Ignorance and Indifference: Decision-Making in the Lab and in the Market*

Sergiy Verstyuk†

This version: 8 October 2024

Abstract

Economic agents live in a dynamic, perpetually changing environment. The ensuing scarcity of information is tackled with invariant ignorance priors, forcing said agents to overweight the chances of less frequent events. Beliefs are updated from diverse information sources, each source assigned an appropriate degree of confidence, and with directly sampled information serving as a canonical metric. Empirically, these mechanisms allow to rationalize and reproduce (i) the prospect theory's probability transformations and the choices from experienced, described, or ambiguous probabilities documented earlier in laboratory experiments (while also accounting for the effect of dynamic belief updating), as well as (ii) the equity premium/risk-free rate levels and equity/dividend volatility scaling observed in financial markets.

Keywords: choice under risk and uncertainty, Allais, Ellsberg, parameterization invariance, Jeffreys prior, maximum entropy, non-ergodicity, changepoint detection, small sample, rare unseen events, shrinkage, belief updating

*Acknowledgements: Pietro Veronesi, George M. Constantinides, Doron Ravid, Matt Taddy; as well as Mohammed Abdellaoui, Jörn Boehnke, Eugene F. Fama, Richard Freeman, Christian Julliard, Leonid Kogan, Scott Duke Kominers, Olivier L'Haridon, Nicholas Polson, Xiao Qiao, Philip J. Reny, Lawrence D. W. Schmidt, Neil Shephard, Andrei Shleifer, Harald Uhlig. This work was supported by the Center of Mathematical Sciences and Applications at Harvard University.

†Contact address: verstyuk@cmsa.fas.harvard.edu.

Contents

1 Introduction

A classical problem originating in statistics and decision theory is accounting for unseen events. Pierre-Simon Laplace (1812) posed it as a question about the chances the sun will rise tomorrow given it always had for the past 5,000 years, and he formulated his rule of succession as a candidate solution to this sunrise problem. Alan Turing and I.J. Good (1953) encountered this issue while decoding the German Enigma machine cypher, later exemplifying it with the problem of estimating the number of yet undiscovered animal species or vocabulary words, thus relating it to computational linguistics more generally. Recent studies from the computer science and machine learning community tackle this problem more rigorously, dealing with optimal coverage adjustment and smoothing of the distributions, identifying their shapes and various statistics.[1] The somewhat older works of Keynes (1921), Jeffreys (1939, 1946) and, very prominently, Jaynes (1957a, 1957b) were concerned with fundamentally related problems.

The need to conduct inference about, to estimate and to account for unobserved classes of contingencies, the very fact of whose existence is unknown, also arises in economics. Consumers, investors, entrepreneurs and job-seekers do not live in a static world, but rather face an evolving stochastic environment that must be learned and responded to in real time. In elementary cases, people regularly encounter new tasks of choice under risk, often being forced to make a decision after only cursory experience (e.g., consider binary lotteries with limited information about the success probabilities). In more complex cases, overall economic regimes change, and what people had learned about the previous regimes is of little or no use (e.g., parameters driving the macro environment are perturbed or completely reset). Turning to jargon, in a non-ergodic[2] world economic agents are permanently struggling with "small samples" (at least in the sense of effective sample size).

This non-ergodic, substantially undersampled environment gives rise to permanent parameter and model uncertainty. The principle of indifference, broadly understood as indifference between equivalent states of knowledge, offers ways of dealing with such ignorance in a disciplined manner. In the context of a decision problem, it is usually implemented by the method of maximum entropy[3] and/or the method of invariance to reparameterization. The first method explicitly focuses on using as little extraneous information as possible about the objects (variables or parameters[4]) involved into analysis. The second requires the analysis to be unaffected by the equivalent transformations of

[1]For example, see Nemenman et al. (2002), Orlitsky et al. (2003), Paninski (2003), Hausser and Strimmer (2009), Valiant and Valiant (2017).

[2]Ergodic stochastic process has its time average equal to ensemble average (e.g., different ensembles may represent different realizations of random parameter initializations).

[3]Entropy of a random variable may be thought of as a measure of the variable's dispersion. This concept plays an important role in the information theory (see MacKay, 2003, for a good introduction).

[4]Note that in a hierarchical analysis like ours, what is treated as a variable upstream may become a fixed parameter downstream.

such objects. Within the Bayesian framework and for the basic univariate cases we consider, they both reduce to a simple mathematical formulation assigning very natural prior probability distributions (known as Jeffreys's approach). They also admit transparent analytical solutions (exploiting the so-called conjugacy).[5]

We apply this indifference-motivated approach to the canonical problems of decision-making under uncertainty, such as primitive binary lotteries as well as the consumption-investment choice. It is standard to assume that lotteries and investment opportunities are driven by parametric Bernoulli and Normal probability distributions, respectively. On top of that, our decision problems are combined with online learning of the unknown parameters. As a consequence, Bayesian updating with invariant ignorance (i.e., indifferent, Jeffreys) priors produces the posterior distributions that: (i) have, respectively, Beta and Student's t forms, and (ii) are biased, or "shrunk", toward quasi-uniformity. That is, the less probable parts of the above posteriors are amplified in comparison to original distributions. Moreover, in a permanently undersampled environment, the amplification effect of such non-informative priors never washes away. From the agents' perspective, these inferred posterior distributions are used in place of unobserved data-generating, or "population", distributions that are responsible for producing observed sample realizations.

This framework naturally delivers an interpretable measure of confidence in a given data-generating mechanism and source of uncertainty (as introduced by Ellsberg, 1961; Fox and Tversky, 1995; also see Abdellaoui et al., 2011a), which boils down to a scalar variable measured in terms of (pseudo-) observations. In the canonical case of observations on random variable draws, data is sampled directly as, say, lottery outcomes (this is tantamount to "experienced" probabilities in the literature). In another case that is common in laboratory settings, information about the probability distribution in question is communicated in the form of declared parameter values, an example is announcing the success probability of a binary lottery to participating agents (providing them with "described" probabilities). Lastly, details about a given probability distribution may remain completely unknown, or ambiguous, and the agents may be informed only about the possible combinations of events in the decision problem considered, such as the total number and possible colors of balls in the urn as well as the colors of favorable outcomes (which gives rise to "judged" probabilities). Furthermore, such calculus of confidence allows to assign a natural scale on the prior distributions—in terms of pseudo-observations—that determines the strength of shrinkage they exert on posterior distributions ultimately used in decision-making.

While the standard setting for discrete probability distributions is lotteries and urns with balls in the lab, the canonical setting for continuous densities is probably asset pricing in financial markets, in which case the non-ergodicity of different lotteries is replaced with

[5]See Jaynes (2003) for an extensive discussion; Kass and Wasserman (1996) provide an excellent review.

evolving macro-financial regimes. In the latter environment, the counterparts to decisions from experience and decisions under conditions of ambiguity are, respectively, decisions under current and under new regime.

On the empirical side, allowing for such regime changes and using the invariant ignorance prior updated with sufficient statistics capturing the historical properties of the data, we successfully replicate the "puzzles"[6] about the levels of the equity risk premium and risk-free rate as well as about the discrepancy between equity and dividend volatilities found in the post-World War II U.S. data, without resorting to implausible levels of risk aversion. Crucially, regime changes are identified on the basis of daily-frequency stock returns, which lets us capture shifts that occur relatively often but have modest magnitudes.

An equally parsimonious model with invariant ignorance prior is able to replicate the results of laboratory experiments with binary lotteries in ambiguity-, description- and experience-based scenarios[7], matching Kahneman and Tversky's (cumulative) prospect theory's "probability distortions" as well as addressing Allais's and Ellsberg's "paradoxes"[8] simultaneously (as long as structural differences in the corresponding sources are properly accounted for, more on which below). However, in contrast to our findings with investments on financial markets, taking the results of laboratory experiments and extracting their subjects' beliefs that fit the data best, we see that such revealed priors are somewhat different from the Jeffreys's invariant prescriptions: while being similarly ignorant, they are "stronger" in the sense of inducing more shrinkage to the center and rigidity to new information (the exact magnitudes differ depending on the experiment, but in all cases they are dominated by the communicated external information, and thus do not depart too far from the invariant alternative).[9]

Importantly, our data on observed draws of lottery outcomes allow to explicitly test a belief updating mechanism, and among the two equally-parameterized model formulations, the one that accounts for the dynamics of information arrival explains subjects' behavior more accurately than the one relying on sufficient statistics only. Also, benefiting from the mutually consistent treatment of different sources of uncertainty, we are able to compare attitudes to them: in the data, experienced and described probabilistic information receive quantitatively similar degrees of confidence, and, notably, so do the judged probabilities. In the latter case, subjects' behavior is consistent with them repeating the approach they take for lotteries under risk at each of the two stages of ambiguous lotteries recursively, and there is no need for an additional ambiguity-attitude parameter(s) in the

[6]Raised by Mehra and Prescott (1985), Weil (1989), Shiller (1981).

[7]For definitions and background, see Hertwig et al. (2004), Abdellaoui et al. (2011a, 2011b).

[8]See Allais (1953), Ellsberg (1961), Kahneman and Tversky (1979, 1992), Camerer and Ho (1994), Wu and Gonzalez (1996), Abdellaoui (2000), Bruhin et al. (2010), Fehr-Duda and Epper (2012), Chew et al. (2017), Halevy (2007).

[9]Stronger priors may be motivated by the desire for robustness to serial correlation in the data, errors in external or internal information processing.

objective function.

The main contributions of the paper can be summarized as follows. First, on the methodological front, we compile—from relatively standard ingredients—a tractable model of decision-making implementing a systematic Bayesian approach with theoretically motivated prior distributions, which quantifies the strength of prior beliefs and confidence in communicated information in terms of (pseudo-) observations, as well as recognizes the fundamental role of dynamics and allows for changes in the environment or the updating of beliefs. We apply this rational approach across the domains of laboratory experiments and public financial markets as well as under conditions of risk and ambiguity.

Second, turning to empirical results, for binary lotteries the proposed model not only fits data equally well if not better than an approach based on flexible probability weighting functions despite the latter's richer parameterization, but—more importantly—in a test of the parsimonious mechanism of belief-updating it shows a superior performance over an equivalent in complexity static alternative. Another important finding is that subjects of laboratory experiments exhibit a surprisingly harmonious behavior across different lottery mechanisms and communication scenarios, in particular their treatment of lotteries under conditions of ambiguity is not different from that of lotteries under risk provided only that we account for the higher-order, compounded nature of the former mechanism.

For investments, the model demonstrates how accounting for a changing environment with indifference priors followed by the learning process allows to fit key moments in macro-financial data under a plausible degree of risk aversion, with regime changes identified using statistical state-of-the-art changepoint detection methods.

When comparing the performance of the invariant ignorance priors across domains, we find that they characterize quite accurately the behavior of financial market participants, but the experimental subjects appear to hold ignorant priors that are stronger than what is prescribed by parameterization invariance, i.e., the subjects have relatively rigid beliefs that are less responsive to the updating information.

In addition to literature references included throughout the text, studies closely related to the present work include the following ones. Probability distortions demonstrated by human subjects in laboratory experiments have been in focus of recent literature that progressed from derivations based on plausible postulates and reduced-form principles (such as Zhang and Maloney, 2011) to proposing detailed mechanisms as in the studies on adjustment to cognitive noise (Steiner and Stewart, 2016; Enke and Graeber, 2019; Khaw, Li, Woodford, 2023, 2021; Frydman and Jin, 2022; Juechems et al., 2021; Netzer et al., 2022; Vieider, 2024). In contrast to these works, the current paper derives probability distortions relying on theoretically motivated indifference priors as a preemptive response to changing stochastic environment (which is also more convincing in application to the aggregate stock market).

A parallel theme in this literature is parameter/state learning and belief updating. Examples include an empirical study by Van de Kuilen and Wakker (2006) as well as a recent methodological contribution by Augenblick and Rabin (2021).[10] A closely related paper is Aydogan (2021), who postulates a similar belief updating rule as a descriptive hypothesis without deriving it from first principles, but then feeds its result into standard prospect theory's probability weighting and value functions modeling the special case of binary choice between ambiguous lotteries; he does not consider a Jeffreys indifferent prior. Moreover, this paper focuses on a fairly minimalist rational Bayesian updating that treats different communication scenarios in a mutually consistent manner.

We exploit the connection between the ideas or problems from (micro) economic theory on the one side and (macro) finance on the other. Within the latter, one relevant line of research is macro-financial regime changes. The closest recent example is Smith and Timmerman (2021), with our differences lying in the identification method used (medium-frequency cross-sectional information there versus higher-frequency time-series information here), and as a result in the number of detections (higher, in our case).[11] In asset pricing this paper is related to the literature on the effects of rare events, both those objectively happening (e.g., see Barro, 2009, or Tsai and Wachter, 2016) and subjectively expected (Weitzman, 2007). The most relevant previous work here is Weitzman (2007), who proposes a Bayesian view on the risks from unobserved rare events ("peso problem"); our paper's novelty is in the more tangible nature of the considered rare events as well as in a higher analytical tractability of the theoretical model. It is also worth mentioning Ghaderi et al. (2022), who calibrate a macro-financial model featuring rare events that fits many data moments without requiring the realization of implausibly large instantaneous shocks, similarly to this paper; but it focuses on negative disaster shocks only, and employs a very different driving mechanism with many more free parameters.

Dynamic aspects take an important role in the macro-finance domain, hence model, parameter and state learning are ubiquitous features there. Besides above-mentioned Weitzman (2007) and Ghaderi et al. (2022), recent examples include Johannes et al. (2016) and Collin-Dufresne et al. (2016) in finance, Kozlowski et al. (2020) and Farmer et al. (2024) in macroeconomics.[12] Broadly speaking, the distinctiveness of our formulation is the consistency with the micro-level decision problem, whereby the risks of regime changes are tackled by adopting the Jeffreys' indifferent ignorance prior.[13]

In addition to the above-mentioned rational approaches, within macro-finance there is also an influential literature that is particularly focused on (arguably, more realistic) microeconomic premises by relying on the boundedly rational, psychological and other

[10] An earlier relevant literature includes, among others, Grether (1980) as well as Hogarth and Einhorn (1992).

[11] Also see Farmer et al. (2023) and Borup et al. (2024), who consider a related question, the time-varying predictability of stock and bond returns.

[12] But this literature goes back even further to Timmermann (1993) and Veronesi (1999).

[13] Earlier works recognizing the limitations of the available sample of macro-financial statistics and/or the evolving nature of the underlying economic parameters include McGrattan and Prescott (2003 and 2005), Dimson et al. (2003), Fama and French (2002).

"behavioral" arguments. For instance, investors' belief updating has been modeled using Kahneman and Tversky's representativeness heuristic, with application to explaining the dynamics of bond (Bordalo et al., 2018) and stock markets (Bordalo et al., 2019).

Another popular strategy in this literature segment is explicitly endowing agents in financial markets with prospect theory-style probability weighting functions. Both the equity market (Barberis and Huang, 2008; De Giorgi and Legg, 2012; Barberis et al., 2021) as well as the options market (Polkovnichenko and Zhao, 2013; Baele et al., 2019) have been fruitfully studied through this lens.[14] In contrast to such studies, this paper maintains the basic premises of rationality (and standard preferences). For example in the case of tail probabilities' amplification, we derive it using Bayesian principles, rather than adopt the probability weighting stipulations from the prospect theory taken as an approximate description of human behavior.

2 Motivation

Laboratory experiments with risky lotteries reveal a number of biases and distortions relative to the standard expected utility benchmark. Accounting for these distortions helps understand the prices of capital assets that are traded on financial markets.

2.1 Probability distortions in lotteries

Consider prospect theory (Kahneman and Tversky, 1979) along with its refinement, the cumulative prospect theory (Tversky and Kahneman, 1992).[15] It applies to choice under risk and captures vast experimental evidence collected on such choices, offering robust empirical predictions. In particular, it is consistent with Allais's (1953) experimental results that challenge the conventional expected utility theory of choice. As the authors themselves emphasize, the cumulative prospect theory (CPT) is not normative and prescriptive, but rather positive and descriptive.

CPT stipulates that (after the initial framing stage) the overall value of an uncertain "prospect" is the sum of values of the outcomes $v(\cdot)$ each multiplied by the decision weight $\pi(\cdot)$. The outcome is defined with respect to a reference point, and the value function potentially differs for positive ("concave for gains") and negative ("convex and steeper for losses") deviations from the reference point. The decision weight for some outcome is a local increment of the probability weighting function $W(\cdot)$, which in turn is defined as a non-linear transformation of the right (left in the case of losses) tail of the original probability distribution $P(\cdot)$ at a given point. The probability weighting function is supposed to overweight the small probabilities in the tails and underweight

[14] More generally, non-expected-utility models have appealing features for understanding macroeconomics and finance, with Ai and Bansal (2018) as well as Beason and Scherindorfer (2022) being among the recent proponents.

[15] For a recent analysis, tests and interpretation, see Peterson et al. (2021) or Gonzalez and Wu (2022).

the moderate and large probabilities in the center of the original distribution (in this regard, the term "distortion" is also commonly used).[16]

We adopt the functional form of the CPT's probability weighting function due to Goldstein and Einhorn (1987),

$$W(P) := \frac{\psi P^\eta}{\psi P^\eta + (1-P)^\eta}, \tag{1}$$

where $\eta \geq 0$ is a parameter largely governing the slope of the weighting function,and $\psi \geq 0$ is a parameter governing the function's elevation, with $\eta = \psi = 1$ corresponding to the standard linear probability weighting.

As another necessary ingredient, we allow for a power form of the value function,

$$v(x) := \mathrm{sign}(x) \times |x|^\varphi, \tag{2}$$

where $x \in \mathbb{R}$ is lottery payoff and $\varphi \geq 0$ is a parameter.

To estimate the parameters, we turn to the work of Bruhin et al. (2010) and use their data on a large number of laboratory experiments with primitive stochastic lotteries conducted in Switzerland and China. The data set includes certainty equivalents for binary lotteries over real monetary gains or losses (mixed lotteries were not considered), comprising 17800 observations in total.

A comprehensive modeling of decision-making involving lotteries and estimation results for the above data set are presented later in sections §3.2 and §4.1. But at this point we are only interested in the measurements of η and ψ parameters. Taking a symmetric approach to gains and losses, their estimates of 0.41 and 0.98, correspondingly, reveal a notable deviation from linearity (as illustrated by the resulting weighting function in Figure 1).[17]

2.2 Investment puzzles

We now verify that CPT-style probability weighting helps rationalizing the prices and outcomes observed on financial markets. In particular, we focus on the magnitudes of equity risk premium and risk-free rate. The standard consumption-based asset pricing equation (see Lucas, 1978) is

$$1 \equiv \mathrm{E}_t^\pi \left[M_{t+1} R_{t+1} \right] := \mathrm{E}_t^\pi \left[\beta \frac{u'(C_{t+1})}{u'(C_t)} R_{t+1} \right], \tag{3}$$

[16]Note that in the case of binary lotteries CPT coincides with the earlier prospect theory, which makes the decision weight-related calculations involving tails and cumulative distribution functions trivial.

