MANNHEIM RESEARCH INSTITUTE FOR THE ECONOMICS OF AGING

BIASED BAYESIAN LEARNING AND THE RISK-FREE RATE PUZZLE

Alexander Ludwig and Alexander Zimper

191-2009

L13, 17_D-68131 Mannheim_Phone +49 621 181-2773/1862_Fax +49 621 181-1863_www.mea.uni-mannheim.de

Biased Bayesian learning and the risk-free rate puzzle*

Alexander Ludwig† Alexander Zimper‡

September 25, 2009

Abstract

According to the risk-free rate puzzle the return on safe assets is much lower than predicted by standard representative agent models of consumption based asset pricing. Based on non-additive probability measures arising in Choquet decision theory we develop a closed-form model of Bayesian learning in which the Choquet estimator of the mean consumption growth rate does not converge to its "true" value. It rather expresses a bias that reflects the agent's ambiguity about his estimator. We calibrate the standard equilibrium conditions of the consumption based asset pricing model to demonstrate that our approach contributes to a resolution of the risk-free rate puzzle when the agent's learning process exhibits a moderate degree of ambiguity that is resolved in a pessimistic way.

Keywords: behavioral macroeconomics; bounded rationality; consumption based asset pricing; risk-free rate puzzle; ambiguity; Bayesian learning; non-additive probability measures

JEL Classification Numbers: C79, D83

---

*We thank Klaus Adam and Peter Wakker for helpful comments and suggestions.

†Mannheim Research Institute for the Economics of Aging (MEA); Universität Mannheim; L13, 17; 68131 Mannheim; Germany; Email: ludwig@mea.uni-mannheim.de.

‡Department of Economics and Econometrics, University of Johannesburg, PO Box 524, Auckland Park, 2006, South Africa. E-mail: azimper@uj.ac.za

1 Introduction

Starting with the seminal contribution of Lucas (1978), theoretical models that determine asset-prices in a rational expectations equilibrium (REE) of a representative consumer economy have become the workhorses of the macroeconomic finance literature. As their main virtue these models derive the relationship between the consumer's uncertainty with respect to future consumption growth and equilibrium asset prices. However, as first demonstrated by Mehra and Prescott (1985), asset returns predicted by theses models are, by some large margin, at odds with actually observed asset returns when structural model parameters are calibrated with realistic values. Three major asset pricing puzzles have emerged from this literature, the "equity premium puzzle", the "risk-free rate puzzle" and the "excess-volatility puzzle". The focus of this paper is on the "risk-free rate puzzle" according to which a realistically calibrated standard consumption based asset pricing model yields a return on risk-free assets of about $5\% - 6\%$ compared to a real-world actual risk-free rate in the range of $1 - 2\%$.

In this paper we address the risk-free rate puzzle under the assumption that the representative agent's long-run estimator for the mean of the consumption growth-rate may be biased because the agent's Bayesian learning process is prone to ambiguity attitudes. In order to focus our analysis, we adopt the preference structure of the original asset pricing model by sticking to the standard assumptions that the representative consumer's expected utility from an infinite consumption stream is additively time-separable and that the consumer's period t utility of consumption is derived from a CRRA (constant relative risk aversion) function. Also in line with standard specifications of the original consumption based asset pricing model, we assume that the agent can observe arbitrarily large sample information drawn from an independently and normally distributed consumption growth rate process.

Unlike standard models of Bayesian learning, however, we consider ambiguous beliefs about the mean of the consumption growth-rate that are formally described as non-additive probability measures (=capacities). Non-additive probability measures arise as generalizations of subjective additive probability measures in *Choquet expected utility* (CEU) theory which relaxes Savage's sure thing principle in order to accommodate for ambiguity attitudes as elicited in Ellsberg (1961) paradoxes (cf. Schmeidler 1986, 1989; Gilboa 1987). In particular, we focus attention on neo-additive capacities in the sense of Chateauneuf, Eichberger and Grant (2007), according to which the decision maker's attitudes towards ambiguity are reflected by decision weights attached to the best, respectively, worst outcome in a given situation. Neo-additive capacities thus formally capture the empirical observation that real-life decision-makers tend to put too much decision-weight on extreme outcomes thereby offering an alternative (=non-additive) way for modelling so-called "fat-tail" phenomena.

As our main technical contribution we develop a closed-form model of Bayesian learning with respect to ambiguous beliefs about the mean of an independently and identically normally distributed stochastic process such that the resulting Choquet Bayesian estimator remains bounded away from the sample mean also in the long-run. As one formal property of this Choquet estimator, we obtain that its bias is the greater the more surprised the agent is by the information he receives. Finally, we use the biased long-run Choquet estimator for the consumption growth rate in the equilibrium conditions of the standard consumption based asset pricing model according to which the equilibrium return of a risk-free asset is given as the inverse of the agent's expectation of the stochastic discount factor.

As our main conceptual result we demonstrate that our approach contributes to a resolution of the risk-free rate puzzle for a sufficiently large degree of ambiguity whereby this ambiguity is resolved in a rather pessimistic way. This possible resolution of the risk-free rate puzzle has a lot of intuitive appeal: the equilibrium return of a riskless asset remains low in the long run because agents will always have a high demand for such assets since, firstly, they will never fully trust their estimation based on the data sample and, secondly, they react to this lack of trust in a rather pessimistic, i.e., cautious, way.

There are several different proposals in the literature on asset return puzzles that also relax the REE assumption according to which the representative agent's subjective estimator must coincide with the corresponding "objective" expected value given as relative frequencies in the economic data. Along this line, Cecchetti, Lam and Mark (2000) consider the implications of rules of thumb with respect to estimates of the consumption growth process and Abel (2002) studies the effects of pessimism and doubt on asset returns. Related to this literature but with less of an *ad hoc* flavor are robust control applications to asset pricing puzzles where pessimism results from an agents' caution in responding to concerns about model misspecification (Hansen, Sargent and Tallarini 1999; Anderson, Hansen, Sargent 2004; Maenhout 2004; Hansen and Sargent 2007). An apparent drawback of both approaches is, however, their incompatibility with standard learning models with Bayesian features or overtones by which subjective beliefs converge to their objective counterparts in the long-run, cf. Barsky and DeLong (1993), Timmermann (1993), Brav and Heaton (2002), Cogley and Sargent (2008) and Adam, Marcet and Nicolini (2008). As our main contribution to this literature, our Choquet Bayesian learning model thus provides a consistent theoretical foundation for why a subjective estimator for the stochastic discount factor may remain biased even in the long-run.

Several authors also obtain long-run biases of subjective beliefs in Bayesian learning models. E.g., Brennan and Xia (2001) and Lewellen and Shanken (2002) consider cases in which the mean of an exogenous dividend process may not be constant over time. Consequently, the consumer can never fully learn the objective parameters of the underlying distribution because observed frequencies do not admit any conclusions about objective probabilities even in the long run. Along the same line, Weitzman (2007) considers a setup in which the variance of the consumption growth rate is a hidden parameter whereas the mean is known. In contrast to these approaches that specify unstable stochastic processes, the representative consumer in our model observes data that is drawn from a stable stochastic process but the posterior distribution is biased away from the objective distribution due to psychological attitudes that do not vanish in the long-run.

Epstein and Schneider (2007) also consider a model of learning under ambiguity sharing with our approach the feature that ambiguity does not necessarily vanish in the long run. Their learning model is based on the *recursive multiple priors* approach (Epstein and Wang 1994; Epstein and Schneider 2003) which, basically, restricts conditional *max min expected utility* (MMEU) preferences of Gilboa and Schmeidler (1989) in such a way that dynamic consistency is satisfied.[1] While MMEU theory is closely related to CEU theory restricted to *convex* capacities (e.g., neo-additive capacities for which the degree of optimism is zero), our learning model differs substantially from Epstein and Schneider's approach. Epstein and Schneider establish long-run ambiguity under the assumption that the decision maker permanently receives ambiguous signals, which they formalize via a multitude of different likelihood functions at each information stage in ad-

[1] Similarly, the aforementioned robust control applications to asset pricing puzzles are also related to the max-min expected (multiple priors) utility theory of Gilboa and Schmeidler (1989), cf. Hansen and Sargent (2001) and Hansen, Sargent, Turmuhambetova and Williams (2006).

dition to the existence of multiple priors. In the case of learning from ambiguous urns without multiple likelihoods, ambiguity obviously vanishes in the learning process (for a formal analysis also see Marinacci 2002). However, the introduction of multiple likelihoods lacks an axiomatic and/or psychological foundation going beyond the mere technical property that long-run ambiguity is sustained by multiple likelihoods. In contrast, our—comparably simple—axiomatically founded model of a Bayesian learner who is prone to psychological attitudes in the interpretation of new information offers a rather straightforward explanation for biased long-run beliefs even when the decision maker receives signals that are not ambiguous. Finally, notice that the restriction of Epstein and Schneider's approach to dynamically consistent preferences excludes preferences that violate Savage's sure-thing principle as elicited in Ellsberg paradoxes. Since our learning model does not exclude dynamically inconsistent decision behavior, it can accommodate a broader notion of ambiguity attitudes.

Our approach—which exclusively focusses on biased Bayesian learning of parameter values—is also related to a literature on asset pricing puzzles that investigates alternative preference structures.[2] For example, Weil (1989), Epstein and Zin (1989, 1991) and others examine the asset pricing implications of general recursive non-expected utility preferences, whereas Campbell and Cochrane (1999), building on Abel (1990), Constantinides (1990) and others, consider preferences with habit formation. As illustrated by Kocherlaokta (1996), realistic calibrations using such alternative preference structures help resolve the risk-free rate puzzle but not the equity premium puzzle. Since we exclusively focus on the risk-free rate puzzle, our own approach is complementary to this strand of the literature. We also regard it as a viable avenue for future research to combine such preference structures with the possibility that subjective beliefs may remain biased in the long-run.

The remainder of our analysis is structured as follows. Section 2 restates the risk-free rate puzzle in the standard model. In Section 3 we describe the benchmark case of Bayesian learning in the absence of ambiguity. Our formal Choquet Bayesian learning model is presented in Section 4. In Section 5 we then apply the resulting estimator of our Choquet Bayesian learning model to

[2] A review of this literature is given in the survey articles by Kocherlakota (1996), Campbell (2003), Mehra and Prescott (2003) and the textbook treatments in Cochrane (2001) and Duffie (2001).

the standard equilibrium conditions of the risk-free rate puzzle framework and present a possible resolution of the puzzle by our approach, both qualitatively and quantitatively. Finally, Section 6 concludes. Our decision theoretic framework of ambiguous beliefs as well as Bayesian updating of ambiguous beliefs is described in more detail in Appendix 1. All proofs are relegated to Appendix 2.

