Log, Stock and Two Simple Lotteries*

Sergiy Verstyuk†

This version: 22 January 2018

Abstract

This paper studies the problem of decision-making under risk by agents whose information processing abilities may be limited. The constructed theoretical framework grounds on findings from economic laboratory experiments, incorporates existing neuroscience knowledge, and is implemented using information-theoretic formalism. Activation of the above information-processing constraints distorts the subjective perception of the objective stochastic environment the agent operates in, and the constrained-optimal decision-making requires appropriate adjustments. In the selected application, a general equilibrium macro-finance model, such biases of subjective perspective as overconfidence, pessimism and categorization emerge endogenously. Categorization manifests itself via dropping from consideration the less important principal dimensions ("dispersion folding") and amplifying the random variables' (dis)similarity ("correlation inflation"). These theoretical results contribute to the understanding of such empirical regularities as the portfolio underdiversification puzzle and style investing phenomenon.

*Acknowledgements: Pietro Veronesi, George M. Constantinides, Doron Ravid, Matt Taddy; as well as Jörn Boehnke, Christian Julliard, Leonid Kogan, Nicholas Polson, Xiao Qiao, Philip J. Reny, Lawrence D. W. Schmidt, Harald Uhlig, Michael Woodford.

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

Contents

The limits of my language mean the limits of my world.
—Ludwig Wittgenstein

1 Introduction

This paper studies the problem of investment under conditions of risk while explicitly accounting for the constraints imposed by the limited cognitive—or information processing—capabilities of decision-makers (but recognizing nevertheless rationality of the latter, and their endeavor to make the most of it). We ask how such limitations bias the subjective perspective on the objective stochastic reality, and how they are manifested in the ultimate investment decisions and observed market prices. For instance, how exactly the value of investment opportunities is assessed and acted upon.

The "complexity-reducing simplification" arising as a result of such constraints generally distorts the agent's perception of the stochastic environment he/she operates in. Applying these simplification mechanics to the problem of investment portfolio choice induces investors to substitute the "complex" original objective probability distribution defining the investment returns with its proxy: the subjective distribution characterized by relatively lower variance ("overconfidence") and decreased mean ("pessimism"). The degree of these biases in the investor's subjective perspective depends on the information processing capacity available to him. If the capacity is sufficiently large, the biases vanish and the subjective perspective coincides with the objective picture, bringing about the benchmark case presumed by conventional "rational expectations". However, even when the capacity limit takes its toll, but the perspective is biased in an optimal way, i.e., the simplification of the subjective distribution is implemented according to a rational adjustment procedure that we specify, then the effects of "overconfidence" and "pessimism" cancel each other out and resulting decisions do not systematically deviate from the optimal ones. That is, investment choices are approximately equal to their unconstrained counterparts and ultimate economic outcomes are close to those achieved under "rational expectations", yet less computational resources are utilized.

Listing the theoretical contributions, in addition to "overconfidence" and "pessimism", within the presented framework endogenously emerges such phenomenon as "categorization", in the sense of differentiation into distinct categories (e.g., different "models of the world", or different asset classes). Specifically, (i) less crucial random variables and data dimensions are "folded" and dropped, simplifying the perceived environment; (ii) correlations are amplified, subjectively clustering covarying random variables together. All being popular mechanisms in theoretical work, they are usually imposed exogenously rather than explicitly derived from first principles.[1]

[1] Finishing the list of theoretical results, the take-aways of a relatively more auxiliary nature and technical flavor include a convenient method for description/encoding of probability distributions (see §F); as well as an introduction of the distinction between effective and physical information processing capacities, which explains some confusing empirical measurements (see §F, §H and §D).

The above theoretical results receive certain empirical support, contributing to understanding of such regularities as the "portfolio underdiversification puzzle" and (variations in) style investing. In particular, the presented framework's theoretical prediction that agents with lower information processing capacity should be more susceptible to categorization is confirmed by observations on style investing behavior in financial markets (that is, decision-making and/or trading in terms of aggregates, i.e. asset classes): indeed, relatively less sophisticated retail investors demonstrate more evidence of such activity than more professional institutional investors do.

In the current paper, information processing is formally modeled as information transmission between (or, equivalently, information storage to/retrieval from) different parts of the brain. Evaluating the merits of available decisions, e.g., by calculating the expected value of a lottery, requires sending a message that describes the corresponding probability distribution from one, perceptive part of agent's brain to another, calculating part, and computing the statistic of interest. However, if the probability distribution is "too complex" relative to the agent's information processing limitations (his individual "bandwidth"), the procedure becomes infeasible. In such a case the distribution has to be "simplified": the decision-maker chooses another probability distribution that roughly approximates the original one but possesses lower "complexity", and uses this subjective simplified distribution in computations of the relevant statistic.

Resorting to technical jargon, the complexity of the distribution is formally quantified by its "Shannon entropy" (a certain measure of the distribution's dispersion), the complexity of the message discribing it is proportional to the complexity of the distribution itself (due to the entropy's relation to the length of efficient communication/compression code); the message is transmitted via the communication channel, and the costs of information transmission are formalized by the capacity limit of this communication channel as measured by the "mutual information" between channel input and output (a specific constraint imposed on the relationship between entropies of ingoing and outgoing messages). Thus, a probability distribution is "too complex" if, loosely speaking, its entropy is larger than the channel capacity.

Before proceeding, it is worth clarifying that the theory of "rational inattention" proposed by Sims (1998, 2003, 2006) also uses the mutual information constraint to formalize the costs of information transmission, but there are fundamental differences with the approach taken here. Conceptually, Sims's theory restricts the agent's external perceptions and focuses on uncertainty about the current state, while this paper deals with constraints on the agent's internal cognition and centers on uncertainty about the future state. Operationally, in rational inattention models the state has already realized but is not yet observed, and the agent, before observing the state and making a decision, chooses an optimal information acquisition strategy that lets him learn the realized state as accurately as his limited capacity allows, which in turn implies that every period nature is sending to an agent the information about one realization of a random variable. While in the present paper the state has not yet realized, and the agent, before making a decision and before realization of the state, chooses an optimal information processing strategy that lets him summarize/represent the space of possible states as accurately as his limited capacity allows, which in turn implies that every period one part of agent's brain is sending to another part of agent's brain the information about the whole probability distribution of a random variable. Thus, the differences are in the timing of events, the object of approximation, the task faced by the agent, the players and interactions involved, as well as in the dimensionality of transmitted messages. See §G for more details about the difference.

Then, the "architecture" of our framework builds from the bottom up: taking simple lotteries as primitives (arguably, lotteries to students of choice under uncertainty are what fruit flies are to geneticists) and completing the construction with the tree model in the style of Lucas (1978) (a structural microfounded general equilibrium workhorse model for stocks in macroeconomics and finance).

The "mortar" binding this construction contains several key ingredients of methodological nature. The process of choice under uncertainty is implemented using information-theoretic formalism (see Appendix §F for details, also see §G). Empirically, it is grounded on experimental evidence: for instance, on the work of Gabaix and Laibson (2000), who study how humans make decisions under risk, and whose results can be interpreted as being consistent with their subjects resorting to complexity/entropy-reducing approximations; together with the overwhelming findings on categorical nature of human visual and auditory perception (e.g., Goldstone and Hendrickson, 2010; Fleming et al., 2013); on other systematic violations of (naïvely unbounded) rational normatives, including demonstrations of "overconfidence" (Camerer, 1995); as well as measurements of implied information processing capacity bounds (see the classical paper by Miller, 1956, whose very title "The Magical Number Seven, Plus or Minus Two" reflects how small the estimates of such capacity in bits are). It is also guided by existing knowledge in neuroscience, for example, explicitly incorporating the concepts such as "working memory" or "bottleneck", as well as the mechanisms such as coordinate transformations and adjustments or recursive processing (see Appendix §H).

Lastly, we touch upon some technical characteristics of the basic framework underlying the analysis. The (macro-) economic setup is deliberately primitive and conventional: a workhorse Lucas tree model, which prevents contamination of the analysis with details not crucial to the main narrative and aids tractability. The fundamental assumptions are very frugal: the strongest one is the log-Normality of the stock returns. The optimization problem is framed in a modular format: there are two segregated sub-problems, which improves transparency as well as facilitates potential modifications and extensions. The formal model admits analytical closed-form solutions: in potential applications, including the one pursued here, the solution is as tractable as the standard canonical case, in spite of being more general.

The present research is connected to several distinct literature clusters, and their exhaustive review is beyond the scope of this paper (an abbreviated overview is offered in Appendix §D). The most closely related works are the following. Gabaix (2014a, 2014b) proposes a similar approach to the simplification of the environment, which envisages discrete and sparse representation of data. The rational inattention theory (Sims, 2003, 2006; Matějka and Sims, 2010; Matějka and McKay, 2014; Ravid, 2016) has successfully introduced information-theoretic methods into economics; it also uses the channel capacity (mutual information) constraint to model the transmission of information, although in a structurally different form and with different aims (see the text above, as well as Appendix §G). Woodford (2012, 2014) departs from the rational inattention theory and uses the channel capacity mechanism while taking related neuroscientific and experimental evidence very seriously, which makes his papers quite closely connected to this paper. Alternative theoretical models that consider simplification and categorization have been formulated, albeit taking the latter as exogenously pre-determined (e.g., Mullainathan, 2002b; Jehiel, 2005). Empirical illustrations in this paper come from the finance area, they are mostly related to the style investing literature such as Barberis and Shleifer (2003) or Peng and Xiong (2006). Other works in this area similarly concerned with the expectation formation process and likewise motivated by existing psychological evidence along with the agents' desire for simplification are Fuster et al. (2012) and Bordalo et al. (2016) (also see an older work by Carroll, 2003; as well as a recent survey by Manski, 2017). Our main differences lie along the dimensions of the area of application, conceptual and methodological approach, parsimony and tractability, experimental and neuroscientific inputs employed.

2 Theory

2.1 Simple lotteries, hard decisions

The aim here is to examine how decisions under risk[2] are made in a setting where the relevant probability distributions are "too complex" to be used in the agent's optimization process.

As a motivating example, consider the following decision problem. Figure 1 presents the payoffs and their corresponding probabilities for two simple lotteries. In principle, from the information given one can calculate all the necessary characteristics of the payoff distributions. For instance, a risk-neutral player cares only about the mean, which is just a probability-weighted sum of each lottery's payoffs — a very simple computation. A player with a mean-variance utility function cares about the first two moments of the payoff distributions — a slightly more involved computation. But what if these arithmetic

[2]In this paper, we deal only with the notion of risk, which defines a situation when probabilities associated with a random variable are known, and only the actual random variable realizations are not known. We ignore the notion of Knightian uncertainty, in which even the probabilities themselves are unknown. The terms "risk" and "uncertainty" are thus used interchangeably throughout.

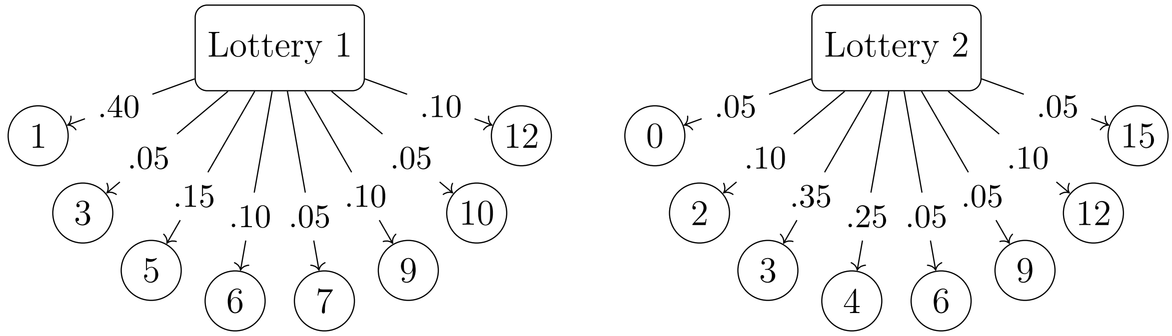

Figure 1: Two lotteries.

operations, however elementary they are, can not be performed in full? Say, because of a time constraint: think of having only 1 second per contingency, i.e., 16 seconds to choose between the two lotteries (in case you tried it, see the footnote[3]).

Formally, the problem looks as follows:

$$\max_{\theta \in \{1,2\}} \mathrm{E}\left[\varphi(\varrho(\boldsymbol{x}|\theta))\right], \tag{1}$$

where the state vector $\boldsymbol{x}$ is idiosyncratically distributed as $\boldsymbol{x} \sim G(\boldsymbol{x})$, the lottery choice $\theta \in \{1, 2\}$, the payoff function $\varrho(\boldsymbol{x}|\theta)$ depends on the chosen lottery, and $\varphi(\varrho)$ is a player's decision criterion (i.e., some kind of felicity function). The player sees all the payoffs as well as the probabilities and has to estimate the expected payoff (or, strictly speaking, expected felicity) under the time constraint.

Gabaix and Laibson (2000) investigate a more complicated version of the same problem in laboratory conditions. They use multinomial recombining trees, each consisting of 10 root nodes and 4 to 9 levels of leaf nodes that are connected by probabilistic edges, with intermediate payoffs in every node; the goal is to select a root node with the highest expected value. Their paper proposes the following heuristic for evaluating the tree payoffs: consider only transit probabilities larger than a certain cutoff probability and calculate the expected payoff ignoring the less probable edges (it is required that the payoffs have a zero mean, and there are no extreme outlier payoffs). The authors interpret this decision rule as simulating the future by identifying typical or representative scenarios. They conduct an experiment where human subjects have to evaluate 12 such trees within 40 minutes, and the proposed algorithm most accurately matches the empirical distribution of choices, in particular, outperforming the fully rational model of behavior.

Effectively, the above method reduces the computational costs of evaluating the expectation in (1) by carefully changing the distribution of the random variable in focus to one with lower "uncertainty" (or entropy) at the cost of some "approximation error".

[3]The lottery on the right has higher mean (5.0 vs. 4.9) and lower variance (14.5 vs. 14.6).

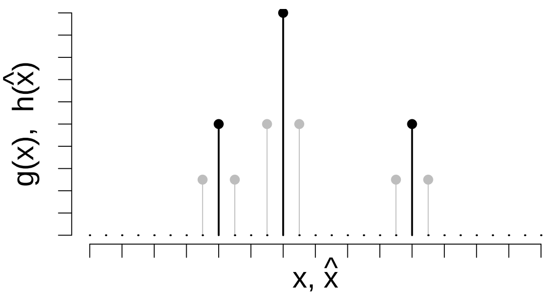

Figure 2: Entropy reduction primer
(higher-entropy PMF $g(\cdot)$ (in grey) vs. lower-entropy $h(\cdot)$ (in black)).