[17]The estimation is conducted on pooled data by maximum likelihood accounting for error heteroscedasticity proportional to lottery range. With the estimate of parameter φ at 1.02, the value function is close to risk-neutrality (hence it is sensible to transfer the resulting probability weighting function to asset pricing setting, which we do next). The subsequent results in §2.2 are very similar if we use instead the CPT estimates from §4.1.

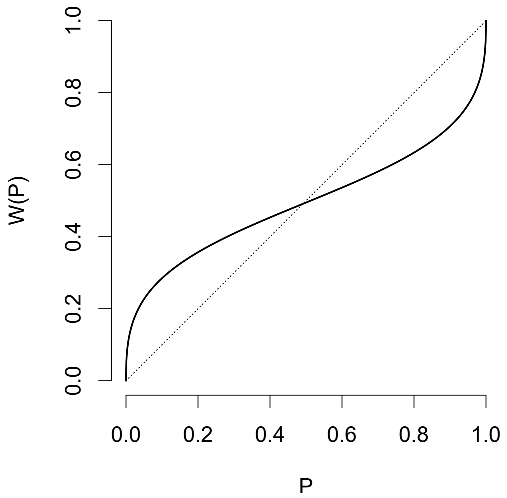


Figure 1: Cumulative prospect theory's probability weighting function, probability weighting (vert. axis) vs. cumulative distribution (horiz.) functions (parameterization used is $W(P)$ from equation (1) with $\hat{\eta} = 0.41$ and $\hat{\psi} = 0.98$).

where $\pi(\cdot)$ is the relevant beliefs-induced probability density function; M_{t+1} is a useful model-induced object called stochastic discount factor (SDF) or pricing kernel, which is defined as the marginal rate of substitution between consumption at time periods t and $t+1$, with β being a subjective time discount rate; $R_{t+1} := (S_{t+1} + D_{t+1})/S_t$ is asset gross return at time $t+1$, with S_{t+1} and D_{t+1} being the corresponding stock price and dividend payment; and C_s is consumption level at time s (and where consumption always equals dividends, $C_s \equiv D_s$ at each time period s, due to the general equilibrium restriction).[18] Intuitively, M_{t+1} discounts the future random payoff R_{t+1} depending both on time preference and state of the world between today and tomorrow, and in expectation the corresponding asset should be valued at 1 (due to normalization). In this context, a commonly used utility function is of the constant relative risk aversion (CRRA) form, $u(C_t) := C_t^{1-\gamma}/(1-\gamma)$, with γ being a RRA coefficient. Then, equation (3) implies the value for the risk-free rate $R_{f,t}$ equal to

$$R_{f,t} := \frac{1}{\mathrm{E}_t^{\pi}[M_{t+1}]}, \tag{4}$$

and for the risk premium on the market return R_{t+1} equal to

$$\mathrm{E}_t^{\pi}[R_{t+1}] - R_{f,t} := -\frac{\mathrm{Cov}_t^{\pi}[R_{t+1}, M_{t+1}]}{\mathrm{E}_t^{\pi}[M_{t+1}]}. \tag{5}$$

Our data set includes several macro-financial time series: stock and government bill returns as well as aggregate consumption and dividends for the United States after World War II.[19] The top row of Table 1 provides selected sample statistics.

[18] Henceforth, we often use the lowercase symbol to denote a natural logarithmic transformation of its uppercase counterpart, for example, $r_t := \ln R_t$.

[19] We abstract away from labor income, although empirically it constitutes a large share of the national income.

Table 1: Calibration Results for Investment Decisions — a Puzzle and a Hint

$\pi(r)$	γ	$\mathrm{E}^{\pi}[r_f]$	$\sqrt{\mathrm{V}^{\pi}[r_f]}$	$\mathrm{E}^{\pi}[r-r_f]$	$\sqrt{\mathrm{V}^{\pi}[r]}$	$\mathrm{E}^{\pi}[\Delta c]$	$\sqrt{\mathrm{V}^{\pi}[\Delta c]}$	$\mathrm{E}^{\pi}[\Delta d]$	$\sqrt{\mathrm{V}^{\pi}[\Delta d]}$
Empirical, $\hat{\pi}$	n/a	0.0066	0.0131	0.0674	0.1628	0.0129	0.0154	0.0277	0.0413
Model, $\mathcal{N}$	1	0.0607	n/a	0.0133	0.1628	0.0640	0.1628	0.0640	0.1628
Model, $\mathcal{N}$	2	0.0607	n/a	0.0133	0.1628	0.0320	0.0815	0.0320	0.0815
Model, $\mathcal{N}$	3	0.0607	n/a	0.0133	0.1628	0.0213	0.0544	0.0213	0.0544
Model, $\mathcal{N}$	4	0.0607	n/a	0.0133	0.1628	0.0160	0.0408	0.0160	0.0408
Model, $\mathcal{N}$	5	0.0607	n/a	0.0133	0.1628	0.0128	0.0326	0.0128	0.0326
Model, CPT_2	1	0.0226	n/a	0.0514	0.3208	0.0640	0.3208	0.0640	0.3208
Model, CPT_2	2	0.0225	n/a	0.0515	0.3208	0.0320	0.1606	0.0320	0.1606
Model, CPT_2	3	0.0225	n/a	0.0515	0.3208	0.0213	0.1071	0.0213	0.1071
Model, CPT_2	4	0.0224	n/a	0.0516	0.3208	0.0160	0.0803	0.0160	0.0803
Model, CPT_2	5	0.0224	n/a	0.0516	0.3208	0.0128	0.0643	0.0128	0.0643

Notes: The columns present the probability density of returns; the coefficient of relative risk aversion; the log risk-free and excess log market returns, the per capita consumption growth (i.e., temporal difference in the log of) as well as the dividend growth (difference in the log of) with their statistical moments. The "Empirical" row: the data sample is U.S. 1947:Q2–2019:Q4, at a quarterly frequency (see Appendix §G for a description of data sources). The "Model" rows: $\beta = 0.99$ per annum; the CRRA utility with a given γ; and a specified probability density of returns $\pi(r)$, which here is either Normal or based on cumulative prospect theory with 2 weighting function parameters (see text for a detailed description). Model inputs are hypothetical $\pi^{CPT}(r)$, γ and β; the rest of the columns are model outputs. The economic variables are measured in real terms, and measurements are converted into annualized values.

We feed the model with a candidate probability density of stock returns $\pi(r_t)$, which investors may possibly be using, and an assumed coefficient of relative risk aversion γ, which parameterizes investors' attitude toward risk, as well as the time discount rate fixed at the value $\beta = 0.99$ per annum. Taking these as given, we use equations (3)–(5) within the Lucas's endowment economy setup to solve for the key statistical moments of the risk-free rate, equity risk premium, as well as the consumption and dividend growth. The target of our exercise is for these key moments, particularly the means of the risk-free rate and risk premium, to match their empirical counterparts from the top row of Table 1. The subsequent rows of Table 1 contain the results.

First, we feed the empirical distribution of log-returns, representing it by the Normal density with mean and variance observed in the sample, $r_t \sim \mathcal{N}(\hat{\mu}, \hat{\sigma})$; see the rows corresponding to "Model, $\mathcal{N}$" in Table 1. The computed solution, however, is not consistent with the observed data. Most striking are the low values obtained for the risk premium $\mathrm{E}^{\pi}[r - r_f]$ and the high values for the risk-free rate r_f under realistic choices of parameter γ, classic facts known, respectively, as the "equity premium puzzle" (Mehra and Prescott, 1985) and the "risk-free rate puzzle" (Weil, 1989).

Now, assuming that investors in financial markets behave similarly to the way human subjects behave in laboratory experiments, we apply the probability weighting function $W(\Pi)$ estimated in §2.1 to the empirical distribution of log-returns $\mathcal{N}(\hat{\mu}, \hat{\sigma})$ and use this distribution instead;[20] see the rows "Model, CPT_2" in Table 1, also see Figure 2 for an illustration of the empirical and the distorted distributions. Importantly, in comparison to the Normal distribution considered earlier, a distribution distorted according to CPT produces a much larger equity premium—e.g., for $\gamma = 4$, we have 5.16% vs. 1.33%, respectively—so that the bulk of the puzzle vanishes. (It is also interesting to note that an order-of-magnitude disparity between empirical volatility of returns and that of dividend growth, posed as the "equity volatility puzzle" (Shiller, 1981), reappears here mechanically and hence also does not contradict the CPT model's restrictions.)[21]

Lastly, CPT-style probability distortions amplifying the tails of empirical distribution of stock returns are also able to address other challenging facts in finance. One such example is the so-called implied risk aversion "smile" pattern and the "non-monotone pricing kernel puzzle" observed in the options segment of financial markets (as shown by Polkovnichenko and Zhao, 2013, as well as Baele et al., 2019).

[20] Mean $\hat{\mu}$ is used as a reference point with respect to which gains or losses and respective probability quantiles are measured.

[21] There are many alternative explanations of these regularities in the existing literature. Broadly speaking, they focus either on investors' preferences, fundamentals of the environment, or investors' beliefs. Among the standard rational approaches, the first class is exemplified by the habit formation model due to Campbell and Cochrane (1999), the second by the long-run risk model of Bansal and Yaron (2004) or rare disaster risk framework from Rietz (1988), Veronesi (2004), Barro (2009), Wachter (2013). The non-standard approaches are represented by, e.g., Epstein and Zin (1990) in the first, Barberis et al. (2001) in the second, and De Giorgi and Legg (2012) or Barberis et al. (2015) in the third class.

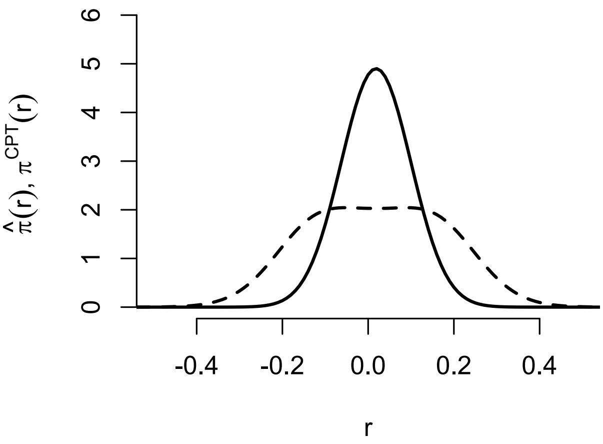


Figure 2: Empirical (solid) vs. hypothetical CPT-based (dashed) probability densities (parameterizations used are $\mathcal{N}(\hat{\mu}, \hat{\sigma}^2)$ with $\hat{\mu} = 0.0185$ and $\hat{\sigma}^2 = 0.0066$ vs. $\text{CPT}(\hat{\eta}, \hat{\psi}; \hat{\mu}, \hat{\sigma}^2)$ with $\hat{\eta} = 0.41$ and $\hat{\psi} = 0.98$ as well as $\hat{\mu}$ and $\hat{\sigma}^2$ as above).

2.3 Interpretation

Why would the CPT weighting function measured in laboratory experiments have any relationship to asset prices formed on competitive financial markets? A promising interpretation is to view the CPT not as a preference theory but as a theory of "default actions" that are optimal only in a constrained, second-best, approximate sense (e.g., see Bossaerts et al., 2008).[22] Next we will argue that such probability distortions are indeed rational responses to a limited information, low confidence, and extensive uncertainty, which can be best understood from a dynamic perspective.

3 Theory

3.1 Conceptual approach

Generally speaking, the problem we are dealing with is about learning, inference and decision in a situation of scarce data. For instance, small samples, rare events and rich stochastic dynamics all present a challenge of this kind.

A necessary step to resolving this problem is to specify what we know and what we do not know, and then to express our ignorance in a precise way. The "principle of indifference" offers a disciplined approach to tackling such ignorance, stipulating that equivalent states of knowledge should be assigned equivalent probabilities. In practice, this principle is usually implemented through the method of invariance or transformation groups (Jeffreys, 1939; 1946) or the method of maximum entropy (Jaynes, 1957a; 1957b).[23]

[22]This is broadly related to the concepts of "ecological rationality" (Simon, 1955, 1956; Todd and Gigerenzer, 2012) and "shortcut heuristics" (Gigerenzer and Goldstein, 1996; Gigerenzer et al., 2011).

[23]A unifying understanding of these two methods based on the "minimum description length principle"

More concretely, taking a Bayesian perspective on probability theory, this approach boils down to specifying carefully a non-informative prior probability distribution that possesses the desired properties. Ultimately, combining the information contained in the prior distribution with the information provided by sample data (if any) defines a stochastic environment that is relevant for a particular decision choice.[24]

A few technical points deserve elaboration. First, the prior distributions we use are characterized by invariance to transformations. The method of transformation groups ensures that the specification of a given problem (in which case specification also includes any prior probabilistic information we might have) is unaffected by equivalent transformations applied to it. In a one-dimensional case, as here, this approach coincides with the Jeffreys method of assigning prior distributions that would be invariant to reparameterization. For instance, it delivers translation invariance for a location parameter (mean in the case of Gaussian distribution) and measurement-units invariance for a scale parameter (variance, respectively).

Second, these prior distributions are characterized by maximal entropy within their corresponding classes of probability distributions. The method of maximum entropy ensures that the prior probability distributions considered reflect no more and no less than the information at hand. This is achieved by maximizing a certain measure of uncertainty (i.e., our lack of knowledge) taking as given the available information, explicitly specified.[25]

Third, the ignorance prior distributions we thus obtain turn out to be conjugate priors, which means that their posterior updates are preserving the distributional form (a concise introduction to conjugate priors is available in DeGroot, 1970).[26]

Formally, we consider an agent solving the following problem:

$$\max_{\boldsymbol{z}_t \in \boldsymbol{B}_t} \int J(\chi_{t+1}, \boldsymbol{z}_t) \pi(\chi_{t+1}) \, \mathrm{d}\chi_{t+1}, \tag{6}$$

where $J(\cdot)$ is a problem-specific objective function; χ_{t+1} is a stochastic payoff at period $t+1$; $\boldsymbol{z}_t$ is a control vector chosen in period t; $\boldsymbol{B}_t$ is a budget set of feasible choices in period t; $\pi(\chi_{t+1})$ is a distribution of the payoff χ_{t+1} constructed as $\pi(\chi_{t+1}) \propto \int_\Theta L(\chi_{t+1}|\boldsymbol{\theta}_t)\pi(\boldsymbol{\theta}_t)\mathrm{d}\boldsymbol{\theta}_t$ with a problem-specific probability distribution of payoffs $L(\chi_{t+1}|\boldsymbol{\theta}_t)$,

has been also proposed (Rissanen, 1983; also see Grünwald, 2007).

[24]At present, both of these methods lack definitive scientific consensus about how to faithfully implement them, which necessitates a certain degree of personal judgment and choice, and can be legitimately criticized for allowing too much flexibility or even arbitrariness. We will highlight the specific instances of this going forward.

[25]Or, equivalently, by minimizing the amount of "residual" information.

[26]This is due to the fact that we are using likelihoods from the exponential family of probability distributions. Those admit conjugate prior distributions that also belong to the exponential family. In turn, the exponential family distributions arise naturally as a result of entropy maximization. Moreover, for the distributions in the exponential family, the Jeffreys priors are optimal in the "minimum description length" sense.

its parameter vector $\boldsymbol{\theta}_t \in \Theta$ and an uninformative prior on parameter values $\pi(\boldsymbol{\theta}_t)$ chosen by the Jeffreys method[27] (we will also consider revealed priors that fit the agents' behavior best, even if they deviate from the Jeffreys approach). The "predictive" distribution of payoffs $\pi(\chi_{t+1})$ combines all information available to a decision-maker (the data-generating distribution, the prior beliefs about its parameters as well as the data realizations observed so far) and in the end is a function of prior's own parameters and relevant updating information.

3.2 Lotteries in laboratory experiments

Consider a simple lottery such that its outcome is represented by a binary random variable $x := x(\iota) \in \{x(0), x(1)\}$ with a probability of successful outcome $x(1)$ being equal to $p \in [0, 1]$. The corresponding Bernoulli probability mass function defines the likelihood of possible outcomes:[28]

$$L(\iota|p) := p^{\iota}(1-p)^{1-\iota}. \tag{7}$$

The postulated decision-making algorithm consists of three steps:

I. Learning from meta-data, potentially followed by "decision under ambiguity";

II. Learning from non-sample data, potentially followed by "decision from description";

III. Learning from sample data, potentially followed by "decision from experience".

We will expand on the details of this process next.

I. As a first step, before facing the lottery mechanism,[29] hence being aware of the primitive structure of the problem (such as how data underlying the lottery draws are generated) but having no information about the specific characteristics of the lottery (which in this example are captured exhaustively by the value of parameter p), the agent formulates a prior distribution that accurately reflects his/her lack of knowledge about the parameter (but does reflect awareness about the algebra of possible events or that $p \in [0, 1]$). The Jeffreys method prescribes a prior proportional to the square root of the

[27]Equivalently, this prior can be viewed as chosen to maximize entropy $\mathcal{E}(\pi(\boldsymbol{\theta}_t))$ subject to constraints on the moments of distribution $\pi(\boldsymbol{\theta}_t)$. Here, the Shannon entropy functional is defined as $\mathcal{E}(\pi(\boldsymbol{\theta})) := -\int \pi(\boldsymbol{\theta}) \ln(\pi(\boldsymbol{\theta}))\, \mathrm{d}\boldsymbol{\theta}$. The moments are chosen in a way that (i) their form ensures conjugacy and (ii) their targeted values lead to the Jeffreys solution. Required moment constraints will be given in the relevant sections.

[28]That is, in equation (6) we set $J(\chi_{t+1}, \boldsymbol{z}_t) := v(\chi_{t+1})$ if $z_t = \mathsf{lottery}$, otherwise $J(\chi_{t+1}, \boldsymbol{z}_t) := v(x_c)$, $x_c \in \mathbb{R}_+$; $\chi_{t+1} := x(\iota)$ with $\iota \in \{0, 1\}$; $\boldsymbol{z}_t \in \{\mathsf{certain}, \mathsf{lottery}\}$; $\boldsymbol{\theta}_t := p$. Since this is a one-period problem, we drop time subscripts; also, integration is replaced with summation.

[29]Say, after reading the general instructions but before taking a seat at the computer desk in the experimental laboratory.