2 The risk-free rate puzzle

According to the risk-free rate puzzle, the return on a risk-free asset derived from a standard Lucas (1978) type asset pricing model is substantially higher than observed in the data. As point of departure we here describe a simple variant of the standard asset-pricing economy. In the macro-finance literature on the equity premium puzzle, it is often assumed that there is one single productive unit which produces the perishable consumption good and that there is accordingly one equity share that is competitively traded. The other security usually considered is a one-period risk-free asset, which pays one unit of the consumption good next period with certainty. As we will later be concerned only with the risk-free rate of return, our setup is even simpler and we start by considering at first one arbitrary asset. Consider a representative period-t agent and let c_s and y_s denote the random variables for the consumption level and dividend payment in period $s > t$ respectively. Conditional on information I_t the representative period-t agent chooses asset holdings z_s for periods $s > t$ as the solution to the maximization problem

$$\max \left(u(c_t) + \sum_{s=t+1}^{\infty} \beta^{s-t} E\left[u(c_s), \pi(c_s \mid I_t)\right] \right) \tag{1}$$

subject to

$$c_s = y_s \cdot z_s + p_s \cdot (z_s - z_{s+1}) \text{ for all } s \tag{2}$$

where $\beta < 1$ is the agent's time-discount factor, p_s is the ex-dividend asset price in period s and $E\left[u(c_s), \pi_t(c_s \mid I_t)\right]$ is the agent's expected utility of period s consumption with respect to the conditional probability measure $\pi_t(c_s \mid I_t)$. The corresponding first order conditions imply for any equilibrium

$$1 = E\left[R^*_{t+1} \cdot M_{t+1}, \pi(c_{t+1} \mid I_t)\right] \text{ for } t = 0, 1, ... \tag{3}$$

whereby

$$R^*_{t+1} = \frac{p^*_{t+1} + y_{t+1}}{p^*_t} \tag{4}$$

denotes the asset's equilibrium *gross-return* in period $t+1$ at equilibrium prices p_t^*, p_{t+1}^*, and

$$M_{t+1} = \beta \cdot \frac{u'(c_{t+1})}{u'(c_t)} \tag{5}$$

denotes the *so-called stochastic discount factor.*

In order to state Mehra and Prescott (1985)'s version of the risk-free puzzle, we resort to a standard utility structure and assume that the per period utility function features constant relative risk aversion (CRRA),

$$u(c_t) = \frac{c_t^{1-\gamma}}{1-\gamma} \text{ for } t = 0, 1, ..., \tag{6}$$

where $\gamma > 0$ is the coefficient of relative risk aversion. With this parametric assumption, the constant equilibrium gross-return of the risk-free asset in (3) becomes

$$R_{t+1}^* = E\left[M_{t+1}, \pi\left(X_{t+1} \mid I_t\right)\right]^{-1} \text{ for } t = 0, 1, ... \tag{7}$$

such that

$$M_{t+1} = \beta \cdot \exp\left(-\gamma X_{t+1}\right) \tag{8}$$

where the random variable $X_{t+1} = \ln c_{t+1} - \ln c_t$ is the stochastic consumption growth, which coincides in equilibrium with the stochastic dividend-payment growth $\ln y_{t+1} - \ln y_t$. Under the assumption that the conditional probability measure $\pi\left(X_{t+1} \mid I_t\right)$ converges for large t to the unconditional "objective" distribution π^* of consumption growth, the long-run equilibrium return of the risk-free asset (7) writes as

$$R^* = E\left[M, \pi^*\right]^{-1} \tag{9}$$

(with probability one), or equivalently stated in logarithmic terms with $r^* = \ln R^*$,

$$r^* = -\ln\beta + \gamma\mu^* - \frac{1}{2}\gamma^2\sigma^2 \tag{10}$$

since

$$\begin{aligned} E\left[M, \pi^*\right] &= E\left[\beta \cdot \exp -\gamma X, \pi^*\right] \\ &= \beta \cdot \exp\left(-\gamma\mu^* + \frac{1}{2}\gamma^2\sigma^2\right) \end{aligned}$$

for the log-normally distributed random variable M.

While (10) provides relevant insights into the qualitative relationship between the agent's uncertainty about consumption growth and the equilibrium price of a risk-free asset, the quantitative implications of the model are strongly at odds with the data as first demonstrated by Mehra and Prescott (1985). For example, calibrating the model with a discount factor of $\beta = 0.98$, a coefficient of relative risk-aversion of $\gamma = 2$ (conventional values of γ range from 1 to 4) and using data for the period 1950-2004[3] to compute the mean and variance of annual consumption growth giving $\mu^* = 2.13\%$ and $\sigma^* = 1.07\%$, results in an equilibrium risk-free rate of 6.25%. This is about 4.06 percentage points higher than our point estimate of the risk-free rate of 2.19%.

It is our aim in this paper to decrease this gap between empirical estimates for the return of a risk-free asset and the equilibrium return as derived from a consumption-based asset pricing model. To this end we consider the equilibrium condition

$$R^* = E\left[M, \pi^{**}\right]^{-1} \Leftrightarrow \tag{11}$$

$$r^* = -\ln \beta + \gamma \mu^{**} - \frac{1}{2}\gamma^2 \sigma^2 \tag{12}$$

where $\pi^{**} = N\left(\mu^{**}, \sigma^2\right)$ such that μ^{**} is not the "true" value of the mean of the consumption growth rate. Quite trivially, condition (11) for the equilibrium return of a risk-free asset would better fit the data than (9) whenever

$$E\left[M, \pi^{**}\right] \gg E\left[M, \pi^*\right] \Leftrightarrow \tag{13}$$

$$\mu^{**} \ll \mu^*. \tag{14}$$

The non-trivial contribution of our paper deals with the question:

> How can Bayesian learners end up in the long run with a biased Bayesian estimator for the stochastic discount factor, i.e., $E\left[M, \pi^{**}\right]$, rather than the standard (unbiased) estimator $E\left[M, \pi^*\right]$?

The standard justification for using an "objective" probability distribution in the statement of the risk-free rate puzzle refers to the fundamental result of the rational Bayesian learning literature that Bayesian estimators are consistent, i.e., will converge in the long-run to the true parameter value. In the following section we develop a specific closed-form model of rational Bayesian learning that will serve at the basis of a closed-form model of biased Bayesian learning, which we present in Section 4. According to this non-rational Bayesian learning model a biased long-run estimator—satisfying, e.g., (14)—will emerge whenever agents have ambiguity attitudes that are reflected in their estimation.

[3]See section 5 for a description of data sources and definitions.

3 Bayesian learning of the consumption growth rate: The rational benchmark case

This section develops our model of Bayesian learning for the benchmark case in which the representative period-n agent's belief is given as an additive probability measure so that this belief does not reflect any ambiguity in our sense. In particular, we consider a consumption growth rate with "objective" normal distribution $\pi^* = N(\mu^*, \sigma^2)$ whereby we assume that the agent's subjective belief about the consumption growth rate is also given by a normal distribution. While we further stipulate that the variance of this distribution, σ^2, is known to the agent, we assume that the agent is uncertain about this distribution's mean. In the absence of ambiguity, we model this uncertainty by a random variable with a truncated normal distribution so that the expected value of this random variable is the agent's best estimator of the mean of the consumption growth-rate distribution. Thereby, we condition the expectation on the period-n agent's information, denoted I_n, given as the n observed past consumption growth rates. The assumption of a truncated rather than an untruncated distribution is, in our opinion, more realistic because no agent would regard any number between minus and plus infinity as a possible expected value of the growth-rate. Furthermore, the truncation ensures that extreme outcomes are well-defined in terms of an *infimum* and *supremum*, respectively, which will be analytically relevant for our concept of ambiguous beliefs.

Formally, we consider a probability space $(\pi, \Omega, \mathcal{F})$ where π denotes a subjective additive probability measure defined on the events in $\mathcal{F}$. In what follows, we describe the construction of the state space, of the event space, and of the additive probability measure. The event space is thereby closely related to our information structure, which we also construct in detail. Finally, we establish consistency of the additive estimator for this closed-form model of Bayesian learning.

Construction of the state space Ω. For some numbers μ and $a > 0$, denote by $\Theta = (\mu - a, \mu + a)$ the *parameter-space* that collects all "values of the mean of the consumption growth rate distribution" that the agent regards as *possible*. We make the following epistemic assumption:

Assumption 1. *The agent always regards the "true" parameter value, denoted μ^*, as possible, i.e., $\mu^* \in (\mu - a, \mu + a)$.*

Denote by $X^{\infty} = \times_{i=1}^{\infty} X_i$ with $X_i = \mathbb{R}$, the *sample-space* that collects all possible infinite sequences of observations about the consumption growth rates in periods $i = 1, 2, ...$ The state space of our model is then defined as

$$\Omega = \Theta \times X^{\infty}, \tag{15}$$

with generic element $\omega = (\theta, x_1, x_2, ...)$.

Construction of the information structure. The information structure of our learning model is described by an infinite sequence of ever *finer* information partitions $\mathcal{P}_1, \mathcal{P}_2, ...$ such that each $\mathcal{P}_n$, $n = 1, 2, ...$, is defined as the collection of information cells

$$\Theta \times \{x_1\} \times ... \times \{x_n\} \times X_{n+1} \times X_{n+2} \times ... \text{ for all } (x_1, ..., x_n) \in \times_{i=1}^{n} X_i. \tag{16}$$

According to this information structure, the agent will never receive any direct information about the true parameter value. That is, regardless of how many observations n the agent makes and regardless of the specific value $(x_1, ..., x_n)$ of observations, he will always regard any parameter value in $(\mu - a, \mu + a)$ as possible. The distinctive feature of any model of Bayesian learning of a parameter value is that learning exclusively follows from (indirect) likelihood considerations but never from (direct) knowledge about the parameter value. Loosely speaking, this feature gives rise to the possibility that the agent of our model may never learn the true parameter value when his likelihood considerations are described by a non-additive probability measure.

Construction of the event space $\mathcal{F}$. Endow Θ with the Euclidean metric and denote by $\mathcal{B}$ the Borel σ-algebra in Θ, i.e., the smallest σ-algebra containing all open subsets of the Euclidean interval $(\mu - a, \mu + a)$. Similarly, endow each X_i with the Euclidean metric and denote by $\mathcal{X}_i$ the Borel σ-algebra in X_i. Our event space $\mathcal{F}$ is then defined as the standard product σ-algebra generated by $\mathcal{B}, \mathcal{X}_1, \mathcal{X}_2, ...$.

In a next step, define by Σ_n the σ-algebra generated by $\mathcal{P}_n$, for $n = 1, 2, ...$ That is, Σ_n is the smallest collection of subsets of Ω that is a σ-algebra and that contains all information cells in $\mathcal{P}_n$. Observe that $\Sigma_1 \subset \Sigma_2 \subset ... \subset \mathcal{F}$ so that the sequence of σ-algebras $(\Sigma_1, \Sigma_2, ..., \mathcal{F})$ constitutes a *filtration*.

Construction of the additive probability measure π. Define by $\tilde{\theta} : \Omega \to \Theta$ such that

$$\tilde{\theta}\left(\theta, x_1, x_2, ...\right) = \theta \tag{17}$$

the $\mathcal{F}$-measurable coordinate random variable that assigns to every state of the world the corresponding "true" parameter value, i.e., the corresponding true mean of the consumption growth rate process. We assume that the agent's prior over $\tilde{\theta}$ is given as a *truncated normal distribution* with support on $(\mu - a, \mu + a)$ such that, for all $A \in \mathcal{B}$,

$$\pi\left(\left\{\omega \in \Omega \mid \tilde{\theta}\left(\omega\right) \in A\right\}\right) = \int\limits_{\theta \in A} \psi_{\tilde{\theta}}\left(\theta\right) d\theta \tag{18}$$

where the density function is given by

$$\psi_{\tilde{\theta}}\left(\theta\right) = \begin{cases} \frac{1}{F(\mu+a)-F(\mu-a)} \cdot \frac{1}{\sqrt{2\pi\tau^2}} \exp\left[-\frac{1}{2}\frac{(\theta-\mu)^2}{\tau^2}\right] & \text{for } \tau < \infty \\ \frac{1}{2a} & \text{for } \tau = \infty, \end{cases} \tag{19}$$

for $\theta \in (\mu - a, \mu + a)$ whereby $F(\cdot)$ denotes the cumulative distribution function of the corresponding untruncated normal distribution with mean μ and variance τ^2, i.e.,

$$F\left(z\right) = \int\limits_{-\infty}^{z} \frac{1}{\sqrt{2\pi\tau^2}} \exp\left[-\frac{1}{2}\frac{\left(\theta - \mu\right)^2}{\tau^2}\right] d\theta. \tag{20}$$

Define by $X_n : \Omega \to \mathbb{R}$, with $n = 1, 2, ...$, such that

$$X_n\left(\theta, x_1, x_2, ...\right) = x_n \tag{21}$$

the Σ_n-measurable coordinate random variable whose value stands in for the consumption growth rate in period n. We assume that, conditional on the parameter-value $\theta \in \Theta$, each X_n is independently and normally distributed with mean θ and variance σ^2 whereby the true variance is known to the agent. That is, for all $\theta \in (\mu - a, \mu + a)$ and all $A \in \mathcal{X}_n$,

$$\pi\left(\left\{\omega \in \Omega \mid \tilde{\theta}\left(\omega\right) = \theta, X_n\left(\omega\right) \in A\right\}\right) = \int\limits_{x \in A} \psi_{X_n}\left(x \mid \theta\right) \cdot \psi_{\tilde{\theta}}\left(\theta\right) dx \tag{22}$$

where the conditional density function is given by

$$\psi_{X_n}\left(x \mid \theta\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\frac{\left(x - \theta\right)^2}{\sigma^2}\right] \text{ for } x \in \mathbb{R}. \tag{23}$$