2.2 Algorithm for decision-making under risk

Our next task is to formalize the idea that the problem of decisions under risk can be simplified via entropy reduction. Appendix §F outlines the algorithm in detail. Here (as well as in §G) we provide an intuitive overview of its key features.

Figure 2 presents the probability mass function (PMF) $g(\cdot)$ for some random variable of interest x, that is a state variable such as next year's real GDP growth rate or a company's revenues. It is rather "complex", with "fine" partitions that allow for "many" (6 here) possible outcomes. In real life, decision-makers usually reduce the space of outcomes to a refined ensemble of just a few possible scenarios: say, a "baseline", "optimistic" and "pessimistic". Such a reduction makes the problem easier to analyze as well as to present and discuss it in business meetings, conference calls, or in written communication. In Figure 2 this is represented by PMF $h(\cdot)$, which is for convenience defined over a new synthetic random variable $\hat{x}$. This new distribution is relatively "simple", with "coarse" partitions that allow for just a "few" (3 here) possible outcomes.

The complexity (or uncertainty, dispersion) of a random variable can be measured by its entropy. Formally, (Shannon) entropy of a discretely distributed random variable x is defined as

$$\mathcal{E}(g(x)) := -\sum_{i=1}^{|\text{supp}(g)|} g(x_i) \log g(x_i), \tag{2}$$

where $|\text{supp}(g)|$ is the cardinality of the support set of $g(\cdot)$.[4,5] Units of measurement are either *bits* (when the logarithms of base 2 are used), or *nats* (for natural logarithms), with 1 nat $:= \log_2 e \approx 1.44$ bits.

For the discrete distributions primer presented in Figure 2, the "complex" distribution $g(\cdot)$ is characterized by higher entropy than the "simple" distribution $h(\cdot)$, so that replacing one with the other in the process of simplification leads to a reduction in the entropy measure. When the space of outcomes is continuous, the so-called differential entropy just invokes integration in place of summation in the above formula (2). The entropy

[4] We use $\log(\cdot)$ for $\log_2(\cdot)$ and $\ln(\cdot)$ for $\log_e(\cdot)$ throughout.

[5] Shannon entropy is a concept originating in the information theory. For a textbook treatment of the information theory, see Cover and Thomas (2006) or MacKay (2003).

reduction can be thought of as a reduction of dispersion, which in the case of continuous distributions is associated with a lower variance and the thinner tails of a distribution.[6]

While from a statistical perspective the entropy of a random variable is some measure of its dispersion, there is another side to the coin. From the information-theoretic perspective, (Shannon) entropy of a random variable is the average length of code that can be used to efficiently carry information about the variable's outcomes. For example, the discrete distribution $h(\hat{x})$ presented in Figure 2 envisages three possible realizations with probabilities $\{0.25, 0.5, 0.25\}$. According to equation (2), this implies the entropy of 1.5 bits, which in turn means that the average codeword length is the same 1.5 bits. Indeed, such a binary alphabet would be $\{00, 1, 01\}$ for the corresponding realizations of the random variable $\hat{x}$ (with codeword "1" used on average twice more frequently than the other two). This would be a more compressed representation than what could be achieved for the higher-entropy distribution $g(x)$ that entails six different realizations with probabilities $\{0.125, 0.125, 0.25, 0.25, 0.125, 0.125\}$, which implies the entropy as well as the average codeword length of 2.5 bits on the basis of alphabet $\{000, 001, 10, 11, 010, 011\}$. Moreover, the above allows to deduce that the length of the code that summarizes the distribution itself is proportional to its entropy.

The instrumental value of this dispersion-code duality is that such a code permits us to (constructively) quantify the demands on the computational or, more broadly, information processing capacity. Consider a simple lottery represented again by distribution $g(\cdot)$ or $h(\cdot)$ from Figure 2, whose numerical payoff probabilities are listed in the preceding paragraph, and with some arbitrary payoff values, e.g., as those in Appendix §F (or §G).

Intuitively, one can think of the task of evaluating this lottery as (a) learning and importing the complex probability distribution $g(\cdot)$ into the external-perceptive part of the brain; (b) drawing realizations from it in a Monte Carlo-type experiment; and (c) transmitting these values via a communication channel to the internal-cognitive part of the brain that calculates the statistics of interest using the transmitted random draws. A communication channel that has a limited capacity would require shorter codes, which are made possible by the simple distribution $h(\cdot)$. As the number of draws increases, the simulated statistics converge to their exact theoretical values, albeit those corresponding to distribution $h(\cdot)$ rather than $g(\cdot)$. Below we provide a slightly more technical account of the mechanisms involved, but skipping to part §2.3 should not preclude understanding the rest of the paper.

Algorithm: Structurally speaking, in our framework the process of mental evaluation of a lottery comprises (i) the summarization of the given information about the lottery

[6]At first pass, it may seem counterintuitive that such a simplifying approximation unambiguously leads to a variance reduction, rather than the perturbations potentially going either way. The reasons are of the technical, information/measure-theoretic nature, and should be clear from the detailed exposition in Appendix §F. However, the unambiguous direction of the change in the variance is an important feature of the landscape indeed, as demonstrated in the rest of this paper.

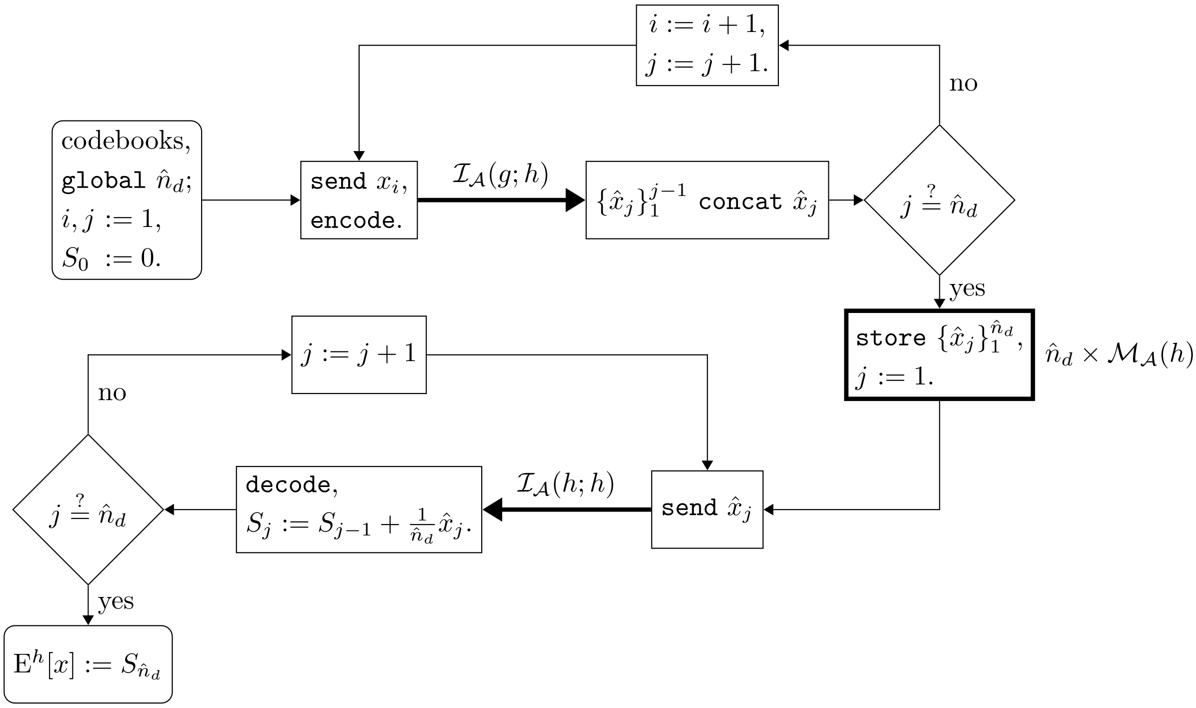

Figure 3: Information processing algorithm primer (potential bottlenecks shown in bold).

distribution (via a communication channel transmitting such a description), (ii) the loading and providing access to this information within a working memory (allocated in the memory storage), (iii) the calculation of lottery's value (via a communication channel transmitting intermediate iterations of such computations). Potential information processing "bottlenecks" may arise at any of the three milestones above, depending on which of the capacity constraints is binding: description channel capacity, storage memory capacity, or computation channel capacity. Such bottlenecks on the way of information flow can preclude the procedure's smooth completion that is necessary for making an optimal decision, with complex distribution $g(\cdot)$ being more susceptible to this predicament than simple distribution $h(\cdot)$. Figure 3 offers a sneak preview of the formal algorithm's mechanics at this particular juncture.

The leading illustration is probably memory and its capacity, which is relatively well studied in neuroscience (see Appendix §H) as well as recognized as an important concept in the economic literature (see Appendix §D for examples). Specifically, cognitive psychology and neuroscience define working memory as a limited capacity system that temporarily maintains and stores information to support human thought processes by providing an interface between perception, long-term memory and action (Baddeley, 2003). A probability distribution that is characterized by high entropy may require too complex a code to represent it, which puts unrealistic demands on the working memory's capacity and prohibits the utilization of a given distribution in mental information processing. Thus, appealing to neuroscience permits us to (structurally) motivate the above quantification of information-processing costs.

The exposition in Appendix §F makes the above statements formal and expands them, demonstrating how general the postulated framework is (not only providing for alternative information processing "bottlenecks", but also allowing for alternative deterministic or probabilistic Monte Carlo-type information processing schemes). In terms of functionality, the algorithm in Appendix §F starts with an evaluation of a simple lottery (formalized as ancillary procedure $\mathcal{P}_{\int}$) and at the end provides a recipe for the evaluation of expectations of arbitrary functions of random variables that may be required for decision-making under risk.

2.3 Information constraint and informational problem

Now we are going to introduce an important object for subsequent exploration. Available effective capacity (from now on, information processing capacity) κ serves as a bound in the information processing capacity/mutual information constraint (in short, information constraint):

$$\begin{aligned}\kappa \geq \mathcal{I}(g(\boldsymbol{x}); h(\hat{\boldsymbol{x}})) = \mathcal{E}(g(\boldsymbol{x})) - \mathcal{E}(f(\boldsymbol{x}|\hat{\boldsymbol{x}})) = \mathcal{E}(h(\hat{\boldsymbol{x}})) - \mathcal{E}(f(\hat{\boldsymbol{x}}|\boldsymbol{x})) = \\ = \mathcal{E}(g(\boldsymbol{x})) + \mathcal{E}(h(\hat{\boldsymbol{x}})) - \mathcal{E}(f(\boldsymbol{x}, \hat{\boldsymbol{x}}))\end{aligned} \tag{3}$$

(note the switch to random vectors, typed in boldface).

Information constraint is the instrumental take-away of the abstract algorithm presented in Appendix §F as well as in part §2.2 above. It quantifies and puts a bound on the costs (in bits of information processing capacity) to utilize the probability density $h(\hat{\boldsymbol{x}})$, for example, to calculate its expected value in the process of solving a maximization problem. The above density is a simplified proxy for a given original probability density $g(\boldsymbol{x})$, where the copy can be exact if the original density is already simple enough. To clarify the notation, $f(\boldsymbol{x}, \hat{\boldsymbol{x}})$ above is an ancillary function that captures the overall structure of stochastic interrelationships in the forthcoming optimization problem; it is a joint multivariate probability density function of $\boldsymbol{x}$, which is distributed according to its marginal density $g(\boldsymbol{x})$, and of $\hat{\boldsymbol{x}}$, which in turn is distributed according to marginal density $h(\hat{\boldsymbol{x}})$.

Intuitively, with the complexity of a random variable $\boldsymbol{x}$ fixed, one can reduce the information-processing costs by letting $\boldsymbol{x}$ being only imperfectly represented by a variable $\hat{\boldsymbol{x}}$ and carry some random "approximation" errors via raising the "uncertainty"/entropy of the conditional distribution of $\boldsymbol{x}$ given $\hat{\boldsymbol{x}}$ (refer to the RHS of the first equality in the top row of information constraint equation 3); e.g., in Figure 2 knowing the realization of $\hat{\boldsymbol{x}}$ leaves a half-half chance of guessing the realization of $\boldsymbol{x}$. Equivalently, if a random variable $\hat{\boldsymbol{x}}$ is simple enough, then there is no need for further entropy reduction, and mutual information may be equated with entropy of $\hat{\boldsymbol{x}}$ by letting the distribution of $\hat{\boldsymbol{x}}$ conditionally on $\boldsymbol{x}$ being degenerate with zero entropy, in the sense that a variable $\boldsymbol{x}$ contains all the information about a variable $\hat{\boldsymbol{x}}$ (the RHS of the second equality in the top row of equation 3); e.g., in Figure 2 knowing $\boldsymbol{x}$ makes $\hat{\boldsymbol{x}}$ a certainty.

Alternatively, information constraint (3) trades-off how fine [coarse] probability measure $h(\hat{\boldsymbol{x}})$ is (middle term of equation in the last row) versus how [in]accurate approximation of $\boldsymbol{x}$ by $\hat{\boldsymbol{x}}$ is (rightmost term), taking $g(\boldsymbol{x})$ (leftmost term) as given. Put differently, with $g(\boldsymbol{x})$ fixed, to satisfy the information constraint one can either (i) reduce the entropy of $h(\hat{\boldsymbol{x}})$ by making it a coarser probability measure, or (ii) increase the entropy of $f(\boldsymbol{x}, \hat{\boldsymbol{x}})$ by making $\hat{\boldsymbol{x}}$ a less accurate approximation of $\boldsymbol{x}$. Note that this is exactly what the heuristic of Gabaix and Laibson (2000) amounts to in the end: a simplified distribution characterized by lower "uncertainty"/entropy at the cost of some "approximation error". In addition to the experimental findings of Gabaix and Laibson (2000), such behavior is also in agreement with neuroscientific evidence on humans' categorical perception (Goldstone and Hendrickson, 2010; Fleming et al., 2013). It is also worth mentioning that the information constraint has the same form as in Sims (2003, 2006), but here it has a different motivation and interpretation (more on this in §1 and §G).

Furthermore, it is important to emphasize that κ is a measure of effective rather than available full physical capacity $\mathcal{K}^*$, which in principle may be orders of magnitude larger. Recognizing the practically unavoidable inefficiencies in capacity utilization (in particular, due to suboptimal information encoding) may explain—at first sight, implausibly—low empirical measurements of implied information processing capacity (going all the way back to Miller, 1956).