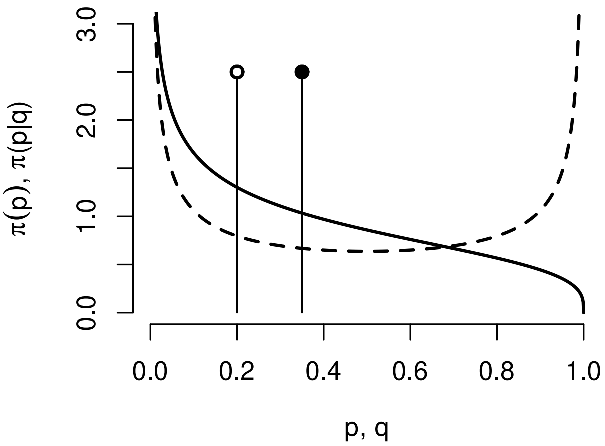


Figure 3: Posterior (solid) vs. prior (dashed) probability densities (parameterizations used are as follows: prior $\pi(p)$ is $\mathcal{B}(\alpha,\beta)$ with $\alpha = 1/2$ and $\beta = 1/2$, posterior $\pi(p|q)$ is $\mathcal{B}(\alpha,\beta)$ with $\alpha = 1/2 + n \times q$ and $\beta = 1/2 + n \times (1-q)$, declared value q = 0.20 (spike with empty bullet on top), parameter $n = 1$, posterior mean $\mathrm{E}[p|q] = 0.35$ (spike with filled bullet on top)).

determinant of the Fisher information matrix[30], which in our case yields:

$$\pi(p) \propto \sqrt{\mathrm{E}^L\left[\left(\frac{\partial}{\partial p}\ln L(\iota|p)\right)^2\right]} = \frac{1}{p^{1/2}(1-p)^{1/2}}. \tag{8}$$

This constitutes the kernel of a Beta probability distribution $\mathcal{B}(\alpha,\beta)$ with parameters $\alpha = 1/2$ and $\beta = 1/2$.[31] The corresponding symmetric U-shaped distribution (as plotted in Figure 3) has most of its probability mass concentrated in the poles on each end of the $[0,1]$ interval. Parameterization invariance implies that, rather than p itself, a transformed parameter $\arcsin(p^{1/2})$ has a uniform distribution on the $[0,\pi/2]$ interval.[32]

Thus, we obtained the Beta–Bernoulli conjugate system. Intuitively, using the insight that any conjugate prior can be viewed as the posterior for a pseudo data set, as well as a standard interpretation of the Beta distribution's parameters α and β as pseudo-observations or pseudo-counts of Bernoulli trials, values $\alpha = 1/2$ and $\beta = 1/2$ represent 1 pseudo-observation equally split between a success and a failure.[33] The number of pseudo-observations in a prior may be thought of as prior's strength, for the effect it produces on posterior distribution.

[30]This ensures the parameterization invariance rather mechanically by preempting the appearance of terms that arise from reparameterization using the standard change of variables procedure.

[31]In this paper $\mathcal{B}(\alpha,\beta)$, with shapes α and β, is parameterized as $f(x|\alpha,\beta) := \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$. Here, $\mathrm{B}(\cdot,\cdot)$ denotes the Beta function, and $\Gamma(\cdot)$ denotes the Gamma function.

[32]This is different from the naïve non-informative prior that implies p uniformly distributed on the real $[0,1]$ interval and that is obtained from the so-called Bayes–Laplace prior $\mathcal{B}(1,1)$.

[33]The naïve Beta–Laplace prior represents 1 successful and 1 unsuccessful pseudo-counts.

Turning to the maximum entropy approach, the Beta distribution maximizes entropy on the $[0, 1]$ interval subject to two symmetric constraints on the logarithm of the geometric mean of the random variable p. Thus, the previous values of α and β may be interpreted alternatively as a specific restriction in such entropy maximization problem.[34]

II. As a second step, now facing the lottery mechanism,[35] the agent learns something about the gamble: he is told that the probability of success is q.

Note that there are no sample data yet as the agent still has not experienced any lottery trials. He only has some meta- and non-sample information (possibly vague or non-credible) that can be used to modify the prior distribution—for example, to bias it one way or another, to make it sharper or more diffuse, etc. This is in the spirit of "no-data problem" following Chernoff and Moses (1959). In such cases, the prior to a large extent determines the conclusion (inference or decision).

Then the agent updates the initial prior with the sufficient statistic that corresponds to the communicated statement that the probability of success p is equal to the declared value q; and does so using the standard mechanism of Bayesian updating, that is through adding new (pseudo-) observations. For the Bernoulli distribution, the sufficient statistic is the total number of successes. So, for the announced probability q, the value of such statistic is going to be q multiplied by the number of pseudo-observations $n \in \mathbb{R}_+$ (for actual observations it would be $n \in \mathbb{N}$).

Consequently, the posterior distribution (the updated prior, really) is

$$\pi(p|q) = \frac{p^{\alpha+nq-1}(1-p)^{\beta+n(1-q)-1}}{\mathrm{B}(\alpha+nq, \beta+n(1-q))}, \tag{9}$$

which, due to conjugacy, is also of the Beta distributional form.

If the agent has full confidence in the lottery mechanism (i.e., in the lottery organizers, or in the device itself), then the number of pseudo-observations n tends to infinity, the prior is washed out, and the posterior distribution becomes concentrated at the declared value q. This is the standard case considered in textbooks. Thus, the sharpness of the updated prior depends on n, which effectively reflects the relative degree of confidence in the lottery mechanism.

However, there is no a priori rationale for ruling out the less-than-full confidence case of $n < \infty$. Indeed, taking this non-dogmatic perspective, n can be understood as degree of confidence in (or credibility of) communicated probabilities and distributions (Ellsberg, 1961) as well as the attitude towards a source of uncertainty (Fox and Tversky, 1995; Abdellaoui et al., 2011a).

[34] Geometric mean of a positive random variable p can be defined as $\exp(\mathrm{E}[\ln(p)])$. The two constraints are $\mathrm{E}^{\pi}[\ln(p)] \equiv \psi(\alpha)-\psi(\alpha+\beta)$ and $\mathrm{E}^{\pi}[\ln(1-p)] \equiv \psi(\beta)-\psi(\alpha+\beta)$, where $\psi(\cdot)$ is the digamma function. Seeming like an instance of those flexible choices mentioned earlier, the restriction is $\exp(\psi(\alpha) - \psi(\alpha + \beta)) \equiv 1/4$ for the probability of success p and $\exp(\psi(\beta) - \psi(\alpha+\beta)) \equiv 1/4$ for the probability of failure $(1-p)$; means do not have to add up to 1. (In the Bayes–Laplace case, the corresponding numerical values are $1/e$ and $1/e$.)

[35] Say, after taking a seat at the computer desk in the lab and seeing the characteristics of the lottery.

Then, if there is no reason to weight one type of prior information higher than the other, by the principle of indifference between sources of pseudo-observations, $n := 1$ is the value that puts two sources on equal footing and treats them equivalently.[36] Intuitively, such an approach can be interpreted as extending the existing 1 pseudo-observation from the initial prior (itself composed of α pseudo-counts of successes and β pseudo-counts of failures) with 1 additional pseudo-observation from the external source (composed of q pseudo-counts of successes and $(1-q)$ pseudo-counts of failures). (Later we will consider cases when n is known or can be estimated, sparing us from the above assumption.)

As a consequence, we obtain the following result, formulated in terms of expected probability $\mathrm{E}[p|q]$.[37]

Proposition 1 (Overweighting Low Probabilities and Underweighting High Probabilities, Discrete Case)**.** *The posterior distribution of* p*, defined by equation (9) with* $0 < n < \infty$*, is not centered at the declared value* q *but instead is biased away from the nearby extremity of the* $[0,1]$ *interval toward the point* $\alpha/(\alpha+\beta)$*. In particular, the mean of the posterior distribution is a ratio*[38]

$$\mathrm{E}[p|q] := \mathrm{E}^{\pi(p|q)}[p] = \frac{\alpha + nq}{\alpha + nq + \beta + n(1-q)} = \frac{\alpha + nq}{\alpha + \beta + n}. \tag{10}$$

(All proofs are relegated to Appendix §F.)

For example, under $n = 1$, a declared value of q that is equal to 0.20 results in a posterior mean of p being 0.35 (see Figure 3 for the posterior shape and mean). This is reminiscent of the probability weighting discussed in section §2. (Note that, in contrast to CPT, our approach is disciplined by Bayesian formalism and does not require to discriminate between probability weights and decisions weights.)

Indeed, juxtaposing the full range of declared probabilities s against the corresponding posterior means of p (as presented in Figure 4) is broadly consistent with the pattern of probability weights we have seen in §2 (as shown in Figure 1). Specifically, low probabilities are increased, high probabilities are decreased, and the emerging distortions are larger for more extreme probabilities.

Intuitively, as confidence measure n increases, the posterior mean of p approaches linearity in q: $\partial(\mathrm{E}[p|q] - q)/\partial n < 0$ for $q < 1/2$ (assuming $\alpha = \beta$).

As can also be seen from the above Proposition, absolute scales of α, β and n cancel out and only their relative proportion matters. Fortunately, we will be able to pin down

[36] The argument here is heuristic, but it can be stated more formally, for example, probabilistically by the method of maximum entropy that stipulates a uniform distribution of discrete probability masses (in short, representing the relative allocation of pseudo-observations by a probability distribution, the solution of $\max_{\{\omega_{\mathrm{J}},\omega_{\mathrm{ext}}\}}\{-\sum_i \omega_i \ln \omega_i + \lambda(\sum_i \omega_i - 1)\}$ gives $\omega_{\mathrm{J}} \equiv \omega_{\mathrm{ext}} := 1/2$), or by the invariance utilizing the symmetry of information sources.

[37] In the binary choice setting, taking $\pi(\chi_{t+1}) := \pi(x(\iota)) = \int_0^1 L(\iota|p)\pi(p|q)\mathrm{d}p$ and substituting from equations (7) and (9), the objective function in equation (6) becomes $v(x_1)\mathrm{E}[p|q] + v(x_0)(1 - \mathrm{E}[p|q])$ if $\boldsymbol{z}_t = \mathsf{lottery}$ is chosen. Hence our focus on the $\mathrm{E}[p|q]$ quantity.

[38] Cf. the usual number-of-observations–weighted average in canonical Bayesian updating.

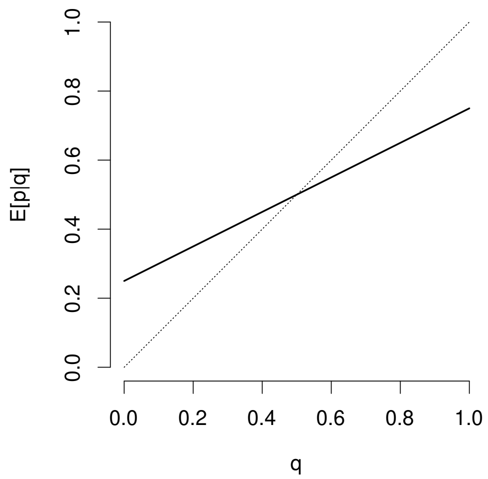


Figure 4: Posterior mean (vertical axis) vs. declared (horizontal axis) probabilities (parameterization used is E$[p|q]$ from equation (10) with $\alpha = 1/2$, $\beta = 1/2$ as well as $n = 1$).

their scales later with the help of observed data on whole sequences of lottery realizations (exploiting the fact that directly sampled observations in our framework represent a canonical type of information that can serve as a gauge for other quantities).

With the updated prior as described, the agent is optimally equipped to make the decision based on the description of payoffs and probabilities of the proposed lottery.

III. As a third step, the agent is now ready to learn from his own observations when more than one trial of the same lottery is run, leading him to make decisions from experience.[39] Clearly, given such an ergodic and stationary environment, the posterior distribution will converge to a probability mass of 1 at q, and eventually, he will learn the true lottery parameter perfectly (thus exhibiting linear probability weighting).

Henceforth, we separate the measures of confidence in communicated probabilities in different informational scenarios denoted by n into: (i) n_a for the judged probability in decisions under ambiguity; (ii) n_d for the declared, or described probability in decisions from description; and (iii) n_e for the drawn, or experienced probability in decisions from experience.

Now, after elaborating on all three steps of the algorithm, let us clarify the intricacies of step **I**, particularly relevant for the case with ambiguity.

The most straightforward approach is to use equation (9) with q reflecting the implicit deeper structure of the problem. But this reduced-form approach can be extended by constructing an explicit compound distribution.

Now the probability p is not a fixed constant, but a more complex object that is explicitly constructed. Thus, Bayesian formulation of the lottery with ambiguous probability of success—focusing on how the data underlying the lottery draws are generated—can be

[39]Say, by pressing a computer key in the lab and observing the realizations.

expressed in terms of an urn with unknown composition of differently-colored balls:

$$L(\iota|p_2)L(p_2|p_1) := \left(\frac{k^*}{k}\right)^{\iota}\left(1-\frac{k^*}{k}\right)^{1-\iota} \times \frac{k!}{k^*!(k-k^*)!}p_1^{k^*}(1-p_1)^{k-k^*}, \quad (11)$$

where two probability distributions, Bernoulli $L(\iota|p_2)$ and Binomial $L(p_2|p_1)$, represent, respectively, the data-generating process for the lottery outcome on the basis of a drawn ball from the given urn, and the data-generating process for how the composition of balls in this urn arises to begin with; k is the total number of balls in the urn (treated as given); k^* is the number of balls with favorable colors;[40] p_1 is the "upstream" probability that the ball put into the urn will have a favorable color; $p_2 := {}^{k^*}/_k$ is the "downstream" probability of drawing from the urn a ball of favorable color that will determine the lottery's final outcome (similarly to p in a non-ambiguous case).

The Jeffreys prior for p_1 is that same from equation (8), $\pi(p_1) \propto p_1^{-1/2}(1-p_1)^{-1/2}$.

Details about the ideal structure of the data-generating mechanism imply that the probability of a favorable ball being in the urn equals some value q.[41] However, in general the true probability p_1 may still differ from the deduced probability q, so we update the initial prior above weighting available information with appropriate number of pseudo-observations. This leads to the posterior distribution that, after integrating it with respect to p_2, gives the same expression for the mean as we saw in equation (10), $\mathrm{E}[p_2|q] = (\alpha + nq)/(\alpha + \beta + n)$.

The above perfectly Bayesian solution can be expressed in the following, more insightful way.

Proposition 2 (Symmetric Source Treatment Under Ambiguity)**.** *The mean of the posterior distribution of the probability of interest p_2 can be equivalently computed in two symmetric stages, one per data-generating mechanism:*

$$\mathrm{E}[p_1|q] = \frac{\alpha + n_1 q}{\alpha + \beta + n_1}$$
$$\mathrm{E}[p_2|\mathrm{E}[p_1|q]] = \frac{\alpha + n_2 \mathrm{E}[p_1|q]}{\alpha + \beta + n_2}.$$

The value of $n := n_1 \equiv n_2$, which we equalize for parsimony and easier identification requirements, in this formulation corresponds to $n + \sqrt{n(\alpha+\beta+n)}$ from the one-stage formulation.

The results in two formulations are numerically equal up to rescaling of n. However, the symmetric two-stage formulation explicitly associates confidence measure n with a data-generating (sub) mechanism, rather than with the lowest-level structural statistic

[40] For example, it is the number of red and black balls if drawing either of them will imply the successful outcome $x(1)$.

[41] For example, by the principle of indifference, it is $^2/_8$ if red or black ball is sought out of 8 balls including red, black and 6 other colors.

q in the more standard reduced-form one-stage approach.[42] This aligns better with the structure of the problem under ambiguity, and allows a more intuitive interpretation. Now, both the process of putting colored balls into the urn as well as the subsequent process of drawing one of them from it to define the lottery outcome have their own degrees of confidence, since both may in principle be corrupted and may legitimately lack full confidence.[43]

Lastly, let us summarize the derivations so far by providing the general recipe when more than one source of data is used.

Proposition 3 (Belief Updating, Discrete Case)**.** *Learning new information related to parameter of interest p (p_2 in the ambiguous case) and updating the corresponding beliefs is implemented as follows:*

$$\mathrm{E}[p|\boldsymbol{q}] = \frac{\alpha + \sum_{s\in\{a,d,e\}} n_s q_s}{\alpha + \beta + \sum_{s\in\{a,d,e\}} n_s}, \tag{12}$$

where q_s and n_s for $s \in \{a, d, e\}$ are, respectively, the communicated information and corresponding number of (pseudo-) observations in ambiguity-, description- and experience-based scenarios (adopting the convention that $n_s := 0$ whenever q_s is unavailable).

Therefore, each source is associated with its own degree of confidence, and the relevant sufficient statistics can be added with appropriate confidence weights.

Remark 1: One may notice that effective probability weighting function derived here (as plotted in Figure 4) differs somewhat from the common views in the existing literature (following Kahneman and Tversky, 1979, 1992; Camerer and Ho, 1994; for an illustration, see Figure 1). Appendix §C shows how a probability weighting function exhibiting a more conventional inverse-S-shaped pattern can be obtained from a mixture of prior distributions that combines the Jeffreys prior with the so-called Haldane prior $\mathcal{B}(0, 0)$. The latter implies a uniform distribution for the logarithm of the odds ratio, $\ln(p/(1 - p))$.

Remark 2: Our approach easily generalizes to the case of non-binary lotteries: the Bernoulli, Beta and Binomial distributions above are just replaced with, respectively, the categorical (generalized Bernoulli), Dirichlet and Multinomial distributions. Moreover,

[42] The latter, two-stage lottery construction with probability weightings applied at each stage is somewhat close to the approach of Segal (1987, 1990). However, we dispense with the rank-dependent utility assumption, and do not rely on transformations into certainty equivalents.

[43] Note that our approach is general enough to treat ambiguous and fixed probabilities in a "continuous", nested way: e.g., when the composition of the urn is fixed and the respective probability of success (i.e., some constant value $^{k^*}/_k$) is announced, n_1 goes to infinity and the scenario naturally turns into that with a described probability. Additionally, notice that the two-stage formulation with symmetric recursive processing of information aligns better with the current understanding of the architecture of biological and artificial neural networks, while the one-stage reduced formulation is closer to canonical hierarchical Bayesian models.

a case with continuous random variables is also amenable, see the next subsection that instead of lotteries deals with stock prices.

Recap: To summarize, we considered decision-making in a stochastic environment that regularly offers fresh lotteries driven by parameters which are potentially unknown. In this setting, decisions must be made in real time, simultaneously with online learning of the relevant parameters; while such learning process itself can not rely solely on observed data but must allow for unseen contingencies in order to account for small samples and infrequent events. In the context of Bernoulli lottery-generating model, the processing of new information and updating of success probability is handled very naturally, and the undersampling-motivated usage of invariant ignorance prior distributions reshapes the lottery success probabilities' posterior Beta distributions in a way that resembles the probability distortions from section §2.1.