Consider now the Σ_n-measurable random variable $\bar{X}_n$ which denotes the average of the consumption growth rates up to period n, i.e.,

$$\bar{X}_n = \frac{1}{n}\left(X_1 + ... + X_n\right). \tag{24}$$

Under the above distributional assumptions, $\bar{X}_n$ is—conditional on parameter θ—normally distributed so that, for all $\theta \in (\mu - a, \mu + a)$ and all Borel-subsets A of the Euclidean line,

$$\pi\left(\left\{\omega \in \Omega \mid \tilde{\theta}(\omega) = \theta, \bar{X}_n(\omega) \in A\right\}\right) = \int\limits_{x \in A} \psi_{\bar{X}_n}(x \mid \theta) \cdot \psi_{\tilde{\theta}}(\theta)\, dx \tag{25}$$

with conditional density function

$$\psi_{\bar{X}_n}(x \mid \theta) = \frac{1}{\sqrt{2\pi\sigma^2/n}} \exp\left[-\frac{1}{2}\frac{(x-\theta)^2}{\sigma^2/n}\right] \text{ for } x \in \mathbb{R}. \tag{26}$$

Bayesian estimation. Let ψ denote the joint density function of $\tilde{\theta}$ and $\bar{X}_n$. Then, by Bayes' rule, the posterior density function of $\tilde{\theta}$ conditional on observation $\bar{x}_n \in \bar{X}_n(\Omega)$ is given by

$$\psi_{\tilde{\theta}}(\theta \mid \bar{x}_n) = \frac{\psi(\theta, \bar{x}_n)}{\psi_{\bar{X}_n}(\bar{x}_n)} \tag{27}$$

$$= \frac{\psi_{\bar{X}_n}(\bar{x}_n \mid \theta) \cdot \psi_{\tilde{\theta}}(\theta)}{\int_{\mu-a}^{\mu+a} \psi_{\bar{X}_n}(\bar{x}_n \mid \theta) \cdot \psi_{\tilde{\theta}}(\theta)\, d\theta}. \tag{28}$$

The expected value of $\tilde{\theta}$ with respect to this posterior density function, i.e., $E\left[\tilde{\theta}, \psi_{\tilde{\theta}}(\theta \mid \bar{x}_n)\right]$, is then the agent's estimator of the mean of the consumption growth rate in the light of new information $I_n = \left\{\omega \in \Omega \mid \bar{X}_n(\omega) = \bar{x}_n\right\}$. Before presenting our first main result, which characterizes the Bayesian estimator for the additive benchmark case, we introduce the following additional objects. Consider the normal density function

$$f_n(x) = \frac{1}{\sqrt{2\pi\rho_n^2}} \exp\left[-\frac{1}{2}\frac{(x-\mu_n^*)^2}{\rho_n^2}\right] \text{ for } x \in \mathbb{R}. \tag{29}$$

and the corresponding cumulative distribution function

$$F_n(z) = \int\limits_{-\infty}^{z} f_n(x)\, dx \text{ for } z \in \mathbb{R}. \tag{30}$$

where

$$\mu_n^* = \frac{\sigma^2/n}{\sigma^2/n + \tau^2} \cdot \mu + \frac{\tau^2}{\sigma^2/n + \tau^2} \cdot \bar{x}_n \tag{31}$$

and

$$\rho_n^2 = \frac{\tau^2 \sigma^2/n}{\tau^2 + \sigma^2/n}. \tag{32}$$

Proposition 1.

(i) *Conditional on observation* $\bar{x}_n \in \bar{X}_n(\Omega)$, *the agent's estimator of the mean of the consumption growth-rate* $X = \ln c_{n+1} - \ln c_n$, *is given as*

$$E\left[\tilde{\theta}, \psi_{\tilde{\theta}}(\theta \mid \bar{x}_n)\right] = \mu_n^* - \rho_n \frac{f_n(\mu + a) - f_n(\mu - a)}{F_n(\mu + a) - F_n(\mu - a)}. \tag{33}$$

(ii) *Let* $\mu^* \in (\mu - a, \mu + a)$. *Then the estimator (33) converges in probability to the true value*

$$E[X, \pi^*] = \mu^* \tag{34}$$

when n *approaches infinity. That is, for every* $c > 0$,

$$\lim_{n\to\infty} prob\left(\left|E\left[\tilde{\theta}, \psi_{\tilde{\theta}}\left(\theta \mid \bar{X}_n\right)\right] - \mu^*\right| < c\right) = 1. \tag{35}$$

Part (ii) of proposition 1 demonstrates that the consumer's subjective estimator converges to the true mean of the growth rate distribution in the long run if his beliefs about the mean are described by a probability distribution as in (19). This convergence result for the closed-form Bayesian estimator (33) can be regarded as a special case of more general results on the *consistency* of (additive) Bayesian estimates, in particular Doob's consistency theorem (Doob 1949; for extensions see Breiman, LeCam, and Schwartz 1964; Lijoi, Pruenster, and Walker 2004). Roughly speaking, Doob's consistency theorem ensures us that only for some subset of parameter values with measure zero an additive Bayesian estimator for an $\mathcal{F}$-measurable $\tilde{\theta}$ may not converge to the true parameter value.

4 Learning in the case of ambiguity

In this Section we develop a Bayesian framework in which the estimator is not derived from additive but rather from non-additive probability measures which arise in Choquet decision theory.[4] Because a non-additive probability measure may additionally reflect ambiguity attitudes, the corresponding non-additive Bayesian estimator will, in general, not converge to the true parameter value. Recall the learning situation as described in Section 3 but assume that the representative agent resolves his uncertainty by an ambiguous rather than an additive belief. That is, instead of the additive probability space $(\pi, \Omega, \mathcal{F})$ we consider the non-additive probability space $(\nu(\cdot \mid \cdot), \Omega, \mathcal{F})$ with conditional neo-additive capacity $\nu(\cdot \mid \cdot)$ so that, for all $A, I \in \mathcal{F}$ with $A \notin \{\emptyset, \Omega\}$,

$$\nu(A \mid I) = \delta_I \cdot \lambda + (1 - \delta_I) \cdot \pi(A \mid I) \tag{36}$$

with

$$\delta_I = \frac{\delta}{\delta + (1 - \delta) \cdot \pi(I)} \tag{37}$$

and additive probability measure π as defined in Section 3. The ambiguity parameter $\delta \in [0, 1]$ is thereby naturally interpreted as degree of the agent's doubt in his additive belief π whereby the optimism parameter $\lambda \in [0, 1]$ determines whether this doubt is resolved in a rather optimistic (high λ) or pessimistic (low λ) way (cf. Appendix 1). In particular, we have for all $A \in \mathcal{B}$ that

$$\nu\left(\left\{\omega \in \Omega \mid \tilde{\theta}(\omega) \in A\right\}\right) = \int_A \chi_{\tilde{\theta}}(\theta)\, d\theta \tag{38}$$

where the non-additive density function is given by

$$\chi_{\tilde{\theta}}(\theta) = \delta\lambda + (1 - \delta) \cdot \psi_{\tilde{\theta}}(\theta) \text{ for } \theta \in (\mu - a, \mu + a) \tag{39}$$

with $\psi_{\tilde{\theta}}$ defined by (19). Because of $\mu = E\left[\tilde{\theta}, \psi_{\tilde{\theta}}(\theta)\right]$, the agent's prior estimator for the mean of the growth rate is, by observation 2, then given as the Choquet expected value (cf. Appendix 1)

$$E\left[\tilde{\theta}, \chi_{\tilde{\theta}}(\theta)\right] = \delta(\lambda(\mu + a) + (1 - \lambda)(\mu - a)) + (1 - \delta)\mu. \tag{40}$$

Conditional on new information $\bar{x}_n$ about the average growth rate the agent forms the following conditional non-additive density function in accordance with (71):

$$\chi_{\tilde{\theta}}(\theta \mid \bar{x}_n) = \delta_{\bar{x}_n} \cdot \lambda + (1 - \delta_{\bar{x}_n}) \cdot \psi_{\tilde{\theta}}(\theta \mid \bar{x}_n) \tag{41}$$

[4] For the decision-theoretic foundations as well as the corresponding mathematical concepts, e.g., non-additive probability measures, Choquet integration, Bayesian update rules for non-additive probaility measures, we refer the interested reader to Appendix 2.

whereby

$$\delta_{\bar{x}_n} = \frac{\delta}{\delta + (1-\delta) \cdot \psi_{\bar{X}_n}(\bar{x}_n)}. \tag{42}$$

The Choquet expected value of $\tilde{\theta}$ with respect to $\chi_{\tilde{\theta}}\left(\theta \mid \bar{X}_n\right)$, i.e., $E\left[\tilde{\theta}, \chi_{\tilde{\theta}}\left(\theta \mid \bar{X}_n\right)\right]$, then represents the agent's posterior estimator for the mean of the stochastic consumption growth rate.

Proposition 2.

(i) *Conditional on observation $\bar{x}_n \in \bar{X}_n(\Omega)$, the agent's Choquet estimator of the mean of the consumption growth-rate $X = \ln C_{n+1} - \ln C_n$ is given as*

$$\begin{aligned} & E\left[\tilde{\theta}, \chi_{\tilde{\theta}}(\theta \mid \bar{x}_n)\right] \qquad (43) \\ = \;& \delta_{\bar{x}_n} \cdot (\lambda(\mu+a) + (1-\lambda)(\mu-a)) + (1-\delta_{\bar{x}_n}) \cdot E\left[\tilde{\theta}, \psi_{\tilde{\theta}}(\theta \mid \bar{x}_n)\right] \end{aligned}$$

with

$$\delta_{\bar{x}_n} = \frac{\delta}{\delta + (1-\delta) \cdot \psi_{\bar{X}_n}(\bar{x}_n)} \tag{44}$$

and $E\left[\tilde{\theta}, \psi_{\tilde{\theta}}(\theta \mid \bar{x}_n)\right]$ defined as (33).

(ii) *Let $\mu^* \in (\mu - a, \mu + a)$. Then the Choquet estimator (43) converges in probability to*

$$E[X, \nu^*] = \delta_{\mu^*} \cdot (\lambda(\mu+a) + (1-\lambda)(\mu-a)) + (1-\delta_{\mu^*}) \cdot \mu^* \tag{45}$$

with

$$\delta_{\mu^*} = \frac{\delta}{\delta + (1-\delta) \cdot \psi_{\tilde{\theta}}(\mu^*)} \tag{46}$$

whereby $\psi_{\tilde{\theta}}$ is given by (19).

Observe that the Choquet Bayesian estimator (45) is in general biased if $\delta > 0$. Thus, whenever there is some ambiguity in our model, the process of Bayesian learning does not converge towards the true parameter value but to a weighted average between this true value and the agent's attitudes towards optimism versus pessimism under ambiguity as expressed by the parameter λ. Furthermore, observe that the Choquet Bayesian limit estimator (43) puts the more weight

δ_{μ^*} on the ambiguity part the smaller the value of $\psi_{\tilde{\theta}}(\mu^*)$, i.e., the prior density function of the unknown parameter evaluated at the true parameter value.[5] That is, the long-run degree of ambiguity is the larger the less prior additive probability had been attached by the agent to some neighborhood around the true value μ^* of the consumption growth rate. This means that the more surprising the information is for an agent with ambiguous beliefs, the more decision weight is attached to the ambiguity part of his limit beliefs. More precisely, we have that $\delta_{\mu^*} \geq \delta$ ($\delta_{\mu^*} < \delta$)—that is, the ambiguity receives more (less) weight in the course of the learning process—if $\psi_{\tilde{\theta}} \leq 1$ ($\psi_{\tilde{\theta}} > 1$).

5 The risk-free rate puzzle revisited

To illustrate how our model contributes to a resolution of the risk-free rate puzzle, we first summarize a number of qualitative insights form our model in subsection 5.1. These insights provide guidance for a calibrated version of our model which follows in subsection 5.2.

