteilungen/abt_ewf/LS_Ludwig/SubjBeliefs_SuppApp.pdf.

not hinge on recalibrating the discount rate. Results are virtually unchanged when we hold it constant at the value calibrated for the RE model.

The intuition for these quantitative findings is as follows: The calibrated model gives rise to sufficient underestimation at young age so that naive CEU agents save less than their RE counterparts. At the same time, naive CEU households only moderately overestimate their future survival chances so that they end up saving less in each period than originally planned in the past. As agents get older, overestimation of future survival beliefs eventually dominates so that the speed of asset decumulation is lower to the effect that the level of old age asset holdings is eventually higher than for RE agents. Finally, sophisticated agents correctly anticipate the more optimistic beliefs of their future selves. For reasons of consumption smoothing they therefore save more which leads them to have higher old-age asset holdings 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), 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 as in the rational expectations model. In contrast, over-estimating beliefs for low probabilities in our CEU model implies lower long-run effective discount rates which leads to higher old-age asset holdings. In this respect our work relates to Halevy (2008) as well as Epper et al. (2011) who argue that hyperbolic time discounting is actually generated by ambiguous survival beliefs. Motivated by this insight, we show in our companion paper (Groneck et al. 2014) that quasi-hyperbolic time-discounting over the life-cycle is formally equivalent to a static CEU life-cycle model in which agents hold globally under-estimating neo-additive survival beliefs that are not subject to Bayesian learning. The CEU model of the present paper is thus formally different from any quasi-hyperbolic time-discounting model. First, it considers neo-additive survival beliefs that can express both, underestimation of large as well as overestimation of low probabilities (cf., the inverse S-shaped probability weighting function of CPT). Second, these neo-additive survival beliefs are not superimposed ad hoc but are derived from a model of Bayesian learning over the life-cycle.

Similarly, standard explanations for insufficient old-age asset decumulation such as a bequest motive (Hurd 1989; Lockwood 2013) and precautionary savings behavior (Palumbo 1999; De Nardi et al. 2010) cannot generate undersaving at young ages. Our model of ambiguous survival beliefs therefore adds to existing explanations for saving behavior by simultaneously generating all three stylized findings: (i) time inconsistency, (ii) undersaving at young and (iii) high asset holdings at old age.

The remainder of our paper is organized as follows. Section 2 constructs our model of Choquet Bayesian learning over the agent's life-cycle. In Section 3 we use the resulting age-dependent Choquet estimators to construct neo-additive probability spaces that characterize, for all ages, the representative agent's ambiguous survival beliefs. Section 4 combines our notion of ambiguous survival beliefs with a multi-period stochastic life-cycle model. Calibration is outlined in Section 5 and 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.

2 Bayesian Learning of Survival Beliefs

Standard life-cycle models with rational expectations use objective survival probabilities, denoted $\psi_{k,t}$ with $k < t$, to model the representative agent's beliefs to survive from age k to age t. These beliefs are independent of the agent's age because there is no learning of survival beliefs over the life-cycle: the rational expectations agent always already knows her true survival chances. Figure 1 demonstrates that real life people do not know their true survival chances. In the absence of such knowledge, it is plausible that some learning of survival beliefs happens over the representative agent's life-cycle.

Instead of a rational expectations agent let us therefore consider a Bayesian decision maker who updates her estimator of her chance to survive from k to t by observing statistical information as she grows 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 read increasingly many health studies/articles.

We start out with the formal description of a classical Bayesian learner whose uncertainty is captured by a unique additive probability measure. Because the estimators of this classical Bayesian learning model converge towards objective survival probabilities, we argue that this learning model cannot plausibly explain the HRS data on subjective survival beliefs (cf. also Ludwig and Zimper (2013)).

In a next step, we construct a model of Choquet Bayesian learning which nests the classical model of Bayesian learning as a special case. Whenever the agent's uncertainty about the joint parameter and sample space cannot be described by some additive probability measure, the resulting Choquet estimators will not converge to objective survival probabilities. As a consequence, the appropriately calibrated age-dependent Choquet estimators will be able to capture the persistent biases between subjective survival beliefs and objective survival probabilities depicted in Figure 1.

2.1 Classical Bayesian Learning

Let us consider a classical Bayesian decision maker who satisfies Savage's (1954) axioms such 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-sigma 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. 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 sigma-algebra $\Sigma(X^\infty)$ and denote by $\Sigma(\Theta \times X^\infty)$ the product sigma-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 $\Sigma(X_{e(h)})$-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

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

8

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[5]

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

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

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

---

[5]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).

where $K_{\alpha,\beta} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$ is a normalizing constant.[6] Given the Binomial distribution (1), we obtain by Bayes' rule the following conditional distribution of $\tilde{\theta}$

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

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

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

That is, the updated estimator $E\left[\tilde{\theta}, \mu\left(\tilde{\theta} \mid \tilde{I}_{e(h)}\right)\right]$ 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.

To sum up: Classical Bayesian learning models imply convergence of all subjective survival beliefs to objective survival probabilities as the agent gains more experience when growing older. However, the age-specific pattern of the biases in Figure 1 suggests that such converging learning behavior might, in reality, not happen over the life cycle: instead of convergence to objective survival probabilities, strong underestimation of objective survival probabilities is persistent for low target ages whereas strong overestimation is persistent for high target ages as the representative agent grows 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 now consider a Choquet decision maker who

---

[6]The gamma function is defined as $\Gamma(y) = \int_0^\infty x^{y-1}e^{-x}dx$ for $y > 0$.

satisfies the axioms of Choquet expected utility theory (e.g., Schmeidler 1989; Gilboa 1987) so that her uncertainty is resolved by a unique *capacity* (=not necessarily additive probability measure) κ rather than by the additive probability measure μ. Formally, $\kappa : \Sigma\left(\Theta \times X^\infty\right) \to [0,1]$ satisfies

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

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

In Choquet decision theory random variables are integrated via the Choquet integral. Formally, the Choquet integral of a bounded $\Sigma\left(\Theta \times X^\infty\right)$-measurable function $f : \Theta \times X^\infty \to \mathbb{R}$ with respect to the capacity κ is defined as the following Riemann integral (cf. Schmeidler 1986)[7]:

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

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

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

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

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

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

---

[7]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

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

which is the familiar method of integrating up some utility function f in *Rank Dependent Utility Theory* or in CPT applied to gains.

At this point, however, we run into two difficulties. First, for Choquet decision makers there exists a multitude of alternative update rules for non-additive probability measures so that there exists, unlike for additive probability measures, no unique definition of a conditional capacity $\kappa\left(\cdot\mid\cdot\right)$. Second, even if we have settled for a specific update rule, it is difficult to work with the Choquet estimator (7), if we do not impose further restrictions on the class of admissible capacities. To address both difficulties, we restrict attention to the class of *neo-additive capacities* (Chateauneuf et al. 2007) which are updated in accordance with the *Generalized Bayesian update rule* (Pires 2002; Eichberger et al. 2007).

Neo-additive Capacities

Neo-additive capacities are an analytically very tractable class of non-additive probability measures which are used in the literature[8] to approximate inverse S-shaped probability weighting functions as typically elicited for CPT (cf., e.g., Tversky and Kahneman 1992; Wu and Gonzalez 1996; 1999).

Recall that the set of *null events*, denoted $\mathcal{N}$, collects all events that the decision maker deems impossible.

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

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

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

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

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

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

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

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

---

[8]See, e.g., Wakker (2010), Abdellaoui et al. (2011), and Ludwig and Zimper (2013).

for all $A \neq \emptyset, \Theta \times X^\infty$. The parameter $\delta \in [0, 1]$ is interpreted as a degree of ambiguity. If there is no ambiguity ($\delta = 0$), ν reduces to the additive probability measure μ. If there is ambiguity ($\delta > 0$), the parameter $\lambda \in [0, 1]$ measures in how far the agent resolves this ambiguity about an event A through over- (high values of λ) versus under-estimation (low values of λ) with respect to the additive probability $\mu(A)$.

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

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

Obviously, if there is no ambiguity, 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

CEU theory has been developed in order to accommodate paradoxes of the Ellsberg (1961) type which show that real-life decision-makers violate Savage's (1954) sure thing principle. Abandoning the sure thing principle implies that there exist several perceivable Bayesian update rules for non-additive probability measures (cf., e.g., Gilboa and Schmeidler 1993; Epstein and Le Breton 1993; Ghirardato 2002; Siniscalchi 2011).

In the present paper we assume that the representative agent forms conditional capacities in accordance with the *Generalized Bayesian* (=GB) update rule. 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*.[9] Moreover, the GB update rule is analytically very tractable whereby it avoids

---

[9]For formal definitions and discussions of these axioms, see, e.g., Pires (2002), Eichberger et al. (2007), and Siniscalchi (2011).

the extreme updating behavior of alternative update rules with axiomatic foundations such as, e.g., the *optimistic* or the *pessimistic* update rule (cf. Gilboa and Schmeidler 1993). According to the GB rule we formally define, for all non-null $A, B \in \Sigma (\Theta \times X^{\infty})$,

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

An application of this update rule to a neo-additive capacity ν gives rise to the following observation.

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

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

such that

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

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

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

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

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

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., $\delta = 0$.

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 $e\left(h\right)$,*

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

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 $e\left(h\right)$ 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 $e\left(h\right)$ 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 $e\left(h\right)$.

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

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

such that

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

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 $\psi_{k,t}$.[10]

3 Ambiguous Survival Beliefs

So far we have been concerned with the 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 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 neo-additive capacity which describes the h-old agent's ambiguous beliefs to survive from a fixed age k to any given age t.

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

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

---

[10]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.

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[11]

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

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

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 $(\Omega, \mathcal{F}^h)$. This observation gives rise to the central definition of our paper, which translates our notion of Choquet Bayesian estimators of survival chances into a neo-additive probability space for survival events.

Definition 2. *Fix some age $h = 1, ..., T$ and some $k \geq h$. Define the neo-additive probability space $(\Omega, \mathcal{F}, \nu_k^h)$ 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 ambiguity 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.

3.2   Discussion

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

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

such that the conditional neo-additive capacity $\nu(\cdot \mid \cdot)$ is characterized by the initial parameters δ and λ combined with an application of the Generalized Bayesian update

---

[11]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 $\lambda \geq \frac{1}{2}$.

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 parameters δ_h and λ_h, is defined as the h-old agent's Choquet estimator of the underlying learning model. Consequently, the parameter values δ_h and λ_h are comprehensively pinned down through equations (18) and (19) by the values of the parameters δ, λ and the agent's age-dependent experience $e\left(h\right)$.

To see how the age-conditional ambiguous survival beliefs ν_k^h depend on the specification of the underlying Choquet Bayesian learning model let us consider three different scenarios. First, suppose that there is no initial ambiguity 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 all values of the experience function $e\left(h\right)$, ν_k^h does not reduce to an additive probability measure because we have a strictly positive ambiguity parameter*

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

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

Hence, the RE model is nested as a special case for $\delta = 0$ and $e(h) \rightarrow \infty$.

As a second scenario, suppose now that there is initial ambiguity in the Choquet Bayesian learning model but that there is no age-dependent learning. In this "static" scenario, the agent's ambiguous survival beliefs thus remain constant over all ages so that, for all h, $e\left(h\right) = n$ 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.

We have already argued in Section 2.2, that the convergence behavior of classical Bayesian learning towards rational expectations is at odds with the data. Similar, 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 ambiguity with Choquet Bayesian learning over the life-cycle such that the agent's experience function $e(h)$ strictly increases in her age h. In this scenario, the age-dependent ambiguous survival beliefs (10) 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.,*

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

Because ambiguous survival beliefs do not converge to objective probabilities whenever there is ambiguity in the Choquet Bayesian learning model, our notion of ambiguous survival beliefs will be able to replicate the age-dependent bias patterns of Figure 1.

Remark. The ambiguity parameter δ_h, given by (18), first decreases in the sample sizes $e(h)$ whereas it increases for all $e(h)$ such that

$$e(h) \geq \sqrt{\frac{1}{2\delta}}.$$

That is, for any $\delta > 0$, the agent's ambiguity with respect to her survival chances will eventually increase in the amount of statistical information that she receives where

$$\lim_{e(h)\to\infty} \delta_h = 1.$$

This feature might seem to be counter-intuitive because one possible interpretation of ambiguity is the lack of sufficient statistical information to form a unique additive belief. Although an in-depth discussion of the ongoing (and fascinating) research on Bayesian learning under ambiguity is beyond the scope of this paper, we briefly discuss the plausibility of this feature in the Supplementary Appendix.

4   Quantitative Life-Cycle Model

This section merges our notion of ambiguous survival beliefs with a life-cycle model. One model period corresponds to one age year. We model a realistic life-cycle income

profile including stochastic and age-specific labor productivity. In addition, a PAYG pension system is modeled with a fixed date of retirement. We assume no annuity markets and a self-imposed borrowing constraint (because there is always a small positive probability of drawing zero income). These elements are included only in order to generate realistic endogenous life-cycle consumption profiles. (Self-imposed) borrowing constraints, stochastic labor income in combination with impatience give a hump-shaped consumption profile, as in the data. Positive pension income implies that savings for retirement are not too large.