3.3 Assets in financial markets

As to the case with continuous distributions, we can get a sense of what would happen by taking the Bernoulli distribution from section §3.2 to the limit. According to the classical de Moivre–Laplace theorem, as n grows larger, the distribution of the number of successes in the Bernoulli trials nx will converge to the Normal distribution with mean np and variance $np(1-p)$. Take, without loss of generality, $p = q > 1/2$. By the earlier derivations, the posterior mean of p would be biased toward $1/2$, which unambiguously increases the corresponding Normal distribution's variance (due to the latter's concavity in p).[44]

Thus, we now focus on a non-ergodic environment with continuous, unbounded distributions.[45] In such a case, even a large number of random variables' realizations would not allow perfect learning of the forever evolving true parameters, and the prior could not possibly become dominated by the sample observations.

Consider stock price returns whose logarithms r_t are Normally distributed with mean μ and variance σ^2. The corresponding Gaussian probability density function defines the

[44]Provided we ignore the effect of learning from different trials: say, by treating each partition of the probability space (every cell of the domain) as a genuinely new lottery.

[45]In terms of lotteries from the previous section §3.2, the non-ergodicity corresponds to running different, non-repeated lottery trials. Also, while the case with lotteries could be criticized for treating probabilities and payoffs differently by taking the payoff values as certain, here with continuous probability distributions this is irrelevant as we can take, without loss of generality, either the domain (payoffs) or the range (probabilities) of a distribution function as a preset gauge.

likelihood of possible returns:[46]

$$L(r_t|\mu,\sigma^2) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(r_t-\mu)^2\right). \tag{13}$$

We follow the same basic algorithm given in §3.2, the only substantial difference is that now external information is either "recycled" from the previous stage of parameter evolution or sampled from the current stage. Note that we explain below the details of each step in a non-consecutive but easier to understand order.

I. In this section, the parameters themselves are (unobservable) random variables. To obtain the corresponding ignorance prior that is invariant under reparameterization, the Jeffreys method yields:[47]

$$\pi(\sigma^2|\cdot) \propto \sqrt{\mathrm{E}^L\left[\left(\frac{\partial}{\partial\sigma^2}\ln L(r_t|\mu,\sigma^2)\right)^2\right]} = \frac{1}{\sigma^2}, \tag{14}$$

$$\pi(\mu|\cdot) \propto \sqrt{\mathrm{E}^L\left[\left(\frac{\partial}{\partial\mu}\ln L(r_t|\mu,\sigma^2)\right)^2\right]} = \frac{1}{\sigma}. \tag{15}$$

Thus, the prior corresponding to the mean is found to be conditional on the variance. Indeed, we can always write $\pi(\mu,\sigma^2) = \pi(\mu|\sigma^2)\pi(\sigma^2)$.

Equations (14) and (15) are consistent with several different probability distributions, but it will prove convenient to proceed in the following way. Assuming that the variance σ^2 follows an Inverse-Gamma distribution[48] and the mean μ is Normally distributed (conditionally on the variance), equations (14) and (15) can be viewed as the limiting cases of the same two probability distributions. That is, an Inverse-Gamma $\mathcal{IG}(\alpha_0,\beta_0)$ with parameters $\alpha_0 = 0$ and $\beta_0 = 0$ as well as a (conditional) Gaussian $\mathcal{N}(\ell_0,\sigma^2/\lambda_0)$ with parameters $\ell_0 \in \mathbb{R}$ (say, $\ell_0 = 0$) and $\lambda_0 \to 0$. Hence, parameterization invariance implies that a transformed scale parameter $\ln\sigma^2$ has a uniform distribution on the real line, leaving the scale to be multiplication-invariant, whereas a location parameter μ has a (conditionally) uniform distribution on the real line, being addition-invariant.

Thus, we obtained the Normal–Inverse-Gamma–Normal conjugate system. Intuitively, the Normal prior distribution's parameter $\lambda_0 \to 0$ and Inverse-Gamma's parameter $\alpha_0 = 0$ can be interpreted as 0 pseudo-observations of the Normal realizations of r_t (hence these priors are improper).

[46] I.e., in equation (6) set $J(\chi_{t+1},\boldsymbol{z}_t) := v(C(\chi_t,\boldsymbol{z}_{t-1})) + \beta V(\chi_{t+1},\boldsymbol{z}_t)$, where $v(\cdot)$ is a consumption-value function and $V(\cdot)$ is a state-value function from the Bellman equation corresponding to the problem; $\chi_{t+1} := r_{t+1}$; $\boldsymbol{z}_t \in \mathbb{R}^2$; $\boldsymbol{\theta}_t := \{\mu,\sigma^2\}$. Note that $\boldsymbol{z}_t$ is substituted out in equilibrium and does not enter our empirical exercises, see Lucas (1978).

[47] In our derivations, mean and variance parameters are treated separably, but treating them as a single vector does not make a difference here. In general, however, this seems like another instance of flexible choices. Substantial disagreement is ongoing in the existing Bayesian literature about how to deal with invariance for location-scale parameters in the exponential-family distributions; relevant sources include DeGroot (1970), Kass and Wassermann (1996), Minka (2001), Gelman et al. (2004), Murphy (2012).

[48] Here, $\mathcal{IG}(h,s)$, with shape h and inverse scale s, is parameterized as $f(x|h,s) := \frac{s^h}{\Gamma(h)}x^{-h-1}\exp\left(-\frac{s}{x}\right)$.

By the maximum entropy approach, the Inverse-Gamma distribution maximizes entropy on the $(0, \infty)$ interval subject to constraints on the arithmetic mean and on the logarithm of the geometric mean of the random variable $1/\sigma^2$. The Normal distribution maximizes entropy on the $(-\infty, \infty)$ interval subject to constraints on the arithmetic mean and on the variance of the random variable μ. Thus, in general the earlier given values of α_0 and β_0 also may be interpreted as specific restrictions on the mean (but not point) values of σ^2, whereas the values of ℓ_0 and λ_0 as restrictions on the first and the second (central) moments of μ in entropy maximizations.[49]

III.A. In order to specify how the initial prior can be enriched with any external information, it is helpful to begin with an ergodic environment in which parameters σ^2 and μ are realized at time $\tau = 0$ (in the subsequent treatment it will be convenient to have an additional time-counting variable) and remain constant thereafter. In such a situation, whenever new observations on returns r_τ are realized, the standard mechanism of Bayesian updating stipulates the procedure below.

Lemma 1 (Belief Updating, Continuous Case with Constant Parameters)**.** *Learning new information related to parameters of interest μ and σ^2 and updating the corresponding beliefs is implemented as follows:*

$$\alpha_\tau := \alpha_0 + \frac{1}{2}\tau, \tag{16}$$

$$\beta_\tau := \beta_0 + \frac{1}{2}\left(\frac{\lambda_0 \tau}{\lambda_0 + \tau}(\bar{r}_\tau - \ell_0)^2 + \tau \bar{s}_\tau^2\right), \tag{17}$$

$$\ell_\tau := \frac{1}{\lambda_0 + \tau}(\lambda_0 \ell_0 + \tau \bar{r}_\tau), \tag{18}$$

$$\lambda_\tau := \lambda_0 + \tau, \tag{19}$$

where the empirical mean is defined as $\bar{r}_\tau := (1/\tau)\sum_{j=1}^{\tau} r_j$ and empirical variance as $\bar{s}_\tau^2 := (1/\tau)\sum_{j=1}^{\tau}(r_j - \bar{r}_\tau)^2$.[50]

Then, the (marginal) posterior distributions for the variance and mean are, respectively,

$$\pi(\sigma^2|\alpha_\tau, \beta_\tau) = \frac{\beta_\tau^{\alpha_\tau}}{\Gamma(\alpha_\tau)}(\sigma^2)^{-\alpha_\tau - 1}\exp\left(-\frac{\beta_\tau}{\sigma^2}\right), \tag{20}$$

$$\pi(\mu|\ell_\tau, \lambda_\tau, \alpha_\tau, \beta_\tau) = \frac{\Gamma(\frac{2\alpha_\tau + 1}{2})}{\Gamma(\frac{2\alpha_\tau}{2})\sqrt{2\pi\beta_\tau/\lambda_\tau}}\left(1 + \frac{(\mu - \ell_\tau)^2}{2\beta_\tau/\lambda_\tau}\right)^{-\frac{2\alpha_\tau + 1}{2}}, \tag{21}$$

[49]The two constraints are $\mathrm{E}^\pi[1/\sigma^2] \equiv \alpha_0\beta_0$ and $\mathrm{E}^\pi[\ln(1/\sigma^2)] \equiv \psi(\alpha_0) + \ln\beta_0$ for the Inverse-Gamma distribution, as well as $\mathrm{E}^\pi[\mu] \equiv \ell_0$ and $\mathrm{E}^\pi[(\mu - \ell_0)^2] \equiv \sigma^2/\lambda_0$ for the Normal distribution. Seeming like another instance of flexible choices, the restrictions are ∞ variance σ^2 and a free (unrestricted) mean μ, respectively.

[50]Subsequently, recursive formulas for calculating empirical mean and variance will be required, these are $\bar{r}_\tau := (1/\tau)((\tau - 1)\bar{r}_{\tau-1} + r_\tau)$ and $\bar{s}_\tau^2 := (1/\tau)((\tau - 1)\bar{s}_{\tau-1}^2 + (r_\tau - \bar{r}_{\tau-1})(r_\tau - \bar{r}_\tau))$.

being, due to conjugacy, of the Inverse Gamma and Student's t form. The latter distribution is the non-standardized Student's t density[51] *with degrees of freedom* $2\alpha_\tau$*, location* ℓ_τ *and scale* $\beta_\tau/(\alpha_\tau\lambda_\tau)$*; in short,* $\mathcal{T}_{2\alpha_\tau}(\ell_\tau, \beta_\tau/(\alpha_\tau\lambda_\tau))$*.*

Note how the Student's t probability density (21) is being born from the Normal density that was conditional on the variance.

In turn, for the next period's return we have the following.

Lemma 2 (Overweighting Low Probabilities and Underweighting High Probabilities, Continuous Case with Constant Parameters)**.** *Conditionally on unknown but fixed parameters* σ^2 *and* μ*, the posterior predictive distribution for next period's return* $r_{\tau+1}$ *is*

$$\pi(r_{\tau+1}|\ell_\tau, \lambda_\tau, \alpha_\tau, \beta_\tau) = \frac{\Gamma(\frac{2\alpha_\tau+1}{2})}{\Gamma(\frac{2\alpha_\tau}{2})\sqrt{2\pi\beta_\tau(1+1/\lambda_\tau)}}\left(1+\frac{(r_{\tau+1}-\ell_\tau)^2}{2\beta_\tau(1+1/\lambda_\tau)}\right)^{-\frac{2\alpha_\tau+1}{2}}, \tag{22}$$

which is a Student's t density parameterized as $\mathcal{T}_{2\alpha_\tau}(\ell_\tau, \beta_\tau(1+1/\lambda_\tau)/\alpha_\tau)$*. For finite* α_τ*, its tails are heavier than the Gaussian tails of the data-generating process (13).*

Therefore, in investor's decision-making the uncertainty about parameter values amplifies the tail probabilities and the overall riskiness of the stock.[52]

II. and III.B. Now, turning to the setting we are primarily interested in, consider a non-ergodic environment in which σ^2 and μ are not constant but at any time may change within $\mathbb{R}_+ \times \mathbb{R}$. More formally, every period t with some probability $p \in [0,1]$ the latent parameter evolution process $\{\ln\sigma_t^2, \mu_t\} \in \mathbb{R}^2$ jumps from its previous period's value $\{\ln\sigma_{t-1}^2, \mu_{t-1}\}$ with a magnitude of change distributed independently and symmetrically on $\mathbb{R}^2$, consequently our unknown parameters of interest $\{\ln\sigma^2, \mu\}$ are reinitialized as $\{\ln\sigma^2, \mu\} := \{\ln\sigma_t^2, \mu_t\}$ and the new regime $k \in \mathbb{N}$ starts. Otherwise, with probability $(1-p)$ the process $\{\ln\sigma_t^2, \mu_t\}$ and the values $\{\ln\sigma^2, \mu\}$ are unchanged. For an ease of exposition, assume that (i) probability p is non-stochastic and is common knowledge; and that (ii) the time of change is known ex post but not ex ante. Such an evolution of the underlying parameter values can be understood as changes in (latent) macro-financial regimes. Effectively, this setting allows for an infinite number of regimes.[53]

Let us consider the dynamics of the system as physical time t passes. While staying within the same regime k, the learning dynamics follow the recursive formulas (16)–(19). As the time since the last parameter change τ_k increases, by equations (20) and (21), the

[51]Here, $\mathcal{T}_d(l,s)$, with degrees of freedom d, location l and scale s, is parameterized as $f(x|d,l,s) := \frac{\Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})\sqrt{d\pi s}}\left(1+\frac{(x-l)^2}{ds}\right)^{-\frac{d+1}{2}}$.

[52]These derivations, which pertain to our ergodic case, are standard in Bayesian analysis (for example, see Murphy, 2012) and are provided in the interest of clarifying our chosen notation.

[53]The methods for inferring regimes and updating the probability of change are widely studied in economics and finance (a recent contribution is Smith and Timmerman, 2021; extensive reviews can be found in Ang and Timmermann, 2012, or Hamilton, 2016). We will just emphasize that identifying the structural breaks is a challenging problem, especially when done in real time or under low-frequency sampling.

parameters eventually would be learned. By equation (22), the posterior Student's t distribution of returns would converge to the true distribution from equation (13). However, these eventual target distributions will not be achieved in finite time as long as probability of regime change p is greater than 0: when at some physical time t the values $\{\ln\sigma_t^2, \mu_t\}$ change, the learned information becomes obsolete.

After the regime change, the learning dynamics will restart. First, formulas (16)–(22) will be re-initialized and the variable counting time since the latest parameter change will set to $\tau_{k+1} := \tau^\star := 0$ (the object $\tau^\star$ is just a constant).[54] Since initially nothing will be known about the new regime, the prior parameter values will be reset, too. For the variance, the values will be set to $\alpha_0 = \beta_0 = 0$, implying the logarithmic prior introduced earlier. For the mean, they will be set to $\lambda_0 \to 0$, which implies infinite variance and effectively no information about the location (in the same vein, ℓ_0 remains unrestricted, but this is of no effect given $\lambda_0 \to 0$).

Second, it is reasonable for an agent to incorporate some empirical information into the above prior. Given that the latent parameter process evolves symmetrically around its most recent position, we would update the initial prior with the sufficient statistics that correspond to the latest available regime, that is, the current one. For the Normal distribution, the sufficient statistic is vector $[\bar{r}_{\tau,k+1}; \bar{s}^2_{\tau,k+1}]$. So, in updating equations (18) and (17), empirical mean $\bar{r}_{\tau,k+1}$ will be replaced with $\bar{r}_{\tau k}$, and empirical variance $\bar{s}^2_{\tau,k+1}$ will be replaced with $\bar{s}^2_{\tau k}$ from the current regime.

Furthermore, by the principle of indifference between sources of pseudo-observations, similarly to $n := 1$ in section §3.2, setting $\tau_{k+1} := \tau^\star := 3$ would put two sources of prior information on equal footing, to the extent possible (3 is the minimum sample size required for our sufficient statistics to be acceptably informative about their corresponding parameters[55]). Intuitively, such an approach can be interpreted as extending the existing 0 pseudo-observations from the initial prior with 3 pseudo-observations from the external source, the latest regime.[56] (However, note that, in general, smaller τ and related quantities such as α can be understood as lower confidence in this information, similarly to what is the case with, e.g., n_a for the judged probability in lottery choice considered in section §3.2. Also, it may be useful to realize that considerations of current and new regimes here parallel the treatment of, respectively, decisions from experience and decisions under conditions of ambiguity in §3.2.)

[54]Note that the vector of all times t thus has a one-to-one correspondence to a concatenation of τ_k's values within each regime as well as over all regimes.

[55]It can be shown that μ and σ^2 from the current, indexed by τ_k, regime—i.e., the parameters which empirical mean and empirical variance ultimately characterize—have well-defined and finite posterior means as long as $\tau_k > 2$: from equations (21) and (20), $\mathrm{E}^\pi[\mu|\cdot] = \bar{r}_{\tau k}$ and $\mathrm{E}^\pi[\sigma^2|\cdot] = (\tau_k/(\tau_k - 2))\bar{s}^2_{\tau k}$. Hence, this is the minimum amount of empirical information needed to characterize μ and σ^2 of vintage τ_k. In turn, this same amount of information is then used to update the prior on μ and σ^2 of vintage τ_{k+1}.

[56]Inclusion of $\bar{r}_{\tau k}$ and $\bar{s}^2_{\tau k}$ into calculations of $\bar{r}_{\tau,k+1}$ and $\bar{s}^2_{\tau,k+1}$ can be done using recursive formulas provided earlier.

Thus, in the contingency of future regime change, the prior predictive[57] distribution, defined by equation (22) with the just described values for τ_{k+1} as well as α_3, β_3, λ_3 and ℓ_3, becomes a Student's t with 3 degrees of freedom, $\mathcal{T}_3(\bar{r}_\tau, 4\bar{s}_\tau^2/3)$. (From the previous sentence onwards, a generic τ is understood as the current regime's τ_k.)

Allowing for both contingencies, we obtain the following result.

Proposition 4 (Overweighting Low Probabilities and Underweighting High Probabilities, Continuous Case with Changing Parameters)**.** *The posterior predictive distribution for next period's return $r_{\tau+1}$ is a p-mixture of two Student's t densities:*

$$\pi^M(r_{\tau+1}|\ell_\tau, \lambda_\tau, \alpha_\tau, \beta_\tau; \bar{r}_\tau, \bar{s}_\tau^2) = (1-p) \times \pi(r_{\tau+1}|\ell_\tau, \lambda_\tau, \alpha_\tau, \beta_\tau) + \\ + p \times \pi(r_{\tau+1}|\bar{r}_\tau, 3, 3/2, 3\bar{s}_\tau^2/2), \tag{23}$$

where $\pi(r_{\tau+1}|\ell_\tau, \lambda_\tau, \alpha_\tau, \beta_\tau)$ is $\mathcal{T}_{2\alpha_\tau}(\ell_\tau, \beta_\tau(1+1/\lambda_\tau)/\alpha_\tau)$ and $\pi(r_{\tau+1}|\bar{r}_\tau, 3, 3/2, 3\bar{s}_\tau^2/2)$ is $\mathcal{T}_3(\bar{r}_\tau, 4\bar{s}_\tau^2/3)$. For finite α_τ or $p=1$, the mixture's tails are heavier than the Gaussian tails of the data-generating process (13).