5.1 Qualitative analysis

If the representative agent is a biased Bayesian learner in our sense, the economy's risk-free interest rate will converge (in probability) to

$$R^{**} = \left(E\left[M, \pi^{**}\right]\right)^{-1} = \left(\beta \cdot \exp\left(-\gamma\mu^{**} + \frac{\gamma^2\sigma^2}{2}\right)\right)^{-1} \tag{47}$$

whereby we have for the biased mean

$$\mu^{**} = E\left[X, \nu^*\right] \tag{48}$$

with $E[X, \nu^*]$ given by (45). In case there is no ambiguity, i.e., $\delta = 0$, the equilibrium condition (47) coincides with the standard equilibrium condition (9) of the additive model which is subject to the risk-free rate puzzle. In case there is some ambiguity, i.e., $\delta > 0$, however, (47) can the better fit the data than (9) the more the true mean exceeds its biased estimator. That is, our approach

[5] Formally, the density function $\psi_{\tilde{\theta}}$ in (46) results as the pointwise limit of the density function $\psi_{\bar{X}_n}$ in (44), and μ^* results as limit (in probability) of the sequence $(\bar{x}_n)_{n\in\mathbb{N}}$ (cf. the proof of proposition 2).

contributes towards a resolution of the risk-free rate puzzle whenever we have

$$\begin{aligned} \mu^* - \mu^{**} &\gg 0 \Leftrightarrow \\ \delta_{\mu^*}\left[\mu^* - \left(\lambda\left(\mu + a\right) + \left(1-\lambda\right)\left(\mu - a\right)\right)\right] &\gg 0 \Leftrightarrow \\ \delta_{\mu^*}\left[\mu^* - \mu + a(1-2\lambda)\right] &\gg 0. \end{aligned} \tag{49}$$

Throughout, we further make the following behavioral assumption:

Assumption 2. *The agent resolves his ambiguity in a weakly pessimistic way, that is,* $\lambda \in [0, 0.5]$.

Under assumption 2, equation (49) then highlights four key channels through which our model contributes towards a resolution of the risk-free rate puzzle. The risk-free rate is the lower

1. the higher the final degree of *ambiguity*, δ_{μ^*},
2. the stronger the initial *downward bias*[6] of the agent's belief with regard to the consumption growth rate as measured by the distance $\mu^* - \mu$,
3. the bigger the truncation range as measured by a—which can be interpreted as a measure of *surprise* because the larger a the bigger is the uncertainty and the more surprised the agent will be to observe information μ^*—and
4. the higher the degree of *pessimism* with which the agent resolves his ambiguity, that is, the smaller λ.

Next, observe from (46) that the final degree of ambiguity, δ_{μ^*}, is thereby increasing in the initial degree of ambiguity, δ, and decreasing in the value of the truncated density function of the agent's prior evaluated at the true parameter value, $\psi_{\tilde{\theta}}(\mu^*)$, cf. (19). With regard to this interaction between δ_{μ^*} and $\psi_{\tilde{\theta}}(\mu^*)$, three noteworthy extreme cases emerge in our learning model:

1. Consider the case in which $\mu = \mu^*$ so that $\psi_{\tilde{\theta}}(\mu^*)$ monotonically decreases in τ, cf. equation (19). Assume now the situation of an agent who is ex-ante very certain that the true parameter value is actually given as his

[6] We refer to a situation of initial pessimism in which $\mu < \mu^*$ as *downward bias* in order to not confuse terminology with the degree of pessimism, λ, with which the agent resolves his ambiguity.

information μ^*—i.e., the possible parameter range $\Theta = (\mu - a, \mu + a)$ is either very small (that is, $diam\,(\Theta) \simeq 0$) or τ is very small (or both). We refer to such a situation as *confidence* of the agent in the correctness of his prior belief. As $\mu = \mu^*$, the empirical observations accordingly confirm the correctness of his belief. Therefore, δ_{μ^*} is close to zero because $\psi_{\tilde{\theta}}\,(\mu^*)$ is a large number (it converges to infinity in the limit) so that the Choquet Bayesian estimator (45) is very close to the true parameter value.

2. Continue to assume that $\mu = \mu^*$ but now consider the converse case in which the agent is ex ante very uncertain that μ is indeed the true value—i.e., the possible parameter range $\Theta = (\mu - a, \mu + a)$ is very large or τ is very large (or both). Assume the extreme case in which $a \to \infty$, i.e., there is no truncation, then $\psi_{\tilde{\theta}}\,(\mu^*) \to 0$ which readily implies that $\delta_{\mu^*} \to 1$. Such a situation characterizes extreme *non-confidence* of the agent in the correctness of his prior belief. Observe that the non-confident agent's Choquet Bayesian estimator (45) then (almost) completely ignores the observed sample information and exclusively expresses his ambiguity attitudes regardless of the fact that the sample information actually confirmed his additive prior.

3. Now assume that $\mu < \mu^*$, i.e., the agent's prior belief features a *downward bias* with regard to the mean consumption growth rate. Then, two offsetting effects of τ on $\psi_{\tilde{\theta}}(\mu^*)$ are at work so that the density function peaks at $\bar{\tau} = \mu^* - \mu$, cf. equation (19). For $\tau \geq \bar{\tau}$, $\psi_{\tilde{\theta}}(\mu^*)$ behaves as in cases 1 and 2. For $\tau < \bar{\tau}$, $\psi_{\tilde{\theta}}(\mu^*)$ monotonically increases in τ. For low τ we then have the situation in which the agent is very *convinced* in his false belief and the actual observations do not confirm this belief. Accordingly, the agent will be *surprised* to observe the information μ^*. In the extreme, $\tau \to 0$ and hence $\psi_{\tilde{\theta}}\,(\mu^*) \to 0$ which readily implies that $\delta_{\mu^*} \to 1$. As in case 2, the agent's Choquet Bayesian estimator (45) (almost) completely ignores the observed sample information and exclusively expresses his ambiguity attitudes. Observe that an interesting knife-edge case is where $\psi_{\tilde{\theta}}(\mu^*) = 1$ so that the degree of ambiguity remains unaltered in the course of the learning process, i.e., $\delta_{\mu^*} = \delta$. This knife-edge case will play a role in our calibration below.

To further focus our analysis, we make the following behavioral assumption:

Assumption 3. *The truncation parameter satisfies* $a < \frac{1}{2}$.

By this assumption the agent regards extremely high or low values of the mean of the consumption growth rate as unrealistic. In fact, our calibration below will consider a value for a which is substantially smaller than $\frac{1}{2}$. This implies that, for $\tau \to \infty$, the value of the truncated density is bounded from below by $\frac{1}{2a} > 1$ for any $0 < a < \frac{1}{2}$, cf. equation (19). Consequently, the form of extreme non-confidence discussed above in case 2 is ruled out. On the other hand, assumption 3 allows for the situation discussed in case 3 in which *confidence* about a wrong belief is paired with *surprise*. In our opinion, this last situation has the most intuitive appeal as a description of learning behavior for the situation at hand.

5.2 Quantitative analysis

While it is relatively straightforward to calibrate the new[7] parameters δ_{μ^*}, λ, μ and a such that the standard formulation of the risk-free rate puzzle can be avoided altogether, the question arises whether such parameter-values are actually plausible. For example, while Mehra and Prescott (1985) observe that the standard model can only avoid the equity premium puzzle for an unrealistically high degree of relative risk-aversion, it might be the case that our model can only avoid the risk-free rate puzzle for unrealistically high degrees of ambiguity and pessimism. It is therefore the purpose of our quantitative exercise to analyze the scope of our model to contribute to a resolution of the risk-free rate puzzle under a reasonable parametrization.

To illustrate the quantitative implications of our model, we take as data updated versions of those studied, e.g., by Shiller (1981) and Mehra and Prescott (1985).[8] The data are annual and we focus on the postwar period 1950 to 2004. The risk-free rate is computed as government bond yields and consumption is real per capita consumption of non-durables and services. Both series are inflation-adjusted by the annual consumer price index (CPI). The resulting moments are $\mu^* = 2.13\%$, and $\sigma = 1.08\%$. As already discussed in section 3, further setting $\beta = 0.98$ and $\gamma = 2$, these moments imply a risk-free rate of about 6.25% under the standard rational expectations model. This exceeds our point estimate of the risk-free rate of 2.19% by about 4.06 percentage points.

We then consider the effects of psychological biases on the risk-free rate under

[7]"New" in the sense that these parameters are relevant to our Bayesian learning model but not to the standard Bayesian learning model with additive beliefs.

[8]We take the data from Robert Shiller's website http://www.econ.yale.edu/shiller/data/chapt26.xls.

three alternative scenarios in which we compute the risk-free rate for the Choquet limit beliefs as expressed in equations (45) and (46). These scenarios are distinguished by different parameterizations of the agents prior belief as captured by the truncated density function, $\psi_{\tilde{\theta}}(\theta)$, cf. equation (19) which we evaluate at $\theta = \mu^*$, cf. equation (46). In all these scenarios, we assume that the agent only regards a relatively narrow range of mean consumption growth rates as possible and set the truncation width, a to $a = \mu^* = 0.0213$. We summarize all calibration parameters in table 1.

Table 1: Calibration parameters

Per capita consumption			
Average growth rate, μ^*		0.0213	
S.D. of growth rate, σ		1.08	
Preferences			
Discount factor, β		0.98	
Coeff. of relative risk aversion, γ		2.0	
Parameters in prior distribution			
Scenario	S1	S2	S3
Mean, μ	$\mu^* = 0.0213$	0.015	0.015
Standard deviation, τ	∞	∞	0.002
Truncation parameter, a		$a = \mu^* = 0.0213$	
Truncation range, $[\mu - a, \mu + a]$	$[0.0, 0.043]$	$[-0.006, 0.036]$	$[-0.006, 0.036]$

In our first scenario (S1), we assume that μ coincides with the true mean per capita consumption growth rate as estimated from the data so that $\mu = \mu^* = 0.0213$. Consequently, the truncation range in scenario S1 is given by $[0.0, 0.043]$ so that the minimum *real* consumption growth rate that the agent perceives as possible is zero. In this scenario, we also assume a flat prior and accordingly set $\tau = \infty$. This implies that $\psi_{\tilde{\theta}}(\mu^*) = \frac{1}{2a} = 23.49$ so that $\delta^* < \delta$, cf. equation (46).

In our second scenario (S2) we stick to the flat prior assumption but now consider a significant *downward bias* of the mean consumption growth rate in the agents prior by setting $\mu = 0.015$. This coincides with the long-run estimate of per capita income growth in OECD countries (Maddison 2007). As we hold the truncation parameter a constant across all scenarios, we consequently assume that the truncation range is $[-0.006, 0.036]$ so that the minimum per capita consumption growth rate that the agent perceives as possible is negative at -0.6%. This may result from a combination of, e.g., a perceived minimum *nominal* growth rate of 1% and an inflation rate of 1.6%.

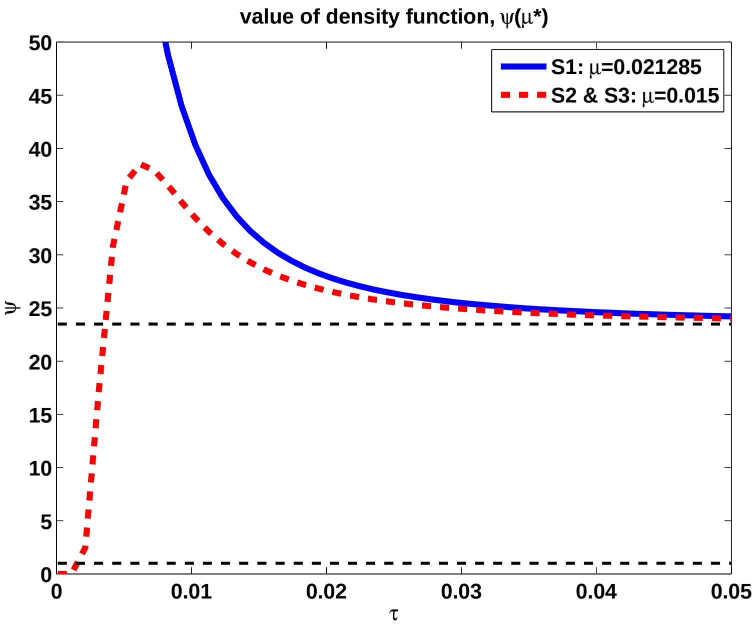
Finally, in scenario 3 (S3) we combine *downward bias* with *confidence*, that is, we assume that the agent is convinced that his wrong prior belief of $\mu = 0.015$ is correct. We thereby depart from the flat prior assumption underlying scenarios S1 and S2. While we continue to assume that the range of possible values is $[-0.006, 0.036]$ we shift almost all probability weight to the center of the distribution. To keep the analysis as parsimonious as possible, we here consider the knife-edge case discussed in subsection 5.1 and accordingly set τ such that $\psi_{\tilde{\theta}}(\mu^*) = 1$ which implies that $\delta = \delta^*$, cf. equation (46). The resulting standard deviation in the agents prior is $\tau = 0.002$.