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. The number of households of age t is denoted by N_t. Population is stationary and we 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 $h \le t$ as the planning age of the household, i.e., the age when households make their consumption and saving plans for the future. At ages $h = 1, \ldots, T$, households face objective risk to survive to some future period t. We denote corresponding objective survival probabilities for all in-between periods k, $h \le k < t$, by $\psi_{k,t}$ where $\psi_{k,t} \in (0,1)$ for all $t \le T$ and $\psi_{k,t} = 0$ for $t = T+1$. We think of survival risk as an idiosyncratic risk that washes out at the aggregate level. Total population is therefore constant and dynamics of the population are correspondingly 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 $t = 0, 1, \ldots, t_r - 1$ 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. For sake of comparability, 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) and references therein. By $\eta^t = (\eta_1, \ldots, \eta_t)$ we denote a history of shocks and $\eta^t \mid \eta^h$ with $h \le t$ is the history $(\eta_1, \ldots, \eta_h, \ldots, \eta_t)$. Let E be the powerset of the finite set E. E^{t_r-1} are σ-algebras generated by $\mathsf{E}, \mathsf{E}, \ldots$. 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 fact, we obtain almost identical numerical results for such a model.

In addition, we assume productivity to vary by age. Accordingly, ϕ_t denotes age-specific productivity which is estimated from the data and results in a 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 continuous 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. $\mathsf{E}^{r\,T-t_r+1}$ are σ-algebras generated by $\mathsf{E}^r, \mathsf{E}^r, ...$ and $\left(\times_{t=t_r}^{T} \mathsf{E}^r, \pi^r \right)$ is the objective probability space in the retirement period.

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

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

We abstract from private annuity markets.[12] The interest rate, r, is assumed to be fixed. With cash-on-hand given as $x_t \equiv a_t \left(1 + r \right) + y_t$ the budget constraint writes as

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

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

---

[12]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, overestimation of survival rates only sets in after the age of 70 and the average underestimation in our total sample is around 27 percentage points.

4.3 Government

We assume a pure PAYG public social security system. Denote by χ the net pension benefit level, i.e., the ratio of pensions to net wages. The government budget is assumed to be balanced each period and is given by

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

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 CEU Preferences

Households face two dimensions of uncertainty, respectively risk, about period t consumption. First, due to our assumption of productivity shocks, agents face a risky labor income. Second, agents are uncertain with respect to their life expectancy. While we model income risk in the standard objective EU way, we model uncertainty about life-expectancy in terms of a CEU agent who holds ambiguous survival beliefs as stated in Definition 2.

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 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 then 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 $\beta = \frac{1}{1+\rho}$.[13] 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 then defined as

---

[13]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.

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

We model the h-old agent as a Choquet decision maker whose survival uncertainty is expressed through the ambiguous 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 agent's Choquet expected utility from consumption plan $\mathbf{c}$ with respect to ν_h^h is given as (cf. Observation 1)

$$\mathbb{E}\left[U\left(\mathbf{c}\right), \nu_h^h\right] = \delta_h\left[\lambda_h \sup_{s\in\{h,h+1,\dots\}} U\left(\mathbf{c}\left(s\right)\right) + (1-\lambda_h) \inf_{s\in\{h,h+1,\dots\}} U\left(\mathbf{c}\left(s\right)\right)\right] \tag{26}$$

$$+ (1-\delta_h) \cdot \sum_{s=h}^{T} \left[U\left(\mathbf{c}\left(s\right)\right), \psi\left(D_s \mid Z_h\right)\right]$$

where $\psi\left(D_s \mid Z_h\right)$ 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

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

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

Proposition 2. *Consider an agent of age h. The 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 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.

23

Because the parameter λ_h determines how much decision weight is (additionally) put on the best versus worst possible utility scenario in the CEU life cycle model (26), we henceforth call λ_h a "relative optimism" parameter. This motivational interpretation of λ_h is somewhat 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 agent constructs a consumption and saving plan that maximizes her lifetime utility. The age-dependent sequence of neo-additive probability spaces $(\Omega, \mathcal{F}, \nu_h^h)$, $h = 1, ..., T$, violates dynamic consistency of the agent's life-cycle utility maximization problem whenever the ambiguous survival beliefs do not reduce to the limiting case of rational expectations, i.e., for all h, $\nu_h^h = \psi(\cdot \mid Z_h)$.[14] To characterize actual behavior in presence of dynamic inconsistency, we analyze both naive and sophisticated agents, cf. Strotz (1955) or inter alia O'Donoghue and Rabin (1999) for procrastination models.

A naive 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 implement the remaining plan. In contrast, sophisticates fully understand the dynamic inconsistency whereby they incorporate the correctly 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.,

---

[14]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 CEU theory) to Epstein and Le Breton (1993), Ghirardato (2002), Siniscalchi (2011) and the Appendix in Zimper (2012).

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

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

Maximization of the above is subject to (23).

Naive Agents

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

Proposition 3. *The Euler equation for the naive CEU 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

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

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 directly verify that the CEU life-cycle maximization problem is dynamically inconsistent if and only if the ambiguous 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

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

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

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

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

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

which holds for $\delta = 0$, $e(h) \to \infty$ implying

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

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

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

As in the static CPT model of Halevy (2008), the life-cycle maximization problem of naive CEU agents is thus dynamically inconsistent. While dynamic inconsistency in Halevy (2008) results from a fixed non-additive probability weighting function, dynamic inconsistency in our model comes with a sequence of non-additive probability weighting functions. Recall from our discussion of Section 3.2 that the latter are a consequence of both updating of survival beliefs for finite experience and ambiguous survival beliefs.

Sophisticated Agents

Sophisticated 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.

Combining first order conditions of optimality for the CEU agent results in a "generalized Euler equation with adjustment factor":

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

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

where

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

26

and

$$\Lambda_{h+1} \equiv \beta(1+r)\frac{\nu_{h,h+2}^h}{\nu_{h,h+1}^h}\left(1-m_{h+1}\right)\left(\frac{\partial V_{h+2}^h}{\partial x_{h+2}} - \frac{\partial V_{h+2}^{h+1}}{\partial x_{h+2}}\right). \qquad (35)$$

Proof: See Appendix A.

Relative to the naive agent, the FOC of the sophisticated 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 $\beta\nu_{h,h+1}^h\Theta_{h+1}$ as the "effective discount factor". The condition is easiest to interpret by noticing that $\Theta_{h+1} > 1$ iff $\varphi_h \equiv \frac{\nu_{h,h+2}^h}{\nu_{h,h+1}^h \cdot \nu_{h+1,h+2}^{h+1}} > 1$, which holds in our calibration of the CEU model. In this case the marginal propensity to save (MPS) 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 $\beta\nu_{h,h+1}^h$. 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 self $h+1$ values consumption more—by consuming more out of cash on hand—then self h compensates this overconsumption 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 CEU agents the value functions of selves h and $h + 1$ in periods $h + 2$ are age-dependent. A positive difference $\frac{\partial V_{h+2}^h}{\partial x_{h+2}} - \frac{\partial V_{h+2}^{h+1}}{\partial x_{h+2}}$ 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

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

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

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

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

Aggregation gives average (or aggregate)

consumption:	$\bar{c}_t = \int c_t(a_t, \eta_t)\Phi_t(da_t \times d\eta_t),$
assets:	$\bar{a}_t = \int a_t \Phi_t(da_t \times d\eta_t),$
income:	$\bar{y}_t = \int y_t(\eta_t)\Phi_t d\eta_t,$
total income:	$\bar{y}_t^{tot} = \bar{y}_t + r\bar{a}_t,$
saving rate:	$\bar{s}_t = \int s_t(a_t, \eta_t)\Phi_t(da_t \times d\eta_t),$ 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 CEU 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 CEU agents we denote the planned average saving rate with superscript h for the respective planning age and compute

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

for all t. This gives hypothetical average profiles of the saving rate in the population if households would 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 CEU agents that

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

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

5   Calibration

With the exception of the discount rate, all parameters of our baseline scenario are calibrated without using the life-cycle model. We refer to these parameters, summarized in Table 1, as (exogenous) first-stage parameters. The discount rate is accordingly referred to as (endogenous) second-stage parameter, cf. Table 2. The remainder of this section provides the details of our approach.

5.1   Household Age

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).[15] We set the horizon to a maximum biological human lifespan at age 125, hence $T = 105$. This choice is motivated by estimates based on Swedish female life-table data by Weon and Je (2009).

5.2   Objective Cohort Data

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 becomes 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.[16] This is due to the fact 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

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

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. The resulting predicted mortality rate function fits actual data very well, cf. Figure S.3 of the Supplementary Appendix. We use it as objective cohort data in the simulation.

According to our parameter estimates reported in Table 1, the implied average mortality rate converges to a value of 0.57 at ages around 110 ($t = 90$). This is well in line

---

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

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

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.

5.3   Estimated Subjective Survival Beliefs

We follow Ludwig and Zimper (2013) and estimate parameters δ and λ, cf. equations (10) and (18), to match the HRS data. Subjective survival rates are obtained by pooling a sample of HRS waves $\{2000, 2002, 2004\}$. Except for heterogeneity in sex and age, we ignore all other heterogeneity across individuals. Before proceeding with the estimation, the experience function $e(h)$ remains to be specified. Since the experience needs to be a positive integer, a general functional form 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. In our baseline specification, we therefore restrict the learning speed such that $\omega = 1$ and consider an alternative parametrization for sensitivity analysis (see below). With this baseline specification and parametrization we get $\delta = 0.0163$ and $\lambda = 0.413$ and implied values for δ_h and λ_h that lie well within reasonable ranges discussed in the literature (see, e.g., Wakker 2010; Abdellaoui et al. 2011), cf. Figure 3 below.[17] These parameters are estimated with very high precision, also see Ludwig and Zimper (2013).

The predicted subjective survival rates resulting from our model of ambiguous 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 2 in Section 6. The R^2 of the regression is around $0.8 - 0.95$.[18]

As a robustness check, we investigate the relevance of the parametrization of the experience function. In particular, we consider 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. Additionally, we study a calibration where we jointly identify ρ and ω to give the best fit to the data on asset holdings. Further details are provided in Subsection 6.2.3.

---

[17]Estimation results are calculated separately for men and women. We take an equally weighted average of the estimated parameters to get an approximation for λ and δ in the population.

[18]The fit is slightly better for women than for men, cf. Ludwig and Zimper (2013). They further perform sensitivity analyses with regard to the choice of the initial age, the specific form of the experience function and focal point answers. This shows that results do not hinge on these aspects. Finally, they document that biases in beliefs are neither due to cohort effects nor selection biases.

5.4 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. Following the estimates of Carroll (1992) we set the probability of receiving zero labor income to $\zeta = 0.005$. Then 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 $t = 0, \ldots, t_r$. We take as initial probability vector of the Markov chain $\pi_0 = [0.5, 0.5, 0]'$, i.e., households do not draw zero income in their first period of life.

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 $\{\phi_t\}$ of wages is estimated based on PSID data applying the method developed in Huggett et al. (2007), 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 $\tau = 0.124$. The pension benefit level then follows from the social security budget constraint (24).

5.5 Preferences

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. For the intertemporal elasticity of substitution (IES), $1/\theta$, we take a conventional value chosen in the literature of $1/3$, i.e., $\theta = 3$. 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 violate our assumption that utility from survival is positive thereby exceeding utility from death. We cure this by two additional modifications of the utility function. First, we add an additive preference shifter to the per period utility function, denoted by $\Upsilon > 0$. With this monotone

Table 1: First-Stage Parameters

Parameter		Source
Technology and Prices		
$w = 1$	Gross wage	normalized
$r = 0.042$	Interest rate	Siegel (2002)
$\tau = 0.124$	Social security contribution rate	irs.gov
$\chi = 0.322$	Net pension benefit level	model outcome
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
$\zeta = 0.005$	Probability of zero labor income	Carroll (1992)
$\zeta^r = 0.001$	Probability of zero retirement income	
Preferences		
$\theta = 3$	Coefficient of relative risk aversion	
$\underline{c} = 1.0e - 08$	Minimum consumption level	
Subjective Survival Beliefs		
$\delta = 0.0163$	Initial degree of ambiguity	HRS
$\lambda = 0.413$	Initial degree of over/underestimation	HRS
$\omega = 1$	Speed of the learning process	
Age Limits and Survival Data		
0	Initial model age (age 20)	
$t_r = 42$	Retirement (age 62)	SCF
$T = 105$	Maximum human lifespan (age 125)	Weon and Je (2009)
$\{\psi_{k,t}\}$	Cohort survival rates	Predictions based on 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.

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

Target (Source)	Asset profile (SCF)		
RE	ρ^{RE}	$=$	0.0344
naive CEU	$\rho^{CEU,n}$	$=$	0.0343
sophisticated CEU	$\rho^{CEU,s}$	$=$	0.0426

Notes: Second-stage parameters are calibrated such that asset moments from the model best match corresponding data moments.

transformation we can ensure (via calibration) that utility from survival is always positive. Of course, this does not affect optimal choices. Second, we take a Stone-Geary specification of the utility function and accordingly let $c_t - \underline{c}$ be its argument for some very small $\underline{c} > 0$. To understand this second modification observe that our specification of the income process with a positive zero income probability achieves differentiability of policy functions and positive asset holdings (and hence consumption) throughout but very low consumption levels have positive probability. This makes it very hard to assign values to Υ such that utility from survival is always positive. With a Stone-Geary-CRRA utility function we have that optimal consumption choices satisfy $c_t > \underline{c}$ because of the lower Inada condition. Accordingly, setting $\Upsilon = -u(\underline{c})$ achieves strictly positive utility in case of survival. Of course, for very small $\underline{c}$, the effect of this modification on optimal choices as well as on the IES is negligibly small. Collecting elements, the per-period utility function reads as

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

for $\Upsilon = -\frac{\underline{c}^{1-\theta}}{1-\theta}$.