Henceforth, we refer to the original random variable $\boldsymbol{x}$ and its probability distribution $g(\boldsymbol{x})$ as to "true", objective, unconstrained entities, while labeling the simplified random variable $\hat{\boldsymbol{x}}$ and its distribution $h(\hat{\boldsymbol{x}})$ as "approximating", subjective, constrained counterparts to and versions of the former.

Now, let us formulate what we call the informational problem, $\mathcal{P}_{\mathcal{I}}$:

$$\min_{f(\boldsymbol{x},\hat{\boldsymbol{x}})} \mathrm{E}^f[d(\boldsymbol{x}, \hat{\boldsymbol{x}})] = \int_{\mathrm{supp}(h)} \int_{\mathrm{supp}(g)} d(\boldsymbol{x}, \hat{\boldsymbol{x}})\, f(\boldsymbol{x}, \hat{\boldsymbol{x}})\, d\boldsymbol{x}\, d\hat{\boldsymbol{x}} \qquad \{\mathcal{P}_{\mathcal{I}}\}$$

subject to the information constraint

$$\mathcal{I}(g(\boldsymbol{x}); h(\hat{\boldsymbol{x}})) \leq \kappa, \qquad [\lambda]$$

as well as the necessary technical restrictions (hereafter assumed implicitly)

$$\int_{\mathrm{supp}(h)} f(\boldsymbol{x}, \hat{\boldsymbol{x}})\, d\hat{\boldsymbol{x}} = g(\boldsymbol{x}) \qquad \forall \boldsymbol{x} \in \mathrm{supp}(g), \qquad [\mu(\boldsymbol{x})]$$

$$f(\boldsymbol{x}, \hat{\boldsymbol{x}}) \geq 0 \qquad \forall \boldsymbol{x} \in \mathrm{supp}(g), \hat{\boldsymbol{x}} \in \mathrm{supp}(h). \qquad [\nu(\boldsymbol{x}, \hat{\boldsymbol{x}})]$$

An agent seeks to minimize the expected "distortion" from using the approximating distribution $h(\hat{\boldsymbol{x}})$ instead of the true distribution $g(\boldsymbol{x})$ subject to information constraint, i.e., given a bound $\kappa > 0$ on the mutual information between these two distributions (Lagrange multipliers on each constraint are specified on the right in square brackets). The distortion function $d(\boldsymbol{x}, \hat{\boldsymbol{x}})$ (i.e., a distance metric for its two arguments) is taken as given. (The choice of appropriate distortion function is problem-specific, we will discuss it later in part §3.2.)[7]

Solution to problem $\mathcal{P}_{\mathcal{I}}$ is provided in Proposition 1.

Proposition 1 (General Solution to Informational Problem)**.** *Let $\boldsymbol{x}$ be a random vector distributed according to an absolutely continuous probability distribution function $G(\boldsymbol{x})$ with a probability density function $g(\boldsymbol{x})$, $d(\boldsymbol{x}, \hat{\boldsymbol{x}})$ be a distortion function for vectors $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$ satisfying the condition*

$$\exists\, \hat{\boldsymbol{x}} : \quad \int_{\mathrm{supp}(g)} d(\boldsymbol{x}, \hat{\boldsymbol{x}})\, g(\boldsymbol{x})\, d\boldsymbol{x} < \infty.$$

Then solution to the informational problem specified in $\mathcal{P}_{\mathcal{I}}$ is given by a conditional probability density

$$f(\boldsymbol{x}|\hat{\boldsymbol{x}}) = \exp\left(\frac{1}{\lambda}\nu(\boldsymbol{x}, \hat{\boldsymbol{x}}) - \frac{1}{\lambda}\mu(\boldsymbol{x}) - \frac{1}{\lambda}d(\boldsymbol{x}, \hat{\boldsymbol{x}})\right), \qquad \forall \hat{\boldsymbol{x}} \in \mathrm{supp}(h).$$

Proof. See Appendix §E.1.

□

The form of the solution may seem counterintuitive at first. However, the key is to realize that the conditional distribution of $\boldsymbol{x}$ given $\hat{\boldsymbol{x}}$ that it provides is actually the distribution of "approximation error", i.e., the deviation of the original random variable $\boldsymbol{x}$ from its simplified counterpart $\hat{\boldsymbol{x}}$. Given the knowledge of $g(\boldsymbol{x})$, it is often easier to proceed by just guessing the density $h(\hat{\boldsymbol{x}})$, which we are chiefly interested in, and then verifying that, together with the deduced conditional distribution, the implied joint density

$$f(\boldsymbol{x}, \hat{\boldsymbol{x}}) := f(\boldsymbol{x}|\hat{\boldsymbol{x}})h(\hat{\boldsymbol{x}})$$

satisfies the necessary requirements. Later in §3.3 we will demonstrate how this can be done with an example.

Whenever the information constraint does not bind, the expected distortion can be reduced to zero, because then $f(\boldsymbol{x}|\hat{\boldsymbol{x}})$ becomes the Dirac delta function centered at $\hat{\boldsymbol{x}}$, $\delta(\boldsymbol{x} - \hat{\boldsymbol{x}})$, $\forall \boldsymbol{x} \in \mathrm{supp}(g), \hat{\boldsymbol{x}} \in \mathrm{supp}(h)$; also $\mathrm{supp}(h) := \mathrm{supp}(g)$; and $\hat{\boldsymbol{x}} = \boldsymbol{x}$, $\forall \boldsymbol{x} \in \mathrm{supp}(g)$, almost surely; as well as $h(\boldsymbol{x}) = g(\boldsymbol{x})$, $\forall \boldsymbol{x} \in \mathrm{supp}(g)$ (thereby inducing "rational expectations"). (More on this later.)

Lastly, we claim that the following property holds.

[7] In terms of Appendix §F's algorithm, our understanding is that in practice the informational problem is solved before Generating codebook at the Simplification step of the algorithm. Basically, this is the fundamental source of information-processing cost savings that we ultimately benefit from: by bearing the fixed costs at the Simplification step, the variable costs are saved in the following steps of the algorithm, which may lead to dramatic overall savings as the latter costs accumulate very quickly during the numerous iterations required to execute the remaining steps. This rules out a kind of infinite regress critique.

Proposition 2 (Flexible Mean Property)**.** *Problem $\mathcal{P}_\mathcal{I}$ always admits the solution such that* $\mathrm{E}^h[\hat{\boldsymbol{x}}] + \check{\boldsymbol{\mu}} = \mathrm{E}^g[\boldsymbol{x}]$ *(provided the latter exists) for any bias $\check{\boldsymbol{\mu}} \in \mathbb{G}$, where $\mathbb{G}$ is some sufficient (extension) field for* domain(g)*.*

Proof. Trivially, the information constraint restricts only the mutual information between the random variables $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$, which depends on the range but not the domain of probability density functions.

□

Thus, Proposition 2 allows to treat the mean of the simplified random variable, $\mathrm{E}^h[\hat{\boldsymbol{x}}]$, as a free parameter in the formulation of problem $\mathcal{P}_\mathcal{I}$ and as some exogenously defined control in the corresponding solution of Proposition 1.

In the following part, we apply the presented theoretical constructs to a suitable model economy.

3 Model

Arguably, the limitations on the complexity of the computations that agents are able to undertake, as well as the distortions in perception and the deviations of decisions from the optimal ones that may be induced by such limitations are germane to many situations with uncertain outcomes, and to investment choices in particular.

3.1 Investment portfolio choice problem

Economic setup: Consider the following economic setting, which is essentially a variation of Lucas (1978) tree model (also refer to Breeden, 1979). For an easier exposition, first we formulate a two-period model, and an infinite-horizon extension follows later.

A representative agent with a lifespan of two periods lives in an exchange economy with opportunities to invest competitively in 1 risk-free and K risky assets. Risky assets are composed of one-period-lived "trees". The unit prices and quantities of shares in the risky trees purchased in period t are denoted by $\boldsymbol{P}_t$ and $\boldsymbol{q}_t$, respectively. The investments in them bring stochastic dividends $\boldsymbol{D}_{t+1}$, or "fruits", at the beginning of period $(t+1)$. A K-sized random vector $\boldsymbol{D}_{t+1}$ is distributed, given $\boldsymbol{D}_t$, according to probability density function $g_D(\boldsymbol{D}_{t+1}|\boldsymbol{D}_t)$. The risk-free asset is also composed of a one-period-lived tree. The unit price and quantity of shares in the risk-free tree purchased in period t are denoted by $P_{0,t}$ and $q_{0,t}$, respectively. The investments in it bring deterministic dividends $D_{0,t+1}$, the same type of fruits as above, at the beginning of period $(t+1)$. A constant scalar $D_{0,t+1}$ is normalized to 1 for all periods t. The fruits are perishable, output can not be stored between periods. We denote by C_t the agent's time-t consumption, and by $u(C_t)$ his per-period utility function, assumed to have a constant relative risk-aversion form, that is discounted at subjective rate β. This endowment economy comprises $\widehat{\boldsymbol{q}}$, a strictly positive K-sized constant vector, of risky trees, whose shares are initially owned by the representative agent. A risk-free tree is fictitious, the economy comprises $\widehat{q}_0 = 0$ of them, i.e., it exists in zero net supply and can be thought of as a cash credit technology.

Problem: The consumer-investor is interested in solving the following consumption and portfolio choice problem, $\mathcal{P}_q$:

$$\max_{C_t, \{q_{0,t}, \boldsymbol{q}_t\}} \{u(C_s) + \beta \mathrm{E}^g_t \left[u(C_{t+1})\right]\} = \{u(C_s) + \beta \int_{\mathbb{R}^K_+} u(C_{t+1})\, g_D(\boldsymbol{D}_{t+1}|\boldsymbol{D}_t)\, d\boldsymbol{D}_{t+1}\} \quad \{\mathcal{P}_q\}$$

subject to budget constraints

$$C_t + P_{0,t} q_{0,t} + \boldsymbol{P}_t^\intercal \boldsymbol{q}_t = q_{0,t-1} + (\boldsymbol{P}_t + \boldsymbol{D}_t)^\intercal \boldsymbol{q}_{t-1},$$
$$C_{t+1} = q_{0,t} + \boldsymbol{D}_{t+1}^\intercal \boldsymbol{q}_t,$$

control variables' domain restriction $C_t, \{q_{0,t}, \boldsymbol{q}_t\} \in \mathbb{R}_+ \times \mathbb{R}^{K+1}$, with $\{q_{0,t-1}, \boldsymbol{q}_{t-1}\}$ and $\boldsymbol{D}_t$ given, with $u(C_t) = C_t^{1-\gamma}/(1-\gamma)$, and where

$$g_D(\boldsymbol{D}_{t+1}|\boldsymbol{D}_t) \text{ is given.}$$

In words, the representative agent chooses consumption and investment values that maximize his current and expected future utility that at the same time satisfy the budget constraints. The expectation is taken with respect to a given objective probability density function that defines the distribution of the stochastic fruit-dividends (which are produced by tree-assets the agents invests in, and which are then eaten as consumption goods).

Infeasibility in general: Using the notation introduced previously in §2, finding optimal solution to problem $\mathcal{P}_q$ requires directly maximizing

$$\mathrm{E}^g_t \left[\varphi^\sharp(\boldsymbol{x}|\boldsymbol{\theta})\right] := u(W_t - \{P_{0,t}, \boldsymbol{P}_t\}^\intercal \boldsymbol{\theta}) + \beta \mathrm{E}^g_t \left[u([1\ \boldsymbol{x}^\intercal]\boldsymbol{\theta})\right],$$

with W_t and $\{P_{0,t}, \boldsymbol{P}_t\}$ known, as well as with $\boldsymbol{x} := \boldsymbol{D}_{t+1}$, and $\boldsymbol{\theta} := \{q_{0,t}, \boldsymbol{q}_t\}$.

Our consumer-investor is assumed to know (say, to have learned by time t) the structure of the problem, i.e., the exact specification of the utility function $u(\cdot)$, the felicity function $\varphi^\sharp(\boldsymbol{\theta}, \boldsymbol{x})$, and the (stationary) distribution $g(\boldsymbol{x})$. Thus, the environment in terms of stochasticity is as primitive as possible. Nevertheless, in general the problem $\mathcal{P}_q$ is infeasible to solve: potentially it violates the information processing capacity constraint. Even though the agent obtains $g(\cdot)$ as an input into the algorithm of Appendix §F, he may be unable to execute the full procedure.

Because of his limited information processing capacities—required to deal with random variables and, in particular, to form expectations about the functions of random variables—such an investor will necessarily have to use subjective "distorted" probability distributions instead of objective ones. The former are less costly, but leave room for certain discrepancies in the computations, and hence in the perceived landscape of the stochastic environment and the resulting investment decisions.

Therefore, the agent focuses instead on maximizing

$$\mathrm{E}_t^h\left[\varphi(\hat{\boldsymbol{x}}|\boldsymbol{\theta})\right] := u(W_t - \{P_{0,t}, \boldsymbol{P}_t\}^\intercal\boldsymbol{\theta}) + \beta\mathrm{E}_t^h\left[u([1\ \hat{\boldsymbol{x}}^\intercal]\boldsymbol{\theta})\right],$$

where

$$h(\hat{\boldsymbol{x}}) \text{ solves } \mathcal{P}_{\mathcal{I}} \text{ given } d(\boldsymbol{x}, \hat{\boldsymbol{x}}) \text{ and } \kappa.$$

To start with, note that the non-stochastic time-t objects are unaffected. However, the latest formulation makes it clear that in solving the stochastic optimization problem, we are using the subjective probability density $h(\hat{\boldsymbol{x}})$ in place of the objective density $g(\boldsymbol{x})$, with the discrepancy between the two densities depending on the available information processing capacity κ. While the distortion function $d(\boldsymbol{x}, \hat{\boldsymbol{x}})$ is chosen to be just some reasonable measure of distance between $\varphi^\sharp(\boldsymbol{x}|\boldsymbol{\theta})$ and $\varphi(\hat{\boldsymbol{x}}|\boldsymbol{\theta})$. Of course, a high enough capacity (denote it as $\kappa^\sharp$) allows density $h(\cdot)$ not to diverge from $g(\cdot)$, reducing the expected distortion to zero. More explicit formulations can be found in Appendix §I.[8]

3.2 Feasible investment portfolio choice problem

Problem: A feasible version of our consumption and investment problem would recognize that the information constraint may be binding, in contrast to the previous formulation that effectively assumes the constraint is not binding, or is "slack". That is, the consumption and portfolio choice problem $\mathcal{P}_q$ has to be combined with the informational problem $\mathcal{P}_{\mathcal{I}}$. This is done without loss of generality, but allows to relax the straitjacket discipline of the standard problem's restrictions. The solution to the informational problem would provide an optimal (with respect to the distortion metric used) probability density function $h(\cdot)$, with respect to which the investment problem will in turn be solved.