Therefore, at the outset parameter uncertainty amplifies the underlying Gaussian tails. As time passes but the system stays under the same macro-financial regime and parameters do not change, the tails of the first mixture component—and thus of the mixture overall—gradually compress due to the progress in learning of the underlying parameters. However, with $p > 0$, there is always a risk of regime change and renewed importance of the relatively heavier tails of the second mixture component.

Finally, the agent is prepared to make an optimal investment decision based on his current (as well as recycled from the past) sample data, and also, while staying in the same regime, to collect more information and learn the underlying parameters ever more accurately.

Recap: To summarize, we considered a stylized stochastic environment that exhibits non-ergodicity due to the parameters that condition the probability distributions of interest evolving randomly over time. Investors deal with parameter uncertainty using invariant ignorance priors that explicitly allow for undersampled or even unobserved events, and update such priors with available sample information. As a result, although the true data-generating distribution is Normal (conditionally on stochastic mean and variance), the inferred posterior distribution used for investment decisions has a Student's t form (sometimes its version with very heavy tails), which can help explain the asset pricing puzzles and regularities from section §2.2.[58]

[57] Which in this situation would be a more precise term than posterior predictive.

[58] Moreover, sample and hypothetical data-generating ("population") distributions permanently differ: recognition of non-ergodicity immediately removes the restriction that sample statistics must converge to or at least be informative about population statistics, which has direct implications for the posterior mixture distribution of returns. Indeed, the posterior distribution $\pi^M(\cdot)$ never converges to some uncon-

4 Empirics

In this section we take our theoretical models to data.

4.1 Lotteries in laboratory experiments

Our intention is to test how accurately the model from §3.2 captures the behavior of human subjects in typical laboratory experiments with binary lotteries.

As is standard in experimental economics, agent's value function is assumed to have a power form given by equation (2), and we allow for different curvatures in gain and loss lotteries: i.e., parameter $\varphi := \varphi^+$ if lottery payoff $x \geq 0$, but $\varphi := \varphi^-$ if $x < 0$. The (posterior mean) probability of lottery's tail payoff x^t, which is defined as the payoff further in absolute value from the reference point 0 than the lottery's sure payoff of the same sign x^s, is given by equation (12). Two above-mentioned equations allow to calculate a model-implied certainty equivalent (CE) of the lottery, while experimental data at our disposal provide an observed CE. The observed CE is postulated to reflect a model-implied CE and an additive error term. Assuming that errors are Normally distributed, the estimation is conducted by maximum likelihood.

Decisions from description: We start with experiments that inform human subjects about the setup by announcing the lottery payoffs and probabilities.

The data set used is the same as in §2.1. As a benchmark for our model to target in terms of CE fit, we take equations (1) and (2), parameterized differently for gains and losses, and then replicate and re-estimate on pooled data the purely data-fitting model formulated by the authors of the original data set. The value of the communicated probability q is taken to be the probability that was announced to the subjects. Under these circumstances, we can only identify the relative numbers of pseudo-observations contained in the initial prior and obtained from the external source, forcing us to fix the confidence in the external source value at $n = n_d := 1$).

Full estimation results are summarized in Table 2. The first column presents the benchmark CPT-based model. The second column presents our model, fixing the initial prior's parameters at the values stipulated by the Jeffreys method, that is $\alpha = \beta = 1/2$, which makes the probabilistic side of the decision problem fully determined by the theoretical considerations and lets only the value function's parameters to be estimated. Our proposed approach achieves a similar goodness-of-fit performance, in spite of using far less parameters and, more importantly, being based on decision-theoretic rather than statis-

ditional population distribution $L(r_t)$ (for lack of the latter in a non-ergodic setting). It also does not converge to the population distribution defined conditionally on the current regime parameters, $L(r_t|\mu, \sigma)$, or even to the sample distribution from the current regime, $\hat{L}(r_t|\mu, \sigma)$ (because of the overarching presence of the risk of regime change). Such restrictions and respective convergence are usually associated with a strong (or naïve) form of "rational expectations".

Table 2: Estimation Results for Lottery Decisions — Information from Description

Parameters	CPT_4	$\mathcal{B}_{\text{Jeffreys}}$	$\mathcal{B}$
φ^+	0.95	0.88	0.90
	(.015)	(.007)	(.007)
η^+	0.46		
	(.005)		
ψ^+	0.93		
	(.013)		
φ^-	1.11	1.10	1.10
	(.018)	(.009)	(.009)
η^-	0.49		
	(.005)		
ψ^-	1.00		
	(.016)		
α, β		0.50	0.30
		(n/a)	(.005)
n_d		1.00	1.00
		(n/a)	(n/a)
R^2	0.9067	0.9006	0.9075

Notes: The explained variable is observed certainty equivalents. The rows contain value function parameters φ^+ and φ^-; probability weighting function parameters η^+, ψ^+ and η^-, ψ^-; the initial prior density parameters α and β; the confidence value n_d; as well as the goodness-of-fit measure R^2. The columns present different model specifications: a cumulative prospect theory specification with 4 weighting function parameters; $\mathcal{B}(\alpha + n_d q, \beta + n_d(1-q))$ model under Jeffreys prior; and $\mathcal{B}(\alpha + n_d q, \beta + n_d(1-q))$ model under revealed prior (see text for a detailed description). The estimation is conducted on pooled data by maximum likelihood accounting for individual- and payoff sign-specific error heteroscedasticity proportional to lottery range. Standard errors (in parentheses) are computed using wild bootstrap with 200 replications. Number of observations is 17800. Data source on laboratory experiments is Bruhin et al. (2010).

tical data-fitting grounds. We are not particularly interested in the nuisance parameters φ^+ and φ^-.

Allowing parameters α and β (still assuming $\alpha = \beta$) to adjust so as to maximize our model's fit to data, we can back out putative priors effectively used by the subjects. The third column of Table 2 provides the estimation results for this "oracle" model: with $\hat{\alpha} = \hat{\beta} = 0.30$, the prior that is revealed to fit the subjects' behavior best is fairly close to the prescriptions of the Jeffreys method, the remaining free parameters barely change, and the model's fit improves only slightly (but this time outperforming the CPT benchmark).

We can compare empirical and model-implied CEs visually by plotting the CE-to-payoffs ratios, appropriately standardized, which is done in Figure 5. Both the benchmark and our revealed-prior models perform very similarly, capturing the characteristic pattern of differences in risk aversion and risk seeking behavior: risk seeking (points above the diagonal that corresponds to risk neutrality) at low probabilities and risk aversion (points below the diagonal) at high probabilities in gain lotteries, but risk aversion (above) at

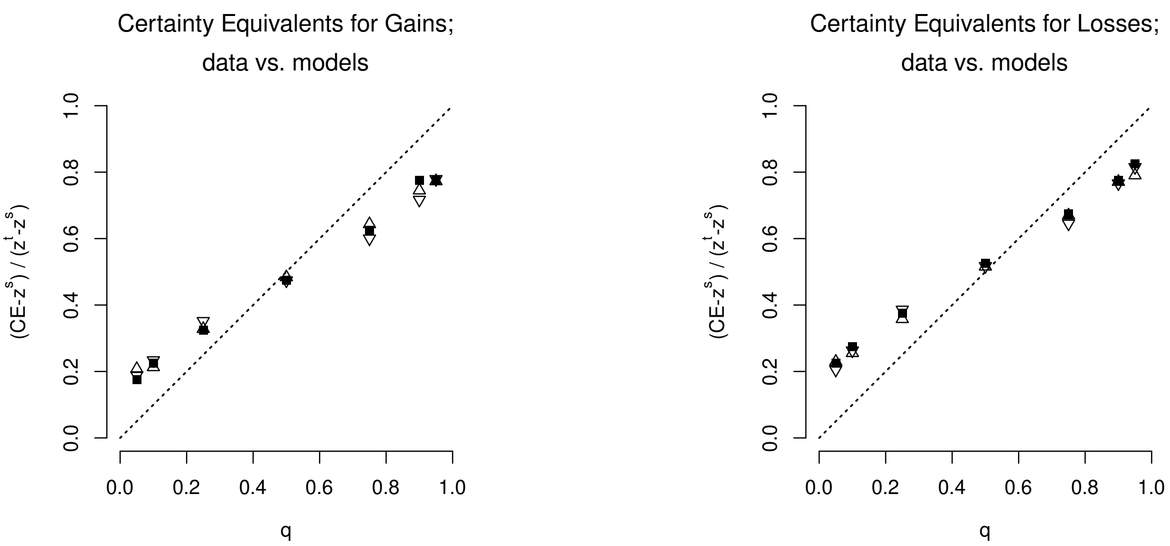


Figure 5: Ratios of CEs to lottery payoffs, standardized, median values (vertical axis) vs. declared true tail probabilities (horizontal axis), where values of CEs are either measured empirically (solid squares), implied by the CPT model (triangles pointing down) or implied by the revealed-prior model (triangles pointing up). Lotteries with gains and with losses, information obtained from description.

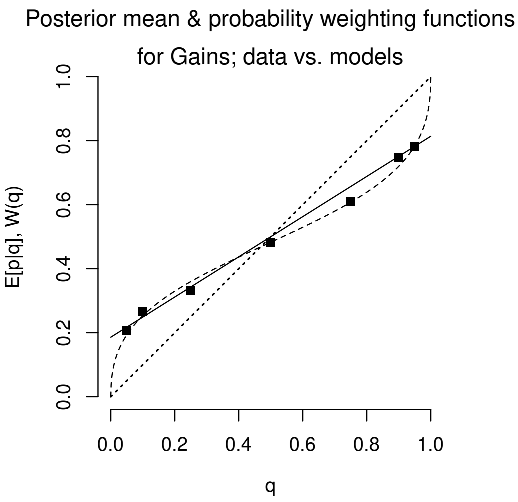
low probabilities and risk seeking (below) at high probabilities in loss lotteries. The estimated posterior mean and probability weighting functions, with empirical weighting function fitted pointwisely for respective quantiles, capture the same pattern, as shown in Figure 6.

Lastly, we can place results on the same interpretable scale: if the number of pseudo-observations in the confidence measure for description-based communication is taken to be 23.24 (basing on Table 3, which will be presented later), the prior revealed to fit subjects in lottery experiment best amounts to 13.94 pseudo-observations, while its counterpart is 1.00 in the Jeffreys prior.

Decisions from experience: Now we consider experiments that communicate the lottery payoffs and probabilities by allowing human subjects to sample the data-generating mechanism by clicking on the computer button.

The data set is from Abdellaoui et al. (2011b), who conducted laboratory experiments eliciting certainty equivalents for binary lotteries in both description-based and experience-based scenarios. The study was based in Paris, its records on gain and loss (but not mixed) lotteries comprises 2928 observations in total.

The estimation approach is the same as before. But this time we are able to identify the absolute numbers of (pseudo-) observations contained in the initial prior and obtained from each of the external source—i.e., either from the declaration or from sampling—which



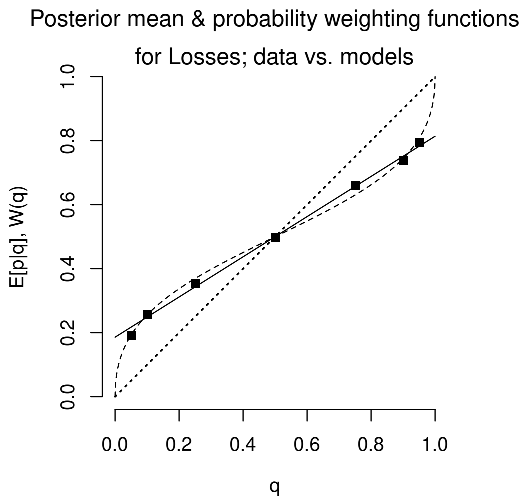


Figure 6: Posterior mean and probability weighting functions (vertical axis) vs. declared true tail probabilities (horizontal axis), with posterior mean function produced by the revealed-prior model from Table 2 (solid) and with probability weighting functions produced by the CPT model from Table 2 (dashes) as well as by agnostic data fitting (solid squares). Lotteries with gains and with losses, information obtained from description.

allows us to quantify and scale the confidence measures n_d and n_e directly.[59]

Estimation results are presented in Table 3. The left panel conducts the same exercise as that in Table 2: despite some numerical differences in parameter estimates, the qualitative picture is unchanged. The middle panel introduces the results for sampling-based lottery choices, with $n = n_e$ variable capturing the number of observations sampled by a given subject: the specification with Jeffreys prior fits the data reasonably well but has room for improvement in comparison to the CPT benchmark, while the oracle model allowing for initial prior parameters α and β (assuming $\alpha = \beta$) to adjust outperforms both specifications (to show the role of variation in the number of sampled draws, in the specification with n_e normalized to 1 we try ignoring this variation, which moderately damages the fit). The right panel offers the encompassing specification that imposes Jeffreys prior while leaving both confidence measures n_d and n_e unrestricted:[60] as before, the Jeffreys prior specification only faintly underperforms the CPT alternative, but both are beaten by the formulation with optimized initial prior parameters α and β that is revealed to fit the participants behavior best (again, the specification with $n_e := 1$ shows a modest deterioration in the goodness of fit).

[59]In our data, the number of sampled observations n_e has the mean of 19.57 and the standard deviation of 8.50.

[60]Values of n_d and n_e are unrestricted for their corresponding data sub-samples, and are set to 0 otherwise.

Table 3: Estimation Results for Lottery Decisions — Information from Description and from Experience

Parameters	Description			Experience				Description & Experience			
	CPT$_4$	$\mathcal{B}_{\text{Jeffreys}}$	$\mathcal{B}$	CPT$_4$	$\mathcal{B}_{\text{Jeffreys}}$	$\mathcal{B}$	$\mathcal{B}$	CPT$_4$	$\mathcal{B}_{\text{Jeffreys}}$	$\mathcal{B}$	$\mathcal{B}$
φ^+	0.71	0.67	0.74	0.68	0.76	0.69	0.67	0.71	0.76	0.71	0.70
	(.023)	(.012)	(.016)	(.026)	(.009)	(.016)	(.014)	(.019)	(.012)	(.011)	(.011)
η^+	0.53			0.50				0.50			
	(.026)			(.025)				(.018)			
ψ^+	0.94			0.90				0.90			
	(.040)			(.047)				(.033)			
φ^-	0.92	0.76	0.81	1.04	1.06	0.96	0.96	0.99	0.91	0.88	0.88
	(.028)	(.013)	(.014)	(.040)	(.009)	(.018)	(.018)	(.022)	(.013)	(.011)	(.011)
η^-	0.71			0.71				0.70			
	(.020)			(.025)				(.016)			
ψ^-	0.82			0.89				0.83			
	(.034)			(.049)				(.026)			
α,β		0.50	0.17		0.50	3.77	0.22		0.50	3.95	0.23
		(n/a)	(.011)		(n/a)	(.249)	(.014)		(n/a)	(.246)	(.013)
n_d		1.00	1.00						3.60	23.24	1.31
		(n/a)	(n/a)						(.262)	(2.005)	(.107)
n_e					data	data	1.00		data	data	1.00
							(n/a)				(n/a)
R^2	0.8653	0.8300	0.8632	0.8080	0.7758	0.8370	0.8324	0.8326	0.8122	0.8473	0.8447
Obs.	1464	1464	1464	1464	1464	1464	1464	2928	2928	2928	2928

Notes: The explained variable is observed certainty equivalents. The rows contain value function parameters φ^+ and φ^-; probability weighting function parameters η^+, ψ^+ and η^-, ψ^-; the initial prior density parameters α and β; the numbers of (pseudo-) observations in the relevant experiments n_d and n_e; as well as the goodness-of-fit measure R^2 and the number of observations in the data subset (see text for a detailed description). The columns are split into three panels depending on information source (description, experience or both) and present different model specifications: a cumulative prospect theory specification with 4 weighting function parameters; $\mathcal{B}(\alpha+nq, \beta+n(1-q))$ model under Jeffreys prior; and $\mathcal{B}(\alpha+nq, \beta+n(1-q))$ model under revealed prior (see text for a detailed description). The estimation is conducted on pooled data by maximum likelihood accounting for individual- and payoff sign-specific error heteroscedasticity proportional to lottery range. Standard errors (in parentheses) are computed using wild bootstrap with 200 replications. Data source on laboratory experiments is Abdellaoui et al. (2011b).

Crucially, in this last panel our proposed model with unrestricted α, β, n_d and n_e effectively (i) uses the variation in the number of sampled draws to scale initial prior's α and β in terms of the number of observations n_e, and then (ii) uses such α and β for estimating the confidence measure n_d scaled in terms of the absolute number of pseudo-observations. It suggests several points of note.

First, the prior parameters scaled in terms of the number of (pseudo-) observations are $\hat{\alpha} = \hat{\beta} = 3.95$, which seems small relatively to the average number of sampled draws ($\hat{\alpha} + \hat{\beta} = 7.90 < 19.57 = \bar{n}_e$), but it is not negligible — this exceeds the Jeffreys' prescription of $\alpha + \beta = 1.00$, and implies that the prior dominates in equation (12) even after the first several (pseudo-) observations have been communicated from the external source.

Second, the estimated confidence in the declared lottery details parameter $\hat{n}_d = 23.24$ is comparable (and statistically indistinguishable) in magnitude to the confidence obtained via sampling the lottery mechanism as reflected in the average number of draws $\bar{n}_e$.

Third, judging by the difference in the goodness of fit between the best-performing unrestricted specification and its alternative that includes the same number of free parameters but ignores the number of sampled draws and keeps n_e at 1, we see that accounting for the variation in sampled draws did improve the explanatory power of the model — although R^2 decreases only modestly from 0.8473 to 0.8447, Vuong (1989) likelihood-based test strongly favors the former specification producing the test statistic of 2.77 with the p-value below 0.01 level.

Placing things on the same scale, the number of pseudo-observations in the prior revealed to fit subjects in lottery experiments best is 7.90, while in the Jeffreys prior it is 1.00, and, for comparison, it is 23.24 in the confidence measure for description-based communication.

Appendix §A provides empirical and model-implied CEs as well as posterior mean and probability weighting functions illustrating the preceding results.

Decisions under ambiguity: Lastly, we turn to experiments that inform their subjects about the possible events that a given data-generating mechanism operates with, and which lottery payoffs are associated with different combinations of these events, but do not reveal the exact subset of events the lottery will eventually be drawn from.

The data set is due to Abdellaoui et al. (2011a), whose laboratory experiments elicited certainty equivalents for binary lotteries in which probabilities were either communicated by description or left ambiguous. Their mechanism operated with balls of 8 different colors and urns with either known or unknown composition of balls. The experiments were conducted in France, they considered only lotteries for gains, and their records comprise 2112 observations in total.