Figure 1—which plots values of the truncated density function $\psi_{\tilde{\theta}}(\mu^*)$ against $\tau \in [0.0, 0.05]$ for $\mu = \mu^*$ and $\mu = 0.015$—better helps to understand how the choice of τ affects $\psi_{\tilde{\theta}}(\mu^*)$ under assumption 1. While for $\mu = \mu^*$ the density function monotonically decreases in τ to its limit of $1/(2a)$, the two offsetting effects of τ described above are at work for $\mu \neq \mu^*$ so that the density function peaks at $\bar{\tau} = \mu^* - \mu = 0.0063$, cf. equation (19). Consequently, for $\mu = \mu^*$ and $a < \frac{1}{2}$, we always have that $\delta > \delta^*$, cf. equation (46). The converse may only be true for $a < \frac{1}{2}$ whenever $\mu \neq \mu^*$. For sake of simplicity, when looking at the combined effects of an initial downward bias and of a high confidence in this last scenario S3, we here compute the flip point and set τ accordingly.

"Full" resolution

For all three scenarios, we next illustrate how the initial degree of ambiguity, δ, and the degree of pessimism, λ, by which the agent resolves his ambiguity, affects the risk-free rate in our model. To this end, we consider values for $\delta \in [0, 1)$ and $\lambda \in [0, 0.5]$. Figures 2-4 display, for the corresponding scenarios S1-S3, surface plots of the risk-free rate on the grid $\mathcal{G}^{\delta,\lambda} = [0, 1) \otimes [0, 0.5]$. The point where $\delta = 0$ (=no ambiguity) and $\lambda = 0.5$ (=no pessimism or optimism) gives the result under rational expectations. As expected, the risk-free rate increases in δ and λ in all scenarios. However, only for scenario S3, in which the agents prior belief

Figure 1: Truncated density for $\mu = \mu^*$ and $\mu = 0.015$



expresses both, some degree of *downward bias* and a high degree of *confidence*, we observe a strong and about linear decrease of the risk-free interest rate along both dimensions, cf. figure 4. In fact, when combining extreme ambiguity (i.e., δ close to one) with high pessimism (i.e., λ close to zero), our parametrization in scenario S3 would enable us to exactly match the observed risk-free rate.

Partial resolution

While we have just demonstrated that our model could—in principle—fully resolve the risk-free rate puzzle, we regard the corresponding parameter combinations as rather unrealistic. For this reason, we now ask how much potential our model has in adding to other theories that were developed in the literature and that contribute to a resolution of the risk-free rate puzzle, cf. our discussion on alternative preference structures in section 1. To this end, we now analyze which parameter constellation of our model may explain one percentage point, that is, about a quarter of the 4.06 percentage point difference between the observed risk-free rate and the predicted rate under the rational expectations hypothesis. Consequently, we take as calibration target a risk-free interest

Figure 2: Risk-free rate in S1: $\mu = \mu^*$, $\tau = \infty$

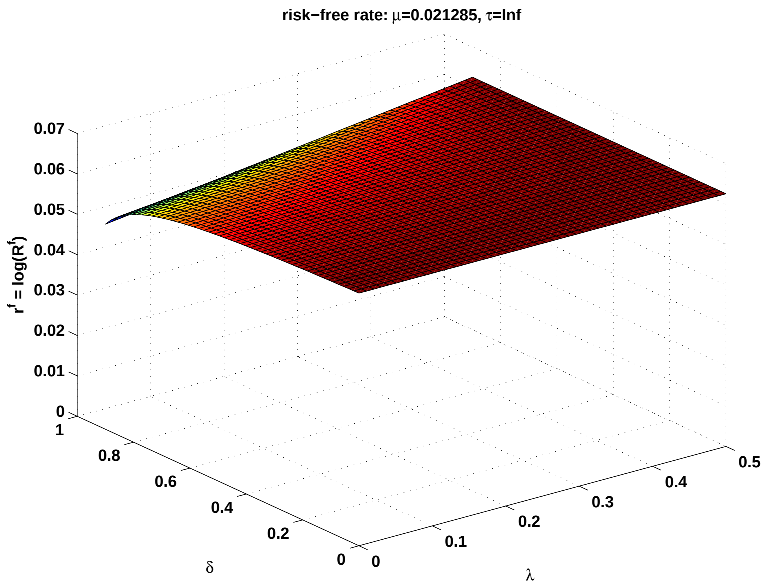


Figure 3: Risk-free rate in S2: $\mu = 0.015$, $\tau = \infty$

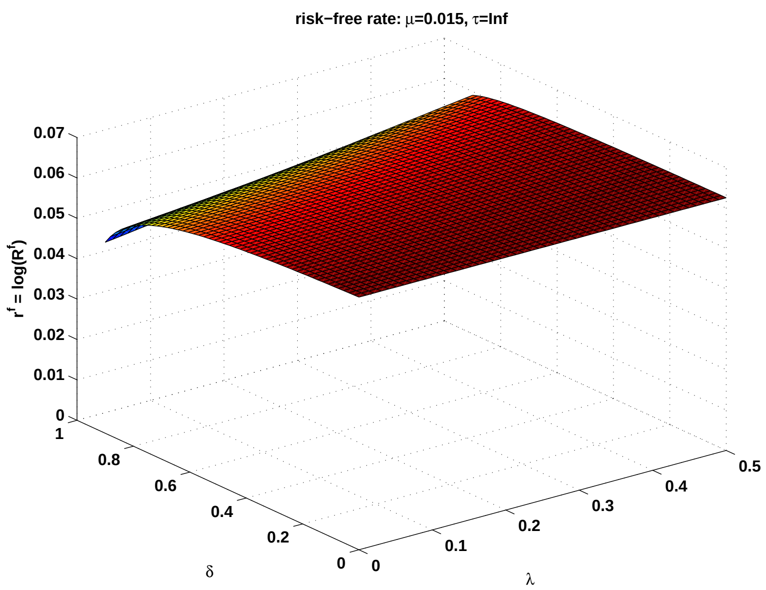
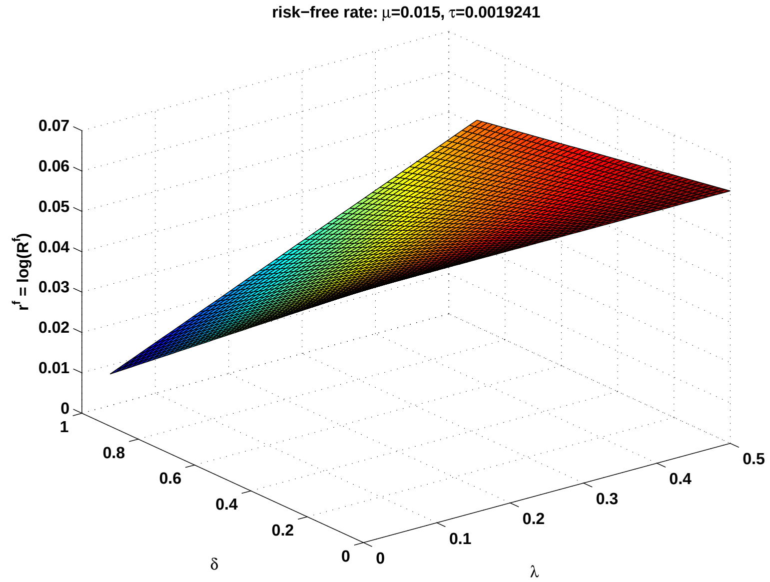


Figure 4: Risk-free rate in S3: $\mu = 0.015$, $\tau = 0.002$



rate of 5.25% (= 6.25% − 1%). Given this target, we then compute from equations (45) and (47) the final degree of ambiguity δ_{μ^*} that would be required to match a risk-free rate of 5.25%. Next, we compute from equation (46) the corresponding initial degree of ambiguity, δ. Results of this experiment for our previous scenarios S1-S3 are displayed in figure 5. According to these results, e.g., a combination of a degree of pessimism of $\lambda = 0.3$ and an initial degree of ambiguity of $\delta = 0.3$ are required in scenario S3 to match the calibration target. These numbers can be regarded as reasonable. In scenarios S1 and S2, however, the required initial degree of ambiguity is still too high.

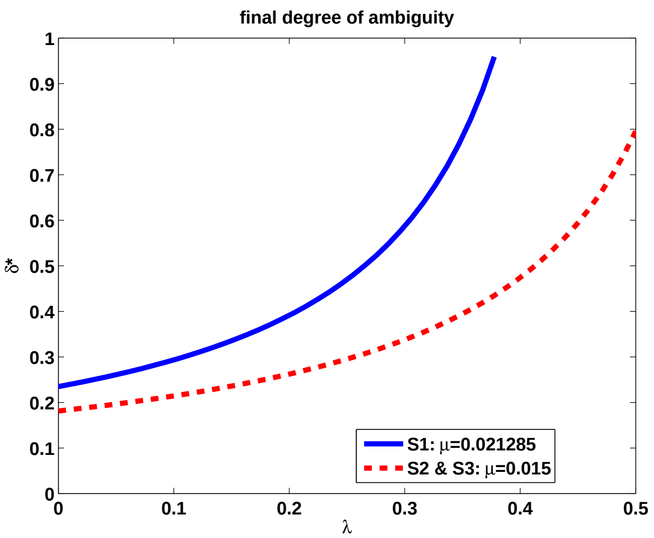
We can therefore conclude that our model can provide a contribution to the resolution of the risk-free rate puzzle if (i) the agent's prior expresses some degree of a downward bias paired with (ii) confidence and if (iii) the agent's belief expresses moderate ambiguity and (iv) the agent resolves this ambiguity in a rather pessimistic way.

6 Conclusion

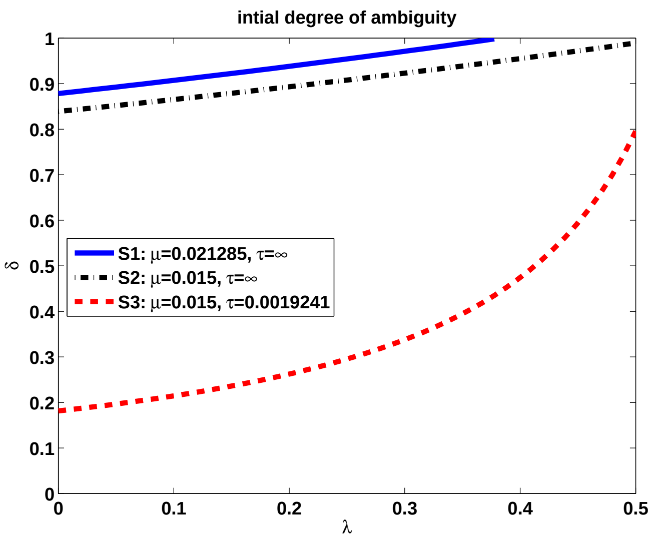
As our main contribution we develop—based on the axiomatic framework of Choquet decision theory—a closed-form model of Bayesian learning with ambiguous

Figure 5: Ambiguity for $\Delta r^f = 0.01$

(a) Final ambiguity, δ_{μ^*}



(b) Initial ambiguity, δ



beliefs for an environment of normally distributed observations combined with an additive part of the prior that is truncated normally distributed over possible parameter values.[9] We apply this framework to the standard consumption based asset pricing model (Lucas 1978; Mehra and Prescott 1985) by assuming that the representative agent has ambiguous beliefs about the mean of the consumption growth rate. Provided that the agent resolves his ambiguity in a pessimistic way, we demonstrate that our approach contributes to a resolution of the risk-free rate puzzle (Weil 1989). In contrast to existing Bayesian explanations of the risk-free rate puzzle, our agent's estimator for the mean of the consumption growth rate will not converge to the sample mean in the long-run. Furthermore, unlike in the learning model of Epstein and Schneider (2007), we offer a straightforward psychological interpretation why limit estimators may still reflect ambiguity.