A key preference parameter of the model is the discount rate which we take as the only second stage parameter in our baseline specification. We calibrate it such that the average asset-to-permanent-income ratio from the model best matches the empirical counterpart. This approach is in the spirit of Gourinchas and Parker (2002) and De Nardi et al. (2010). Data on assets and permanent income is taken from the SCF. Appendix B describes in more detail how the data is constructed.

Denote by $\bar{a}_t^{data}$ average age-specific net-worth and by $\bar{y}p_t^{data}$ average permanent-income constructed by pooling SCF data from 1992 to 2007. As defined in Section 4.6, $\bar{a}_t$ is the model counterpart. Correspondingly, we denote model permanent income by $\bar{y}_t^p$ which is calculated as the constant annuity payment from the net present value of average (labor, respectively retirement) income $\bar{y}_t$ over the life-cycle discounted with the riskfree interest rate $r = 0.042$.

We target the life-cycle profile between ages 30 ($t_0 = 10$) and 90 ($T_0 = 70$). 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. Our choice of the terminal age at 90 is due to data limitations at very high ages. Beyond age 90 there are too few observations on assets in the data so that (smoothed) asset age profiles get rather wiggly. Accordingly we search for ρ to solve

$$\min_{\rho} \frac{1}{2} \sum_{t=t_0}^{T_0} \left(\frac{\bar{a}_t^{data}}{(\bar{y}_t^p)^{data}} - \frac{\bar{a}_t(\rho)}{\bar{y}_t^p(\rho)} \right)^2. \tag{38}$$

For our baseline results, we calibrate a different subjective time discount rate ρ for each of the three models, the RE, the naive and the sophisticated CEU model. Parameter estimates in Table 2 document that the difference between subjective discount factors calibrated for the RE and the naive CEU model is small whereas the difference to the sophisticated CEU model is large. In Section 6.2 we explain the reason for these differences. Importantly, we also investigate how results are affected by recalibration. In these experiments we hold the discount rate constant at its calibrated value for the RE agent.

6 Results

6.1 Ambiguous versus Rational Survival Beliefs

Figure 2 compares predicted subjective survival rates resulting from our model of ambiguous survival beliefs with their empirical counterparts and corresponding objective survival rates for men in Panel (a) and for women in Panel (b). Interview age is shown on the abscissa. Actual subjective survival beliefs are depicted in the figure as a blue solid line and corresponding objective beliefs as a red dashed-dotted line. To understand this figure, recall that actual subjective survival beliefs are elicited in the HRS only for a combination of interview ages and target ages. The step function of corresponding objective beliefs follows from changes in the interview age / target age assignment. For example, a 69 year old person is asked about her subjective assessment to live until age 80 whereas a 70 year old is asked about her probability to reach age 85. The 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. Furthermore, 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.[19] Finally, the figure shows as a green dashed line the predicted subjective survival rates from our model for the parameter estimates of δ and λ as given in Table 1. Overall, we can conclude from this figure that the fit of predicted to actual subjective survival rates is very good. In particular, the model replicates underestimation of survival rates at younger ages and overestimation at older ages.

Figure 2: Objective, Subjective and Predicted Subjective Survival Rates

(a) Women   (b) Men
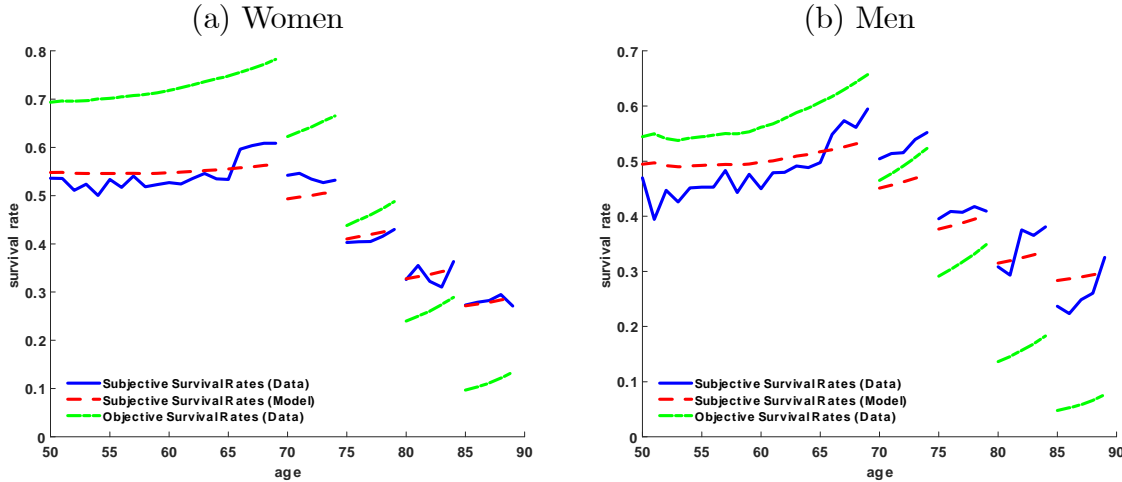
Notes: Unconditional survival probabilities to different specific target ages according to the questions in the HRS. Interview age is on the abscissa. The solid blue line are subjective survival beliefs, the dashed-dotted red line are the corresponding objective survival rates and the dashed green line are simulated subjective survival beliefs from the estimated CEU model.

Figure 3 shows the age-specific degree of ambiguity in Panel (a) and the degree of relative optimism in Panel (b) both as a function of planning age h. The degree of ambiguity, δ_h, is a monotonically increasing and concave function over planning age h. At the same time, relative optimism, λ_h, is a decreasing and convex function, albeit the decrease in relative optimism over age is quantitatively small. Observe that there are two dynamics in the model: first, psychological attitudes are changing over age according to the pattern in Figure 3 and second, objective survival chances decrease in age, cf. Figure A.3 in the Supplementary Appendix. This latter effect in combination with the positive estimates of λ_h and δ_h leads to increasingly optimistic biases of predicted subjective survival beliefs despite the fact that λ_h is slightly decreasing.

---

[19]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 downward sloping within interview age / target age bins. For example, the objective survival rate of a 52 year old man to live to age 80 turns out to be slightly higher than of a 53 year old man because the 52 year old man belongs to a younger cohort.

Figure 3: Degree of Ambiguity and Relative Optimism over the Life Cycle
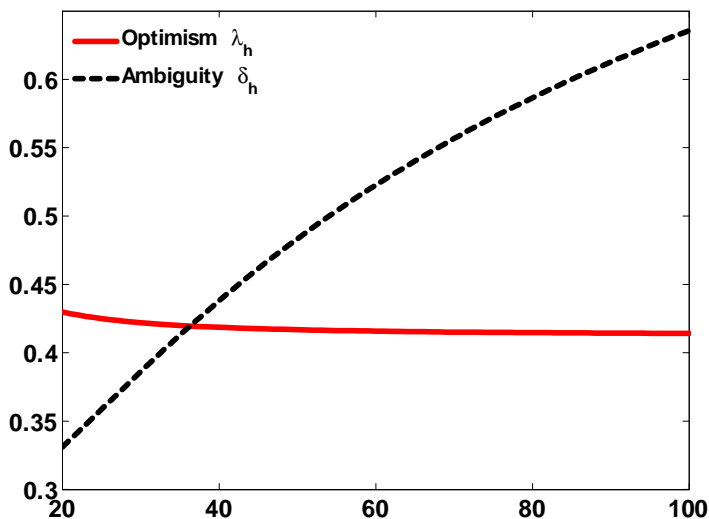
Notes: Degree of ambiguity δ_h and relative optimism λ_h as a function of planning age h.

Figure 4 compares ambiguous subjective survival functions—i.e., the subjective hazard rates—as red solid lines to their objective counterparts as black dashed lines. The four panels of the figure represent different planning ages h. Panel (a) is for planning age 20 ($h = 0$) and Panel (d) for planning age 85, ($h = 65$). In each of the four panels 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 3. The initial point of survival functions 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 (d), the distance of this point to a survival rate of 1 increases. Generally, the subjective survival functions exhibit an initial blip relative to the objective data.[20]

The key observation from Figure 4 is that subjective survival functions are flatter than their objective counterparts which is in line with Hammermesh (1985), Peracchi and Perotti (2010), Elder (2013) and several others. Furthermore, ambiguous survival beliefs match the stylized fact described by Wu et al. (2013): 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 future. Also notice that the overestimation of survival probabilities becomes more pronounced as the agent gets older. I.e., the point at which the subjective and the objective survival curves intersect

---

[20]This initial blip results from the parsimonious structure of our model but otherwise does not affect our results much, cf. Section 6.2.3 for a sensitivity analysis with respect to the size of this initial blip.

moves to the left when moving across the figure from Panel (a) to Panel (d).

Figure 4: Survival Functions
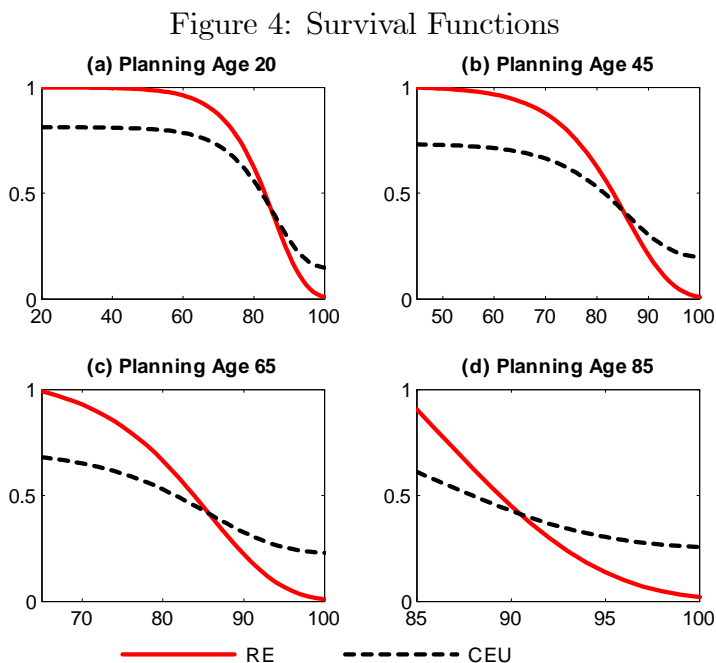
Notes: Unconditional objective and subjective probabilities viewed from different planning ages h. Target age t is depicted on the abscissa.

6.2 Life-Cycle Profiles with Ambiguous Beliefs

6.2.1 Baseline Calibration

To highlight the effects of modeling subjective survival beliefs on life-cycle savings we conveniently compress all information by showing average asset holdings of CEU agents compared to RE agents who use objective survival data. We focus on the average asset-to-permanent income ratio as described in the calibration section. We scale assets with the same annuity value as the one used for estimating preference parameters, cf. Section 5.5.

Figure 5 shows our results by displaying average asset holdings over the life-cycle for the three types of agents, RE agents as the black dotted line, naive CEU agents as the blue dashed line and sophisticated CEU agents as the red dashed-dotted line. The profiles of our calibrated models are compared to the data, shown as a gray solid line. Assets steadily increase until retirement entry and fall thereafter. This implies positive saving rates during working life while agents dissave during retirement.

The overall shape of life-cycle asset holdings is explained as follows Households save for life-cycle and precautionary motives. As to the latter, 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). Assets are accumulated in order to finance retirement consumption. 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 5: Assets-to-Permanent Income, CEU, RE and Data
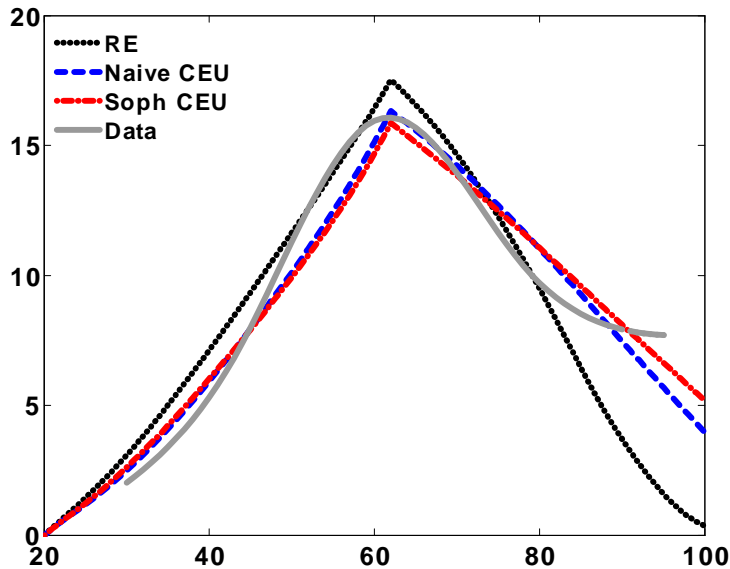
Notes: Average asset-to-permanent-income ratios from SCF data and for CEU and RE agents using recalibrated preference parameters ρ for all models. The data covers ages 30 and 95. Details on data are provided in Appendix B.

With regard to differences in asset accumulation across types, first focus at the RE type. Relative to the data, the dynamically consistent RE model features higher saving and therefore stronger asset accumulation on average until retirement and a faster speed of asset decumulation thereafter. Accordingly, through the lens of the RE model the data are puzzling: households save too little until retirement in the data relative to the RE model and have asset holdings in old age that are too high (in the data relative to the RE model). Any attempt to improve the fit of the RE 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 CEU model gives rise to less saving during the accumulation phase and a much slower speed of asset decumulation than for the RE

model, moving it close to the data. The driving force for undersaving (relative to the RE 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 4.