Thus, a feasible version of the consumption and portfolio choice problem, $\mathcal{P}_{q\mathcal{I}}$, is formulated as follows:

$$\max_{C_t, \{q_{0,t}, \boldsymbol{q}_t\}} \{u(C_s) + \beta\mathrm{E}_t^h\left[u(C_{t+1})\right]\} = \{u(C_s) + \beta\int_{\mathbb{R}_+^K} u(C_{t+1})\, h_D(\hat{\boldsymbol{D}}_{t+1}|\hat{\boldsymbol{D}}_t)\, d\hat{\boldsymbol{D}}_{t+1}\} \quad \{\mathcal{P}_{q\mathcal{I}}\}$$

subject to budget constraints

$$\begin{aligned} C_t + P_{0,t}q_{0,t} + \boldsymbol{P}_t^\intercal\boldsymbol{q}_t &= q_{0,t-1} + (\boldsymbol{P}_t + \hat{\boldsymbol{D}}_t)^\intercal\boldsymbol{q}_{t-1}, \\ C_{t+1} &= q_{0,t} + \hat{\boldsymbol{D}}_{t+1}^\intercal\boldsymbol{q}_t, \end{aligned}$$

[8]Our understanding is that the solution to the informational problem $\mathcal{P}_{\mathcal{I}}$ has already been learned by time t (e.g., calculated and made available in some effective analogue of abstract formal symbolic-like representation) or is just computationally easy relative to the solution of the unconstrained problem $\mathcal{P}_q$.

control variables' domain restriction $C_t, \{q_{0,t}, \boldsymbol{q}_t\} \in \mathbb{R}_+ \times \mathbb{R}^{K+1}$, with $\{q_{0,t-1}, \boldsymbol{q}_{t-1}\}$ and $\hat{\boldsymbol{D}}_t$ given, with $u(C_t) = C_t^{1-\gamma}/(1-\gamma)$, and where

$$h_D(\hat{\boldsymbol{D}}_{t+1}|\hat{\boldsymbol{D}}_t) \text{ solves } \mathcal{P}_{\mathcal{I}} \text{ given } d(\boldsymbol{D}_{t+1}, \hat{\boldsymbol{D}}_{t+1}) \text{ and } \kappa,$$
$$g_D(\boldsymbol{D}_{t+1}|\boldsymbol{D}_t) \text{ is given.}$$

The crucial difference from before is that in the feasible formulation of the consumption and portfolio choice problem the expectation is now taken with respect to the endogenous subjective probability density function for stochastic fruit-dividends, which itself has to be obtained as an optimal solution to the auxiliary informational problem. Also, note that in time t, the subjective random variables coincide with their objective counterparts (almost surely), but we signify them with hats nevertheless to preserve notational succession across the time periods.

Extension to infinite horizon: The economic setting is modified so that a representative agent has an infinite lifespan. Risky assets are composed of infinitely-lived trees. They bring stochastic dividends $\boldsymbol{D}_{t+1}$ at the beginning of period $(t+1)$. A K-sized random vector $\boldsymbol{D}_s$ follows a time-homogeneous (stationary) Markov chain defined by transition probability density function $g_D(\boldsymbol{D}_{s+1}|\boldsymbol{D}_s) = g_D(\boldsymbol{D}_{t+1}|\boldsymbol{D}_t)$ for all periods $s > t$. The risk-free asset is still composed of a one-period-lived tree. Its deterministic dividends $D_{0,t+1}$ are normalized to 1 for all periods t.

The formulations in terms of a sequence problem for both the infeasible as well as the feasible versions of this dynamic programming problem are available in Appendix §A. The Bellman equation corresponding to the infinite-horizon feasible problem, $\mathcal{P}_{\mathcal{QI}}$, is

$$\{\mathcal{P}_{\mathcal{QI}}\}$$

$$v(\{q_{0,t-1}, \boldsymbol{q}_{t-1}\}, \hat{\boldsymbol{D}}_t) = \max_{C_t, \{q_{0,t}, \boldsymbol{q}_t\}} \left\{u(C_t) + \beta \mathrm{E}_t^h \left[v(\{q_{0,t}, \boldsymbol{q}_t\}, \hat{\boldsymbol{D}}_{t+1})\right]\right\} \tag{PQI-1}$$

subject to

$$C_t + P_{0,t} q_{0,t} + \boldsymbol{P}_t^{\intercal} \boldsymbol{q}_t = q_{0,t-1} + (\boldsymbol{P}_t + \hat{\boldsymbol{D}}_t)^{\intercal} \boldsymbol{q}_{t-1}, \tag{PQI-2}$$

domain restriction $C_t, \{q_{0,t}, \boldsymbol{q}_t\} \in \mathbb{R}_+ \times \mathbb{R}^{K+1}$, with the same utility function specification, and where (spelling out the relationship to $\mathcal{P}_{\mathcal{I}}$ explicitly)

$$h_D(\hat{\boldsymbol{D}}_{t+1}|\hat{\boldsymbol{D}}_t) := \int_{\mathrm{supp}(g_D)} f(\boldsymbol{D}_{t+1}, \hat{\boldsymbol{D}}_{t+1}|\boldsymbol{D}_t, \hat{\boldsymbol{D}}_t)\, d\boldsymbol{D}_{t+1}, \tag{PQI-3}$$

$$f_D(\boldsymbol{D}_{t+1}, \hat{\boldsymbol{D}}_{t+1}|\boldsymbol{D}_t, \hat{\boldsymbol{D}}_t) := \arg\left\{\min_{f(\cdot,\cdot)} \mathrm{E}^f \left[d(v^\sharp(\{q_{0,t}, \boldsymbol{q}_t\}, \boldsymbol{D}_{t+1}), v(\{q_{0,t}, \boldsymbol{q}_t\}, \hat{\boldsymbol{D}}_{t+1}))\right] \right.$$
$$\left. \text{s.t. } \mathcal{I}(g_D(\boldsymbol{D}_{t+1}|\boldsymbol{D}_t); h_D(\hat{\boldsymbol{D}}_{t+1}|\hat{\boldsymbol{D}}_t)) \leq \kappa \right\}, \tag{PQI-4}$$

$$g_D(\boldsymbol{D}_{t+1}|\boldsymbol{D}_t) \text{ is given.} \tag{PQI-5}$$

Above we denote the maximum value function as $v^\sharp(\{q_{0,t-1}, \boldsymbol{q}_{t-1}\}, \boldsymbol{D}_t)$ for the infeasible/unconstrained case, and as $v(\{q_{0,t-1}, \boldsymbol{q}_{t-1}\}, \hat{\boldsymbol{D}}_t)$ for the feasible/constrained case. Using the notation introduced earlier, now $\varphi^\sharp(\boldsymbol{x}|\boldsymbol{\theta}) := v^\sharp(\boldsymbol{\theta}, \boldsymbol{x})$ and $\varphi(\hat{\boldsymbol{x}}|\boldsymbol{\theta}) := v(\boldsymbol{\theta}, \hat{\boldsymbol{x}})$. Because the objective function incorporates an additional constraint, for a given $\boldsymbol{\theta}$-parameter, $\varphi(\hat{\boldsymbol{x}}|\boldsymbol{\theta})$ lies weakly below the unconstrained $\varphi^\sharp(\boldsymbol{x}|\boldsymbol{\theta})$, as does $v(\boldsymbol{\theta}, \hat{\boldsymbol{x}})$ relative to $v^\sharp(\boldsymbol{\theta}, \boldsymbol{x})$.

The Bellman equation's formulation is standard except that the probability density function $h_D(\cdot)$ with respect to which it is defined stems from (PQI-4), the solution to auxiliary sub-problem $\mathcal{P}_\mathcal{I}$.

Lastly, note that the environment in terms of dynamics and stochasticity is as primitive as possible, the agent solves the same problem again and again.[9]

Appropriate distortion function: The solution to $\mathcal{P}_\mathcal{I}$, referred to in (PQI-4), requires choosing some appropriate distortion function $d(\cdot,\cdot)$. Specifying one that is both reasonable in terms of the objective of larger problem $\mathcal{P}_{\mathcal{QI}}$, and convenient to work with analytically, is a laborous task that we attend to next. Besides somewhat technical arguments, an important intermediate result we show below is a pessimistic downward adjustment of the mean of the approximating distribution, which is required in order to compensate for the latter's lower entropy (and variance).

Define the new variables, W_{t+1} (for wealth, cum dividend), $R_{0,t+1}$ and $\boldsymbol{R}_{t+1}$ (gross returns), as well as, to avoid any confusion, $y_{0,t}$ and $\boldsymbol{y}_t$ (nominal investments):

$$W_{t+1} := q_{0,t} + (\boldsymbol{P}_{t+1} + \boldsymbol{D}_{t+1})^\intercal \boldsymbol{q}_t =: \tag{4}$$

$$=: P_{0,t} R_{0,t+1} q_{0,t} + (\operatorname{diag}(\boldsymbol{P}_t)\, \boldsymbol{R}_{t+1})^\intercal\, \boldsymbol{q}_t =: \tag{5}$$

$$=: R_{0,t+1} y_{0,t} + \boldsymbol{R}_{t+1}^\intercal \boldsymbol{y}_t; \tag{6}$$

and similarly $\hat{W}_{t+1}$, $\hat{\boldsymbol{R}}_{t+1}$.

Now, restrict the stochastic processes considered to independent identically distributed random variables (rather than Markovian), that is $g_R(\boldsymbol{R}_{t+1}|\boldsymbol{R}_t) := g_R(\boldsymbol{R}_{t+1})$. Furthermore, assume the log-Normality of the returns, i.e., $g_R(\boldsymbol{R}_{t+1})$ is $\log\mathcal{N}(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r)$; for log-returns $\boldsymbol{r}_{t+1} := \ln \boldsymbol{R}_{t+1}$ (also reserving $r_{0,t+1} := \ln R_{0,t+1}$) this means

$$g_r(\boldsymbol{r}_{t+1}) \text{ is } \mathcal{N}(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r) \tag{7}$$

(we will "reverse-engineer" later what $g_D(\boldsymbol{D}_{t+1}|\boldsymbol{D}_t)$ this requires).[10]

Realize that the value function has the following form:

$$v^\sharp(\{q_{0,t}, \boldsymbol{q}_t\}, \boldsymbol{D}_{t+1}) = A\,(W_{t+1})^{1-\gamma}\,, \tag{8}$$

[9] Implicitly, in terms of Bayesian inference with regard to $\boldsymbol{x}$ and $g(\cdot)$, we are in a degenerate situation with a "flat", never-updated prior. The task of embedding non-trivial Bayesian updating into our framework, particularly into algorithm of Appendix §F, is outside the scope of this paper.

[10] A parametric log-Normal probability distribution is assumed here for analytical convenience, it is just a theoretical proxy for some non-parametric distribution dealt with in practically relevant problems.

where $A := (1-\beta)^{-\gamma}/(1-\gamma)$; and analogously $v(\{q_{0,t}, \boldsymbol{q}_t\}, \hat{\boldsymbol{D}}_{t+1})$.[11]

Thus, a reasonable choice for the distortion function in the sense of L^2 norm (squared) would be $d(\boldsymbol{r}_{t+1}, \hat{\boldsymbol{r}}_{t+1}) := ||v^\sharp(\{q_{0,t}, \boldsymbol{q}_t\}, \boldsymbol{D}_{t+1}) - v(\{q_{0,t}, \boldsymbol{q}_t\}, \hat{\boldsymbol{D}}_{t+1})||_2^2$. However, given the CRRA functional form, we modify it to

$$\begin{aligned} d(\boldsymbol{r}_{t+1}, \hat{\boldsymbol{r}}_{t+1}) &:= \frac{1}{(1-\gamma)^2} || \ln |v^\sharp(\{q_{0,t}, \boldsymbol{q}_t\}, \boldsymbol{D}_{t+1})| - \ln |v(\{q_{0,t}, \boldsymbol{q}_t\}, \hat{\boldsymbol{D}}_{t+1})| \, ||_2^2 = \\ &= (\ln W_{t+1} - \ln \hat{W}_{t+1})^2. \end{aligned} \tag{9}$$

It is worth saying explicitly that a distortion function so formulated is appropriate with regard to its economic grounding (a well-defined distance measure in felicity terms), but it also reflects some discretion with regard to the norm (L^2, squared) and the transformation (logarithmic) used.

Next, we produce an adaptation of the above formulation of the distortion function that is, for our purposes, more convenient to use.

Proposition 3 (Distortion Function)**.** *The above distortion function, in the context of problem $\mathcal{P}_{\mathcal{QI}}$ and given the distributional assumptions, can be reformulated as follows:*

$$\begin{aligned} d(\boldsymbol{r}_{t+1}, \hat{\boldsymbol{r}}_{t+1}) = (\ln W_{t+1} - \ln \hat{W}_{t+1})^2 \approx \\ \approx \left(\boldsymbol{\omega}_t^\intercal (\boldsymbol{r}_{t+1} - \hat{\boldsymbol{r}}_{t+1} + \check{\boldsymbol{\mu}}_r)\right)^2, \end{aligned} \tag{P3-1}$$

where

$$\boldsymbol{\omega}_t := \mathrm{diag}(\boldsymbol{P}_t)\boldsymbol{q}_t \,/\, W_t$$

is a K-vector of the shares of wealth invested in risky assets, while the mean $\hat{\boldsymbol{\mu}}_r$ of the simplified random variable $\hat{\mathbf{r}}_{t+1}$ equals

$$\hat{\boldsymbol{\mu}}_r := \boldsymbol{\mu}_r + \check{\boldsymbol{\mu}}_r, \tag{P3-2}$$

and where

$$\check{\boldsymbol{\mu}}_r := \frac{1}{2}\mathrm{diag}^{-1}(\boldsymbol{\Sigma}_r - \hat{\boldsymbol{\Sigma}}_r) - \frac{1}{2}(\boldsymbol{\Sigma}_r - \hat{\boldsymbol{\Sigma}}_r)\boldsymbol{\omega}_t, \tag{P3-3}$$

is a bias term, with $\hat{\boldsymbol{\Sigma}}_r$ denoting the variance-covariance matrix for $\hat{\mathbf{r}}_{t+1}$.