The estimation approach is similar to that used for decisions from experience, except that now confidence measures n_d and n_a can only be scaled in relative terms. In the ambiguous case, the communicated ("judged") value of q is associated with the event partitioning that corresponds to characteristics of the lottery under consideration (e.g., to ⅛ for a bet on drawing a red-colored ball).

Estimation results are available in Table 4. We see that although in all cases specifications under Jeffreys prior fare competitively, they do underperform the CPT benchmark; but the models with initial prior parameters α and β allowed to change so as to fit the data perform as well as the benchmark.

The most interesting findings are on the right panel with a prior revealed to fit the data best. First, we see that the confidence in ambiguous information q about lottery probability parameter p that is contained merely in the primitives of the mechanism itself, relatively to information communicated by description, is 0.42 to 1. Assuming the confidence measure for decisions from description is the same 23.24 pseudo-observations that we estimated earlier in Table 3, our relative measure above amounts to $23.24 \times 0.42 = 9.76$ pseudo-observations. This is less than the corresponding value for decisions from description, though it exceeds the weight of information contained in the revealed prior: $9.76 > 3.25 = 23.24 \times (0.07 + 0.07)$.

Furthermore, the two rightmost columns of Table 4, which, following Proposition 2, present the results of the two-stage symmetric treatment of data-generating (sub) mechanisms underlying lotteries under ambiguity, allow a more intuitive interpretation of the confidence degrees demonstrated in experiments under ambiguity: the per-stage value $\hat{n}_a = 0.91$ is now much closer to the value of 1 corresponding to information from description, n_d, and translates into $23.24 \times 0.91 = 21.15$ pseudo-observations. Importantly, this closeness between the two-stage n_a and n_d comes together with the fact that revealed prior parameters in this specification are the same as in the purely description-based lottery choices: $\hat{\alpha} = \hat{\beta} = 0.07$ both here and in the left panel. In this sense, there is no sizeable misalignment between ambiguity- and description-based decision scenarios.

Placing things on the same scale again, the number of pseudo-observations in the prior revealed to fit subjects in lottery experiments best is 3.25, while in the Jeffreys prior it is 1.00, and, for comparison, it is 23.24 (by assumption) in the confidence measure for description-based communication.

Appendix §B provides empirical and model-implied CEs as well as posterior mean and probability weighting functions illustrating the preceding results.

4.2 Assets in financial markets

Now we are going to test the performance of the model from §3.3 in capturing the decisions of investors in financial markets and the resulting asset prices.

Table 4: Estimation Results for Lottery Decisions — Information from Description and under Ambiguity

Parameters	Description			Ambiguity			Description & Ambiguity				
	CPT_2	$\mathcal{B}_{\text{Jeffreys}}$	$\mathcal{B}$	CPT_2	$\mathcal{B}_{\text{Jeffreys}}$	$\mathcal{B}$	CPT_2	$\mathcal{B}_{\text{Jeffreys}}$	$\mathcal{B}$	$\mathcal{B}^2_{\text{Jeffreys}}$	$\mathcal{B}^2$
φ^+	1.10	1.13	1.11	1.14	1.05	1.11	1.15	1.15	1.15	1.15	1.15
	(.035)	(.021)	(.017)	(.044)	(.012)	(.016)	(.032)	(.014)	(.012)	(.012)	(.012)
η^+	0.83			0.66			0.73				
	(.035)			(.027)			(.021)				
ψ^+	1.01			0.96			1.00				
	(.036)			(.046)			(.033)				
α,β		0.50	0.07		0.50	0.19		0.50	0.07	0.50	0.07
		(n/a)	(.017)		(n/a)	(.020)		(n/a)	(.020)	(n/a)	(.019)
n_d		1.00	1.00					1.00	1.00	1.00	1.00
		(n/a)	(n/a)					(n/a)	(n/a)	(n/a)	(n/a)
n_a					1.00	1.00		2.63	0.42	5.72	0.91
					(n/a)	(n/a)		(.268)	(.134)	(.487)	(.262)
R^2	0.7396	0.6872	0.7397	0.5612	0.5419	0.5616	0.6424	0.6223	0.6444	0.6223	0.6444
Obs.	858	858	858	1254	1254	1254	2112	2112	2112	2112	2112

Notes: The explained variable is observed certainty equivalents. The rows contain value function parameters φ^+; probability weighting function parameters η^+, ψ^+; the initial prior density parameters α and β; the numbers of (pseudo-) observations in the relevant experiments n_d and n_a; as well as the goodness-of-fit measure R^2 and the number of observations in the data subset (see text for a detailed description). The columns are split into three panels depending on information source (description, ambiguity or both) and present different model specifications: a cumulative prospect theory specification with 2 weighting function parameters; $\mathcal{B}(\alpha+nq, \beta+n(1-q))$ model under Jeffreys prior; and $\mathcal{B}(\alpha+nq, \beta+n(1-q))$ model under revealed prior; as well as the last two models split into two stages for the ambiguous case in the right panel (see text for a detailed description). The estimation is conducted on pooled data by maximum likelihood accounting for individual-specific error heteroscedasticity proportional to lottery range. Standard errors (in parentheses) are computed using wild bootstrap with 200 replications. Data source on laboratory experiments is Abdellaoui et al. (2011a).

Our testing approach is analogous to that of section §2.2: the candidate probability density of stock returns $\pi^M(r)$ from equation (23), together with an assumption about the probability of regime change p, CRRA coefficient γ and the time discount rate β, are the models' only inputs, while the levels and volatility of equity premium and risk-free rate, as well as the key statistical moments of consumption and dividend growth are reverse-engineered from returns' distribution using equations (3)–(5) under the general equilibrium structure of Lucas (1978).[61] The target of this test is for the model-implied quantities to match their empirical counterparts.[62] Recall that this requires a high equity premium, low risk-free rate, and high volatility of stock returns relative to that of dividend growth.

The data set used is the same as in §2.2. The empirical approach is that of calibration rather than estimation.

Our choices of γ and β are similar to those used earlier. The choice of the probability of regime change p is more involved.

Regime changes: We calibrate the regime change probability p basing on statistical properties of the time series of stock returns employing formal changepoint detection algorithms. However, in order to identify higher than quarterly-frequency changes as well as alleviate difficulties posed by short samples, for this particular task (and in contrast to the actual asset pricing application) we use daily returns data; moreover, since detection results are sensitive to test procedure settings, we run several different such procedures (considering only model-free approaches).[63] Thus, Table 5 summarizes the results for (i) a parametric Gaussian mean-variance based approach in a frequentist spirit due to Killick et al. (2012), (ii) a similar parametric approach but in Bayesian tradition suggested by Adams and MacKay (2007), and (iii) a non-parametric approach relying on Cramer–von-Mises test statistic proposed in Ross and Adams (2012). Evidently, different procedures produce very disparate results, even the number of detected regime changes varies dramatically. Appendix §D reports on findings of some straightforward statistical associations between regime changes and lags/leads of daily/monthly/quarterly real GDP growth, GDP gap, unemployment rate, NBER recessions indicator, inflation rate, federal funds rate, market returns and volatility variables. Guided by two extremes of the quarterly-frequency results in Table 5, we consider two options: $p = 1$, implying a regime change every period (a quarter, in our application); and $p = 0.2$, implying an expected

[61]That is, we solve numerically for the distribution of consumption (and dividend) growth that is consistent with above equations as well as with the $C \equiv D$ restriction (and satisfies the unimodality requirement).

[62]Strictly speaking, the properties of the distribution of consumption (dividend) growth, which is implied by the model and has to be consistent with the posterior distribution of returns considered, do not have to closely match the empirical moments of the realized consumption growth. But they should be within the range of sensibility for the model to be trustworthy.

[63]The informativeness of accounting for higher-frequency information has recently been shown in Ai and Bansal (2018), Schorfheide et al. (2018), Farmer et al. (2023), Borup et al. (2024).

Table 5: Detection results for regime changes

	P-F	P-B	NP	Observations
Daily	52	2269	145	18474
Monthly	44	613	123	873
Quarterly	43	271	93	291

Notes: The input data are log market returns. The rows present the number of detected changepoints for different frequencies, with the daily frequency results aggregated to monthly and monthly aggregated to quarterly. The columns contain results of three different model-free changepoint detection procedures: parametric frequentist method (P-F), parametric Bayesian method (P-B) and non-parametric method (NP); the last column shows the number of observations. The data sample is U.S. 1947:M4:D1–2019:M12:D31, at a daily frequency (see Appendix §G for a description of data sources).

regime duration of 5 periods (quarters).

Model performance: Full calibration results are summarized in Table 6. The top row of the Table provides the empirical moments that our model is targeting.[64] Starting with the updated Jeffreys prior developed in §3.3, we calculate the model's output under condition that the probability of regime change p is either 1.0 or 0.2; see the rows corresponding to "Model, $\mathcal{T}_{2\alpha}$ and $\mathcal{T}_3$" in Table 6 with column p containing 1.0 or 0.2, respectively. While in the latter case the model clearly underperforms, in the former case of $p = 1.0$ it is able to hit the target: for γ between 4 and 5, the equity risk premium is of the order of 6%–8%, the risk-free rate is no more than 1%. Additionally, the mean and standard deviation of consumption (and dividend) growth are at low and medium single-digit values, respectively, both of which look plausible. Also note that dividend growth volatility is 4–5 times smaller than returns volatility in this case, which replicates the "excess" equity volatility—or, given the logic of our exercise, "insufficient" dividend volatility—relationship observed in the data. In particular, setting $\gamma = 4.4$ with $p = 1.0$ yields $\mathrm{E}^{\pi}[r_f] = 0.0064$, $\mathrm{E}^{\pi}[r - r_f] = 0.0677$, $\sqrt{\mathrm{V}^{\pi}[r]} = 0.3124$, $\mathrm{E}^{\pi}[\Delta c] = \mathrm{E}^{\pi}[\Delta d] = 0.0146$ and $\sqrt{\mathrm{V}^{\pi}[\Delta c]} = \sqrt{\mathrm{V}^{\pi}[\Delta d]} = 0.0733$. Therefore, our model's investors are in agreement with the observed market values on such tradable instruments as levels of equity premium and risk-free rate or relative magnitudes of return and dividend volatilities.[65]

[64]Because we work under assumption of a non-ergodic environment in general, each empirical moment used (as given in the top row of Table 6) should be interpreted not as an estimate of a theoretical moment (e.g., mean value) and not even as a sample moment (a sample mean), but as a time average of several sample moments (an average of sample means) that investors encountered over the historical period considered. The question is whether, given such a history of macro-financial statistics, investors' behavior as reflected in market prices was, on average, consistent with the proposed model.

[65]The mean consumption (dividend) growth obtained in our calibration is 1.46%, which is not too far from empirical values of 1.29% (2.77%), adding plausibility. The corresponding standard deviation is 7.33%, which is naturally larger than 1.54% (4.13%) observed in the sample, but it is comparable to alternative results. Specifically, it is lower than the 17% standard deviation of welfare equivalent consumption growth in Weitzman (2007); it falls in between the 4.24% and 14.84% standard deviations

Table 6: Calibration Results for Investment Decisions — Model

$\pi(r)$	p	γ	$\mathrm{E}^{\pi}[r_f]$	$\sqrt{\mathrm{V}^{\pi}[r_f]}$	$\mathrm{E}^{\pi}[r-r_f]$	$\sqrt{\mathrm{V}^{\pi}[r]}$	$\mathrm{E}^{\pi}[\Delta c]$	$\sqrt{\mathrm{V}^{\pi}[\Delta c]}$	$\mathrm{E}^{\pi}[\Delta d]$	$\sqrt{\mathrm{V}^{\pi}[\Delta d]}$
Empirical, $\hat{\pi}$	n/a	n/a	0.0066	0.0131	0.0674	0.1628	0.0129	0.0154	0.0277	0.0413
Model, $\mathcal{T}_{2\alpha}$ and $\mathcal{T}_3$	1.0	1	0.0286	n/a	0.0470	0.3124	0.0656	0.3124	0.0656	0.3124
Model, $\mathcal{T}_{2\alpha}$ and $\mathcal{T}_3$	1.0	2	0.0218	n/a	0.0526	0.3124	0.0322	0.1588	0.0322	0.1588
Model, $\mathcal{T}_{2\alpha}$ and $\mathcal{T}_3$	1.0	3	0.0176	n/a	0.0566	0.3124	0.0214	0.1069	0.0214	0.1069
Model, $\mathcal{T}_{2\alpha}$ and $\mathcal{T}_3$	1.0	4	0.0112	n/a	0.0629	0.3124	0.0160	0.0805	0.0160	0.0805
Model, $\mathcal{T}_{2\alpha}$ and $\mathcal{T}_3$	1.0	5	-0.0077	n/a	0.0818	0.3124	0.0128	0.0645	0.0128	0.0645
Model, $\mathcal{T}_{2\alpha}$ and $\mathcal{T}_3$	0.2	1	0.0490	n/a	0.0253	0.2256	0.0642	0.2256	0.0642	0.2256
Model, $\mathcal{T}_{2\alpha}$ and $\mathcal{T}_3$	0.2	2	0.0476	n/a	0.0265	0.2256	0.0320	0.1141	0.0320	0.1141
Model, $\mathcal{T}_{2\alpha}$ and $\mathcal{T}_3$	0.2	3	0.0467	n/a	0.0273	0.2256	0.0213	0.0764	0.0213	0.0764
Model, $\mathcal{T}_{2\alpha}$ and $\mathcal{T}_3$	0.2	4	0.0454	n/a	0.0286	0.2256	0.0160	0.0574	0.0160	0.0574
Model, $\mathcal{T}_{2\alpha}$ and $\mathcal{T}_3$	0.2	5	0.0416	n/a	0.0324	0.2256	0.0128	0.0458	0.0128	0.0458
Model, $\mathcal{T}_{2\alpha}$ and $\mathcal{T}_2$	0.2	1	0.0462	n/a	0.0321	0.2966	0.0682	0.2966	0.0682	0.2966
Model, $\mathcal{T}_{2\alpha}$ and $\mathcal{T}_2$	0.2	2	0.0353	n/a	0.0406	0.2966	0.0329	0.1508	0.0329	0.1508
Model, $\mathcal{T}_{2\alpha}$ and $\mathcal{T}_2$	0.2	3	0.0268	n/a	0.0484	0.2966	0.0217	0.0989	0.0217	0.0989
Model, $\mathcal{T}_{2\alpha}$ and $\mathcal{T}_2$	0.2	4	0.0134	n/a	0.0615	0.2966	0.0162	0.0761	0.0162	0.0761
Model, $\mathcal{T}_{2\alpha}$ and $\mathcal{T}_2$	0.2	5	-0.0259	n/a	0.1007	0.2966	0.0129	0.0617	0.0129	0.0617

Notes: The columns present the probability density of returns with the probability of regime change; the coefficient of relative risk aversion; the log risk-free and excess log market returns, the per capita consumption growth (i.e., temporal difference in the log of) as well as the dividend growth (difference in the log of) with their statistical moments. The "Empirical" row: the data sample is U.S. 1947:Q2–2019:Q4, at a quarterly frequency (see Appendix §G for a description of data sources). The "Model" rows: $\beta = 0.99$ per annum; the CRRA utility with a given γ; and a specified probability density of returns $\pi(r)$, which here is a $(1-p,\, p)$–mixture of either $\mathcal{T}_{2\alpha_t}(\ell_t, \beta_t(1+1/\lambda_t)/\alpha_t)$ and $\mathcal{T}_3(\bar{r}_t, 4\bar{s}_t^2/3)$ or $\mathcal{T}_{2\alpha_t}(\ell_t, \beta_t(1+1/\lambda_t)/\alpha_t)$ and $\mathcal{T}_2(\bar{r}_t, 3\bar{s}_t^2/2)$ (see text for a detailed description). Model inputs are p and hypothetical $\pi^M(r)$ as well as γ and β; the rest of the columns are model outputs. The economic variables are measured in real terms, and measurements are converted into annualized values.

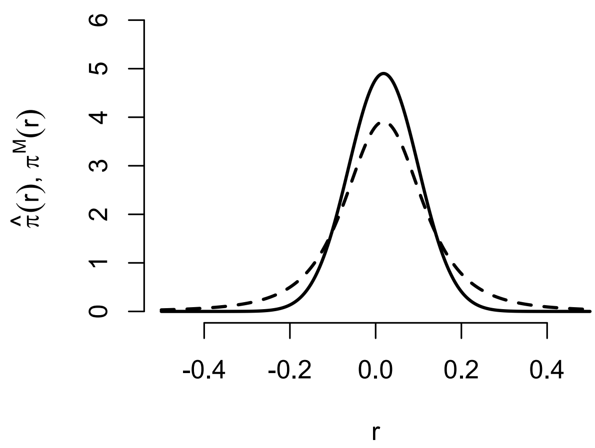


Figure 7: Empirical (solid) vs. hypothetical data-generating (dashed) probability densities (parameterizations used are $\mathcal{N}(\hat{\mu}, \hat{\sigma}^2)$ with $\hat{\mu} = 0.0185$ and $\hat{\sigma}^2 = 0.0066$ vs. $\mathcal{T}_3(\bar{r}_t, 4\bar{s}_t^2/3)$ with $\bar{r}_t = 0.0185$ and $\bar{s}_t^2 = 0.0066$ corresponding to $\pi^M(r|\cdot)$ from equation (23) with $p = 1.0$).

Importantly, although the mean level of the stock market's return effectively, via the density $\pi^M(r)$, was one of the exogenous data inputs to the model, the ultimate split within this given level into the layer of the risk-free return and a risk premium on top is the endogenous result dictated by the theoretical model. Moreover, the consumption and dividend growth statistics were in no way among the data inputs to the model, and they are entirely model-dictated.

The observed empirical probability density function $\hat{\pi}(r)$ and the calibrated hypothetical data-generating probability density $\pi^M(r)$ are plotted in Figure 7. The result is somewhat different from what we saw earlier in Figure 2: even though there is more probability mass in the central area of the distribution $\pi^M(r)$ than in the distribution $\pi^{CPT}(r)$, the remaining probability mass is allocated disproportionately more toward the far tails in the former distribution than in the latter one due to the role of heavy-tailed $\mathcal{T}_3(r)$ in $\pi^M(r)$. Still, the hypothetical "population" distribution "spreads out" in comparison to the observed sample, which is consistent with the pattern of probability weighting from section §2 (Figure 1).[66]

Additionally, in order to verify the sensitivity of our results, we consider an alternative implementation of the initial prior update. Specifically, when updating the prior with the sufficient statistics from the current regime, we will set $\tau^\star := 2$ instead of 3 that was used

of, respectively, consumption and dividend growth over 600,000 years of simulated data featuring rare booms and disasters in Tsai and Wachter (2016); it is also consistent with the finding in Malloy et al. (2009) that consumption growth of stockholders exhibits a sensitivity of factor about 3 to 4 relatively to aggregate consumption growth.