We also employ a calibrated version of our model to ask whether our model has interesting quantitative implications. Under rational expectations, the risk-free rate in our model exceeds the observed risk-free rate by 4.06 percentage points. The key question we ask in the quantitative analysis is which parameter constellation in our extended model with psychological biases would explain one percentage point of this difference. With this calibration target we show that our model can indeed contribute to a resolution of the risk-free rate puzzle. This holds for moderate degrees of ambiguity and pessimism if the agent's prior belief features a downward bias of the mean consumption growth rate which comes along with a high degree of confidence in the accuracy of this biased prior belief.

Evidently, these calculations are mainly illustrative. By restricting attention to learning of the mean consumption growth rate, our learning model does not give rise to sufficient moment conditions to identify parameter values of our model from the data. In our future research, we will extend our analysis to include learning of the variance of the consumption growth rate, which would enable us to also address other asset pricing puzzles. We also plan to use the moment conditions derived from such an extended framework to estimate the free parameters of our model. In this respect, our parsimonious representation of biased beliefs will be particularly useful.

Despite such limitations, the main virtue of the present paper is to represent a flexible and axiomatically founded alternative to existing approaches in the

[9] For closed-form models of biased Bayesian learning within an environment of Bernoulli trials and Beta-distributed priors see Zimper (2009) and Zimper and Ludwig (2009).

asset pricing literature with biased beliefs such as the recursive multiple priors approach (Epstein and Wang 1994) as well as to robust control applications (Hansen, Sargent and Tallarini 1999; Hansen and Sargent 2007). As shown in Hansen and Sargent (2001) and Hansen, Sargent, Turmuhambetova and Williams (2006), robust decision rules—while incompatible with Bayesian learning—are related to the max-min expected (multiple priors) utility theory of Gilboa and Schmeidler (1989). As in the recursive multiple priors approach (Epstein and Wang 1994; Epstein and Schneider 2003, 2007) such "robust" decision makers are therefore purely pessimistic with respect to the moves of nature which, in our framework, corresponds to the special case of neo-additive beliefs with a zero degree of optimism. However, the decision maker of our approach may also express optimistic attitudes with respect to the moves of nature in that he does not only have the "worst" but also the "best" possible outcome in mind. In this respect, our framework is more general and more in line with recent psychological evidence on decision making under risk and uncertainty than the above cited literature (see, e.g., Wakker 2004 and the cited literature therein). In our future research we plan to further exploit this flexibility of our approach.

To focus our analysis, we here stick to the standard preference structure of the consumption based asset pricing model and only alter its model of estimators for the mean of the consumption growth-rate. We use this simplified framework to establish a relationship between long-run pessimistic beliefs and low returns on risk-free assets. As our approach is in line with intuition and empirical observations that pessimistic rather than optimistic agents choose secure assets as their favorite saving device (Puri and Robinson 2007), the established relationship is a viable contribution to the long list of possible explanations for the low risk-free rate. Moreover, pessimism in our sense is a hard-wired property of an agent's personality that does not simply vanish through new information but rather affects the interpretation of new information. We thereby provide a decision theoretically sound foundation of Abel (2002)'s concept of pessimism.

7 Appendix 1: Decision-theoretic foundations

7.1 Ambiguous beliefs

We consider a measurable space $(\Omega, \mathcal{F})$ with $\mathcal{F}$ denoting a σ-algebra on the state space Ω. As a generalization of the concept of additive probability measures, however, we now consider non-additive probability measures, i.e., capacities, that are used for modeling ambiguous beliefs within the decision-theoretic framework of Choquet expected utility (CEU) theory.[10] In contrast to an additive probability measure, a capacity $\nu : \mathcal{F} \to [0,1]$ must only satisfy the conditions of normalization and monotonicity w.r.t. set-inclusion, i.e.,

(i) $\nu(\emptyset) = 0$, $\nu(\Omega) = 1$

(ii) $A \subset B \Rightarrow \nu(A) \leq \nu(B)$ for all $A, B \in \mathcal{F}$.

Additional properties of capacities are used in the literature for formal definitions of, e.g., *ambiguity* and *uncertainty attitudes* (Schmeidler 1989; Epstein 1999; Ghirardato and Marinacchi 2002), *pessimism* and *optimism* (Eichberger and Kelsey 1999; Wakker 2001), as well as *sensitivity to changes in likelihood* (Wakker, 2004). The Choquet expected value of a bounded random variable $Y : \Omega \to \mathbb{R}$ with respect to capacity ν is formally defined as the following Riemann integral extended to domain Ω (Schmeidler 1986):

$$E[Y, \nu] = \int_{-\infty}^{0} (\nu(\{\omega \in \Omega \mid Y(\omega) \geq z\}) - 1)\, dz + \int_{0}^{+\infty} \nu(\{\omega \in \Omega \mid Y(\omega) \geq z\})\, dz. \tag{50}$$

Our own approach focuses on non-additive beliefs that are defined as *neo-additive capacities* in the sense of Chateauneuf, Eichberger and Grant (2007).

Definition. *For a given measurable space* $(\Omega, \mathcal{F})$ *the neo-additive capacity,* ν*, is defined, for some* $\delta, \lambda \in [0,1]$ *by*

$$\nu(A) = \delta \cdot (\lambda \cdot \omega^{o}(A) + (1-\lambda) \cdot \omega^{p}(A)) + (1-\delta) \cdot \pi(A) \tag{51}$$

for all $A \in \mathcal{F}$ *such that* π *is some additive probability measure and we have*

[10] Schmeidler (1986, 1989) provides a decision-theoretic axiomatization of CEU theory within the Anscombe-Aumann (1963) framework. Gilboa (1987) provides an according axiomatic foundation within the Savage (1954) framework.

for the non-additive capacities ω^o

$$\omega^o(A) = 1 \text{ if } A \neq \emptyset \tag{52}$$
$$\omega^o(A) = 0 \text{ if } A = \emptyset \tag{53}$$

and ω^p *respectively*

$$\omega^p(A) = 0 \text{ if } A \neq \Omega \tag{54}$$
$$\omega^p(A) = 1 \text{ if } A = \Omega. \tag{55}$$

Observe that for non-degenerate events, i.e., $A \notin \{\emptyset, \Omega\}$, the neo-additive capacity ν in (51), simplifies to

$$\nu(A) = \delta \cdot \lambda + (1 - \delta) \cdot \pi(A). \tag{56}$$

Neo-additive capacities can thus be interpreted as non-additive beliefs that stand in for deviations from additive beliefs such that a parameter δ (*degree of ambiguity*) measures the lack of confidence the decision maker has in some subjective additive probability distribution π. The following proposition extends a result (Lemma 3.1) of Chateauneuf, Eichberger and Grant (2007) for finite random variables to the more general case of random variables with a bounded range.

Observation 1. *Let* Y *be a random variable with bounded range, i.e.,* $diam Y(\Omega) < \infty$*. Then the Choquet expected value (50) of* Y *with respect to a neo-additive capacity (51) is given by*

$$E[Y, \nu] = \delta(\lambda \sup Y + (1 - \lambda) \inf Y) + (1 - \delta) E[Y, \pi]. \tag{57}$$

Proof. By an argument in Schmeidler (1986), it suffices to restrict attention to a non-negative valued random variable Y so that

$$E[Y, \nu] = \int_0^{+\infty} \nu(\{\omega \in \Omega \mid Y(\omega) \geq z\}) \, dz, \tag{58}$$

which is equivalent to

$$E[Y, \nu] = \int_{\inf Y}^{\sup Y} \nu(\{\omega \in \Omega \mid Y(\omega) \geq z\}) \, dz$$

since Y is closed and bounded. We consider a partition P_n, $n = 1, 2, ...,$ of Ω with members

$$A_n^k = \{\omega \in \Omega \mid a_{k,n} < X(\omega) \leq b_{k,n}\} \text{ for } k = 1, ..., 2^n$$

such that

$$\begin{aligned} a_{k,n} &= [\sup Y - \inf Y] \cdot \frac{(k-1)}{2^n} + \inf Y \\ b_{k,n} &= [\sup Y - \inf Y] \cdot \frac{k}{2^n} + \inf Y. \end{aligned}$$

Define the step functions $a_n : \Omega \rightarrow \mathbb{R}$ and $b_n : \Omega \rightarrow \mathbb{R}$ such that, for $\omega \in A_n^k$, $k = 1, ..., 2^n$,

$$\begin{aligned} a_n(\omega) &= a_{k,n} \\ b_n(\omega) &= b_{k,n}. \end{aligned}$$

Obviously,

$$E[a_n, \nu] \leq E[Y, \nu] \leq E[b_n, \nu]$$

for all n and

$$\lim_{n \rightarrow \infty} E[b_n, \nu] - E[a_n, \nu] = 0.$$

That is, $E[a_n, \nu]$ and $E[b_n, \nu]$ converge to $E(Y, \nu)$ for $n \rightarrow \infty$. Furthermore, observe that

$$\begin{aligned} \inf a_n &= \inf Y \text{ for all } n \text{, and} \\ \sup b_n &= \sup Y \text{ for all } n. \end{aligned}$$

Assume next that

$$E[b_n, \nu] = \delta(\lambda \sup b_n + (1 - \lambda) \inf b_n) + (1 - \delta) E[b_n, \pi]. \tag{59}$$

for all n. Since $\lim_{n \rightarrow \infty} \inf b_n = \lim_{n \rightarrow \infty} \inf a_n$ and $E[b_n, \pi]$ is continuous in n, we have

$$\begin{aligned} \lim_{n \rightarrow \infty} E[b_n, \nu] &= \delta\left(\lambda \lim_{n \rightarrow \infty} \sup b_n + (1 - \lambda) \lim_{n \rightarrow \infty} \inf b_n\right) + (1 - \delta) \lim_{n \rightarrow \infty} E[b_n, \pi] \\ &= \delta(\lambda \sup Y + (1 - \lambda) \inf Y) + (1 - \delta) E(Y, \pi). \end{aligned}$$

In order to prove proposition 3, it therefore remains to be shown that (59) holds for all n. Since b_n is a step function, (58) becomes

$$\begin{aligned} E[b_n, \nu] &= \sum_{A_n^k \in P_n} \nu\left(A_n^{2^n} \cup ... \cup A_n^k\right) \cdot (b_{k,n} - b_{k-1,n}) \\ &= \sum_{A_n^k \in P_n} b_{k,n} \cdot \left[\nu\left(A_n^{2^n} \cup ... \cup A_n^k\right) - \nu\left(A_n^{2^n} \cup ... \cup A_n^{k-1}\right)\right], \end{aligned}$$

implying for a neo-additive capacity

$$\begin{aligned}
E\left[b_n,\nu\right] &= \sup b_n \left[\delta\lambda + (1-\delta)\,\pi\left(A_n^{2^n}\right)\right] + \sum_{k=2}^{2^n-1} b_{k,n}\,(1-\delta)\,\pi\left(A_n^k\right) \\
&\quad + \inf b_n \left[1-\delta\lambda-(1-\delta)\sum_{k=2}^{2^n}\pi\left(A_n^k\right)\right] \\
&= \delta\lambda \sup b_n + (1-\delta)\sum_{k=1}^{2^n} b_{k,n}\pi\left(A_n^k\right) + \inf b_n\left[\delta-\delta\lambda\right] \\
&= \delta\left(\lambda \sup b_n + (1-\lambda)\inf b_n\right) + (1-\delta)\,E\left[b_n,\pi\right].
\end{aligned}$$

□

According to observation 1, the Choquet expected value of a random variable Y with respect to a neo-additive capacity is a convex combination of the expected value of Y with respect to some additive probability measure π and an ambiguity part. If there is no ambiguity, i.e., $\delta = 0$, then the Choquet expected value in (57) reduces to the standard expected value of a random variable with respect to an additive probability measure. In case there is some ambiguity, however, the second parameter λ measures how much weight the decision maker puts on the best possible outcome of Y when resolving his ambiguity. Conversely, $(1-\lambda)$ is the weight he puts on the worst possible outcome of Y. As a consequence, we interpret λ as an "optimism under ambiguity" parameter whereby $\lambda = 1$, respectively $\lambda = 0$, corresponds to extreme optimism, respectively extreme pessimism, with respect to resolving ambiguity in the decision maker's belief.