Under an additional assumption about the timing of an update of vector $\boldsymbol{\omega}_t$ (the requirement to minimize maximum loss), the following refinement can be made:

$$\begin{aligned} \left(\boldsymbol{\omega}_t^\intercal (\boldsymbol{r}_{t+1} - \hat{\boldsymbol{r}}_{t+1} + \check{\boldsymbol{\mu}}_r)\right)^2 \propto (\boldsymbol{r}_{t+1} - \hat{\boldsymbol{r}}_{t+1} + \check{\boldsymbol{\mu}}_r(\widehat{\omega}_t))^\intercal (\boldsymbol{r}_{t+1} - \hat{\boldsymbol{r}}_{t+1} + \check{\boldsymbol{\mu}}_r(\widehat{\omega}_t)) =: \\ =: d(\boldsymbol{r}_{t+1}, \hat{\boldsymbol{r}}_{t+1}), \end{aligned} \tag{P3-4}$$

where

$$\begin{aligned} \check{\boldsymbol{\mu}}_r(\widehat{\omega}_t) &:= \frac{1}{2}\mathrm{diag}^{-1}(\boldsymbol{\Sigma}_r - \hat{\boldsymbol{\Sigma}}_r)(1-\widehat{\omega}_t) = \\ &= \frac{1}{2}\mathrm{diag}(\sigma_{r,1}^2 - \hat{\sigma}_{r,1}^2, \cdots, \sigma_{r,K}^2 - \hat{\sigma}_{r,K}^2)\mathbf{1}(1-\widehat{\omega}_t) \end{aligned} \tag{P3-5}$$

[11]We abstract away from the special fact that in the general equilibrium of this particular exchange economy, $v(\cdot)$ actually attains and equals $v^\sharp(\cdot)$ identically.

is another bias term, with $\widehat{\omega}_t$ denoting the total share of wealth invested in risky assets, i.e.,

$$\widehat{\omega}_t := \mathbf{1}^\intercal \boldsymbol{\omega}_t.$$

Proof. See Appendix §E.2.

□

The function $\mathrm{diag}^{-1}(\cdot)$ above is an inverse of the function $\mathrm{diag}(\cdot)$, where the latter takes as an argument a vector and returns a diagonal matrix with a given vector's elements on the main diagonal, hence the former takes a diagonal (or just square) matrix and returns a column vector with the main diagonal's elements of a given matrix. Also, from now on the elements of a matrix are denoted with the same letter as the matrix but in lowercase.

An important intermediate result stated in the Proposition above is that in order to cancel out the effect on the wealth dynamics (our ultimate benchmark) of replacing the original variance, $\boldsymbol{\Sigma}_r$, with the simplified variance, $\hat{\boldsymbol{\Sigma}}_r$, and thus to ensure the expected growth rate of the simplified log-wealth, $\ln \hat{W}_{t+1}$, equals that of the original log-wealth, $\ln W_{t+1}$, we also must adjust the mean of the simplified random variable, $\hat{\boldsymbol{\mu}}_r$, as shown in equation (P3-2). The corresponding bias term $\check{\boldsymbol{\mu}}_r$ consists of two sub-terms (equation P3-3).

The first (on the RHS of equation P3-3), rather technical one, is an artefact of the continuous-time approximation based on the (geometric) Brownian motion. It ensures that the expected return on the simplified and original wealth would coincide if the simplified wealth indeed followed the dynamics captured by the simplified variance; hence it can be viewed as a term setting the "origin" point.

The second one (on the RHS of equation P3-3) is more intuitive. It adjusts the mean downward from the origin (for positive risky investments) with the magnitude of the adjustment increasing in the share of wealth invested as well as in the size of the discrepancy between the original and simplified variances (a non-negative quantity, as shown later in §3.3); hence it can be viewed as a term imposing "pessimism". While also being a byproduct of the continuous-time approximation undertaken in the derivations above, the second adjustment term entails a downward shift in the expectation of log-return $\hat{\boldsymbol{r}}_{t+1}$ and effectively adopts a conservative view at the potential investment opportunities.[12]

Remark 1 (Decision Rule- vs. Subjective Perception-Adjustment)**.** *A more direct way of correcting for the difference between the original and simplified vairances is to adjust the decision rules (i.e., policy functions dictating the choice of control variables $C_t, \{q_{0,t}, \boldsymbol{q}_t\}$) rather than the subjective perception (i.e., the mean of the simplified distribution $\hat{\boldsymbol{\mu}}_r$). Adjusting the decision rules shifts control variables $C_t, \{q_{0,t}, \boldsymbol{q}_t\}$ directly; which takes extra*

[12]The above "pessimism" adjustment is not to be confused with the (quasi) certainty discount implied by the certainty-equivalent counterpart to a risky asset under expected utility theory. For instance, note that we have isolated out the risk-aversion parameter in equation (9), immunizing the distortion function from the effect of risk attitude.

$(K+1)$ *adjustment parameters. Adjusting the mean shifts control variables* $C_t, \{q_{0,t}, \boldsymbol{q}_t\}$ *indirectly, by affecting the state variables* $\{q_{0,t-1}, \boldsymbol{q}_{t-1}\}, \hat{\boldsymbol{D}}_t$ *and the chosen subjective probability density* $h_D(\hat{\boldsymbol{D}}_{t+1}|\hat{\boldsymbol{D}}_t)$*, via* $h_r(\hat{\boldsymbol{r}}_{t+1})$*, that control variables are functionally dependent on; which takes extra* K *adjustment parameters. Reasoning in terms of degrees of freedom, the latter option, with its lower number of adjustment parameters, is (weakly) more restrictive.*[13]

Therefore, we are dealing with a simplified distribution $h_r(\hat{\boldsymbol{r}}_{t+1})$ that is biased, but biased in an optimal, expected distortion-minimizing way (see Appendix §E.2.1). For example, in the case of one risky asset that is the sole constituent of the investment portfolio (i.e., $\omega_t = 1$), the term setting the origin places the mean as if the risky asset is driven by a stochastic process that is determined by the simplified rather than the true variance (effectively matching the expected values of simplified and true returns, $\mathrm{E}_t^h[\hat{R}_{t+1}] = \mathrm{E}_t^g[R_{t+1}]$), while the pessimism term guarantees that the simplified mean does not depart from the true one (thus, matching the means $\hat{\mu}_r = \mu_r$). This results in the subjective expected return on the simplified portfolio being pessimistically biased and undershooting its objectively expected level (that is, $\mathrm{E}_t^h[\hat{W}_{t+1}/W_t] = \mathrm{E}_t^h[\hat{R}_{t+1}] = \exp(\hat{\mu}_r + 0.5\hat{\Sigma}_r) < \exp(\mu_r + 0.5\Sigma_r) = \mathrm{E}_t^g[R_{t+1}] = \mathrm{E}_t^g[W_{t+1}/W_t]$), but it actually ensures optimality in the end.

Remark 2 (Computational Benefits of Mean Adjustment)**.** *It may seem unreasonable to reduce information processing costs by using an approximating random variable with a simplified variance instead of the original one only to increase the burden down the line by necessitating the manipulations with the bias term (another kind of infinite regress critique). The reason the proposed approach works is because (conditionally on* $\boldsymbol{\Sigma}_r$*,* $\hat{\boldsymbol{\Sigma}}_r$ *and* $\boldsymbol{\omega}_t$*) the bias term* $\check{\boldsymbol{\mu}}_r$ *is a non-stochastic object and possesses zero (in discrete case,* $-\infty$ *in continuous case) entropy, hence manipulating it is less computationally intensive than it is in the case of stochastic objects; e.g., consider the summation operations required to implement integration. (It can also be understood from a measure-theoretic standpoint as an issue of dimensionality: the stochastic objects (i.e., random variables as measurable functions from a space of outcomes to a measurable space) are characterized by a non-trivial profile on the corresponding measurable space, while zero-entropy objects (i.e., fixed constants) have a flat profile—if singletons, otherwise flat except a single atom—and the corresponding space is in some sense degenerate.)*

An additional assumption, required only for $K > 1$ cases, that disciplines the optimization-related information-processing and prohibits the dependence of $d(\boldsymbol{r}_{t+1}, \hat{\boldsymbol{r}}_{t+1})$ on $\boldsymbol{\omega}_t$ (by replacing the dependence on exact $\boldsymbol{\omega}_t$ with agnostic motive of minimization of maximum loss for any possible $\boldsymbol{\omega}_t$) allows to deduce a reasonable yet operationally applicable distortion function given in equation (P3-4).[14] It produces a convenient sum-of-squares

[13] The author thanks Michael Woodford for raising this issue.

[14] Note that for any vector $\check{\boldsymbol{\mu}}_r(\widehat{\omega}_t)$, the refined distortion function does not depend on the scalar $\widehat{\omega}_t$;

formulation for suitably translated (bias-corrected) differences between the true and approximated log-returns. This finalizes the operational definition of the distortion function we are going to use for the informational sub-problem of problem $\mathcal{P}_{\mathcal{QI}}$.

Decorrelation via coordinate change: Part of our solution method relies on random variables being uncorrelated. This is implemented by transforming coordinates to the principal axes of the variance-covariance matrix, represented by its eigenvectors.

Thus, define the transformed random variables and parameters as

$$\boldsymbol{x} := \boldsymbol{\Xi}^\intercal \boldsymbol{r}_{t+1}, \quad (10) \qquad \hat{\boldsymbol{x}} := \boldsymbol{\Xi}^\intercal \hat{\boldsymbol{r}}_{t+1}, \quad (11) \qquad \check{\boldsymbol{\mu}}(\widehat{\omega}_t) := \boldsymbol{\Xi}^\intercal \check{\boldsymbol{\mu}}_r(\widehat{\omega}_t), \quad (12)$$

where $\boldsymbol{\Xi}$ is a square matrix with eigenvectors in its columns that is obtained from eigendecomposition (or diagonalization) of matrix $\boldsymbol{\Sigma}_r$.[15] The eigendecomposition procedure, $\mathcal{P}_{\boxslash}$, implies the following relationships:

$$\{\boldsymbol{\Sigma}, \boldsymbol{\Xi}\} := \text{eigendecompose}(\boldsymbol{\Sigma}_r); \qquad \{\mathcal{P}_{\boxslash}\}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_K^2 \end{bmatrix}, \qquad \boldsymbol{\Sigma}_r = \boldsymbol{\Xi}\boldsymbol{\Sigma}\boldsymbol{\Xi}^{-1}, \qquad \boldsymbol{\Xi}^\intercal = \boldsymbol{\Xi}^{-1}.$$

Proposition B.1 in Appendix §B states that the decorrelating transformation which facilitates our subsequent solution is innocuous.

Equilibrium: An equilibrium of the economy formed by the exogenously given economic setting described in §3.1 and the endogenously chosen optimal solutions to the feasible problem $\mathcal{P}_{\mathcal{QI}}$ is a collection of a continuous price function $\{P_0(\boldsymbol{D}_t), \boldsymbol{P}(\boldsymbol{D}_t)\} : \mathbb{R}_+^K \mapsto \mathbb{R}_+^{K+1}$, a continuous and bounded value function $v(\{q_{0,t-1}, \boldsymbol{q}_{t-1}\}, \boldsymbol{D}_t) : \mathbb{R}_+^{K+1} \times \mathbb{R}_+^K \mapsto \mathbb{R}$, and an absolutely continuous joint probability distribution function $F_D(\boldsymbol{D}_{t+1}, \hat{\boldsymbol{D}}_{t+1} | \boldsymbol{D}_t, \hat{\boldsymbol{D}}_t) : \mathbb{R}_+^K \times \mathbb{R}_+^K \mapsto [0, 1]$ such that:

(i) [consumption and investment optimality] Bellman equation (PQI-1) subject to the budget constraint (PQI-2), control variable's domain restriction, no-Ponzi-schemes constraint and with given utility function specification is satisfied;

(ii) [consumption and investment coherence] goods and asset markets clear, i.e.,

$$C_t = \widehat{\boldsymbol{q}}^\intercal \boldsymbol{D}_t, \qquad \boldsymbol{q}_t = \widehat{\boldsymbol{q}}, \qquad q_{0,t} = \widehat{q}_0;$$

in other words, given the bias term $\check{\boldsymbol{\mu}}_r(\widehat{\omega}_t)$, the solution to the informational problem is invariant to the chosen value of the bound $\widehat{\omega}_t$. This fact will be put to work later.

[15] In terms of Appendix §F's algorithm, our understanding is that this transformation is performed when solving the informational problem before the Generating codebook step of the algorithm presented in Appendix §F (the inverse transformation may be conducted either before or after maximizing the objective function in the process of solving the consumption and portfolio choice problem).

(iii) [informational optimality] joint probability density function $f_r(\boldsymbol{r}_{t+1}, \hat{\boldsymbol{r}}_{t+1})$ solves (utilizing the decorrelating transformation) the informational problem $\mathcal{P}_{\mathcal{I}}$ with $g_r(\boldsymbol{r}_{t+1})$ as the true density and κ as the information processing capacity;

(iv) [informational coherence] probability density functions for dividends $f_D(\boldsymbol{D}_{t+1}, \hat{\boldsymbol{D}}_{t+1} | \boldsymbol{D}_t, \hat{\boldsymbol{D}}_t)$ (referred to in equation PQI-4), $g_D(\boldsymbol{D}_{t+1} | \boldsymbol{D}_t)$ (referred to in PQI-5) and $h_D(\hat{\boldsymbol{D}}_{t+1} | \hat{\boldsymbol{D}}_t)$ (referred to in PQI-3) are consistent with the densities for returns $f_r(\boldsymbol{r}_{t+1}, \hat{\boldsymbol{r}}_{t+1})$, $g_r(\boldsymbol{r}_{t+1})$ and $h_r(\hat{\boldsymbol{r}}_{t+1})$, also the correlated random variables' densities $f_r(\boldsymbol{r}_{t+1}, \hat{\boldsymbol{r}}_{t+1})$, $g_r(\boldsymbol{r}_{t+1})$ and $h_r(\hat{\boldsymbol{r}}_{t+1})$ are consistent with the decorrelated variables' densities $f(\boldsymbol{x}, \hat{\boldsymbol{x}})$, $g(\boldsymbol{x})$ and $h(\hat{\boldsymbol{x}})$, i.e., $\forall \boldsymbol{D}_{t+1}, \hat{\boldsymbol{D}}_{t+1} \in \mathbb{R}_+^K$:

$$\begin{aligned} f_D(\boldsymbol{D}_{t+1}, \hat{\boldsymbol{D}}_{t+1} | \boldsymbol{D}_t, \hat{\boldsymbol{D}}_t) &= f_r(\boldsymbol{r}_{t+1}, \hat{\boldsymbol{r}}_{t+1}) = f_r(\Xi \boldsymbol{x}, \Xi \hat{\boldsymbol{x}}) = f(\boldsymbol{x}, \hat{\boldsymbol{x}}), \\ g_D(\boldsymbol{D}_{t+1} | \boldsymbol{D}_t) &= \quad g_r(\boldsymbol{r}_{t+1}) \quad = \quad g_r(\Xi \boldsymbol{x}) \quad = g(\boldsymbol{x}), \\ h_D(\hat{\boldsymbol{D}}_{t+1} | \hat{\boldsymbol{D}}_t) &= \quad h_r(\hat{\boldsymbol{r}}_{t+1}) \quad = \quad h_r(\Xi \hat{\boldsymbol{x}}) \quad = h(\hat{\boldsymbol{x}}). \end{aligned}$$

A policy function determining the optimal investment in tree shares $\boldsymbol{q}(\{q_{0,t-1}, \boldsymbol{q}_{t-1}\}, \hat{\boldsymbol{D}}_t)$ could be added to the list of equilibrium objects, but it is a constant function that is identically equal to $\widehat{\boldsymbol{q}}$ because the considered economy is an autarky. Although not made explicit, the informational coherence equations subsume the Jacobians of the transformations.