The sophisticated CEU model generates very similar results compared to the naive CEU model: on average saving rates during the working period are almost identical and so is asset accumulation. Old-age asset holdings of sophisticated agents are slightly higher than those of naive agents. The reason is that sophisticates, by foreseeing the optimistic biases of their own future selves, decumulate assets at a lower speed for reasons of consumption smoothing. The close similarities between the two CEU agents only occur because we recalibrate the discount rate. It is almost one percentage point higher for the sophisticated agent, cf. Table 2. We discuss this in detail below in Section 6.2.2, where we also provide additional interpretation for our findings.

Table 3 comprises our results by reporting summary statistics for all three agent types and the data. As a summary statistic for the goodness of fit of the three models we report the R^2s from the non-linear regressions in (38). While R^2 looses its usual interpretation in non-linear models as a measure of the fraction of the overall variation explained by the model, it is still a useful summary statistic of goodness fit. It is bounded from above by 1 and a value closer to 1 indicates better fit. Results on the R^2s confirm the visual impression gained from Figure 5, i.e., the fit of the naive CEU model is best and the one of the RE model is worst.

The average saving rate of both the naive and sophisticated CEU agents during the prime saving years, ages 25-54, is about 9.4%. The corresponding average saving rate in the US is 9.5%.[21] On the contrary, RE agents save on average 13.5%, exceeding the relevant data by 4 percentage points.

Comparing plans and realizations for naive CEU agents we observe that, initially, CEU agents plan to save more and consume less during working life which would result in higher assets. The planned average saving rate of naive CEU agents at age 20 for ages 25-54 is 15.1%, compared to the average realized saving rate for that age bin of 9.4%. The fact that actual saving behavior deviates from plans naturally follows from time inconsistency. That saving is lower than planned means that households moderately overestimate their future survival rates, leading us back to the predictions of the simple 3-period model, cf. the Supplementary Appendix. If overestimation was stronger, then they would actually save more than originally planned. The patterns we find are qualitatively consistent with findings in the literature on undersaving: Barsky

---

[21]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. Thus, we chose CES data as reported by Bosworth et al. (1991).

Table 3: Summary Statistics

	RE	CEU		Data[1]
		Naive	Soph.	
R^2	0.756	0.946	0.937	
Saving rate[2]	13.5%	9.4%	9.4%	9.5%
Saving rate, planned[2]	–	15.1%	–	
Assets at age 75 relative to 62[3]	70.0%	77.7%	78.7%	72.4%
Assets at age 85 relative to 62[3]	37.0%	56.8%	60.5%	53.0%
Assets at age 95 relative to 62[3]	*9.1%*	*34.8%*	*41.8%*	*47.9%*

[1] The data for asset decumulation is calculated from SCF data. Due to small sample sizes, SCF data on average asset holdings at age 95 cannot be measured reliably and are thus reported in italics. 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.

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.

Finally, Table 3 also summarizes the sizable differences in old-age asset holdings between RE and CEU agents. For naive CEU agents, average asset holdings at ages 75, 85 and 95 relative to those at retirement entry are 77.7%, 56.8% and 34.8% compared to 72.4%, 53.0% and 47.9% in the data. Recall that the last data point, i.e., asset holdings at age 95, has to be looked at with care because of few observations. On the contrary, these values are only at 70.0%, 37.0%, and 9.1% for RE agents. Sophisticated CEU agents have even higher assets at old age relative to assets at retirement entry. Also notice that the overall fit of the sophisticated CEU model to the data is worse than for naive agents.

We can therefore conclude that the combination of ambiguous survival beliefs with the assumption of naivety has to be considered as a candidate explanation for the joint occurrence of low retirement savings, time inconsistent saving behavior and high old-age asset holdings.

6.2.2 The Effects of Discounting

As described in Section 5, the calibrated discount rate varies across all model variants in our baseline results, cf. Table 2. This section documents how our main findings are affected by this approach to calibration. To this end, we hold constant the value of the discount rate calibrated for the RE model of 3.4% and use it in the two variants of the CEU model. With this strategy we single out the pure effects of ambiguous survival beliefs. Results on asset holdings are displayed in Figure 6 and corresponding summary statistics are provided in Table 4.

Figure 6: Assets-to-Permanent Income, CEU, RE and Data: Constant ρ
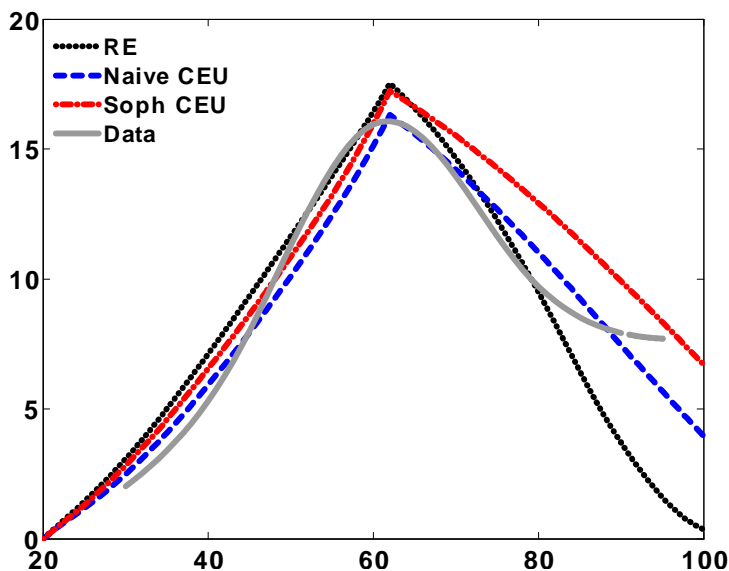
Notes: Average asset-to-permanent-income ratios from SCF data and for CEU and RE agents using $\rho^{RE} = 0.034$ as a preference parameter for all models. The data covers ages 30 and 95. Details on data are provided in Appendix B.

As the difference of calibrated discount rates between the RE and the naive CEU models is not large, cf. Table 2, our results do not change much for naive CEU agents. With the lower RE-model discount rate, saving increases slightly and hence the average asset decumulation speed also decreases. The R^2, as a summary statistic for the goodness of fit, decreases very mildly to 0.945.

Significant changes occur for the sophisticated CEU model where the calibrated discount rate is almost one percentage point higher in our baseline calibration. Relative to this, the R^2 strongly decreases to 0.823. The saving rate during the working period goes up to 11.6% which is—although still lower than for the RE agent—more than observed in the data.

Table 4: Summary Statistics: Constant ρ [1)]

	Naive CEU		Soph. CEU		Data
	Baseline	ρ^{RE}	Baseline	ρ^{RE}	
R^2	0.946	0.945	0.937	0.823	
Saving rate	9.4%	9.3%	9.4%	11.6%	9.5%
Saving rate, planned	15.1%	15.0%	–	–	
Assets at age 75 relative to 62	77.9%	77.7%	79.0%	82.8%	72.4%
Assets at age 85 relative to 62	57.1%	56.8%	60.8%	66.5%	53.0%
Assets at age 95 relative to 62	*34.9%*	*34.8%*	*42.1%*	*48.3%*	*47.9%*

Notes: See Table 3 for a description of how the statistics are constructed.

We complement this picture by displaying life-cycle consumption relative to permanent income across agent types (again holding ρ constant) in Figure 7. At younger ages, both the naive and the sophisticated households consume more than RE agents. In the middle stages of the life-cycle optimism starts to dominate their survival belief formation. This increases the consumption growth rate so that both CEU agents consume less at middle and more at old age than RE agents. Finally, observe from the figure that sophisticates indeed consume less at young ages than do naifs leading to higher asset holdings over the life-cycle.

Again, the simple three-period model from the Supplementary Appendix provides guidance for understanding these results. Sophisticates foresee the increasing optimism of their own future selves. Given their relatively low inter-temporal elasticity of substitution of 1/3 they therefore place a high value on the marginal utility from saving, give up on consumption when young and build up higher asset positions during the working period than their naive counterparts. In consequence, they also decumulate assets at a lower speed in old age.

6.2.3 The Effects of Experience

We next analyze the importance of our assumed experience function for life-cycle asset holdings. First, 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, cf. Section 3.2. 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 $\bar{\lambda} = 0.424$. Second, we determine ρ and ω jointly to give the best fit on life-cycle asset holdings by minimizing function (38). This gives $\omega = 127$, $\delta = 0.00013$ and $\lambda = 0.412$.

Figure 7: Life-Cycle Consumption of CEU and RE: Constant ρ
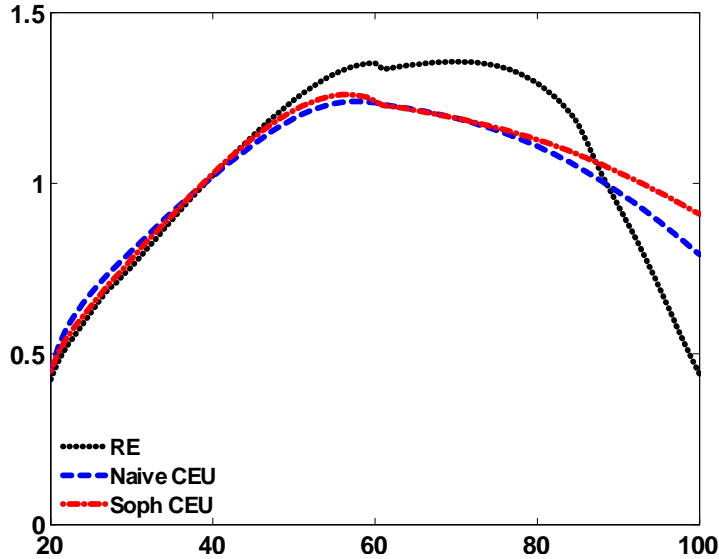
Notes: Average consumption relative to permanent income as described in the calibration section over the life-cycle for CEU and RE agents using $\rho^{RE} = 0.034$ as a preference parameter for all models.

Our results show that the key quantitative implications of our model are little affected by the experience function, even when assuming constant experience. For this scenario, the average saving rate decreases for both CEU types, cf. Table 5. The reason is the initial blip of subjective survival beliefs at younger ages, cf. Figure 4, which now is larger. Hence, the initial underestimation of survival beliefs is more pronounced yielding lower saving rates.[22] For naive CEU agents, the difference between planned and realized saving rates increases. The asset decumulation speed during the retirement period, however, is little affected. Finally, notice that the R^2s decrease compared to our baseline scenario. The data thus supports our notion that Bayesian learning of ambiguous survival beliefs happens over the life cycle.

The model with the "best fit" speed of learning process, $\omega = 127$, gives very similar results as our baseline specification. Relative to that, the R^2 increases only mildly and summary statistics are little affected. The reason for these small changes is that an increase of $e(h)$ at all ages mainly affects the level of our estimate for δ (the point estimate decreases from 0.0163 in our baseline specification to 0.00013) whereas λ is basically unchanged and λ_h is now virtually flat. Because λ_h is already relatively flat in our baseline parametrization with $\omega = 1$, cf. Figure 3, and because the level change

_____

[22]In our baseline specification, the subjective belief of a 20-year old to survive to age 21 is 0.811, cf. Figure 4, in the model variant with constant experience it is only 0.675.

Table 5: Summary Statistics: The Effect of Experience

	Naive CEU			Sophisticated CEU		
	baseline	$e(h) = n$	$\omega = 127$	baseline	$e(h) = n$	$\omega = 127$
Discount rate ρ	0.0343	0.0323	0.0344	0.0462	0.0421	0.0428
R^2	0.946	0.937	0.9464	0.937	0.926	0.939
Saving rate	9.4%	8.6%	9.7%	9.4%	8.9%	9.4%
Saving rate, planned	15.1%	16.6%	14.7%	–	–	–
Assets at 75 rel. to 62	77.9%	79.0%	77.6%	79.0%	79.9%	78.5%
Assets at 85 rel. to 62	57.1%	58.5%	56.4%	60.8%	62.1%	60.1%
Assets at 95 rel. to 62	*34.9%*	*36.0%*	*34.3%*	*42.1%*	*43.2%*	*41.3%*

Notes: Results for the CEU model with constant experience, $e(h) = n$, and with the "best fit" learning speed, $\omega = 127$. For a description of how the statistics are constructed see Table 3.

of δ does not affect the dynamics much, results are not affected much by increasing ω beyond our baseline level of 1. It is important to emphasize that our estimation is unrestricted, i.e., we also allow for negative values of ω despite our theoretical restriction that $\omega \in \mathbb{N}$. Hence, this sensitivity analysis also shows that increasing experience—and thereby increasing δ_h and decreasing λ_h—is not only plausible on a priori grounds—cf. our discussion in Section 3.2—but also supported by the data.

7 Concluding Remarks

This paper constructs a model of Choquet Bayesian learning of ambiguous survival beliefs. In a next step, it studies implications of ambiguous survival beliefs for consumption and saving behavior. Point of departure of our analysis is that people make mistakes in assessing their chances to survive into the future: "young" people tend to underestimate whereas "old" people tend to overestimate their survival probabilities. We construct and parametrize a model of Bayesian learning of ambiguous survival beliefs which replicates these patterns. The resulting conditional neo-additive survival beliefs are merged into a stochastic life-cycle model with CEU (=Choquet expected utility) agents to study life-cycle consequences compared to agents with rational expectations (RE).