7.2 Bayesian updating of ambiguous beliefs

In contrast to Bayesian updating of additive probability measures, there exist several perceivable Bayesian update rules for non-additive probability measures (cf. Gilboa and Schmeidler 1993; Sarin and Wakker 1998; Eichberger, Grant and Kelsey 2006; Siniscalchi 2001, 2006). Since Bayesian updating of neo-additive capacities is central to our formal model of learning under ambiguity, this section describes in detail how such update rules arise in the subjective probability framework from preferences over Savage acts conditional on the fact that some event has occurred. We thereby focus attention on the so-called full (or generalized) Bayesian update rule.

Define the Savage-act $f_I h : \Omega \rightarrow \mathbb{R}$ such that

$$f_I h(\omega) = \begin{cases} f(\omega) & \text{for } \omega \in I \\ h(\omega) & \text{for } \omega \in \neg I \end{cases} \tag{60}$$

where I is some event which we interpret as "new information". Recall that Savage's sure-thing principle claims that, for all acts f, g, h, h' and all events I,

$$f_I h \succeq g_I h \text{ implies } f_I h' \succeq g_I h'. \tag{61}$$

Let us interpret event I as new information received by the agent. The sure-thing principle then implies a straightforward way for deriving preferences $\succeq_I$, conditional on the new information I, from the agent's original preferences $\succeq$ over Savage-acts. Namely, we have

$$f \succeq_I g \text{ if and only if } f_I h \succeq g_I h \text{ for any } h, \tag{62}$$

implying for a subjective expected utility maximizer

$$f \succeq_I g \text{ if and only if } E[u(f), \pi(\cdot \mid I)] \geq E[u(g), \pi(\cdot \mid I)] \tag{63}$$

where $u(f(\omega))$, $\omega \in \Omega$, are von Neumann-Morgenstern utility indices for the consequences realized by act f and $\pi(\cdot \mid I)$ is the standard additive conditional probability measure defined on $(\Omega, \mathcal{F})$.

In order to accommodate ambiguity attitudes as elicited in Ellsberg paradoxes (Ellsberg 1961), CEU theory drops the sure-thing principle. As a consequence, conditional CEU preferences are no longer derivable from (62) since the specification of the act h is now relevant. As one possible specification of h, let us consider conditional CEU preferences satisfying, for all acts f, g,

$$f \succeq_I g \text{ if and only if } f_I h \succeq g_I h \tag{64}$$

where h is the so-called conditional certainty equivalent of g, i.e., h is the constant act such that $g \sim_I h$. The corresponding Bayesian update rule for the non-additive beliefs of a CEU decision maker is the so-called full Bayesian update rule which is given as follows (Eichberger, Grant, and Kelsey 2007):

$$\nu(A \mid I) = \frac{\nu(A \cap I)}{\nu(A \cap I) + 1 - \nu(A \cup \neg I)} \tag{65}$$

where $\nu(A \mid I)$ denotes the conditional capacity for event $A \in \mathcal{F}$ given information $I \in \mathcal{F}$.

Observation 2. *Suppose that* $\pi(I) > 0$. *Then an application of the full Bayesian update rule (65) to a prior neo-additive capacity (56) results in the posterior neo-additive capacity*

$$\nu(A \mid I) = \delta_I \cdot \lambda + (1 - \delta_I) \cdot \pi(A \mid I) \tag{66}$$

such that

$$\delta_I = \frac{\delta}{\delta + (1-\delta) \cdot \pi(I)} \tag{67}$$

and

$$\pi(A \mid I) = \frac{\pi(A \cap I)}{\pi(I)}. \tag{68}$$

Proof. Observe that

$$\begin{aligned}
\nu(A \mid I) &= \frac{\delta \cdot \lambda + (1-\delta) \cdot \pi(A \cap I)}{\delta \cdot \lambda + (1-\delta) \cdot \pi(A \cap I) + 1 - (\delta \cdot \lambda + (1-\delta) \cdot \pi(A \cup \neg I))} \\
&= \frac{\delta \cdot \lambda + (1-\delta) \cdot \pi(A \cap I)}{1 + (1-\delta) \cdot (\pi(A \cap I) - \pi(A \cup \neg I))} \\
&= \frac{\delta \cdot \lambda + (1-\delta) \cdot \pi(A \cap I)}{1 + (1-\delta) \cdot (\pi(A \cap I) - \pi(A) - \pi(\neg I) + \pi(A \cap \neg I))} \\
&= \frac{\delta \cdot \lambda + (1-\delta) \cdot \pi(A \cap I)}{1 + (1-\delta) \cdot (-\pi(\neg I))} \\
&= \frac{\delta \cdot \lambda + (1-\delta) \cdot \pi(A \cap I)}{\delta + (1-\delta) \cdot \pi(I)} \\
&= \frac{\delta \cdot \lambda}{\delta + (1-\delta) \cdot \pi(I)} + \frac{(1-\delta) \cdot \pi(I)}{\delta + (1-\delta) \cdot \pi(I)} \pi(A \mid I) \text{ for } \pi(I) > 0 \\
&= \delta_I \cdot \lambda + (1 - \delta_I) \cdot \pi(A \mid I)
\end{aligned}$$

with δ_I given by (67).□

Observe that we have, by observation 2, the following characterization of a conditional neo-additive capacity with respect to the full Bayesian update rule

$$\nu(A \mid I) \cdot (\delta + (1-\delta) \cdot \pi(I)) = \delta \cdot \lambda + (1-\delta) \cdot \pi(A \cap I). \tag{69}$$

Consider now the case that the information I is an event in which a continuously distributed random variable X with density function ψ takes on some value in the Euclidean Borel set B, i.e., $I = \{\omega \mid X(\omega) \in B\}$. Then $\pi(I) = \int_{x \in B} \psi(x)\, dx =$

0 whenever $I = \{\omega \mid X(\omega) = x\}$ so that (68) is not well-defined for "point" information. In that case—relevant to our framework—we can nevertheless use (69) for the characterization of the conditional neo-additive capacity $\nu(A \mid x)$ as follows: for all $I \in \mathcal{F}$, $\nu(A \mid x)$ has to satisfy

$$\int_{x \in B} \nu(A \mid x) \cdot (\delta + (1-\delta) \cdot \psi(x))\, dx = \delta \cdot \lambda + (1-\delta) \cdot \pi(A \cap I). \tag{70}$$

Obviously, one version of $\nu(A \mid x)$ is then given as

$$\nu(A \mid x) = \delta_x \cdot \lambda + (1-\delta_x) \cdot \pi(A \mid x) \tag{71}$$

where

$$\delta_x = \frac{\delta}{\delta + (1-\delta) \cdot \psi(x)} \tag{72}$$

and $\pi(A \mid x)$ satisfies

$$\int_{x \in B} \pi(A \mid x) \cdot \psi(x)\, dx = \pi(A \cap I). \tag{73}$$

By (66) with (67) and (68) we now possess a formal rule that describes a plausible way of how a Bayesian decision maker revises his ambiguous beliefs in the light of new information. Moreover, (71) with (72) and (73) characterizes this update rule whenever the information is drawn from a continuous probability distribution. In the next section we will apply this update rule to a model of Choquet Bayesian learning of the average consumption growth rate.++++

Appendix 2: Proofs of the propositions

Proof of proposition 1

Standard calculation (cf., e.g., Berger (1985, p.127)) shows that, for $\theta \in (\mu - a, \mu + a)$,

$$\begin{aligned}\psi\left(\theta, \bar{x}_n\right) &= \frac{1}{F\left(\mu + a\right) - F\left(\mu - a\right)} \qquad (74)\\ &\cdot \frac{1}{\sqrt{2\pi\rho_n^2}} \exp\left[-\frac{1}{2}\frac{\left(\theta - \mu_n^*\right)^2}{\rho_n^2}\right] \cdot \frac{1}{\sqrt{2\pi\left(\tau^2 + \sigma^2/n\right)}} \exp\left[-\frac{1}{2}\frac{\left(\bar{x}_n - \mu\right)^2}{\tau^2 + \sigma^2/n}\right]\end{aligned}$$

where

$$\mu_n^* = \frac{\sigma^2/n}{\sigma^2/n + \tau^2} \cdot \mu + \frac{\tau^2}{\sigma^2/n + \tau^2} \cdot \bar{x}_n \qquad (75)$$

and

$$\rho_n^2 = \frac{\tau^2\sigma^2/n}{\tau^2 + \sigma^2/n}. \qquad (76)$$

Consequently,

$$\begin{aligned}\psi_{\bar{X}_n}\left(\bar{x}_n\right) &= \int_{\mu - a_\varepsilon}^{\mu + a_\varepsilon} \psi\left(\theta, \bar{x}_n\right) d\theta \qquad (77)\\ &= \frac{1}{F\left(\mu + a\right) - F\left(\mu - a\right)} \cdot \\ &\quad \frac{1}{\sqrt{2\pi\left(\tau^2 + \sigma^2/n\right)}} \exp\left[-\frac{1}{2}\frac{\left(\bar{x}_n - \mu\right)^2}{\tau^2 + \sigma^2/n}\right] \cdot \int_{\mu - a_\varepsilon}^{\mu + a_\varepsilon} \frac{1}{\sqrt{2\pi\rho_n^2}} \exp\left[-\frac{1}{2}\frac{\left(\theta - \mu_n^*\right)^2}{\rho_n^2}\right] d\theta,\end{aligned}$$

so that, by an application of Bayes' rule,

$$\begin{aligned}\psi_{\tilde{\theta}}\left(\theta \mid \bar{x}_n\right) &= \frac{\psi\left(\theta, \bar{x}_n\right)}{\psi_{\bar{X}_n}\left(\bar{x}_n\right)} \qquad (78)\\ &= \frac{\frac{1}{\sqrt{2\pi\rho_n^2}} \exp\left[-\frac{1}{2}\frac{\left(\theta - \mu_n^*\right)^2}{\rho_n^2}\right]}{\int_{\mu - a}^{\mu + a} \frac{1}{\sqrt{2\pi\rho_n^2}} \exp\left[-\frac{1}{2}\frac{\left(\theta - \mu_n^*\right)^2}{\rho_n^2}\right] d\theta} \qquad (79)\\ &= \frac{1}{F_n\left(\mu + a\right) - F_n\left(\mu - a\right)} \frac{1}{\sqrt{2\pi\rho_n^2}} \exp\left[-\frac{1}{2}\frac{\left(\theta - \mu_n^*\right)^2}{\rho_n^2}\right], \qquad (80)\end{aligned}$$

by the definition of F_n. Observe now that the conditional density function $\psi_{\tilde{\theta}}\left(\theta \mid \bar{x}_n\right)$ corresponds to a truncated normal distribution of $\tilde{\theta}$ with support on $(\mu - a, \mu + a)$. Conditional on $\bar{x}_n$, $\tilde{\theta}$ has therefore the expected value

$$\begin{aligned}E\left[\tilde{\theta}, \psi_{\tilde{\theta}}\left(\theta \mid \bar{x}_n\right)\right] &= \int_{\mu - a}^{\mu + a} \theta \cdot \psi_{\tilde{\theta}}\left(\theta \mid \bar{x}_n\right) d\theta \qquad (81)\\ &= \mu_n^* - \rho_n \frac{f_n\left(\mu + a\right) - f_n\left(\mu - a\right)}{F_n\left(\mu + a\right) - F_n\left(\mu - a\right)}. \qquad (82)\end{aligned}$$

This proves part (i) of proposition 1.

In order to prove part (ii) observe that

$$\lim_{n\to\infty} \rho_n = \lim_{n\to\infty} \frac{\tau^2\sigma^2}{n\cdot\tau^2+\sigma^2} = 0 \tag{83}$$

as well as

$$\begin{aligned}
\lim_{n\to\infty} f_n(\mu+a) &= 0 && (84)\\
\lim_{n\to\infty} f_n(\mu-a) &= 0 && (85)\\
\lim_{n\to\infty} F_n(\mu+a) &= 1 && (86)\\
\lim_{n\to\infty} F_n(\mu-a) &= 0 && (87)
\end{aligned}$$

(in probability) whenever $\mu^* \in (\mu-a, \mu+a)$. The equations (84) - (87) follow thereby from the fact that all probability mass gets eventually concentrated at μ^* with probability one when n approaches infinity. Collecting equations (83) - (87) gives

$$\begin{aligned}
\lim_{n\to\infty} E\left[\tilde{\theta}, \psi_{\tilde{\theta}}(\theta \mid \bar{x}_n)\right] &= \lim_{n\to\infty} \mu_n^* - \lim_{n\to\infty} \rho_n \frac{f_n(\mu+a)-f_n(\mu-a)}{F_n(\mu+a)-F_n(\mu-a)} && (88)\\
&= \lim_{n\to\infty} \mu_n^* - 0, && (89)
\end{aligned}$$

which proves part (ii) of proposition 1.□

Proof of proposition 2

Part (i) follows from propositions 1 and 2 and (71) with (72) and (73).