Note that the conditions (iii) and (iv) replace the traditional rational expectations assumption (Lucas, 1978). Otherwise, the notion of equilibrium is standard. The existence of an equilibrium is proven by constructing its instance and solving the model.

3.3 Solution

The consumption and investment segment of the larger problem is fairly standard, so here we only focus on the crucial elements of the informational part, benefiting from the clearly segregated formulations of these two sub-problems. (The full solution to the feasible consumption and portfolio choice problem $\mathcal{P}_{\mathcal{QI}}$ is available in Appendix §C.)

Taking the general solution to $\mathcal{P}_{\mathcal{I}}$ from Proposition 1, exploiting the flexible mean property due to Proposition 2, using the (refined) distortion function from Proposition 3, and applying decorrelating transformation allowed by Proposition B.1 yields:

$$f(\boldsymbol{x} | \hat{\boldsymbol{x}}) = \exp\left(\frac{1}{\lambda} \nu(\boldsymbol{x}, \hat{\boldsymbol{x}}) - \frac{1}{\lambda} \mu(\boldsymbol{x}) - \frac{1}{\lambda} (\boldsymbol{x} - \hat{\boldsymbol{x}} + \check{\boldsymbol{\mu}}(\widehat{\omega}_t))^\intercal (\boldsymbol{x} - \hat{\boldsymbol{x}} + \check{\boldsymbol{\mu}}(\widehat{\omega}_t)) \right), \quad \forall \hat{\boldsymbol{x}} \in \operatorname{supp}(h). \tag{13}$$

Given our knowledge about the probability distribution of $\boldsymbol{x}$, we can solve for the whole stochastic structure of the relationship between $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$, as shown in the following Proposition. But an attentive reader has already spotted the kernel of a Gaussian probability density function in the last equation, which suggests the direction of our subsequent exploration and explains the labor we put into specification of the distance function.

Proposition 4 (Specific Solution to Informational Problem)**.** *Let the general solution to the informational problem, which is specialized to the chosen distortion function and accounts for the decorrelating transformation, be given by the conditional probability density function $f(\boldsymbol{x}|\hat{\boldsymbol{x}})$ from (13), where the random vector $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with the constant vector $\boldsymbol{\mu}$ implicitly defined via (10) and the matrix $\boldsymbol{\Sigma}$ described in the relationships $\mathcal{P}_{\boxslash}$.*

Then, the specific solution to the informational problem can take one of two forms, depending on the magnitude of information processing capacity κ (i.e., tightness of shadow price/Lagrange multiplier on information constraint λ):

(a) Interior solution ("large" κ, "small" λ).

$$f(\boldsymbol{x}|\hat{\boldsymbol{x}}) = (2\pi)^{-\frac{K}{2}} \left|\frac{\lambda}{2}\boldsymbol{I}_K\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \hat{\boldsymbol{x}} + \check{\boldsymbol{\mu}}(\widehat{\omega}_t))^\intercal \left(\frac{\lambda}{2}\boldsymbol{I}_K\right)^{-1} (\boldsymbol{x} - \hat{\boldsymbol{x}} + \check{\boldsymbol{\mu}}(\widehat{\omega}_t))\right), \quad \forall \hat{\boldsymbol{x}} \in \mathbb{R}^K;$$

$$\boldsymbol{x} = \hat{\boldsymbol{x}} - \check{\boldsymbol{\mu}}(\widehat{\omega}_t) + \boldsymbol{\epsilon},$$

where

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Psi}), \qquad \boldsymbol{\Psi} = \frac{\lambda}{2}\boldsymbol{I}_K,$$

$$\hat{\boldsymbol{x}} \sim \mathcal{N}(\hat{\boldsymbol{\mu}}(\widehat{\omega}_t), \hat{\boldsymbol{\Sigma}}), \qquad \hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} - \boldsymbol{\Psi};$$

$$\lambda = 2\left(e^{-2\kappa}|\boldsymbol{\Sigma}|\right)^{\frac{1}{K}}.$$

Interior solution is valid if the following condition holds: $\sigma_k^2 > \frac{\lambda}{2}$, $\forall k \in \{1, \dots, K\}$.

(b) Boundary solution ("small" κ, "large" λ).

$$f(\boldsymbol{x}|\hat{\boldsymbol{x}}) = (2\pi)^{-\frac{K}{2}} |\boldsymbol{\Psi}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \hat{\boldsymbol{x}} + \check{\boldsymbol{\mu}}(\widehat{\omega}_t))^\intercal \boldsymbol{\Psi}^{-1} (\boldsymbol{x} - \hat{\boldsymbol{x}} + \check{\boldsymbol{\mu}}(\widehat{\omega}_t))\right), \quad \forall \hat{\boldsymbol{x}} \in \operatorname{supp}(h);$$

$$\boldsymbol{x} = \hat{\boldsymbol{x}} - \check{\boldsymbol{\mu}}(\widehat{\omega}_t) + \boldsymbol{\epsilon},$$

where

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Psi}), \qquad \boldsymbol{\Psi} = \begin{bmatrix} \lambda/2 & & 0 & 0 & \cdots & 0 \\ & \ddots & & \vdots & \ddots & \vdots \\ 0 & & \lambda/2 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \sigma_{k^*+1}^2 & & 0 \\ \vdots & \ddots & \vdots & & \ddots & \\ 0 & \cdots & 0 & 0 & & \sigma_K^2 \end{bmatrix},$$

$$\hat{\boldsymbol{x}} \sim \mathcal{N}(\hat{\boldsymbol{\mu}}(\widehat{\omega}_t), \hat{\boldsymbol{\Sigma}}), \qquad \hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} - \boldsymbol{\Psi};$$

$$\{\sigma_k^2\}_1^K := \operatorname{sortdescending}(\{\sigma_k^2\}_1^K),$$

$$k^* := \arg\min_{k \in \{1,\dots,K\}} \{\sigma_k^2 \mid \sigma_k^2 > \frac{\lambda}{2}\},$$

$$\lambda = 2\left(e^{-2\kappa}\sigma_{k^*+1}^{-2} \cdots \sigma_K^{-2}|\boldsymbol{\Sigma}|\right)^{\frac{1}{k^*}}.$$

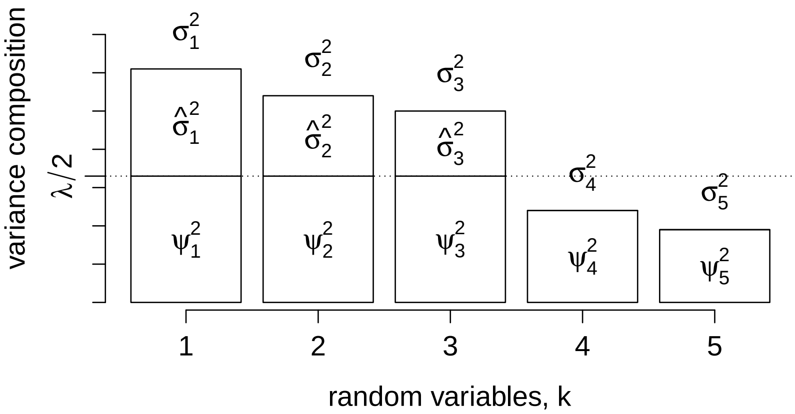

Figure 4: "Reverse water-filling".

(The last $(K-k^*)$ elements of vector $\hat{\boldsymbol{x}}$ are to be understood as deterministic scalars, or alternatively as Dirac delta functions centered at $\{\hat{\mu}_{k^*+1}(\widehat{\omega}_t), \cdots, \hat{\mu}_K(\widehat{\omega}_t)\}$ at a cost of abusing the notation when dealing with their Radon-Nikodym derivatives.)

Boundary solution is valid if the following condition holds: $\exists k \in \{1, \ldots, K\} : \sigma_k^2 \leq \frac{\lambda}{2}$.

Proof. See Appendix §E.4. □

The qualifiers 'interior' and 'boundary' in the formulation of the above Proposition should be understood in relation to the Cartesian product set $\bigtimes_1^K [0, \sigma_k^2]$. Figure 4 illustrates the (relatively more general) "reverse water-filling" logic of the boundary solution of Proposition 4. Consider a situation when information processing capacity κ is large enough so that information constraint does not bind: then the corresponding Lagrange multiplier λ is 0; elements on the diagonal of the variance-covariance matrix for approximation error $\boldsymbol{\Psi}$ are all 0, $\psi_k^2 = 0$, $\forall k \in \{1, \ldots, K\}$; while diagonal elements of the variance-covariance matrices $\hat{\boldsymbol{\Sigma}}$ and $\boldsymbol{\Sigma}$ are equalized, $\hat{\sigma}_k^2 = \sigma_k^2$, $\forall k \in \{1, \ldots, K\}$. If κ decreases and the information constraint starts to bind, initially the following happens: λ increases; ψ_k^2 rises slightly but remains equal for all k; while $\hat{\sigma}_k^2$ equal the difference between σ_k^2 and the level of ψ_k^2 for all k. If κ decreases even further, a qualitative change in the picture occurs, and we move from the interior to the boundary solution case: at some point λ increases enough for ψ_k^2 to catch up with the lowest σ_k^2 and the corresponding $\hat{\sigma}_k^2$ to become 0; then the remaining ψ_l^2 for $l \neq k$ depart from ψ_k^2 and continue to rise, so that the "reverse water-filling" process goes on for the rest of $\hat{\sigma}_l^2$. And so on.

From the information-processing perspective, σ_k^2 represent the total information available for processing, $\hat{\sigma}_k^2$ represent the information that is actually processed, while ψ_k^2 represent the information that is omitted and consitutes the approximation errors. As information processing capacity decreases, the information omissions disproportionally affect the data dimensions with low informational content (small eigenvalues σ_k^2).[16]

[16]Lastly, note that in our case we benefit from the Normality of distribution $g(\cdot)$, among others as-

The above configuration of the solution is not just a technical attribute, but also has deep economic implications.

Corollary 1 (Specific Solution to Informational Problem: Dispersion Folding)**.** *The specific solution to the informational problem, as given in the statement of Proposition 4, is characterized by the folded dispersions (or collapsed randomness/distributions) of the less volatile components of random vector $\hat{\boldsymbol{x}}$ in the boundary solution case. I.e., the corresponding subjective variances become 0:*

$$\hat{\sigma}_k^2 = 0, \quad \forall k > k^*.$$

Proof. Immediate from Proposition 4.

□

Considering (the more revealing case here) of boundary solution, the role of simplification is manifested in dropping some of the random variables' dimensions (or the random variables themselves, if they are uncorrelated) from the agent's approximation. Due to the effects of entropy/variance reduction culminating in its "folding" or "collapse"/"contraction", such random variables' dimensions are replaced with non-stochastic objects, that is by their—sufficiently biased—means (cf. the sparsity logic of Gabaix, 2014a).

Intuitively, in the case of two uncorrelated random variables (alternatively, two random variables' dimensions) x_1 and x_2, according to Corollary 1 as the "folded" random variable (dimension) $\hat{x}_2$ effectively becomes non-stochastic, a simple univariate (one-dimensional) approximating model based on $\hat{x}_1$ emerges subjectively.

For example, consider a vineyard: a garden planted with both mature and young grapevines (think of them as two different "trees", each representing an aggregate of identical vines). The formers' "payoff" is determined by their exposure to light and humidity conditions, while the latters' by their time to first harvest and healthiness of the development; hence, the two are uncorrelated. If the young vines are also few in number, a binding capacity constraint may lead to total disregard of the presence of these less important (and thus barely worth spending capacity on) vines.

Taking a slightly more involved case of correlated random variables, five different grapes may be perfectly characterized by such attributes as their acidity, body, flavor (say, spice), sugar and tannin levels. Here, a binding capacity constraint may result in focusing on the most crucial attributes (say, acidity, body and sugar levels) and ignoring the rest (flavor and tannin levels). (Corollary 2 develops this theme a little further.)

sumptions: then the resulting distribution $h(\cdot)$ turns out to be Normal as well, but is characterized by lower variance. The result is not always as straight-forward: e.g., in the case of distribution $g(\cdot)$ having a bounded support, a discretely distributed solution for $h(\cdot)$ arises in the analysis of Matějka and Sims (2010).

The results of Proposition 4 in economically interesting terms such as returns, i.e., after the inversion of the decorrelating transformation, are presented in the subsequent Proposition 5.

Proposition 5 (Specific Solution to Informational Problem: Representation in Economic Terms)**.** *The specific solution to the informational problem given in the statement of Proposition 4 can be equivalently represented in terms of returns. In particular, the following decomposition is valid:*

$$\boldsymbol{r}_{t+1} = \hat{\boldsymbol{r}}_{t+1} - \check{\boldsymbol{\mu}}_r(\widehat{\omega}_t) + \boldsymbol{\epsilon}_{r,t+1}, \qquad \forall \boldsymbol{r}_{t+1} \in \mathbb{R}^K,$$

also producing

$$\boldsymbol{\Sigma}_r = \hat{\boldsymbol{\Sigma}}_r + \boldsymbol{\Psi}_r, \qquad \forall \boldsymbol{\Sigma}_r \textit{ that is } K \times K \textit{ positive semi-definite},$$

where

$$\begin{aligned} \hat{\boldsymbol{r}}_{t+1} &\sim \mathcal{N}(\hat{\boldsymbol{\mu}}_r(\widehat{\omega}_t), \hat{\boldsymbol{\Sigma}}_r), \\ \boldsymbol{\epsilon}_{r,t+1} &\sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Psi}_r), \end{aligned}$$

with $\hat{\boldsymbol{\mu}}_r(\widehat{\omega}_t)$ *and* $\check{\boldsymbol{\mu}}_r(\widehat{\omega}_t)$ *given by Proposition 3, as well as with* $\hat{\boldsymbol{\Sigma}}_r$ *and* $\boldsymbol{\Psi}_r$ *defined as*

$$\hat{\boldsymbol{\Sigma}}_r := \boldsymbol{\Xi}\hat{\boldsymbol{\Sigma}}\boldsymbol{\Xi}^{-1}, \qquad \boldsymbol{\Psi}_r := \boldsymbol{\Xi}\boldsymbol{\Psi}\boldsymbol{\Xi}^{-1}$$

basing on the results of Proposition 4.