We show that agents of our model behave dynamically inconsistent. As a result, CEU agents save less at younger ages than they actually planned to save. Due to underestimation of survival at young age, CEU agents also save less than RE agents. Despite this tendency to undersave, CEU agents eventually have higher asset holdings after retirement because of the overestimation of survival probabilities in old age. Overall,

the calibrated CEU model provides an accurate quantitative picture of life-cycle asset holdings until about age 85. Furthermore, the assumption of naive CEU agents better fits the data than the assumption of sophisticated CEU agents. Our model of biases in the assessment of survival prospects therefore adds to explanations for three empirical findings: (i) time inconsistency of agents, (ii) undersaving at younger ages and (iii) high asset holdings at old age. Hence, our model hits at—but does not kill—"three birds with one stone".

Our work gives rise to several avenues of future research. First, observe that the ambiguous survival belief functions depicted in Figure 4 closely resemble quasi-hyperbolic time discounting functions, cf., e.g., Laibson (1997). In Groneck et al. (2014) we compare the formal relationship between quasi-hyperbolic time-discounting, on the one hand, and a static CPT/CEU model, on the other hand. As our main finding we show that quasi-hyperbolic time-discounting over the life-cycle is formally equivalent to a static CPT/CEU life cycle model with neo-additive capacities such that (i) the ambiguity parameter is positive whereas (ii) the optimism parameter is zero. Our analysis further implies that a positive optimism parameter rather than Bayesian learning under ambiguity is responsible for the qualitative feature that CPT/CEU agents might—in contrast to quasi-hyperbolic time-discounting agents—oversave in old age.

Second, we plan to combine our notion of CEU agents with bequest motives in order to cover important aspects of life-cycle decisions. The main challenge for this generalizing approach will be to come up with a parsimonious model in which all calibrated behavioral parameters are identified.

Third, we will 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.

Finally, in our current research, cf. Grevenbrock et al. (2015), we identify probability weighting functions by using the full panel dimension of the HRS. Specifically, we construct estimated objective survival rates and compare those to subjective ones at the individual level. We use this data to estimate inverse-S-shaped probability weighting functions which we identify for different age groups. Approximating these functions linearly lends empirical support to the dynamics of our learning model (which we here derive solely on theoretical grounds based on decision theoretic foundations). That is, we indeed find in our extended data analysis that the optimism parameter is decreasing whereas the ambiguity parameter is increasing in age.

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.

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

Andreoni, J. and C. Sprenger (2012). Risk Preferences Are Not Time Preferences. The 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. The 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. The 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.

Bloom, D. E., D. Canning, M. Moore, and Y. Song (2006). The Effect of Subjective Survival Probabilities on Retirement and Wealth in the United States. NBER Working Paper 12688.

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.

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.

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

Ellsberg, D. (1961). Risk, Ambiguity, and the Savage Axioms. Quarterly Journal of Economics 75, 643–669.

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.

Fernandez-Villaverde, J. and D. Krueger (2007). Consumption over the Life Cycle: Facts from Consumer Expenditure Survey Data. The Review of Economics and Statistics 89(3), 552–565.

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.

Grevenbrock, N., Groneck, M., Ludwig, A., and A. Zimper (2015). Biased Survival Beliefs, Psychological and Cognitive Explanations, and the Demand for Life Insurances. Mimeo.

Groneck, M., Ludwig, A., and A. Zimper (2014). Cumulative Prospect Theory versus Quasi-hyperbolic Discounting. Mimeo.

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

Hammermesh, 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.

Huggett, M., G. Ventura, and A. Yaron (2007). Sources of Lifetime Inequality. NBER Working Paper 13224.

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

Hurd, M. and S. Rohwedder (2010). Wealth Dynamics and Active Saving at Older Ages: Do They Add Up? Mimeo.

Hurd, M. D., D. L. McFadden, and L. Gan (1998). Subjective Survival Curves and Life Cycle Behavior. In Inquiries in the Economics of Aging, pp. 259–309. University of Chicago Press.

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.

Lee, R. D. and L. Carter (1992). Modeling and Forecasting U.S. Mortality. Journal of the American Statistical Association 87(419), 659–671.

Lockwood, L. M. (2013). 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. NBER Working Papers 17078.

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.

O'Donoghue, T. and M. Rabin (1999). Doing It Now or Later. The 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. Yzici (2012). Optimal Life-cycle Capital Taxation under Self-Control Problems, Working Paper.

Pavoni, N. and H. Yzici (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.

Pires, C. P. (2002). A Rule for Updating Ambiguous Beliefs. Theory and Decision 53(2), 137–152.

Pollak, R. A. (1968). Consistent Planning. The 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.

Salm, M. (2010). Subjective Mortality Expectations and Consumption and Saving Behaviours among the Elderly. Canadian Journal of Economics 43(3), 1040–1057.

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. The Review of Economic Studies 23(3), 165–180.

Tversky, A. and D. Kahneman (1992). Advances in Prospect Theory: Cumulative Representations of Uncertainty. Journal of Risk and Uncertainty 5(4), 297–323.

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

Wakker, P. P. and A. Tversky (1993). An Axiomatization of Cumulative Prospect Theory. Journal of Risk and Uncertainty 7, 147–176.

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

Wu, G. and R. Gonzalez (1996). Curvature of the Probability Weighting Function. Management Science 42(12), 1676–1690.

Wu, G. and R. Gonzalez (1999). Nonlinear Decision Weights in Choice under Uncertainty. Management Science 45(1), 74–85.

Wu, S., R. Stevens, and S. Thorp (2013). Die Young or Live Long: Model Subjective Survival Probabilities. Centre for the Study of Choice (CenSoC) Working Paper No. 13-004.

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,

$$\mu\left(\tilde{I}_{e(h)} = j\right) = \binom{e(h)}{j} \frac{(\alpha + j - 1) \cdot ... \cdot \alpha \cdot (\beta + e(h) - j - 1) \cdot ... \cdot \beta}{(\alpha + \beta + e(h) - 1) \cdot ... \cdot (\alpha + \beta)}, \quad (39)$$

$$\text{for } j \in \{0, ..., e(h)\}.$$

The uniform distribution is characterized by $\alpha = \beta = 1$, implying for (39) that

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

$$= \frac{1}{1 + e(h)}.$$

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).$\square$

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

$$\sum_{t=h+1}^{T} \psi\left(D_t \mid Z_h\right) \sum_{s=h+1}^{t} \beta^{s-h} \mathbb{E}\left[u\left(c_s\right), \pi\left(\eta_s | \eta_h\right)\right]$$

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

$$= \sum_{t=h+1}^{T} \beta^{t-h} \mathbb{E}\left[u\left(c_t\right), \pi\left(\eta_t | \eta_h\right)\right] \cdot \psi_{h,t}.$$

Consequently, (26) can be equivalently rewritten as

$$
\mathbb{E}\left[U\left(\mathbf{c}\right),\nu_h^h\right] = \delta_h \left(\lambda_h \left(u(c_h) + \sum_{t=h+1}^{T} \beta^{t-h} \mathbb{E}\left[u\left(c_t\right),\pi\left(\eta_t|\eta_h\right)\right]\right) + (1-\lambda_h)u(c_h)\right)
$$

$$
+ (1-\delta_h)\left(u(c_h) + \sum_{t=h+1}^{T} \psi\left(D_t \mid Z_h\right) \sum_{s=h+1}^{t} \beta^{s-h} \mathbb{E}\left[u\left(c_s\right),\pi\left(\eta_s|\eta_h\right)\right]\right)
$$

$$
= u(c_h) + \delta_h \lambda_h \sum_{t=h+1}^{T} \beta^{t-h} \mathbb{E}\left[u\left(c_t\right),\pi\left(\eta_t|\eta_h\right)\right]
$$

$$
+ (1-\delta_h) \sum_{t=h+1}^{T} \psi_{h,t} \cdot \beta^{t-h} \mathbb{E}\left[u\left(c_t\right),\pi\left(\eta_t|\eta_h\right)\right]
$$

$$
= u(c_h) + \sum_{t=h+1}^{T} \nu_{h,t}^h \cdot \beta^{t-h} \mathbb{E}\left[u\left(c_t\right),\pi\left(\eta_t|\eta_h\right)\right],
$$

which proves the proposition. $\square$

A.3   Proof of Proposition 4

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

$$
V_h^h\left(x_h,\eta_h\right) = \max_{c_h,x_{h+1}} \left\{u\left(c_h\right) + \beta\nu_{h,h+1}^h \mathbb{E}_h\left[V_{h+1}^h\left(x_{h+1},\eta_{h+1}\right)\right]\right\}
$$

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

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 given by

$$
\frac{du}{dc_h} = \beta R \nu_{h,h+1}^h \mathbb{E}_h\left[\frac{\partial V_{h+1}^h(\cdot)}{\partial x_{h+1}}\right] \tag{40a}
$$

$$
\frac{du}{dc_{h+1}} = \beta R \nu_{h+1,h+2}^{h+1} \mathbb{E}_{h+1}\left[\frac{\partial V_{h+2}^{h+1}(\cdot)}{\partial x_{h+2}}\right]. \tag{40b}
$$

To get an expression for the derivative of the value function of self h with respect to cash-on-hand in period $h+1$, appearing on the right-hand-side of (40a), notice that the familiar Envelope condition does not hold. This captures the notion that self h correctly

anticipates that future self $h+1$ will deviate from the optimal consumption plan of self h. The respective derivative of the value function writes as

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

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

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

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

Collecting equations, the relevant first-order conditions of self h are (40a) and (41). Condition (40b) is a constraint to the maximization problem of self h, again because self h correctly anticipates optimality of behavior of self $h+1$.

Rewrite (40b) by adding and subtracting terms as

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

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

to get

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

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 (42) in (41) to get

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

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

Using the above in (40a) we finally get

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

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

$$= \beta R\nu_{h,h+1}^h\mathbb{E}_h\left[\frac{du}{dc_{h+1}}\Theta_{h+1} + \Lambda_{h+1}\right].$$

☐

B Appendix: Details on 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 26-95. Our total sample contains 21.560 respondents.

To construct the average life-cycle profile of the—appropriately smoothed (see below)—asset-to-permanent-income ratio we proceed as follows.[23] Define assets as net worth including housing wealth, but excluding implicit pension and social security wealth. We deflate assets and income to 1992 Dollars. To approximate permanent income we first compute gross labor and social security income by excluding income from capital gains.[24] Using data from Cagetti (2003)—who approximates tax rates for different income percentiles—we next compute after-tax income. Based on the—appropriately smoothed (see below)—age-specific averages of net income we compute the net-present value and convert this to annuities using the calibrated interest rate of $r = 0.042$. This gives our permanent-income approximation. Finally, we compute the asset-to-income ratio from these two time series.

Average age-specific assets and net income are both smoothed over age by applying a cubic spline regression. We use robust fitting by three iterations of weighted least squares. Respective weights are computed from previous residuals.

---

[23]To construct the data we adopted 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.

[24]Our 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.

S Supplementary Appendix: A Life-Cycle Model with Ambiguous Survival Beliefs

S.1 Age-increasing Ambiguity. An Explanantory Discussion

Note that age-increasing ambiguity in our Choquet Bayesian learning model is not an ad hoc assumption but rather a formal implication which turns out to be remarkably robust with respect to alternative Bayesian update rules for Choquet decision makers (cf. Gilboa and Schmeidler 1993; Zimper 2011). One reason is that ambiguity here applies to the joint probability of the parameter and sample space so that the updating process itself—and not only the distribution over parameters—is subject to ambiguity. As a consequence, the agent does not perceive the data generating process as i.i.d. as in the standard (non-ambiguous) Bayesian set-up. Furthermore, a large amount of statistical information refers to an event which has an ex ante small likelihood to be ex post observed within this non-i.i.d. environment. Bayesian updating under ambiguity "punishes" small ex ante likelihoods in the sense that the decision maker's ambiguity increases if she observes information which she considered ex ante as unlikely. In what follows we present four reasons why this rather mechanical consequence of the updating process also has some intuitive appeal.

First, an age-increasing δ_h captures the intuitive notion that, as the objective risk of survival becomes less likely, agents attach less and less weight to this objective probability. Our model's convergence property implies that survival rates are overestimated eventually even when the initial degree of ambiguity, δ, is low. Overestimation at old age may result from the fact that people have survived the gamble against death several times before. Consequently, one possible heuristic interpretation of age-increasing δ_h might be that people want to avoid a realistic assessment of their encounter with death.[1]

Second, the concept of *likelihood-insensitivity* (cf., Wakker 2004; 2010; Abdellaoui et al. 2011), may provide an alternative heuristic interpretation for the age-increasing δ_h of our model. These authors interpret δ_h not as an ambiguity but rather as a cognitive parameter which reflects the empirical observation that people do not sufficiently distinguish between non-degenerate probabilities. For instance, an extreme example for likelihood insensitivity are "fifty-fifty" probability assessments for any uncertain event and its complement. Under this cognitive interpretation, likelihood insensitivity—and

---

[1]This interpretation is consistent with the observation of Kastenbaum (2000) who summarizes the insights of psychological research on the reflection about personal death as follows: "There are divergent theories and somewhat discordant findings, but general agreement that most of us prefer to minimize even our cognitive encounters with death."

not necessarily ambiguity—would increase with age. Given that old people increasingly suffer from cognitive impairments, this alternative interpretation has some intuitive appeal.