Part (ii). Combining the characterization (77) of $\psi_{\bar{X}_n}$ with the definition (30) of F_n gives

$$\begin{aligned}
\psi_{\bar{X}_n}(\bar{x}_n) = &\frac{1}{F(\mu+a)-F(\mu-a)}\cdot\\
&\frac{1}{\sqrt{2\pi(\tau^2+\sigma^2/n)}}\exp\left[-\frac{1}{2}\frac{(\bar{x}_n-\mu)^2}{\tau^2+\sigma^2/n}\right]\cdot\left[F_n(\mu+a)-F_n(\mu-a)\right].
\end{aligned}$$

By (86) and (87), we have

$$\lim_{n\to\infty}\psi_{\bar{X}_n}(\bar{x}_n) = \frac{1}{F(\mu+a)-F(\mu-a)}\cdot\frac{1}{\sqrt{2\pi\tau^2}}\exp\left[-\frac{1}{2}\frac{(\bar{x}_n-\mu)^2}{\tau^2}\right]$$

whenever $\mu^* \in (\mu-a, \mu+a)$, whereby $\bar{x}_n$ coincides with probability one with μ^*. Consequently, $\psi_{\bar{X}_n}(\bar{x}_n)$ converges in probability to $\psi_{\tilde{\theta}}(\mu^*)$ when n gets large, which proves (46). Finally, observe that the additive estimator $E\left[\tilde{\theta}, \psi_{\tilde{\theta}}(\theta \mid \bar{x}_n)\right]$ converges in probability to μ^* whenever $\mu^* \in (\mu-a, \mu+a)$.□

References

Abel, A.B. (1990), "Asset Prices under Habit Formation and Catching Up with the Joneses", *American Economic Review Papers and Proceedings* **80**, 38-42.

Abel, A.B. (2002), "An Exploration of the Effects of Pessimism and Doubt on Asset Returns", *Journal of Economic Dynamics and Control* **26**, 1075-1092.

Adam, K., A. Marcet and J.P. Nicolini (2008), "Stock Market Volatility and Learning", ECB Working Paper **862**, European Central Bank.

Anderson, E.W., Sargent, T.J. and L.P. Hansen (2003), "A Quartet of Semi-groups for Model Specification, Robustness, Prices of Risk, and Model Detection", *Journal of the European Economic Association* **1**, 68-123.

Anscombe, F.J., and R.J. Aumann (1963), "A Definition of Subjective Probability", *Annals of American Statistics* **34**, 199-205.

Barsky, R.B. and J.B. DeLong (1993), "Why Does the Stock Market Fluctuate?", *Quarterly Journal of Economics* **108**, 291-311.

Berger, J.O. (1985), *Statistical Decision Theory and Bayesian Analysis*, Springer: Berlin.

Brav, A. and J.B. Heaton (2002), "Competing Theories of Financial Anomalies", *Review of Financial Studies* **15**, 575-606.

Breiman, D., LeCam, L. and L. Schwartz (1964), "Consistent Estimates and Zero-One Sets", *The Annals of Mathematical Statistics* **35**, 157-161.

Brennan, M.J. and Y. Xia (2001), "Stock Price Volatility and Equity Premium", *Journal of Monetary Economics* **47**, 249-283.

Campbell, J.Y. (2003), "Consumption Based Asset Pricing", in: G.M. Constantinides, M. Harris and R.M. Stulz, eds., *Handbook of the Economics and Finance*. Amsterdam: Elsevier, 808-887.

Campbell, J.Y. and J.Cochrane (1999), "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior ", *Journal of Political Economy* **107**, 205-251.

Cecchetti, S.G, Lam, P., and C.M. Nelson (2000), "Asset Pricing with Distorted Beliefs: Are Equity Returns Too Good to Be True?", *American Economic Review* **90**, 787-805.

Chateauneuf, A., Eichberger, J., and S. Grant (2007), "Choice under Uncertainty with the Best and Worst in Mind: Neo-additive Capacities", *Journal of Economic Theory* **127**, 538-567.

Cochrane, J.H. (2001), *Asset Pricing*, Princeton University Press: Princeton.

Constantinides, G.M. (1990), "Habit Formation: A Resolution of the Equity Premium Puzzle", *Journal of Political Economy* **98**, 519-543.

Cogley, T. and T.J. Sargent (2008), "The market price of risk and the equity premium: A legacy of the Great Depression?", *Journal of Monetary Economics* **55(3)**, 454-476.

Doob, J.L. (1949), "Application of the Theory of Martingales" in In Le Calcul des Probabilite´s et ses Applications, Colloques Internationaux du Centre National de la Recherche Scientifique **13**, 23-27.

Duffie, D. (2001), *Dynamic Asset Pricing Theory*, Princeton University Press: Princeton.

Eichberger, J., and D. Kelsey (1999), "E-Capacities and the Ellsberg Paradox", *Theory and Decision* **46**, 107-140.

Eichberger, J., Grant, S., and D. Kelsey (2006), "Updating Choquet Beliefs", *Journal of Mathematical Economics* **43**, 888-899.

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

Epstein, L.G. (1999), "A Definition of Uncertainty Aversion", *Review of Economic Studies* **66**, 579-608.

Epstein, L.G. and M. Schneider (2003), "Recursive Multiple-Priors", *Journal of Economic Theory* **113**, 1-31.

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

Epstein, L.G. and T. Wang (1994), “Intertemporal Asset Pricing Under Knightian Uncertainty”, *Econometrica* **62**, 283-322.

Epstein, L.G. and S. Zin (1989), “Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework”, *Econometrica* **57**, 937-968.

Epstein, L.G. and S. Zin (1991), “Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Investigation”, *Econometrica* **99**, 263-286.

Ghirardato, P., and M. Marinacci (2002), “Ambiguity Made Precise: A Comparative Foundation”, *Journal of Economic Theory* **102**, 251-289.

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

Gilboa, I. and D. Schmeidler (1989), “Maxmin Expected Utility with Non-Unique Priors”, *Journal of Mathematical Economics* **18**, 141-153.

Gilboa, I., and D. Schmeidler (1993), “Updating Ambiguous Beliefs”, *Journal of Economic Theory* **59**, 33-49.

Hansen, L.P. and T.J. Sargent (2001), “Robust Control and Model Uncertainty”, *American Economy Review Papers and Proceedings* **91**, 60-66.

Hansen, L.P. and T.J. Sargent (2007), “Robustness”, Princeton University Press: Princeton.

Hansen, L.P., Sargent, T.J., and T.D. Tallarini (1999), “Robust Permanent Income and Pricing ”, *The Review of Economic Studies* **66**, 873-907.

Hansen, L.P., Sargent, T.J., Turmuhambetova, G. and N. Williams (2006), “Robust Control and Model Misspecification”, *Journal of Economic Theory* **128**, 45-90.

Kocherlakota, N. (1996), “The Equity Premium: It’s Still a Puzzle”, *Jounral of Economic Literature* **34**, 42-71.

Lewellen, J. and J. Shanken (2000), “Learning, Asset-Pricing Tests, and Market Efficiency”, *Journal of Finance* **57**, 1113-1145.

Lijoi, A., Pruenster, I. and S.O. Walker (2004), "Extending Doob's Consistency Theorem to Nonparametric Densities", *Bernoulli*, **10**, 651-663.

Lucas, R.E., Jr. (1978), "Asset Prices in an Exchange Economy", *Econometrica* **46**, 1429-1446.

Maddison, A. (2007), *Contours of the World Economy, 1-2030 AD*, Oxford University Press: Oxford.

Maenhout, P.J. (2004), "Robust Portfolio Rules and Asset Pricing", *The Review of Financial Studies* **17**, 951-983.

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

Mehra, R., and E.C. Prescott (1985), "The Equity Premium: A Puzzle", *Journal of Monetary Economics* **15**, 145-161.

Mehra, R., and E.C. Prescott (2003), "The Equity Premium in Retrospect", in: G.M. Constantinides, M. Harris and R.M. Stulz, eds., *Handbook of the Economics and Finance.* Amsterdam: Elsevier, 808-887.

Puri, M. and D.T. Robinson (2007), "Optimism and Economic Choice", *Journal of Financial Economics* **86**, 71-99.

Sarin, R., and P.P. Wakker (1998), "Revealed Likelihood and Knightian Uncertainty", *Journal of Risk and Uncertainty* **16**, 223-250.

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

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**, 571-587.

Shiller, R.J., (1981), "Do stock prices move too much to be justified by subsequent changes in dividends?", *American Economic Review* **71**, 421-436.

Siniscalchi, M. (2001), "Bayesian Updating for General Maxmin Expected Utility Preferences", mimeo.

Siniscalchi, M. (2006), "Dynamic Choice under Ambiguity", mimeo.

Timmermann, A. (1993), "How Learning in Financial Markets Generates Excess Volatility and Predictability in Stock Prices", *Quarterly Journal of Economics* **108**, 1135-1145.

Wakker, P.P. (2001), "Testing and Characterizing Properties of Nonadditive Measures through Violations of the Sure-Thing Principle", *Econometrica* **69**, 1039-1059.

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

Weil, P. (1989), "The Equity Premium Puzzle and the Riskfree Rate Puzzle", *Journal of Monetary Economics* **24**, 401-421.

Weitzman, M. (2007), "Subjective Expectations and Asset-Return Puzzles", *American Economic Review*, **79(4)**, 1102-1130.

Zimper, A. (2009) "Half Empty, Half Full and Why We Can Agree to Disagree Forever", *Journal of Economic Behavior and Organization*, **71**, 283-299.

Zimper, A., and A. Ludwig (2009), "On Attitude Polarization Under Bayesian Learning With Non-Additive Beliefs", *Journal of Risk and Uncertainty* **39**, 181-212.

Discussion Paper Series

Mannheim Research Institute for the Economics of Aging, Universität Mannheim

To order copies, please direct your request to the author of the title in question.

Nr.	Autoren	Titel	Jahr
179-09	Alexander Ludwig, Edgar Vogel	Mortality, Fertility, Education and Capital Accumulation in a simple OLG Economy	09
180-09	Edgar Vogel	From Malthus to Modern Growth: Child Labor, Schooling and Human Capital	09
181-09	Steffen Reinhold, Hendrik Jürges	Secondary School Fees and the Causal Effect of Schooling on Health Behaviour	09
182-09	Steffen Reinhold, Kevin Thom	Temporary Migration and Skill Upgrading: Evidence from Mexican Migrants	09
183-09	Hendrik Jürges, Eberhard Kruk, Steffen Reinhold	The effect of compulsory schooling on health – evidence from biomarkers	09
184-09	Nicola Fuchs-Schündeln, Dirk Krüger, Mathias Sommer	Inequality Trends for Germany in the Last Two Decades: A Tale of Two Countries	09
185-09	Francesco Cinnirella, Joachim Winter	Size Matters! Body Height and Labor Market Discrimination: A Cross-European Analysis	09
186-09	Hendrik Jürges, Steffen Reinhold, Martin Salm	Does Schooling Affect Health Behavior? Evidence from Educational Expansion in Western Germany	09
187-09	Michael Ziegelmeyer	Das Altersvorsorge-Verhalten von Selbständigen – eine Analyse auf Basis der SAVE-Daten	09
188-09	Beatrice Scheubel, Daniel Schunk, Joachim Winter	Don't raise the retirement age! An experiment on opposition to pension reforms and East-West differences in Germany	09
189-09	Martin Gasche	Die sozialversicherungspflichtig Beschäftigten im deutschen Sozialversicherungssystem: Eigenschaften, Beitragsleistungen und Leistungsbezug	09
190-09	Martin Gasche	Implizite Besteuerung im deutschen Sozialversicherungssystem	09
191-09	Alexander Ludwig, Alexander Zimper	Biased Bayesian learning and the risk-free rate puzzle	09