Proof. See Appendix §E.5.

□

Moreover, arguments in Appendix §C.3 allow to replace $\check{\boldsymbol{\mu}}_r(\widehat{\omega}_t)$ and $\hat{\boldsymbol{\mu}}_r(\widehat{\omega}_t)$ with, respectively, $\check{\boldsymbol{\mu}}_r$ and $\hat{\boldsymbol{\mu}}_r$ in the statement of Proposition 5 (as well as in Proposition 4).[17]

To sum up, Proposition 3 demonstrates the appropriateness of our approximation procedure and allows to derive the normality of $\hat{\boldsymbol{r}}_{t+1}$ with the subsequent decomposition equations of Propositions 4–5, providing us with great analytical convenience; while Appendix §C.3 ensures that approximation accuracy result from Proposition 3 has not been lost in the process.

There are two main economic results in Proposition 5: (i) due to simplification $\hat{\boldsymbol{\Sigma}}_r$, the variance-covariance matrix of the simplified log-returns $\hat{\boldsymbol{r}}_{t+1}$, is smaller than $\boldsymbol{\Sigma}_r$, its counterpart for the original log-returns $\boldsymbol{r}_{t+1}$ (i.e., the difference $\hat{\boldsymbol{\Sigma}}_r - \boldsymbol{\Sigma}_r$ is a negative

[17] There are also a couple of minor technical details of note. To aid subsequent analysis, it may be worthwhile highlighting that in the interior solution case the variance-covariance matrix of approximation errors for returns is again diagonal and remains unchanged: $\boldsymbol{\Psi}_r = \boldsymbol{\Psi}$. Another revealing result in the interior solution case is that the optimal mean bias term defined in Proposition 3.1, $\check{\boldsymbol{\mu}}_r$, takes the form conformable with the bias term from Proposition 3.3, $\check{\boldsymbol{\mu}}_r(\widehat{\omega}_t)$: $\check{\boldsymbol{\mu}}_r = \frac{1}{2}\mathrm{diag}^{-1}(\boldsymbol{\Sigma}_r - \hat{\boldsymbol{\Sigma}}_r) \odot (\boldsymbol{1} - \boldsymbol{\omega}_t)$, where $\odot$ denotes the Hadamard product.

semi-definite matrix); (ii) as compensation for the simplification above, $\hat{\boldsymbol{\mu}}_r$, the mean of $\hat{\boldsymbol{r}}_{t+1}$, is biased toward pessimism in comparison to $\boldsymbol{\mu}_r$, its counterpart for $\boldsymbol{r}_{t+1}$. Both results are obtained endogenously and follow from §3.2.

In addition to the claim that $\hat{\boldsymbol{\Sigma}}_r$ is smaller than $\boldsymbol{\Sigma}_r$, there are more facets to this attribute of the solution.

Corollary 2 (Specific Solution to Informational Problem: Correlation Inflation)**.** *The specific solution to the informational problem represented in economic terms, as given in the statement of Proposition 5, is characterized by:*

(a) The inflated correlations between the elements of $\hat{\boldsymbol{r}}_{t+1}$ *relative to those for the elements of* $\boldsymbol{r}_{t+1}$ *in the interior solution case. I.e., the generic correlation coefficient's subjective version moves away from its objective value towards* 1 *(or* −1*):*

$$|\hat{\rho}_{r,kl}| \geq |\rho_{r,kl}|, \qquad \forall k,l \in \{1,\dots,K\};$$

which can be seen directly in the relationship

$$\hat{\rho}_{r,kl} = \rho_{r,kl} \times \frac{\left(\sum_{m=1}^{K} \xi_{km}^2 \sigma_m^2\right)^{1/2} \left(\sum_{m=1}^{K} \xi_{lm}^2 \sigma_m^2\right)^{1/2}}{\left(\sum_{m=1}^{K} \xi_{km}^2 \sigma_m^2 - \psi_1^2\right)^{1/2} \left(\sum_{m=1}^{K} \xi_{lm}^2 \sigma_m^2 - \psi_1^2\right)^{1/2}}, \quad \forall k,l \in \{1,\dots,K\},$$

where

$$\psi_1^2 := \left(e^{-2\kappa}|\boldsymbol{\Sigma}|\right)^{\frac{1}{K}} < \min_{m\in\{1,\dots,K\}} \sigma_m^2, \qquad \sum_{m=1}^{K} \xi_{km}^2 = 1, \quad \forall k \in \{1,\dots,K\}.$$

(b) Either inflated or shrinking correlations between the elements of $\hat{\boldsymbol{r}}_{t+1}$ *relative to those for the elements of* $\boldsymbol{r}_{t+1}$ *in the boundary solution case. I.e., the generic correlation coefficient's subjective version may move away from its objective value either toward* 0 *or* 1 *(*−1*):*

$$|\hat{\rho}_{r,kl}| \gtreqless |\rho_{r,kl}|, \qquad \forall k,l \in \{1,\dots,K\}.$$

Proof. See Appendix §E.7.

□

In spite of the ambiguous result for the specific correlation coefficients—and even that only for the boundary solution case—the region of the parameter space affected by such indeterminacy is (loosely speaking) "small". Because, together with the shrinking diagonal variance terms, inflation of the off-diagonal covariance terms contributes to achieving a variance-covariance matrix $\hat{\boldsymbol{\Sigma}}_r$ smaller than $\boldsymbol{\Sigma}_r$—mechanics that apply both to the interior and boundary solution cases, thus making the "correlation inflation" outcome robust. Furthermore, note that the reduced variances and biased correlations are the reflections of the trade-off allowed by the information processing capacity constraint (3).

Intuitively, Corollary 2 shows how the inflation of the correlations between the elements of $\hat{\boldsymbol{r}}_{t+1}$ as compared to the correlations between the elements of $\boldsymbol{r}_{t+1}$ emerges subjectively; which effectively leads to further attraction of the positively correlated elements and the repulsion of the negatively correlated ones that ultimately results in the pooling of the random vector's components into relatively detached categories.[18]

For example, think of Cabernet Sauvignon, Pinot Noir and Shiraz grapevines being pulled together into one category, Pinot Grigio and Sauvignon Blanc grapevines into another category, and with two categories of plants being pushed apart as very distinct kinds of capital goods—say, "red" vs. "white"—that are characterized by different attributes.

Lastly, we notice the following fact.

Corollary 3 ("Satisficing")**.** *A binding information processing capacity bound κ, such that $\kappa < \kappa^\sharp$, implies a positive expected value-function shortfall:*

$$\mathrm{E}_t^f[v^\sharp(\{q_{0,t}, \boldsymbol{q}_t\}, \boldsymbol{D}_{t+1}) - v(\{q_{0,t}, \boldsymbol{q}_t\}, \hat{\boldsymbol{D}}_{t+1})] > 0$$

for all $\{q_{0,t}, \boldsymbol{q}_t\} \in \mathbb{R}^{K+1}$; which is operationally (to investor) and observationally (to econometrician) equivalent, in terms of decision outcomes, to "satisficing" (or optimality in a constrained sense that generally does not achieve the first-best level).

Proof. Take $\lambda = 0$ in Propositions 4 and 5, recognizing their premises from Propositions 1 and 3.

□

This is a straight-forward instance of the concept of "satisficing"—as opposed to "maximizing"—introduced by Simon (1997). Here, the constrained (i.e., a second-best) value function $v(\cdot)$ serves the role of the (endogenous) "aspiration level" that is sought-after.[19]

4 Discussion

4.1 Theoretical results

The information processing capacity κ that is low enough to make the information constraint (3) binding induces a subjective probability measure $h(\cdot)$ that is different from the objective measure $g(\cdot)$. Using the former in place of the latter for making decisions under

[18]In computational cognitive science, as neural nework models undergo supervised learning to perform categorization tasks, they demonstrate an emergent property of categorical perception: the latter is characterized by within-category compression and between-category separation, similarly to the "correlation inflation" effect above. For details, see Tijsseling and Harnad (1997), Damper and Harnad (2000).

[19]Also note how decision optimality depends on recognition of information-processing constraints, and thus may look differently from the perspective of investors, who are inside actors in the model economy, and econometricians, who tend to be outside observers of the economic activities. For the importance of such differentiation, see, e.g., Hansen (2014).

risk is less computationally burdensome, but at the same time biases the decision-making environment in a certain, predictable direction. Although this discrepancy may give rise to decision outcomes deviating from the unconstrained, "rational expectations" alternative, constrained optimality (offered by the solution to the feasible problem $\mathcal{P}_{\mathcal{QI}}$) is still within reach as long as optimal adjustments are made. Figure 5 illustrates the differences between the objective landscape of the stochastic environment and the subjective perspective on this stochastic landscape for a case with two risky assets (under condition that the necessary adjustments are indeed undertaken).

One key result can be viewed as the effective **"overconfidence"**. Because of the constraint on utilized information processing capacity, the subjective probability measure $h(\cdot)$ is characterized by lower entropy than the objective probability measure $g(\cdot)$, i.e., the former is a coarser version of the latter. In our case (with log-normal payoffs), the entropy reduction is achieved solely by decreasing the variance of the relevant random variables. Specifically, the variance decomposition equation from Proposition 5 implies that $\hat{\boldsymbol{\Sigma}}_r$, the variance-covariance matrix of $\hat{\boldsymbol{r}}_{t+1}$, is smaller than $\boldsymbol{\Sigma}_r$, its counterpart for $\boldsymbol{r}_{t+1}$ (the difference is a negative semi-definite matrix). This is reflected in Figure 5 by the relatively more peaked probability densities in the right panel.

An additional important result can be viewed as necessary **"pessimism"**. In order to compensate for this entropy-reducing (hence, variance-decreasing in our case) simplification, the means of approximating random variables have to be adjusted. Thereby, a biased second moment necessitates a compensating bias in the first moment. Effectively, the required adjustment amounts to adopting a subjectively pessimistic view at the future state of the world (i.e., for a positive investment its expected value is biased downward). The exact bias term $\breve{\boldsymbol{\mu}}_r$, as given by Proposition 3, depends on the shares of wealth invested in different risky assets, $\boldsymbol{\omega}_t$, as well as on the discrepancy between objective and subjective variances, $(\boldsymbol{\Sigma}_r - \hat{\boldsymbol{\Sigma}}_r)$. Failure to adopt such a compensating adjustment, or using an incorrect adjustment term leads to suboptimal decision outcomes and subsequent losses.[20] For instance, ignoring the pessimistic adjustment and setting the mean $\hat{\boldsymbol{\mu}}_r$ of log-returns $\hat{\boldsymbol{r}}_{t+1}$ so as to match the subjectively expected level of returns $\mathrm{E}_t^h[\hat{\boldsymbol{R}}_{t+1}]$ to the objectively expected one $\mathrm{E}_t^g[\boldsymbol{R}_{t+1}]$ for a risk-averse agent would result in relative "overinvestment" into risky assets in terms of the wealth share. Figure 5 demonstrates the optimal result with the expected value of the top right panel's probability distribution moved slightly to the west (although the bottom right panel's distribution is actually shifted to the east).

[20]Note that this result is not necessarily at odds with our initial motivating experiment due to Gabaix and Laibson (2000), which implies "overconfidence" without "pessimism". If the differences of simplified variances from true variances are about the same for different root nodes in the choice set, the mean bias term is roughly equalized between available choices, and thus can be ignored in their setting.

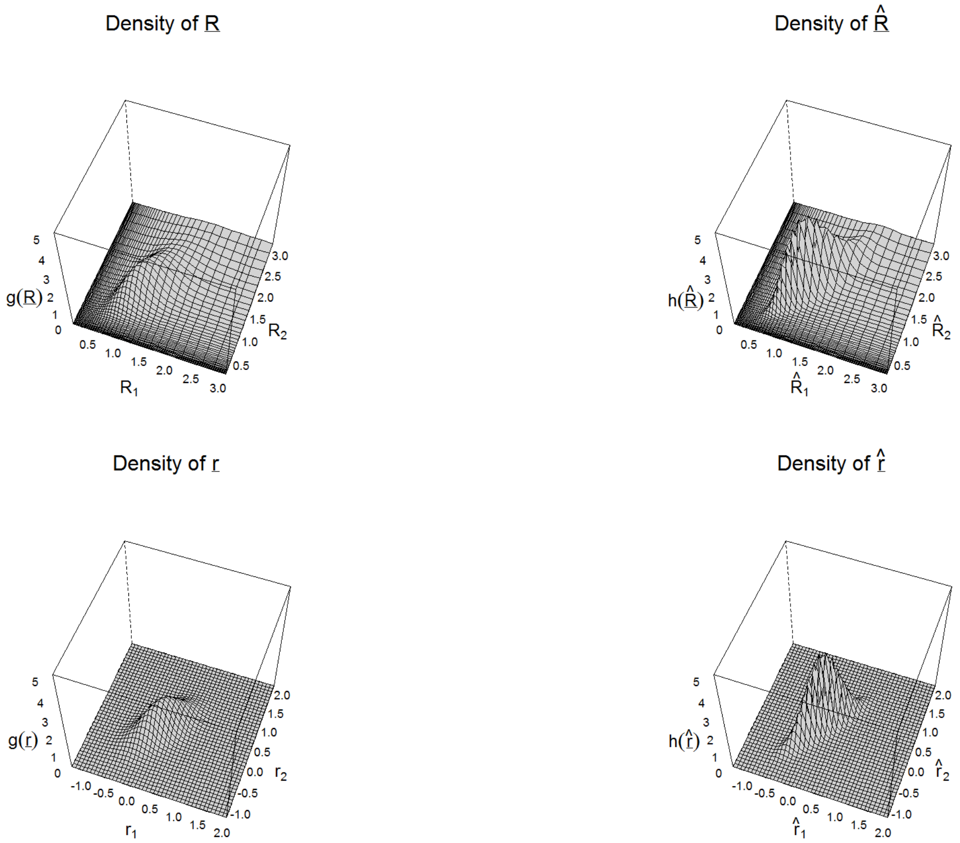

Figure 5: Objective and subjective probability densities, two risky assets (parameterizations used are $\boldsymbol{R}_{t+1} \sim \log \mathcal{N}(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r)$, $\boldsymbol{r}_{t+1} \sim \mathcal{N}(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r)$, $\boldsymbol{\mu}_r = [0.10; 0.20]$, $\boldsymbol{\Sigma}_r = [0.10, 0.08; 0.08, 0.16]$, $\mathrm{E}_t^g[\boldsymbol{R}_{t+1}] = [1.16; 1.32]$ (left panel); and, for $\kappa = 1$ nat ≈ 1.44 bits with $\boldsymbol{\omega}_t = [0.5; 0.5]$, $\hat{\boldsymbol{R}}_{t+1} \sim \log \mathcal{N}(\hat{\boldsymbol{\mu}}_r, \hat{\boldsymbol{\Sigma}}_r)$, $\hat{\boldsymbol{r}}_{t+1} \sim \mathcal{N}(\hat{\boldsymbol{\mu}}_r, \hat{\boldsymbol{\Sigma}}_r)$, $\hat{\boldsymbol{\mu}}_r = [0.11; 0.21]$, $\hat{\boldsymbol{\Sigma}}_r = [0.06, 0.08; 0.08, 0.12]$, $\mathrm{E}_t^h[\hat{\boldsymbol{R}}_{t+1}] = [1.15; 1.31]$ (right panel)).