Third, Nicholls et al. (2014) investigate whether violations of Savage's (1954) sure-thing principle (STP), typically interpreted as the expression of ambiguity attitudes, decrease or increase if the subjects receive an increasing amount of statistical information. As their main finding, these authors conclude that "[...] statistical learning has, at best, no impact on STP violations. At worst, it might even be causing STP violations to increase." (p. 14). This empirical finding suggests that conventional wisdom about Bayesian learning might not be adequate for situations with ambiguity.

Finally, there exist alternative models of Bayesian learning under ambiguity such that ambiguity might decrease in the amount of statistical information.[2] Within a multiple-priors set-up, Marinacci (2002) restricts ambiguity to the parameter space whereas Bayesian updating happens with respect to a standard (i.e., non-ambiguous) i.i.d. data-generating process. The convergence behavior of the Bayesian learning process in the Marinacci (2002) model crucially depends on the support of the priors held by the decision maker. If not all priors have the same support, ambiguity does not necessarily vanish when an unlimited amount of statistical information becomes available. Epstein and Schneider (2007) consider two dimensions of ambiguity. First (as in Marinacci 2002), ambiguity with respect to prior beliefs is expressed through multiple priors; second, ambiguity with respect to the updating process is expressed through *multiple likelihoods*. Furthermore, these authors impose a specific *expected maximum likelihood criterion* as a prior-selection rule. This may reject initially plausible priors in light of new information. Ambiguity with respect to posterior beliefs vanishes in the original Epstein and Schneider (2007) model if, and only if, there is no ambiguity with respect to the updating process. However, even if there is no ambiguity with respect to the updating process, ambiguity might not vanish in a modified—and ad hoc equally plausible—version of the Epstein and Schneider (2007) model. This model would consider—instead of the expected maximum likelihood criterion—some alternative prior selection rule such as the *minimal Kullback-Leibler divergence criterion* (cf. Zimper and Ma 2014). The analysis in Marinacci (2002) and in Epstein and Schneider (2007) thus suggests that it requires quite strong ad hoc assumptions on the priors' support, on the updating process as well as on the prior-selection rule for ambiguity to vanish in alternative theoretical models of Bayesian learning under ambiguity.

---

[2]For a detailed discussion of these models we refer the interested reader to Zimper and Ma (2014).

S.2 A Three-Period Model

We provide the intuition for how ambiguous survival beliefs affect consumption and saving behavior in a simple three-period model ($T = 2$) without income risk which can be solved analytically. In this simple model we abstract from borrowing constraints, hence $a_{t+1} < 0$, for $t < T$ is possible. The no-Ponzi condition $a_{T+1} \geq 0$ is of course assumed. To simplify the analysis we assume the discount factor β to be one and an interest rate r of zero.

As shown in Section 4 or the paper, lifetime utility for $T = 2$ with ambiguous survival beliefs is expressed as

$$U_0^0 = u(c_0) + \nu_{0,1}^0 \left(u(c_1) + \frac{\nu_{0,2}^0}{\nu_{0,1}^0} u(c_2) \right),$$

where $\nu_{k,t}^h$ is the subjective survival belief from Proposition 1 of the paper. Recall that superscripts denote the respective planning age.

As in the paper, we normalize the utility from death to zero and assume a $CRRA$ per-period utility function with preference shifter $\Upsilon \geq 0$ for the utility from survival. Since we here ignore income risk the additional Stone-Geary parameter $\bar{c}$ is not required. Also recall that lifetime utility of CEU agents reduces to the standard rational expectations case if and only if there is no initial ambiguity, i.e., iff $\delta = 0$.

We define by $x_t \equiv a_t + y_t$ cash-on-hand as the sum of financial assets a_t and income y_t. In addition, define the present value of future income, $h_t \equiv \sum_{s=t+1}^T y_s$, as human wealth. Finally, let total wealth be $w_t \equiv x_t + h_t$. The budget constraint is then given by

$$w_{t+1} = w_t - c_t.$$

In light of the data on subjective beliefs displayed in Figure 1 of the paper we interpret period 0 of the simple model as the period when survival probabilities are underestimated, i.e., up to actual age of about 70. Period 1 then reflects the period when there is overestimation in the data. Correspondingly, we make the following assumption:

Assumption S1. *We assume for some $\delta > 0$ that*

$$\psi_{0,1} > \nu_{0,1}^0 = \delta_0 \lambda_0 + (1 - \delta_0)\psi_{0,1} \tag{1}$$

i.e., that $\lambda_0 < \psi_{0,1}$, as well as

$$\psi_{1,2} < \nu_{1,2}^1 = \delta_1 \lambda_1 + (1 - \delta_1)\psi_{1,2} \tag{2}$$

3

i.e., that $\lambda_1 > \psi_{1,2}$.[3]

We now turn to the complete inter-temporal household solution to analyze how consumption and saving decisions are altered by biases in subjective survival beliefs.

S.2.1 Rational Expectations

The reference model is the standard solution to the rational expectations model (where $\delta_0 = \delta_1 = 0$). Here, lifetime utility does not depend on the planing period, i.e., $U_1^0 = U_1^1$, and in period 0 it is given by $U^0 = u(c_0) + \psi_{0,1}\left(u(c_1) + \psi_{1,2}u(c_2)\right)$.

Observation S1. *Policy functions of the rational expectations solution are linear in total wealth,* $c_t = m_t w_t$, *where*

$$
m_t = \begin{cases} \dfrac{1}{1+\dfrac{\left(\psi_{t,t+1}\right)^{\frac{1}{\theta}}}{m_{t+1}}} & \text{for } t < T \\[4ex] 1 & \text{for } t = T. \end{cases}
$$

Hence:

$$
m_0 = \frac{1}{1 + \psi_{0,1}^{\frac{1}{\theta}} + \psi_{0,2}^{\frac{1}{\theta}}}, \qquad m_1 = \frac{1}{1 + \psi_{1,2}^{\frac{1}{\theta}}}.
$$

Proof. See, e.g., Deaton (1992) $\square$.

S.2.2 Naive CEU Households

To draw a distinction between RE and CEU households, we use superscript n to denote policy functions (in terms of marginal propensities to consume) of *n*aive CEU households. Given that the household consumes all outstanding wealth in the final period 2 (i.e., $m_2^n = 1$) the solution of the household's problem for all other periods are as follows:

---

[3]Notice that, despite equation (2), we may have that the household in period 0 underestimates the probability to survive from period 1 to 2, hence we may have that

$$
\psi_{1,2} > \nu_{1,2}^0 = \delta_0\lambda + (1 - \delta_0)\psi_{1,2}.
$$

This is so because $\delta_0 < \delta_1$ and therefore less weight is put on the relative optimism parameter λ.

Proposition S1. *For the naive CEU household we get:*

- *The solution to the problem in period 1 is:*

$$c_1^n = m_1^n w_1 \quad where \quad m_1^n = \frac{1}{1 + \left(\nu_{1,2}^1\right)^{\frac{1}{\theta}}}. \tag{3}$$

- *The plan in period 0 for period 1 is:*

$$c_1^{0,n} = m_1^{0,n} w_1 \quad where \quad m_1^{0,n} = \frac{1}{1 + \left(\frac{\nu_{0,2}^0}{\nu_{0,1}^0}\right)^{\frac{1}{\theta}}}, \tag{4}$$

  where we denote the planning period as a superscript.

- *The solution in period 0 is:*

$$c_0^n = m_0^n w_0 \quad where \quad m_0^n = \frac{1}{1 + \frac{(\nu_{0,1}^0)^{\frac{1}{\theta}}}{m_1^{0,n}}} = \frac{1}{1 + (\nu_{0,1}^0)^{\frac{1}{\theta}} + (\nu_{0,2}^0)^{\frac{1}{\theta}}}. \tag{5}$$

Proof. The first-order condition in period 1 is:

$$u_c(c_1) = \nu_{1,2}^1 u_c(c_2)$$

which directly yields (3). Analogously, the first-order condition for period 1 from the perspective of period 0 is given by:

$$u_c(c_1) = \frac{\nu_{0,2}^0}{\nu_{0,1}^0} u_c(c_2)$$

which yields (4). Finally, the first-order condition in period 0 is:

$$u_c(c_0) = \nu_{0,1}^0 u_c(c_1)$$

yielding

$$m_0^n = \frac{1}{1 + \frac{(\nu_{0,1}^0)^{\frac{1}{\theta}}}{m_1^{0,n}}}.$$

Notice that

$$(\nu_{0,1}^0)^{-\frac{1}{\theta}} m_1^{0,n} = \frac{(\nu_{0,1}^0)^{-\frac{1}{\theta}}}{1 + \left(\frac{\nu_{0,2}^0}{\nu_{0,1}^0}\right)^{\frac{1}{\theta}}} = \frac{1}{(\nu_{0,1}^0)^{\frac{1}{\theta}} + (\nu_{0,2}^0)^{\frac{1}{\theta}}}.$$

Using this in the above gives the last term in (5). $\square$

Comparing the policy functions of the RE agent (cf. Observation S1) and the naive agent (cf. Proposition S1) yields the following Proposition S2 which highlights the consequences of ambiguous survival beliefs for life-cycle consumption and asset accumulation.

Proposition S2. *Comparing the naive CEU agent to the RE agent we get the following implications:*

- *There is undersaving in a sense that*

$$m_0^n > m_0 \quad \Leftrightarrow \quad c_0^n > c_0 \quad \Leftrightarrow \quad w_1^n < w_1$$

if and only if there is sufficient underestimation (SU) *of survival risk by the naive CEU agent in a sense that*

$$(\nu_{0,1}^0)^{\frac{1}{\theta}} + (\nu_{0,2}^0)^{\frac{1}{\theta}} < \psi_{0,1}^{\frac{1}{\theta}} + \psi_{0,2}^{\frac{1}{\theta}}. \tag{SU}$$

- *Naive CEU agents save less in period 1 than originally planned, i.e.,*

$$m_1^{0,n} < m_1^n,$$

if and only if there is moderate overestimation (MO) *of survival risk in a sense that*

$$\nu_{1,2}^1 < \frac{\nu_{0,2}^0}{\nu_{0,1}^0}. \tag{MO}$$

- *There is oversaving in the sense that*

$$m_1^n < m_1 \quad \Leftrightarrow \quad \frac{c_2^n}{c_1^n} > \frac{c_2}{c_1} \quad \Leftrightarrow \quad \frac{w_2^n}{w_1^n} < \frac{w_2}{w_1},$$

by Assumption S1, equation (2) (i.e., by overestimation). Combined with condition (SU) this implies that

$$c_1^n < c_1.$$

- *Naive CEU agents have higher wealth than RE agents*

$$w_2^n > w_2 \quad \Leftrightarrow \quad c_2^n > c_2,$$

if and only if there is sufficient overestimation (SO) *of survival beliefs in period 1 in a sense that*

$$\nu_{1,2}^1 > \psi_{1,2} \left(\frac{1 + (\nu_{0,1}^0)^{\frac{1}{\theta}} + (\nu_{0,2}^0)^{\frac{1}{\theta}}}{1 + \psi_{0,1}^{\frac{1}{\theta}} + \psi_{0,2}^{\frac{1}{\theta}}} \cdot \frac{\psi_{0,1}^{\frac{1}{\theta}} + \psi_{0,2}^{\frac{1}{\theta}}}{(\nu_{0,1}^0)^{\frac{1}{\theta}} + (\nu_{0,2}^0)^{\frac{1}{\theta}}} \cdot \frac{1 + (\nu_{1,2}^1)^{\frac{1}{\theta}}}{1 + \psi_{1,2}^{\frac{1}{\theta}}} \right)^{\theta} > \psi_{1,2}. \tag{SO1}$$

Proof.

- That $m_0^n > m_0$ under condition (SU) immediately follows from the expressions for the respective marginal propensities in Observation S1 and Proposition S1.

- That $m_1^{0,n} < m_1^n$ under condition (MO) immediately follows from comparing the respective marginal propensities given in Proposition S1.

- That $m_1^n < m_1$ under Assumption S1, equation (2), again immediately follows from comparison of the respective expressions for marginal propensities in Observation S1 and Proposition S1.