[66]The critical role of (negative) tail risks in a related context is advocated by Kozlowski et al. (2019). The importance of milder downside risks in argued in Beason and Scherindorfer (2022): for instance, compare their Figure 2 with our Figure 7.

earlier (with the motivation that 2 is the minimum sample size required for calculating 2 empirical moments, i.e., the mean $\bar{r}_\tau$ and variance $\bar{s}_\tau^2$). This modifies the mixture distribution from equation (23): the first component becomes $\mathcal{T}_{2\alpha_\tau}(\ell_\tau, \beta_\tau(1 + 1/\lambda_\tau)/\alpha_\tau)$ with a reduced value of α_τ, and the second component turns into $\mathcal{T}_2(\bar{r}_\tau, 3\bar{s}_\tau^2/2)$. As a result, the tails of the mixture acquire more probability mass.

The interesting case here is when $p = 0.2$. For these alternative calibration results, see in Table 6 the rows corresponding to "Model, $\mathcal{T}_{2\alpha}$ and $\mathcal{T}_2$" with column p containing 0.2.[67] This version of the model performs very similarly to our preferred formulation described in section §4.2 (the rows corresponding to "Model, $\mathcal{T}_{2\alpha}$ and $\mathcal{T}_3$" with column p containing 1.0 in the same Table), in particular it is able to match the equity premium and risk-free rate targets equally well. Therefore, invariant ignorance priors prove to be important and empirically successful, although specific dynamic structure of the economy embodied in the frequency of regime changes is more debatable.

5 Discussion

"Rationalization" of puzzling facts: The prospect theory's probability weighting transformations, when applied to lottery choices as well as when transferred directly into investment decisions in section §2, may simply be useful approximations of the (means of) Beta and (a mixture of) the Student's t densities that arise from a conventional Bayesian updating procedure under certain discipline about prior assumptions, according to our arguments from section §3 and calculations from §4. Specifically, this discipline stipulates combining an initial invariant ignorance prior with external prior information (judged/declared lottery probability or statistics on the latest regime) and sample information. We then use the same conceptual approach to model both the lottery choices in the lab experiments as well as the capital asset investments in financial markets, successfully matching the corresponding performance metrics. For instance, in the former case we achieve a similar or better goodness of fit as the CPT[68] with equal or lower number of free parameters, while at the same time appealing to decision-theoretic rather than data-fitting rationale. In the latter case, we match the levels of the equity risk premium and risk-free rate as well as the discrepancy between equity and dividend volatilities relying on

[67]There is a caveat to these calculations. One of the mixture density's components, a Student's t distribution with 2 degrees of freedom $\mathcal{T}_2(\bar{r}_\tau, 3\bar{s}_\tau^2/2)$, has tails that are heavy enough for causing problems with integral convergence, in particular it does not have a finite variance (while a Student's t distribution with just more than 2 degrees of freedom does). To circumvent the problem, in our numerical calculations we utilize the relationship $\int f(x)(1-p)\pi_1(x)\,\mathrm{d}x + \int f(x)p\pi_2(x)\,\mathrm{d}x = \int f(x)\left((1-p)\pi_1(x) + p\pi_2(x)\right)\mathrm{d}x$, and work directly with the integrand $f(x)\left((1-p)\pi_1(x) + p\pi_2(x)\right)$. Using the routines from the standard numerical integration package (Piessens et al., 1983), the integral converges for a low enough mixture weight p (as long as $f(x)$ does not contain too high moments of x). Effectively, such a procedure evaluates numerically a convergent integral that is a small perturbation away from our "razor-edge" case, and the result should be viewed as an approximation.

[68]Which is a very high bar to reach, according to Fudenberg et al. (2022).

plausible levels of risk-aversion, while appealing to the same decision-theoretic grounds as above (and with an addition of the regime-change probability, which is calibrated separately).

In other words, our theoretical results suggest a "rationalization" of the stylized empirical facts across two different domains, that is Kahneman–Tversky's probability distortions (along with Allais's and Ellsberg's paradoxes) in experiments with lotteries as well as the equity premium and risk-free rate puzzles in asset prices.

Alternative explanations: Alternative models aimed at rationalizing these empirical facts have been proposed, but they are usually not applicable to both micro-level (i.e., probability distortions in laboratory experiments with lottery choices) and macro-level phenomena (i.e., risk premia in asset prices on financial markets).

Micro-level theories, such as models based on rank-dependent utility (as in Epstein and Zin, 1990) or on the noise in neural information processing (e.g., Steiner and Stewart, 2016), rely on specific constraints stemming from individual preferences and costs. However, in competitive markets idiosyncratic effects usually would be "arbitraged away" by professional institutional agents, so these micro-level theories must be combined with arguments about a sufficiently large size of such "behavioral" investors for them to be marginal market participants and on the limits of arbitrage (Shleifer and Vishny, 1997). Additionally, most of the empirical evidence underlying such theories comes from experiments with small stakes, and also the magnitudes of deviations from the classical theory is known to significantly diminish with experience (see Fox et al., 1996; List, 2004; List and Haigh, 2005; Van de Kuilen and Wakker, 2006; Van de Kuilen, 2009).[69] (But, of course, the above does not say anything against the validity of such theories within their usual domain, in fact we view them as complementary to the approach proposed here.)

Conversely, macro-level theories, which emphasize the role of uncertainty about an economy's evolving structure (Weitzman, 2007), habit formation (Campbell and Cochrane, 1999) and so on, do not directly translate to static stochastic choice setups.

Unifying framework: The recognition that learning and inference from the limited available information are inherent constituents of decision-making under risk and uncertainty offers a rational unifying perspective on a theoretical normative axiomatic prescription for and empirical positive statistical description of the optimal choice in this setting. Indeed, rational behavior of our decision-makers entails that their optimal decisions, made effectively under the unobserved but inferred data-generating distribution, satisfy the standard von Neumann–Morgenstern axioms. Viewed by experimenters or econometricians under an observed but inadequate sample distribution, however, these

[69]Moreover, in the case of cognitive noise models the measures of cognitive uncertainty are increasing in the sample size, which amplifies probability distortions, thus hampering rather than helping professional investors.

same decisions seem to deviate from rationality and exhibit phenomena such at the Allais and Ellsberg paradoxes or the equity premium puzzle.

Revealed priors: Our approach to decision-making rests on ignorance prior distributions, and within this class we particularly emphasize the so-called Jeffreys priors. We found that both in decisions on binary lotteries and in decisions on capital asset investments ignorant Jeffreys priors align well with experimental results; but in the case of lottery choice the best-fitting, or revealed, priors, though being ignorant, deviate somewhat from the Jeffreys prescriptions. Specifically, the number of pseudo-observations contained in priors (i.e., their strength) revealed to fit the behavior in lottery experiments best is larger than in the Jeffreys prior: 3.25–13.94 (depending on the experiment) as opposed to 1.00. As a result, in the former case there is a more pronounced nonlinearity and overweighting of tail probabilities (their shrinkage to the center of distribution) of $\pi(p|q)$, and a relatively higher number of new (pseudo-) observations is needed to move inferred probabilities p from their initial prior to communicated values q. However, in these terms revealed priors are still weaker than the degree of confidence in description-based communication, pinned down at 23.24 pseudo-observations. Thus, the role of ignorant priors is secondary to communicated information. In this sense, revealed priors are fairly close to the Jeffreys solution.

It is worth noting that higher number of prior pseudo-observations is beneficial in the environments when effective sample size of any new observations is less than nominal and data should not be taken at face value. For example, in the presence of serially correlated data, which is famously not the case with asset returns but prevalent in many other natural processes. A related point is that data-generating mechanism for new observations may be more trustworthy in the case of large public markets than in small laboratories. In addition to the above, robustness to any information-processing noise may warrant priors with higher numbers of pseudo-observations. While in equilibrium several models and priors each tuned to a different environment (cf. Gerd Gigerenzer's adaptive toolbox) or different degree of agents' sophistication (as shown by Bruhin et al., 2010) are likely to coexist, it is probably not surprising that a more professional, larger-stake, competitive and repeated nature of public markets results in the aggregate prior there being more attuned with the actual environment as well as with the fully rational behavior than what has been found in laboratory experiments.

Source confidence: Our confidence measure (represented by n_a, n_d, n_e, as well as $\tau^\star$) quantifies the weight of the communicated information about a given data-generating mechanism (or a source of uncertainty) in terms of (pseudo-) observations, with directly sampled information serving as a gauge for informativeness of the intermediary sources. It provides a mutually consistent treatment of ambiguous, described and experienced information in ergodic environments as well as information from the current and past regimes in non-ergodic ones. Taking the same prior distribution in judgement-, description- and experience-based cases, we are able to compare confidence measures across contexts: (a) empirically, confidence measures in description- and experience-based scenarios are very close — 23.24 and 19.57 (pseudo-) observations, respectively, which suggests that after all they represent very similar sources of uncertainty[70]; (b) and even though on the surface it would appear to be much lower under ambiguity, an explicit approach recognizing the two-part nature of the underlying data-generating mechanism finds the degree of confidence that aligns closely with the other two communication scenarios — 21.15 pseudo-observations separately for the upstream and for the downstream sources of uncertainty, which suggests that in all three cases data-generating (sub) mechanisms are treated about equally except for the need to compound them in the more complex judgement-based case. Admittedly, there is a large modern literature measuring experimental differences between lottery choices in ambiguity-, description- and experience-based scenarios (Abdellaoui et al., 2011a; Hertwig et al., 2004; Abdellaoui et al., 2011b), and the advantage of our approach is reducing them to a single-scalar and readily interpretable measure.

Instrumentally for a decision-maker the role of confidence measure is to modulate the relative influence of communicated information and (ignorant) prior, with a higher degree of confidence leading to a more linear probability weighting and thinner tails of the posterior distribution.

Shrinkage: We obtained a manifestation of the well-known shrinkage phenomenon. Originally, shrinkage was proposed for better estimation of the mean of a multivariate Normal distribution from an ultra-small sample of a single vector observation (Stein, 1956; James and Stein, 1961).[71] The concept of Stein-type shrinkage stimulated the reinterpretation of old as well as the development of new methods of statistical analysis, such as the Good–Turing frequency estimation in computational linguistics and machine learning (Good, 1953); least absolute shrinkage and selection operator, or LASSO, estimation in statistics and machine learning (Tibshirani, 1996); and improved estimates of the variance-covariance matrix (Ledoit and Wolf, 2004a; Jagannathan and Ma, 2003) and of the mean vector (Jorion, 1985; 1986) of portfolio returns in finance.

In our case, it amounts to deforming ("shrinking") the distribution of interest in the direction toward the invariant ignorance prior represented by a quasi-uniform distribution ("shrinkage target"). This quasi-uniform prior distribution serves as an "Occam factor" that favors simple models. Effectively, it countervails overfitting the observed data, which is a major concern in the situation of undersampling. In particular, distributions of the lotteries' (transformed) probabilities are shrunk toward the uniform distribution on the $[0, \pi/2]$ interval, thus biasing the small probabilities upward and the large probabilities

[70]Quantitatively, comparable magnitudes arise if, for instance, confidence n_d is high enough, while subjects are able to draw enough observations for n_e to catch up with it.

[71]In this connection, also see the regularization literature, for example, Chen and Haykin (2002), Bickel and Li (2006).

downward. Distributions of the asset returns' (log-) variances and (conditional) means are both shrunk toward the uniform distributions on $\mathbb{R}$, prudently amplifying the uncertainty and skepticism about these returns-driving parameters. (See Appendix §E for more details.)

Dynamics: The reason our agents rely on ignorance priors that lead to overweighting of tail probabilities is the constantly evolving environment they operate in and the ensuing scarcity of information. This problem is less severe in the case of different lotteries in the lab (e.g., in most experiments subjects do not have to guess when a new lottery is played). In the case of financial markets, temporally aggregated quarterly-frequency data available to consumption-based asset pricing modelers may be concealing the changes in the underlying returns-generating probability distributions that happen at higher frequencies. Applying model-free changepoint detection algorithms to daily-frequency data, we pin down the quarterly probability of such regime changes at $p = 1$ in our preferred calibration or $p = 0.2$ in the alternative specification. What we then get is a story of perpetual changes that are (a) more frequent than "rare disasters" (cf. Veronesi, 2004; Barro, 2009) or even than not-so-rare parameter/model breaks (Smith and Timmermann, 2021; and even Farmer et al., 2023; Borup et al., 2024), (b) usually fairly moderate in magnitude (see Beason and Scherindorfer, 2022), as well as (c) symmetric and so can be good or bad (as in Tsai and Wachter, 2016). Since changes of this objective frequency and magnitude lead to substantial amplification of subjective tail probabilities, their effect is similar (at least in temporally aggregated sense) to that of extrapolative beliefs suggested in, e.g., Bordalo et al. (2019).[72]

Within the same environment, agents update their beliefs about the underlying parameters and future random variable realizations by digesting new sample data. For example, as more observations of successes or failures from the same lottery are accumulated, deviations from linearity in probability weights and relative overweighting of small probabilities gradually vanish: effectively, n_e value increases so that posterior mean of p approaches q in equation (12) — this is "learning". Importantly, an explicit test confirms that accounting for the number of sampled draws does improve the explanatory performance of our model for the results of laboratory experiments with decisions from experience. Moreover, as rational agents accumulate familiarity with a given class of data-generating mechanisms, they will learn more accurately the optimal values of relevant confidence parameters. As a human subject participates in more laboratory experiments offering a similar kind of lottery choices (but with different values of q), α and β should decrease relatively to n_e if data prove to be accurate, serially uncorrelated, and the subject learns how to better process them, although it is an empirical question how closely they will get to the Jeffreys solution — one can think of this as "adaptation"; separately, n_a or n_d should increase

[72] Also note that our periodic learning re-initializations in response to parameter changes can be alternatively interpreted as a discrete version of the continuous discounting of past data.

relatively to α and β as the subject verifies the reliability of a given source of uncertainty — this can be interpeted as "confidence accumulation" (now we are talking about the steps beyond the third step of the algorithm in §3.2). In the case of financial markets, depending on their own investment history agents may analogously consider that some regime changes are less dramatic than others, and may correspondingly adapt their priors, but it is again an empirical question how they will move relatively to the Jeffreys values.

6 Conclusion

This paper in motivated by some stylized facts about the paradoxical choice behavior humans exhibit in laboratory experiments as well as about the asset pricing puzzles found in financial markets, drawing empirical parallels between these two decision-making domains. Arguably, these stylized facts reflect the human aspiration to account for unobserved contingencies when dealing with only small samples of relevant data, a situation that arises routinely in practice in the form of new gambles or new economic regimes.

Indeed, we can rationalize such decision-making behavior by taking an explicitly dynamic perspective and appealing to Bayesian updating under the requirements of parameterization invariance and maximum entropy. Effectively, this formulates rational optimality-motivated foundations behind the probability distortions captured by the CPT of Kahneman and Tversky in microeconomics and decision theory, as well as behind the equity premium/risk-free rate and some other puzzles in macro-finance. Our theoretical results found empirical support, they are consistent both with the observations from laboratory experiments and with the data on financial market prices.

In future research, it would be interesting to conduct laboratory experiments with lottery decisions shedding further light on the factors explaining the variation in the strength of ignorance priors revealed by human subjects, as well as on the relationship between specific characteristics of different sources of uncertainty and the confidence amounts that decision-makers assign to them, with a particular focus on how the behavior changes with time. It would also be helpful to perform a deeper analysis of regime changes in macro-financial time series, notwithstanding the role of data sampling frequency as well as robustness to detection procedures used and maintained model hypotheses. On the theoretical side, the search for more definitive implementations of the principle of indifference as well as the methods of maximum entropy and invariance to reparameterization in canonical decision problems remains an important challenge lying ahead.

References

[1] Abdellaoui, Mohammed. (2000) "Parameter-Free Elicitation of Utility and Probability Weighting Functions", *Management Science*, 46(11): 1497–1512.

[2] Abdellaoui, Mohammed, Aurélien Baillon, Laetitia Placido and Peter P. Wakker. (2011a) "The Rich Domain of Uncertainty: Source Functions and Their Experimental Implementation", *American Economic Review*, 101(2): 695–723.

[3] Abdellaoui, Mohammed, Olivier L'Haridon and Corina Paraschiv. (2011b) "Experienced vs. Described Uncertainty: Do We Need Two Prospect Theory Specifications?" *Management Science*, 57(10): 1879–1895.

[4] Adams, Ryan Prescott and David J.C. MacKay. (2007) "Bayesian Online Changepoint Detection" *arXiv*, 0710.3742.

[5] Ai, Hengjie and Ravi Bansal. (2018) "Risk Preferences and the Macroeconomic Announcement Premium", *Econometrica*, 86(4): 1383–1430.

[6] Allais, Maurice. (1953) "Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Americaine", *Econometrica*, 21(4): 503–546.

[7] Ang, Andrew and Allan Timmermann. (2012) "Regime Changes and Financial Markets", *Annual Review of Financial Economics*, 4: 313–337.

[8] Augenblick, Ned and Matthew Rabin. (2021) "Belief Movement, Uncertainty Reduction, and Rational Updating", *The Quarterly Journal of Economics*, 136(2): 933–985.

[9] Aydogan, Ilke. (2021). "Prior Beliefs and Ambiguity Attitudes in Decision from Experience", *Management Science* 67(11): 6934–6945.

[10] Baele, Lieven, Joost Driessen, Sebastian Ebert, Juan M. Londono, Oliver G. Spalt. (2019) "Cumulative Prospect Theory, Option Returns, and the Variance Premium", *The Review of Financial Studies*, 32(9): 3667–3723.

[11] Bansal, Ravi and Amir Yaron. (2004) "Risks for the Long-Run: A Potential Resolution of Asset Pricing Puzzles." *Journal of Finance*, 59(4), 1481–1509.

[12] Barberis, Nicholas, Robin Greenwood, Lawrence Jin, and Andrei Shleifer. (2015) "X-CAPM: An extrapolative capital asset pricing model", *Journal of Financial Economics*, 115(1): 1–24.

[13] Barberis, Nicholas, Ming Huang, Tano Santos. (2001) "Prospect Theory and Asset Prices", *The Quarterly Journal of Economics*, 116(1): 1–53 .

[14] Barberis, Nicholas and Ming Huang. (2008) "Stocks as Lotteries: The Implications of Probability Weighting for Security Prices", *American Economic Review*, 98(5): 2066–2100.

[15] Barberis, Nicholas, Lawrence J. Jin and Baolian Wang. (2021). "Prospect Theory and Stock Market Anomalies", *The Journal of Finance*, 76(5): 2639–2687.

[16] Barro, Robert J. (2009). "Rare Disasters, Asset Prices, and Welfare Costs." *American Economic Review*, 99(1): 243–264.