- By the respective expressions for marginal propensities in Observation S1 and Proposition S1, the inequality $w_2^n > w_2$ and therefore $c_2^n > c_2$ holds iff

$$(1 - m_0^n)(1 - m_1^n) > (1 - m_0)(1 - m_1)$$

$$\Leftrightarrow \quad \frac{(\nu_{0,1}^0)^{\frac{1}{\theta}} + (\nu_{0,2}^0)^{\frac{1}{\theta}}}{1 + (\nu_{0,1}^0)^{\frac{1}{\theta}} + (\nu_{0,2}^0)^{\frac{1}{\theta}}} \cdot \frac{(\nu_{1,2}^1)^{\frac{1}{\theta}}}{1 + (\nu_{1,2}^1)^{\frac{1}{\theta}}} > \frac{\psi_{0,1}^{\frac{1}{\theta}} + \psi_{0,2}^{\frac{1}{\theta}}}{1 + \psi_{0,1}^{\frac{1}{\theta}} + \psi_{0,2}^{\frac{1}{\theta}}} \frac{\psi_{1,2}^{\frac{1}{\theta}}}{1 + \psi_{1,2}^{\frac{1}{\theta}}}$$

$$\Leftrightarrow \quad \left(\frac{\nu_{1,2}^1}{\psi_{1,2}}\right)^{\frac{1}{\theta}} > \underbrace{\frac{1 + (\nu_{0,1}^0)^{\frac{1}{\theta}} + (\nu_{0,2}^0)^{\frac{1}{\theta}}}{1 + \psi_{0,1}^{\frac{1}{\theta}} + \psi_{0,2}^{\frac{1}{\theta}}} \cdot \frac{\psi_{0,1}^{\frac{1}{\theta}} + \psi_{0,2}^{\frac{1}{\theta}}}{(\nu_{0,1}^0)^{\frac{1}{\theta}} + (\nu_{0,2}^0)^{\frac{1}{\theta}}}}_{>1,\ \text{by condition (SU)}} \cdot \underbrace{\frac{1 + (\nu_{1,2}^1)^{\frac{1}{\theta}}}{1 + \psi_{1,2}^{\frac{1}{\theta}}}}_{>1,\ \text{by Assumption S1, equation (2)}}$$

$$(6)$$

from which condition (SO1) readily follows. To see that the first term on the RHS of (6) exceeds one under condition (SU), notice that condition (SU) implies

$$\frac{\psi_{0,1}^{\frac{1}{\theta}} + \psi_{0,2}^{\frac{1}{\theta}}}{(\nu_{0,1}^0)^{\frac{1}{\theta}} + (\nu_{0,2}^0)^{\frac{1}{\theta}}} > 1$$

$$\Leftrightarrow \quad 1 + \frac{1}{(\nu_{0,1}^0)^{\frac{1}{\theta}} + (\nu_{0,2}^0)^{\frac{1}{\theta}}} > 1 + \frac{1}{\psi_{0,1}^{\frac{1}{\theta}} + \psi_{0,2}^{\frac{1}{\theta}}}$$

$$\Leftrightarrow \quad \frac{\psi_{0,1}^{\frac{1}{\theta}} + \psi_{0,2}^{\frac{1}{\theta}}}{1 + \psi_{0,1}^{\frac{1}{\theta}} + \psi_{0,2}^{\frac{1}{\theta}}} \cdot \frac{1 + (\nu_{0,1}^0)^{\frac{1}{\theta}} + (\nu_{0,2}^0)^{\frac{1}{\theta}}}{(\nu_{0,1}^0)^{\frac{1}{\theta}} + (\nu_{0,2}^0)^{\frac{1}{\theta}}} > 1.$$

□

As to the first implication, we require condition (SU) in order for the simple model to give rise to undersaving. Observe that condition (SU) has a straightforward interpretation for log-utility ($\theta = 1$). It can then be rewritten as

$$SLE_0 \equiv 1 + \nu_{0,1}^0 + \nu_{0,2}^0 < LE_0 \equiv 1 + \psi_{0,1} + \psi_{0,2},$$

i.e., subjective life-expectancy at birth, SLE_0, is less than the respective objective life-expectancy, LE_0. The term *sufficient underestimation (SU)* means that underestimation of subjective beliefs in period 0, cf. Assumption S1, equation (1), must be sufficiently strong in order to dominate any overestimation of subjective survival beliefs to occur eventually. Otherwise the naive CEU agent would (weakly) save more than the RE agent, given her forward looking behavior.

As to the second implication, we require *moderate overestimation*, cf. condition (MO), of subjective survival beliefs. In contrast to condition SU, condition MO refers to survival beliefs formed in period 1 for the probability to survive to period 2. Accordingly, it restricts $\nu_{1,2}^1$ by an upper bound which is determined by the ratio of subjective beliefs, $\frac{\nu_{0,2}^0}{\nu_{0,1}^0}$. That is, only if overestimation is not too large, we can expect model households to save less than originally planned. Otherwise the naive CEU agent would revise her plan to (weakly) save more than originally planned.

The third implication states that our assumption on overestimation, cf. Assumption S1, equation (2), immediately gives rise to the implication that the speed of asset decumulation of the naive CEU agent is less than the speed of decumulation of the RE agent. This does not, however, imply that period 2 asset holdings of the naive CEU agent exceed those of the RE agent because the effects of sufficient underestimation in period 0 and overestimation in period 1 work in opposite directions as far as asset holdings are concerned.

This observation readily implies that an additional lower bound on the degree of overestimation is required in order to find that asset holdings in period 2 of the naive CEU agent exceed those of the RE agent. This is stated as *sufficient overestimation* in condition (SO1). Optimism has to be sufficiently strong to dominate the initial underestimation of survival beliefs. As an interpretation of the lower bound observe that the lower bound increases if the initial underestimation of survival belief gets stronger, i.e., if the gap between $\psi_{0,1}$ and $\nu_{0,1}^0$ increases.

The analysis so far clarifies that it is a quantitative question whether the calibrated life-cycle model can generate the three empirical regularities on saving behavior: (i) time inconsistent behavior to the effect that people save less than originally planned (under "moderate overestimation"); (ii) undersaving at young age (under "sufficient underestimation"); (iii) too high old-age asset holdings (under "sufficient overestimation").

S.2.3 Sophisticated CEU Agents

Unlike naive agents, sophisticated agents anticipate the correct lifetime utility for all future selves as additional constraints, i.e., they anticipate that their future selves will not be acting in their interest. The only way to influence future selves behavior is via the

savings decision of current self 0. Thus, sophisticated agents take (over-) consumption of future selves into account when making their current saving plans.

The solution to the problem of the sophisticated CEU agent is as follows:

Proposition S3. *The solution to the sophisticated CEU agent's problem in period 0 is given by*

$$c_0^s = m_0^s w_0 = \frac{1}{1 + \frac{\left(\Theta_0\left(m_1^s\right) \cdot \nu_{0,1}^0\right)^{\frac{1}{\theta}}}{m_1^s}} w_0, \tag{7}$$

where $m_1^s = m_1^n$ and $m_2^s = m_2^n = 1$ and

$$\Theta_0\left(m_1^s\right) \equiv \left(m_1^s + \frac{\nu_{0,2}^0}{\nu_{0,1}^0 \nu_{1,2}^1}\left(1 - m_1^s\right) \right).$$

$\Theta_0\left(m_1^s\right) > 1$ *under condition (MO). The solution for m_0^s is given by*

$$m_0^s = \frac{1}{1 + \left(1 + (\nu_{1,2}^1)^{\frac{1}{\theta}}\right)^{1 - \frac{1}{\theta}} \left(\nu_{0,1}^0 + \nu_{0,2}^0 (\nu_{1,2}^1)^{\frac{1}{\theta} - 1}\right)^{\frac{1}{\theta}}}. \tag{8}$$

Proof. In period 2 we obviously have $m_2^s = m_2^n$. The FOC of the sophisticated agent in period 1 is the same as for the period 1 naive agent, cf. the proof of Proposition S1. From this it follows that $m_1^s = m_1^n$. In period 0 the first-order condition is given by

$$\frac{du}{dc_0^s} = \nu_{0,1}^0 \Theta_0\left(m_1^s\right) \frac{du}{dc_1^s}$$

$$\Leftrightarrow \quad \frac{c_1^s}{c_0^s} = \left(\nu_{0,1}^0 \Theta_0\left(m_1^s\right)\right)^{\frac{1}{\theta}}$$

$$\Leftrightarrow \quad c_0^s = \left(\nu_{0,1}^0\right)^{-\frac{1}{\theta}} \left(\Theta_0\left(m_1^s\right)\right)^{-\frac{1}{\theta}} (w_0 - c_0^s) m_1^s$$

$$\Leftrightarrow \quad c_0^s = \frac{1}{1 + \frac{1}{\left(\Theta_0\left(m_1^s\right) \cdot \nu_{0,1}^0\right)^{-\frac{1}{\theta}} m_1^s}} w_0$$

where

$$\Theta_0\left(m_1^s\right) \equiv m_1^s + \frac{\nu_{0,2}^0}{\nu_{0,1}^0 \nu_{1,2}^1}\left(1 - m_1^s\right).$$

We obviously get that

$$\Theta_0\left(m_1^s\right) > 1 \qquad \Leftrightarrow \qquad \frac{\nu_{0,2}^0}{\nu_{0,1}^0 \nu_{1,2}^1} > 1.$$

The latter term is condition (MO).

To derive m_0^s start from

$$\Theta_0\left(m_1^s\right)\nu_{0,1}^0 = \left(m_1^s + \frac{\nu_{0,2}^0}{\nu_{0,1}^0\nu_{1,2}^1}(1-m_1^s)\right)\nu_{0,1}^0$$

$$= m_1^s\nu_{0,1}^0 + \frac{\nu_{0,2}^0}{\nu_{1,2}^1}(1-m_1^s)$$

$$= \frac{1}{1+(\nu_{1,2}^1)^{\frac{1}{\theta}}}\nu_{0,1}^0 + \frac{\nu_{0,2}^0}{\nu_{1,2}^1}\frac{(\nu_{1,2}^1)^{\frac{1}{\theta}}}{1+(\nu_{1,2}^1)^{\frac{1}{\theta}}}$$

$$= \frac{\nu_{0,1}^0 + \nu_{0,2}^0(\nu_{1,2}^1)^{\frac{1}{\theta}-1}}{1+(\nu_{1,2}^1)^{\frac{1}{\theta}}},$$

therefore

$$\left(\Theta_0\left(m_1^s\right)\nu_{0,1}^0\right)^{-\frac{1}{\theta}}m_1^s = \left(\frac{\nu_{0,1}^0 + \nu_{0,2}^0(\nu_{1,2}^1)^{\frac{1}{\theta}-1}}{1+(\nu_{1,2}^1)^{\frac{1}{\theta}}}\right)^{-\frac{1}{\theta}}\frac{1}{1+(\nu_{1,2}^1)^{\frac{1}{\theta}}}$$

$$= \frac{\left(\nu_{0,1}^0 + \nu_{0,2}^0(\nu_{1,2}^1)^{\frac{1}{\theta}-1}\right)^{-\frac{1}{\theta}}}{\left(1+(\nu_{1,2}^1)^{\frac{1}{\theta}}\right)^{1-\frac{1}{\theta}}}$$

and

$$1 + \frac{1}{\left(\Theta_0\left(m_1^s\right)\nu_{0,1}^0\right)^{-\frac{1}{\theta}}m_1^s} = 1 + \frac{\left(1+(\nu_{1,2}^1)^{\frac{1}{\theta}}\right)^{1-\frac{1}{\theta}}}{\left(\nu_{0,1}^0 + \nu_{0,2}^0(\nu_{1,2}^1)^{\frac{1}{\theta}-1}\right)^{-\frac{1}{\theta}}}$$

$$= \frac{\left(\nu_{0,1}^0 + \nu_{0,2}^0(\nu_{1,2}^1)^{\frac{1}{\theta}-1}\right)^{-\frac{1}{\theta}} + \left(1+(\nu_{1,2}^1)^{\frac{1}{\theta}}\right)^{1-\frac{1}{\theta}}}{\left(\nu_{0,1}^0 + \nu_{0,2}^0(\nu_{1,2}^1)^{\frac{1}{\theta}-1}\right)^{-\frac{1}{\theta}}}.$$

Using the above in (7) gives (8). □

This shows that the solution to the sophisticated agent's problem in terms of policy functions (i.e., in terms of marginal propensities to consume) is identical to the naive agent in periods 1 and 2. This is due to the fact that the marginal propensity in period 2 is known to be $m_2^s = m_2^n = 1$ for all types. Consequently, $\Theta_1(m_2^s) = 1$ and therefore also $m_1^s = m_1^n$.

As $\Theta_0(m_1^s) > 1$ under "moderate overestimation", cf. condition (MO), we find that condition (MO) leads to a higher growth rate of marginal utilities of sophisticates compared to naifs, implying that consumption growth increases. $\Theta_0\left(m_1^s\right) > 1$ reflects the sophisticated agent's high valuation of savings. At the same time, $\Theta_0\left(m_1^s\right)$ depends

negatively on the MPC of the future self 1, m_1^s, implying that self 0's propensity to save decreases when future self 1's MPC increases. Yet, these statements refer only to the change of consumption over time. The level of consumption of sophisticates in period 0 of course also depends negatively on m_1^s. The higher is m_1^s the higher will be m_0^s for reasons of consumption smoothing. Hence, whether c_0^s is lower than c_0^n depends on these offsetting forces. The next proposition makes this explicit:

Proposition S4. *Define Ξ as*

$$\Xi \equiv \frac{\left(1 + (\nu_{1,2}^1)^{\frac{1}{\theta}}\right)^{1-\frac{1}{\theta}} \left(\nu_{0,1}^0 + \nu_{0,2}^0 (\nu_{1,2}^1)^{\frac{1}{\theta}-1}\right)^{\frac{1}{\theta}}}{(\nu_{0,1}^0)^{\frac{1}{\theta}} + (\nu_{0,2}^0)^{\frac{1}{\theta}}}.$$

We have

$$\Xi \begin{cases} > 1 & \Leftrightarrow m_0^n > m_0^s,\ c_0^n > c_0^s,\ c_1^n < c_1^s,\ c_2^n < c_2^s \\ < 1 & \Leftrightarrow m_0^n < m_0^s,\ c_0^n < c_0^s,\ c_1^n > c_1^s,\ c_2^n > c_2^s \\ = 1 & \text{if } \theta = 1 \quad \Leftrightarrow m_0^n = m_0^s,\ c_0^n = c_1^s,\ c_1^n = c_1^s,\ c_2^n = c_2^s. \end{cases}$$