[17] Beason, Tyler and David Schreindorfer. (2022) "Dissecting the Equity Premium", *Journal of Political Economy*, 130(8): 2203–2222.

[18] Bickel, Peter J. and Bo Li (2006) "Regularization in Statistics." *Test*, 15(2): 271-344.

[19] Bordalo, Pedro, Nicola Gennaioli and Andrei Shleifer. (2018) "Diagnostic Expectations and Credit Cycles", *The Journal of Finance*, 73(1): 199–227.

[20] Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer. (2019). "Diagnostic Expectations and Stock Returns", *The Journal of Finance*, 74(6): 2839–2874.

[21] Borup, Daniel, Jonas N. Eriksen, Mads M. Kjær, and Martin Thyrsgaard. (2024) "Predicting Bond Return Predictability", *Management Science*, 70(2): 931–951.

[22] Bossaerts, Peter, Kerstin Preuschoff and Ming Hsu. (2008). "The Neurobiological Foundations of Valuation in Human Decision Making Under Uncertainty." In: Paul W. Glimcher, Colin F. Camerer, Ernst Fehr, Russell A. Poldrack (eds.), *Neuroeconomics: Decision Making and the Brain*, Academic Press.

[23] Bruhin, Adrian, Helga Fehr-Duda and Thomas Epper. (2010) "Risk and Rationality: Uncovering Heterogeneity in Probability Distortion", *Econometrica*, 78(4): 1375–1412.

[24] Camerer, Colin F. and Teck-Hua Ho. (1994) "Violations of the Betweenness Axiom and Nonlinearity in Probability", *Journal of Risk and Uncertainty*, 8(2): 167–196.

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

[26] Chen, Zhe and Simon Haykin. (2002). "On Different Facets of Regularization Theory." *Neural Computation*, 14: 2791–2846.

[27] Chernoff, Herman and Lincoln E. Moses. (1959) *Elementary Decision Theory*, New York: Wiley.

[28] Chew, Soo Hong, Bin Miao and Songfa Zhong. (2017) "Partial Ambiguity", *Econometrica*, 85(4): 1239–1260.

[29] Collin-Dufresne, Pierre, Michael Johannes, and Lars A. Lochstoer. (2016). "Parameter Learning in General Equilibrium: The Asset Pricing Implications", *American Economic Review*, 106(3): 664–698.

[30] De Giorgi, Enrico G. and Shane Legg. (2012) "Dynamic portfolio choice and asset pricing with narrow framing and probability weighting", *Journal of Economic Dynamics and Control*, 36: 951–972.

[31] DeGroot, Morris H. (1970) *Optimal Statistical Decisions*, McGraw-Hill.

[32] Dimson, Elroy, Paul Marsh and Mike Staunton. (2003) "Global Evidence on the Equity Risk Premium". *LBS Institute of Finance and Accounting Working Paper*, No. IFA 385.

[33] Ellsberg, Daniel. (1961) "Risk, Ambiguity, and the Savage Axioms Get access Arrow", *The Quarterly Journal of Economics*, 75(4): 643–669.

[34] Enke, Benjamin and Thomas Graeber. (2019) "Cognitive Uncertainty", *SSRN Working Paper*, No. 3489380.

[35] Epstein, Larry G. and Stanley E. Zin. (1990) "'First-Order' Risk Aversion and the Equity Premium Puzzle." *Journal of Monetary Economics*, 26(3): 387–407.

[36] Fama, Eugene F. and Kenneth R. French. (2002) "The Equity Premium", *The Journal of Finance*, 57: 637–659.

[37] Farmer, Leland E., Lawrence Schmidt, and Allan Timmermann. (2023). "Pockets of Predictability", *The Journal of Finance*, 78(3): 1279–1341.

[38] Farmer, Leland E., Emi Nakamura, and Jón Steinsson. (2024). "Learning about the Long Run", *Journal of Political Economy*, 132(10), 3334–3377.

[39] Fehr-Duda, Helga and Thomas Epper. (2012) "Probability and Risk: Foundations and Economic Implications of Probability-Dependent Risk Preferences", *Annual Review of Economics*, 4: 567–593.

[40] Fox, Craig R., Brett A. Rogers and Amos Tversky. (1996) "Options Traders Exhibit Subadditive Decision Weights", *Journal of Risk and Uncertainty*, 13(1): 5–17.

[41] Fox, Craig R. and Amos Tversky. (1995) "Ambiguity Aversion and Comparative Ignorance", *The Quarterly Journal of Economics*, 110(3): 585–603.

[42] Frydman, Cary, Lawrence J. Jin. (2022) "Efficient Coding and Risky Choice", *The Quarterly Journal of Economics*, 137(1): 161–213.

[43] Fudenberg, Drew, Jon Kleinberg, Annie Liang, and Sendhil Mullainathan. (2022) "Measuring the Completeness of Economic Models", *Journal of Political Economy*, 130(4): 956–990.

[44] Gelman, Andrew, John B. Carlin, Hal S. Stern, Donald B. Rubin. (2004) *Bayesian data analysis*, 2nd edition, Chapman & Hall.

[45] Ghaderi, Mohammad, Mete Kilic, Sang Byung Seo. (2022) "Learning, slowly unfolding disasters, and asset prices", *Journal of Financial Economics*, 143(1): 527–549.

[46] Gigerenzer, Gerd and Daniel G. Goldstein. (1996) "Reasoning the Fast and Frugal Way: Models of Bounded Rationality", *Psychological Review*, 103(4): 650–669.

[47] Gigerenzer, Gerd, Ralph Hertwig and Thorsten Pachur (eds.). (2011) *Heuristics: The Foundations of Adaptive Behavior*, Oxford University Press.

[48] Goldstein, William M., and Hillel J. Einhorn. (1987) "Expression theory and the preference reversal phenomena", *Psychological Review*, 94(2): 236–254.

[49] Gonzalez, Richard, and George Wu. (2022). "Composition rules in original and cumulative prospect theory", *Theory and Decision*, 92: 647–675.

[50] Good, I.J. (1953). "The Population Frequencies of Species and the Estimation of Population Parameters", *Biometrika*, 40(3/4): 237–264.

[51] Grether, David M. (1980). "Bayes Rule as a Descriptive Model: The Representativeness Heuristic", *The Quarterly Journal of Economics*, 95(3): 537–557.

[52] Grünwald, Peter D. (2007) *The Minimum Description Length Principle*, The MIT Press, Cambridge, MA.

[53] Halevy, Yoram. (2007) "Ellsberg Revisited: An Experimental Study", *Econometrica*, 75(2): 503–536.

[54] Hamilton, James D. (2016) "Macroeconomic Regimes and Regime Shifts." In: John B. Taylor and Harald Uhlig (eds.), *Handbook of Macroeconomics*, volume 2, part A, chapter 3, pages 163–201, Elsevier.

[55] Hausser, Jean and Korbinian Strimmer. (2009) "Entropy Inference and the James-Stein Estimator, with Application to Nonlinear Gene Association Networks", *Journal of Machine Learning Research*, 10: 1469–1484.

[56] Hertwig, Ralph, Greg Barron, Elke U. Weber and Ido Erev. (2004) "Decisions from Experience and the Effect of Rare Events in Risky Choice", *Psychological Science*, 15(8): 534–539.

[57] Hogarth, Robin M., Hillel J. Einhorn. (1992) "Order effects in belief updating: The belief-adjustment model", *Cognitive Psychology*, 24(1): 1–55.

[58] Jagannathan, Ravi and Tongshu Ma. (2003) "Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps." *Journal of Finance*, 58(4): 1651–1684.

[59] James, W. and Charles M. Stein. (1961) "Estimation with Quadratic Loss", Proc. Fourth Berkeley Symp. Math. Statist. Prob. 1: 361–379.

[60] Jaynes, Edwin Thompson. (1957a) "Information Theory and Statistical Mechanics", *Physical Review*, 106(4): 620–630.

[61] Jaynes, Edwin Thompson. (1957b) "Information Theory and Statistical Mechanics. II", *Physical Review*, 108(2): 171–190."

[62] Jaynes, Edwin T. (2003) *Probability Theory: The Logic of Science*, Cambridge University Press.

[63] Jeffreys, Harold. (1939) *Theory of Probability*, Oxford University Press.

[64] Jeffreys, Harold. (1946) "An Invariant Form for the Prior Probability in Estimation Problems," *Proceedings of the Royal Society A*, 186(1007): 453–461.

[65] Johannes, Michael, Lars A. Lochstoer, Yigun Mou. (2016). "Learning about Consumption Dynamics", *The Journal of Finance*, 71(2): 551–600.

[66] Jorion, Philippe. (1985) "International Portfolio Diversification with Estimation Risk." *Journal of Business*, 58(3): 259–278.

[67] Jorion, Philippe. (1986) "Bayes-Stein Estimation for Portfolio Analysis." *Journal of Financial and Quantitative Analysis*, 21(3): 279–292.

[68] Juechems, Keno, Jan Balaguer, Bernhard Spitzer, and Christopher Summerfield. (2021) "Optimal utility and probability functions for agents with finite computational precision", *Proceedings of the National Academy of Sciences*, 118(2): e2002232118.

[69] Kahneman, Daniel and Amos Tversky. (1979) "Prospect Theory: An Analysis of Decision under Risk", *Econometrica*, 47(2): 263–291.

[70] Kass, Robert E. and Larry Wassermann. (1996) "The Selection of Prior Distributions by Formal Rules", *Journal of the American Statistical Association*, 91(435): 1343–1370.

[71] Keynes, John Maynard. (1921) *A Treatise on Probability*, London: Macmillan & Co.

[72] Khaw, Mel Win, Ziang Li, and Michael Woodford. (2021) "Cognitive Imprecision and Small-Stakes Risk Aversion", *The Review of Economic Studies*, 88(4): 1979–2013.

[73] Khaw, Mel Win, Ziang Li, and Michael Woodford. (2023) "Cognitive Imprecision and Stake-Dependent Risk Attitudes", *NBER Working Paper*, No. 30417.

[74] Killick, R., P. Fearnhead and I. A. Eckley. (2012) "Optimal Detection of Changepoints With a Linear Computational Cost" *Journal of the American Statistical Association*, 107(500): 1590–1598.

[75] Kozlowski, Julian, Laura Veldkamp and Venky Venkateswaran. (2019) "The Tail That Keeps the Riskless Rate Low", *NBER Macroeconomics Annual 2019*, 33:, 253–283.

[76] Kozlowski, Julian, Laura Veldkamp and Venky Venkateswaran. (2020) "The Tail That Wags the Economy: Beliefs and Persistent Stagnation". *Journal of Political Economy*, 128(8): 2839–2879.

[77] Laplace, Pierre-Simon. (1825) *Philosophical Essay on Probabilities*, London: Chapmand & Hall.

[78] Ledoit, Olivier and Michael Wolf (2004a) "A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices." *Journal of Multivariate Analysis*, 88(2): 365–411.

[79] List, John A. (2004) "Neoclassical Theory Versus Prospect Theory: Evidence from the Marketplace", *Econometrica*, 72(2): 615–625.

[80] List, John A.and Michael S. Haigh. (2005) "A Simple Test of Expected Utility Theory Using Professional Traders", *Proceedings of the National Academy of Sciences*, 102(3): 945–948.

[81] Lucas, Robert E., Jr. (1978) "Asset Prices in an Exchange Economy." *Econometrica*, 46 (6): 1429–1445.

[82] MacKay, David J. C. (2003) *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press.

[83] Malloy, Christopher J., Tobias J. Moskowitz and Annette Vissing-Jørgensen (2009). "Long-Run Stockholder Consumption Risk and Asset Returns." *Journal of Finance*, 64(6): 2427–2479.

[84] McGrattan, Ellen R. and Edward C. Prescott (2003) "Average Debt and Equity Returns: Puzzling?" *American Economic Review*, 93: 392–397.

[85] McGrattan, Ellen R. and Edward C. Prescott (2005) "Taxes, Regulations, and the Value of U.S. and U.K. Corporations", *Review of Economic Studies*, 92: 767–796.

[86] Mehra, Rajnish and Edward C. Prescott. (1985) "The Equity Premium: A Puzzle", *Journal of Monetary Economics*, 15(2): 145–161.

[87] Minka, Thomas P. (2001) "Inferring a Gaussian distribution." Manuscript.

[88] Murphy, Kevin P. (2012) *Machine Learning: A Probabilistic Perspective*, The MIT Press.

[89] Nemenman, Ilya, Fariel Shafee and William Bialek. (2002) "Entropy and Inference, Revisited," In: T. G. Dietterich, S. Becker, and Z. Ghahramani (eds.), *Advances in Neural Information Processing Systems*, 14, MIT Press.

[90] Netzer, Nick, Arthur Robson, Jakub Steiner, Pavel Kocourek. (2021) "Endogenous Risk Attitudes", *SSRN Working Paper*, No. 4024773.

[91] Orlitsky, Alon, Narayana P. Santhanam, Junan Zhang. (2003) "Always Good Turing: Asymptotically Optimal Probability Estimation", *Science*, 302(5644): 427–431.

[92] Paninski, Liam. (2003) "Estimation of Entropy and Mutual Information", *Neural Computation*, 15: 1191–1253.

[93] Peterson, Joshua C., David D. Bourgin, Mayank Agrawal, Daniel Reichman, and Thomas L. Griffiths. (2021). "Using large-scale experiments and machine learning to discover theories of human decision-making", *Science*, 372: 1209–1214.

[94] Piessens, Robert, Elise de Doncker-Kapenga, Christoph W. Überhuber, David Kahaner. (1983) *QUADPACK: A Subroutine Package for Automatic Integration*, Springer-Verlag.

[95] Polkovnichenko, Valery and Feng Zhao. (2013) "Probability Weighting Functions Implied in Options Prices", *Journal of Financial Economics*, 107: 580–609.

[96] Rietz, Thomas. (1988) "The Equity Risk Premium: A Solution." *Journal of Monetary Economics*, 22: 117–131.

[97] Rissanen, Jorma. (1983) "A Universal Prior for Integers and Estimation by Minimum Description Length", *The Annals of Statistics*, 11(2): 416–431.

[98] Ross, Gordon J., Niall M. Adams. (2012) "Two Nonparametric Control Charts for Detecting Arbitary Distribution Changes" *Journal of Quality Technology*, 44: 102–116.

[99] Schorfheide, Frank, Dongho Song and Amir Yaron. (2018) "Identifying Long-Run Risks: A Bayesian Mixed-Frequency Approach", *Econometrica*, 86(2): 617–654.

[100] Segal, Uzi. (1987) "The Ellsberg Paradox and Risk Aversion: An Anticipated Utility Approach", *International Economic Review*, 28(1): 175–202.

[101] Segal, Uzi. (1990) "Two-Stage Lotteries without the Reduction Axiom", *Econometrica*, 58(2): 349–377.

[102] Shiller, Robert J. (1981) "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?", *The American Economic Review*, 71(3): 421–436.

[103] Shleifer, Andrei, and Robert W. Vishny. (1997) "The Limits of Arbitrage", *The Journal of Finance*, 52(1): 35–55.

[104] Simon, Herbert A. (1955) "A Behavioral Model of Rational Choice", *Quarterly Journal of Economics*, 69(1): 99–118.

[105] Simon, Herbert A. (1956) "Rational Choice and the Structure of the Environment", *Psychological Review*, 63(2): 129–138.

[106] Smith, Simon C., Allan Timmermann. (2021) "Break Risk", *The Review of Financial Studies*, 34(4): 2045–2100.

[107] Stein, Charles M. (1956), "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Distribution", Proc. Third Berkeley Symp. Math. Statist. Prob. 1: 197–206.

[108] Steiner, Jakub and Colin Stewart. (2016) "Perceiving Prospects Properly," *American Economic Review*, 106(7): 1601–1631.

[109] Tibshirani, Robert. (1996) "Regression Shrinkage and Selection via the Lasso", *Journal of the Royal Statistical Society. Series B*, 58(1): 267–288.

[110] Timmermann, Allan G. (1993) "How Learning in Financial Markets Generates Excess Volatility and Predictability in Stock Prices", *The Quarterly Journal of Economics*, 108(4): 1135–1145.

[111] Todd, Peter M., Gerd Gigerenzer and the ABC Research Group. (2012) *Ecological Rationality: Intelligence in the World*, Oxford: Oxford University Press.

[112] Tsai, Jerry and Jessica A. Wachter. (2016) "Rare Booms and Disasters in a Multisector Endowment Economy", *Review of Financial Studies*, 29(5): 1113–1169.

[113] Tversky, Amos and Daniel Kahneman. (1992) "Advances in Prospect Theory: Cumulative Representation of Uncertainty", *Journal of Risk and Uncertainty*, 5: 297–323.

[114] Valiant, Gregory and Paul Valiant. (2017) "Estimating the Unseen: Improved Estimators for Entropy and other Properties", *Journal of the ACM*, 64(6)((37)): 1–41.

[115] Van de Kuilen, Gijs, and Peter P. Wakker. (2006)."Learning in the Allais paradox", *Journal of Risk and Uncertainty*, 33, 155–164.

[116] Van de Kuilen, Gijs. (2009) "Subjective Probability Weighting and the Discovered Preference Hypothesis", *Theory and Decision*, 67: 1–22.

[117] Vieider, Ferdinand M. (2024) "Decisions Under Uncertainty as Bayesian Inference on Choice Options", *Management Science*.

[118] Veronesi, Pietro. (2004) "The Peso Problem Hypothesis and Stock Market Returns", *Journal of Economic Dynamics and Control*, 28: 707–725.

[119] Veronesi, Pietro. (1999). "Stock Market Overreactions to Bad News in Good Times: A Rational Expectations Equilibrium Model Get access Arrow", *The Review of Financial Studies*, 12(5): 975–1007.

[120] Vuong, Quang H. (1989) "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses" *Econometrica*, 57(2): 307–333.

[121] Wachter, Jessica A. (2013). "Can Time-Varying Risk of Rare Disasters Explain Aggregate Stock Market Volatility?" *The Journal of Finance*, 68(3): 987–1035.

[122] Weil, Philippe. (1989) "The Equity Premium Puzzle and the Risk-Free Rate Puzzle." *Journal of Monetary Economics*, 24: 401–421.

[123] Weitzman, Martin L. (2007) "Subjective Expectations and Asset-Return Puzzles." *American Economic Review*, 97 (4): 1102–1130.

[124] Wu, George and Richard Gonzalez. (1996) "Curvature of the Probability Weighting Function", *Management Science*, 42(12): 1676–1690.

[125] Zhang, Hang, Laurence T. Maloney. (2012) "Ubiquitous Log Odds: A Common Representation of Probability and Frequency Distortion in Perception, Action, and Cognition", *Frontiers in Neuroscience*, 6(1): 1–14.