Proof. Observe from Propositions 5 and 7 that

$$m_0^n \geq m_0^s$$

$$\Leftrightarrow \quad \frac{1}{1 + (\nu_{0,1}^0)^{\frac{1}{\theta}} + (\nu_{0,2}^0)^{\frac{1}{\theta}}} \geq \frac{1}{1 + \left(1 + (\nu_{1,2}^1)^{\frac{1}{\theta}}\right)^{1-\frac{1}{\theta}} \left(\nu_{0,1}^0 + \nu_{0,2}^0 (\nu_{1,2}^1)^{\frac{1}{\theta}-1}\right)^{\frac{1}{\theta}}}$$

$$\Leftrightarrow \quad \left(1 + (\nu_{1,2}^1)^{\frac{1}{\theta}}\right)^{1-\frac{1}{\theta}} \left(\nu_{0,1}^0 + \nu_{0,2}^0 (\nu_{1,2}^1)^{\frac{1}{\theta}-1}\right)^{\frac{1}{\theta}} \geq (\nu_{0,1}^0)^{\frac{1}{\theta}} + (\nu_{0,2}^0)^{\frac{1}{\theta}}$$

$$\Leftrightarrow \quad \Xi \equiv \frac{\left(1 + (\nu_{1,2}^1)^{\frac{1}{\theta}}\right)^{1-\frac{1}{\theta}} \left(\nu_{0,1}^0 + \nu_{0,2}^0 (\nu_{1,2}^1)^{\frac{1}{\theta}-1}\right)^{\frac{1}{\theta}}}{(\nu_{0,1}^0)^{\frac{1}{\theta}} + (\nu_{0,2}^0)^{\frac{1}{\theta}}} \geq 1.$$

We readily observe from $\Xi = 1 \Leftrightarrow m_0^s = m_0^n$ for $\theta = 1$ (in which case $\nu_{1,2}^1$ does not influence the consumption decision of sophisticates in period 0). Furthermore, if $m_0^s < m_0^n$, then $w_1^s > m_1^n$. Given that the consumption growth rate between periods 1 and 2 of the two agents is identical and given that they both consume the same present value of life-time resources, w_0, this also implies that $c_1^s > c_1^n$ and $c_2^s > c_2^n$ and vice versa for $m_0^s > m_0^n$. $\square$

It would be desirable to further make statements about how Ξ varies with θ. To approach this, observe from Proposition S1 that an increase of θ decreases m_0^n. This is due to the desire for consumption smoothing: increasing θ (decreasing the IES) increases the

11

consumption growth rate which increases savings and reduces the marginal propensity to consume today. For the sophisticated agent the analogous effects are at work but there is an important (at least partially) offsetting one. This is easiest to see by inspection of equation (7) in Proposition S3. We assume for the remainder of the analysis that condition (MO) holds to the effect that $\Theta_0(m_1^s) > 1$. First, increasing θ reduces m_1^s. As seen from Proposition S3 this reduces m_0^s. Second, increasing θ also increases $(\nu_{0,1}^0)^{\frac{1}{\theta}}$ which also contributes to a reduction of m_0^s. Third, increasing θ also indirectly affects $\Theta_0(m_1^s)$: by decreasing m_1^s, $\Theta_0(m_1^s)$ goes up. As $\Theta_0(m_1^s) > 1$—by condition (MO)—this effect further contributes to a reduction of m_0^s. Finally, however, notice that there is a direct effect of increases in θ via term $\Theta_0(m_1^s)^{\frac{1}{\theta}}$. Under condition (MO) the derivative of this term with respect to θ is negative. It is given by $-\frac{1}{\theta^2} \ln\left(\Theta_0(m_1^s)\right) \Theta_0(m_1^s)^{\frac{1}{\theta}}$. Hence, holding $\Theta_0(m_1^s)$ constant this partially offsetting effect is particularly strong for low values of θ (and weak for high values of θ). The effect is also strong when $\nu_{1,2}^1$ is low because then m_1^s is low. We can therefore expect that $\frac{\partial \Xi}{\partial \theta} > 0$ if $\nu_{1,2}^1$ exceeds some critical value. The next sufficient condition establishes this locally at $\theta = 1$:

Proposition S5. $\left.\frac{\partial \Xi}{\partial \theta}\right|_{\theta=1} > 0$ *if*

$$\nu_{1,2}^1 > \nu_{0,1}^0 + \nu_{0,2}^0 - 1. \tag{9}$$

Proof. $\frac{\partial \Xi}{\partial \theta}$ has the same sign as $\frac{\partial \ln \Xi}{\partial \theta}$. We get

$$\ln \Xi = \left(1 - \frac{1}{\theta}\right) \ln\left(1 + (\nu_{1,2}^1)^{\frac{1}{\theta}}\right) + \frac{1}{\theta} \ln\left(\nu_{0,1}^0 + \nu_{0,2}^0 (\nu_{1,2}^1)^{\frac{1}{\theta}-1}\right) - \ln\left((\nu_{0,1}^0)^{\frac{1}{\theta}} + (\nu_{0,2}^0)^{\frac{1}{\theta}}\right)$$
$$= \Xi_1 + \Xi_2 + \Xi_3.$$

We find

$$\frac{\partial \Xi_1}{\partial \theta} = \frac{1}{\theta^2} \ln\left(1 + (\nu_{1,2}^1)^{\frac{1}{\theta}}\right) - \frac{1}{\theta^2}\left(1 - \frac{1}{\theta}\right) \frac{\ln(\nu_{1,2}^1)(\nu_{1,2}^1)^{\frac{1}{\theta}}}{1 + (\nu_{1,2}^1)^{\frac{1}{\theta}}} > 0, \text{ if } \theta \geq 1 \tag{10}$$

$$\frac{\partial \Xi_2}{\partial \theta} = -\frac{1}{\theta^2} \ln\left(\nu_{0,1}^0 + \nu_{0,2}^0 (\nu_{1,2}^1)^{\frac{1}{\theta}-1}\right) - \underbrace{\frac{1}{\theta^3} \frac{\nu_{0,2}^0 \ln(\nu_{1,2}^1)(\nu_{1,2}^1)^{\frac{1}{\theta}-1}}{\nu_{0,1}^0 + \nu_{0,2}^0 (\nu_{1,2}^1)^{\frac{1}{\theta}-1}}}_{<0} \tag{11}$$

$$\frac{\partial \Xi_3}{\partial \theta} = -\frac{1}{\theta^2} \frac{\ln(\nu_{0,1}^0)(\nu_{0,1}^0)^{\frac{1}{\theta}} + \ln(\nu_{0,2}^0)(\nu_{0,2}^0)^{\frac{1}{\theta}}}{(\nu_{0,1}^0)^{\frac{1}{\theta}} + (\nu_{0,2}^0)^{\frac{1}{\theta}}} > 0. \tag{12}$$

Hence, for $\theta \geq 1$ any ambiguity in the sign of $\frac{\partial \Xi}{\partial \theta}$ can only come from the first term in the derivative (11). For $\theta = 1$, the term is obviously positive (so that the overall derivative is positive) if $\nu_{0,1}^0 + \nu_{0,2}^0 < 1$. For the other case, i.e., for $\nu_{0,1}^0 + \nu_{0,2}^0 \geq 1$, we get by comparing for $\theta = 1$ the first term in (10) with the first term in (11) the sufficient condition. $\square$

12

Figures S.1 and S.2 present an illustration for a calibrated version of the simple model. For simplicity, we consider a static model here and choose $\delta_1 = \delta_2 = 0.5$. We also set $\lambda = 0.5$, $\psi_{0,1} = 1$ and $\psi_{1,2} = 0.25$. We consider $\theta \in \{0.5, 1.5\}$. This parametrization is such that Assumption S1 holds. It also gives rise to conditions (MO), (SU), (SO1) and (9). The figures confirm the findings from the previous propositions. Importantly, the relative consumption patterns between RE and CEU households displayed in Figure S.1 and the differences between sophisticates and naifs shown in Panel (a) of Figure S.2 correspond to our findings in the quantitative model (which also features a relatively low IES), cf. Figure 7 in the paper.

Figure S.1: Consumption

(a) $\theta = 1.5$ (b) $\theta = 0.5$
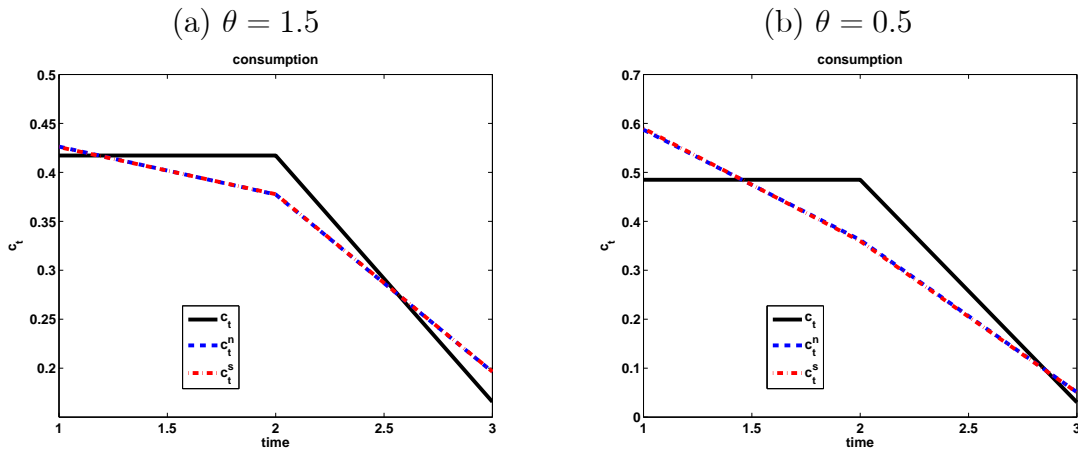
Notes: Consumption for different values of θ.

Figure S.2: Difference in Consumption: Naives & Sophisticates
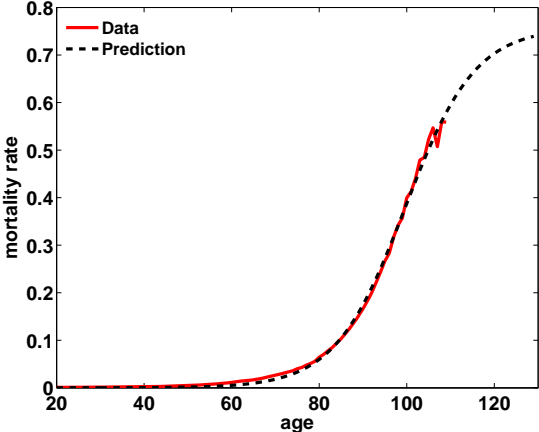
(a) $\theta = 1.5$ (b) $\theta = 0.5$
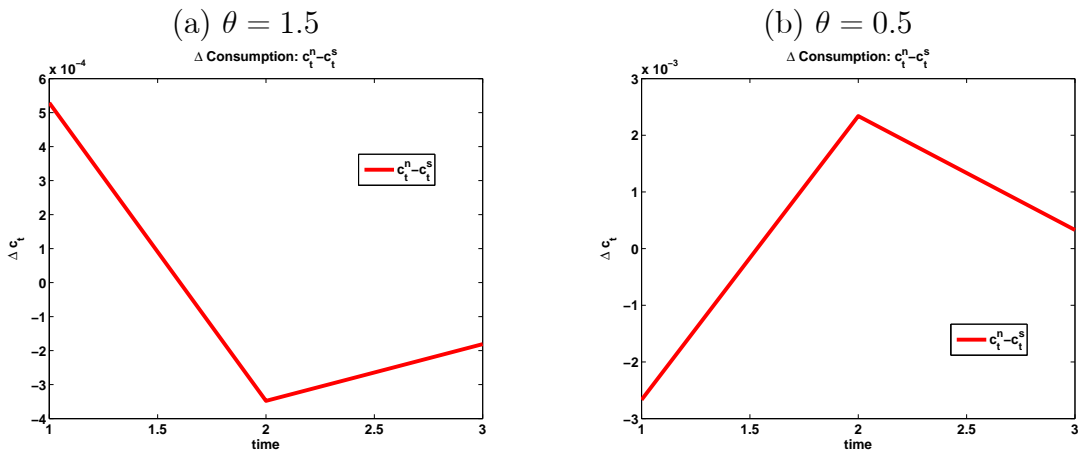
Notes: Difference in consumption between naifs and sophisticates for different values of θ.

13

S.3 Additional Results

Objective Survival Rates

Figure S.3: Mortality Rates: Data vs. Estimation

Notes: Average mortality rates from 2000-2010 using HMD data (red solid line) and predicted mortality rates (black dashed line) using the logistic frailty model given in (36) in the paper. Parameter estimates are provided in Table 1 in the paper.

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.

Deaton, A. (1992). Understanding Consumption. Oxford: Clarendon Press.

Epstein L. G. and M. Schneider (2007). Learning under Ambiguity. Review of Economic Studies 74, 1275-1303.

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

Kastenbaum, R. (2000). The Psychology of Death. New York: Springer Publishing Company.

Marinacci, M. (2002). Learning from Ambiguous Urns. Statistical Papers 43, 145-151.

Nicholls, N., Romm, A. T., and A. Zimper (2014). The Impact of Statistical Learning on Violations of the Sure-Thing Principle. Accepted for publication in Journal of Risk and Uncertainty.

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

Wakker, P. P. (2004). On the Composition of Risk Preference and Belief. Psychological Review 111, 236–241.

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

Zimper, A. (2011). Do Bayesians Learn their Way out of Ambiguity? Decision Analysis 8, 269-285.

Zimper, A. and W. Ma (2014). Bayesian Learning with Multiple Priors: The Role of Prior-Revision Rules. Mimeo.