Another key result is emerging **"categorization"**.[21,22] In our case, the positively correlated elements of $\boldsymbol{r}_{t+1}$ become even more correlated in $\hat{\boldsymbol{r}}_{t+1}$, thus similarly behaved co-moving assets exhibit a sort of attraction. While negatively correlated elements become even more so, leading to the repulsion of the counter-moving assets. These dynamics engender subjective clustering of different assets into relatively distinct categories, "asset classes". The mechanics behind it are the same as before: the covariance terms in $\hat{\boldsymbol{\Sigma}}$, the variance-covariance matrix of decorrelated simplified random variables $\hat{\boldsymbol{x}}$, are unchanged, but the variance terms are reduced. Consequently, leading to the correlation coefficients exceeding those in $\boldsymbol{\Sigma}$, the variance-covariance matrix of decorrelated original random variables $\boldsymbol{x}$. Strictly speaking, the effect on the correlations in $\hat{\boldsymbol{\Sigma}}_r$ is ambiguous and depends on the tightness of the information constraint and the combination of the diagonal elements in $\boldsymbol{\Sigma}$, as well as on decorrelating eigenvectors $\boldsymbol{\Xi}$. However, overall the "correlation inflation" outcome dominates: the entropy of a Gaussian random vector is a one-to-one map with the determinant of its variance-covariance matrix, and so can be reduced only by decreasing the variances of and/or increasing the (absolute magnitudes of) covariances between the random vector's elements.

As a consequence, assets in the same category (i.e., positively correlated ones) would tend to be treated as more similar than they actually are, while assets in different categories (negatively correlated) would seem more different that they really are. For this reason, without proper adjustment of the mean, as say is happening in expected value-matching that results in overall "overinvesment" into risky assets, both "over-" and "un-

[21]It is worthwhile to distinguish two kinds of categorization: bundling of random variables' support set partitions (i.e., bundling of states; e.g., coarsening of the state space through merging of several states into one) and pooling of the random variables/different random vector's elements themselves (i.e., pooling of types; e.g., conditionally on one random variable the other converges to a non-stochastic (Dirac delta) function). (In principle, merging of support set partitions also happens in the process of quantization of a continuous random variable, but we focus on different issues at this point.)

Technically, entropy/variance reduction for uncorrelated random variables leads, at its extreme, to "dispersion folding" or "randomness collapse"/"distribution contraction" (folding and dropping out low-volatility categories, think of it as categorization in vs. categorization away; this formally corresponds to bundling of states). However, for correlated random variables the above pertains to random variables' dimensions, and leads to "correlation inflation" (clustering into similar categories, categorization together vs. categorization apart; this formally corresponds to pooling of types). The former is dealt with by Corrollary 1, the latter by Corrollary 2.

Importantly, now we examine onle one of the two categorization mechanisms: instead of focusing on categorization as "folding" and dropping out (less important) random variables, here we focus on categorization as "correlation inflation" between random variables. We discuss the economic role of the other mechanism in more detail elsewhere, as a part of a separate line of investigation.

[22]In the words of Herbert Simon (1997), "The human being striving for rationality and restricted within the limits of his knowledge has developed some working procedures that partially overcome these difficulties. These procedures consist in assuming that he can isolate from the rest of the world a closed system containing a limited number of variables and a limited range of consequences." In our framework, the "limited number of variables" idea roughly corresponds to the "folding" effect, while the "limited range of consequences" to "correlation inflation" and entropy/variance reduction more generally.

derinvestment" is possible for specific risky assets depending on the relative size and direction of the correlation biases. It can be seen in Figure 5 that the random variables described by the probability densities of the right panel are more aligned along the west-east axis and hence exhibit a higher pairwise correlation ($\hat{\rho}_{r,12} = 0.90 > 0.63 = \rho_{r,12}$).[23]

Nevertheless, the **decision outcomes** (say, prices $P^*_{0,t}$ and $\boldsymbol{P}^*_t$) for agents implementing the approximation procedures correctly by construction, in accordance with Corollary 3, achieve constrained optimality. That is, are in the "neighborhood" of fully optimal unconstrained "rational expectations" outcomes, with the "radius" of the neighborhood inversely related to the magnitude of information processing capacity κ. Technically, decision errors appear due to the approximation to the wealth process, which leads to a positive expected distortion $\mathrm{E}^f_t[d(\boldsymbol{r}_{t+1}, \hat{\boldsymbol{r}}_{t+1})] = \boldsymbol{\Psi}_r \geq \boldsymbol{0}$. The approximation is very accurate in practice though (see the comments on Proposition 3.1 in §E.2). Also note that decision errors are non-systematic/symmetric around the fully optimal levels (see Proposition 5), hence they are likely to cancel out on aggregate; and are relatively smaller for contingencies that impact welfare the most (see §I). At the same time, non-negligible radius of the neighborhood of deviations from full optimality leaves room for positive **trading volumes** even between agents having access to exactly same information and identical in all other respects except different levels of κ (although agent multiplicity is not modeled explicitly here).[24]

4.2 Practice and empirics

Now, we briefly talk about practical illustrations and empirical findings related to the presented theoretical framework. Our main result regarding the subjective ampification of the correlations between different risky assets has a number of practical consequences.

In general, correlation inflation leads to apparent **"underdiversification"**, meaning that from a subjective perspective investment portfolios may look less diversified than they actually are (recall how correlation coefficient rises from the true value of 0.63 to the approximating level of 0.90 in the simple illustration of §4.1). This produces the familiar "portfolio concentration/underdiversification puzzle" (see Blume and Friend (1975), Statman (1987), Kelly (1995) for individual investors' domestic portfolios; French and

[23]One way to look at this categorization result is through the lens of principal component analysis. Propositions 4 and 5 reveal the effective amplification of the relative magnitude of the largest eigenvalues of the subjective variance-covariance matrix that in turn leads to the amplification of the relative share of the subjective random variables' variance captured by their first principal components. As long as any two variables share the same leading principal components—in other words, the absolute magnitude of their pairwise correlation coefficient is high—then an increase in these leading principal components' importance increases (in absolute terms) the correlation between the two variables.

[24]We again abstract away from the fact that in our particular exchange economy, consumption and investment outcomes coincide in constrained and unconstrained cases, and (as long as competitive equilibrium is unique/markets are complete) so do prices, hence there is actually no room for decision errors as well as trade.

Poterba (1991) for country-level international portfolios) in the situation of "undersampling" of the complex population distribution, that is when using a sample variance (i.e., a lower-entropy approximation) in place of a population variance featuring unobserved "rare events" (i.e., a higher-entropy true counterpart).[25]

In the case of positively correlated assets, their subjective clustering into **asset classes** emerges endogenously, since correlation inflation entails categorization. For example, stocks in the Australian mining company BHP Billiton and Chicago Mercantile Exchange futures contracts on crude WTI oil may have a "true" correlation of returns well below 1, and yet be subjectively viewed by investors as correlated more tightly than that and treated as a single asset class "commodities"; while shares of U.S. companies with very different business fundamentals may be mechanically merged into an asset class "small value" or "technology" stocks. Such effects engender the (self-reinforcing) popularity of operating in terms of aggregated asset classes instead of disaggregated assets among investors and econometricians alike, fueling the interest in "asset allocation", "asset comovement" and "style investing" (see Sharpe, 1992; Brinson et al., 1986 and 1991; Doeswijk et al., 2014; Fama and French, 1993; Barberis and Shleifer, 2003; Barberis et al., 2005).

A straightforward theoretical prediction of our approach is that agents with lower information processing capacity will be relatively more predisposed to such clustering, manifesting stronger real effects. In turn, more sophisticated investors tend to be professional market participants who, due to competitive forces in the labor market, are supposed to be less capacity-constrained and less prone to suboptimal decisions, thus enabling empirical identification. The style investing phenomenon is very well studied in the literature, it provides enough evidence and variation to verify this claim. Indeed, there are ample confirmations of such clustering, it has a real effect on demand and outcomes, and that effect goes over and above the class member's true characteristics, its underlying fundamentals (e.g., Pindyck and Rotemberg, 1990; Chan et al., 2000; Teo and Woo, 2004; Froot and Teo, 2008; Choi and Sias, 2009). Moreover, unsystematic findings scattered in several style investing papers provide empirical evidence that correlation, or herding, of investment decisions is stronger within less sophisticated retail investors than it is among more sophisticated institutional investors. An explicit comparison confirming this result is conducted in Kumar and Lee (2006) as well as, more prominently, in Jame and Tong (2014), with the latter work reporting the values for a popular measure of such herding in two market constituencies we are concerned about at 4.01% and 2.09%, respectively. This is consistent with the above theoretical prediction.

[25] In spite of being a subjective problem of perspective, it may have real consequences for an investor that relies on the simplified variance $\hat{\boldsymbol{\Sigma}}_r$ but the matched expected return $\mathrm{E}^h[\hat{\boldsymbol{R}}_{t+1}] = \mathrm{E}^g[\boldsymbol{R}_{t+1}]$, and maximizes portfolio return subject to a constraint on the accepted level of portoflio variance. In general, he will underappreciate the benefits of diversification: in a simple example, expanding a portfolio from one asset with variance $\hat{\sigma}_r$ to a portfolio split equally between two assets with equal variances $\hat{\sigma}_{r,1} = \hat{\sigma}_{r,2} := \hat{\sigma}_r$ and correlation coefficient $\hat{\rho}_{r,12}$ lowers the portfolio variance by $0.5\hat{\sigma}_r(1 - \hat{\rho}_{r,12})$, with the latter quantity contracting as $\hat{\sigma}_r$ falls and $\hat{\rho}_{r,12}$ rises further.

In the case of negatively correlated assets, they may endogenously form subjective **hedging instruments**. For instance, portfolios of government bonds and portfolios of stocks may have a slightly negative "true" correlation of returns in some regimes/time periods, usually involving so-called "flight-to-quality" episodes (Li, 2002; Connolly et al., 2005; Guidolin and Timmermann, 2007; Andersson et al., 2008; Yang et al., 2009), but subjective amplification of such correlations could be responsible for an often-held view of bonds serving the role of a hedge for stocks (for examples, see Canner et al., 1997). Such hedging motives are the main focus of the "strategic asset allocation" literature, e.g., see Brennan et al. (1997), Campbell and Viceira (2002b), also refer to Wachter (2010) for a more recent review paper.

5 Conclusion

We conclude with a concise summary as well as make several closing remarks that discuss additional aspects of our work and put it into a broader perspective.

The ambition of the current paper is to improve our understanding of how in the stochastic, risky environment economic decisions are made by real people, whose information-processing abilities may be limited, as opposed to fictitious entities endowed with unbounded computational resources. The paper develops a positive (rather than normative) theoretical framework for decision-making under risk, which presumes rational optimizing behavior of the agents, builds from first principles (and yet is very tractable analytically), follows the discipline of information theory, is consistent with theoretical and empirical findings in neuroscience as well as with the results of economic laboratory experiments. In particular, neuroscientific and information-theoretic arguments allow us to structurally motivate and constructively quantify the costs of information processing. The selected application is an investment problem in the context of a general equilibrium Lucas tree model.

In the constructed model, agent's constrained-optimal behavior exhibits such effects as "overconfidence", "pessimism" and "categorization". "Categorization" is implied by "dispersion folding" (pruning and simplifying the perceived environment) and embodied in "correlation inflation" (subjectively clustering covarying random variables together). The latter effect may be responsible for some well-known empirical regularities including portfolio underdiversification puzzle and style investing phenomenon.

This paper develops an approach to evaluating expectations of stochastic objects that explicitly accounts for constraints imposed by the available information processing capacity. This is done without loss of generality, as traditional "rational expectations" are nested within and emerge as a special case when the information constraint is not binding. Such generalization effectively allows to explore the information processing demands of the rational expectation formation. It turns out that, from a technical standpoint, as a process of computing an integral with respect to some probability measure its requirements are not prohibitive, because various adjustments can reduce the computational costs dramatically without substantial efficiency losses, thus producing the decision outcomes approaching the “rational expectations” benchmark quite closely (at least given the assumptions made here). However, from a more conceptual standpoint, as an equilibrium notion it imposes strong restrictions, and forcing the objective and subjective distributions to coincide is not innocuous and may lead to very misleading, even puzzling results. In short, “rational expectations” provide a sufficiently robust benchmark, but may be too inflexible to be applied blindly.

Lastly, we should emphasize that our treatment is more general than it may seem at first: it also accounts for information processing performed with the aid of machines, which is relevant for any empirical parallels going beyond toy examples. Appendix §K fleshes out this point in more detail.

References

[1] Abel, Andrew B. (2002) “An Exploration of the Effects of Pessimism and Doubt on Asset Returns.” *Journal of Economic Dynamics and Control*, 26 (7–8): 1075–1092.

[2] Alvarez, Fernando and Urban J. Jermann. (2005) “Using Asset Prices to Measure the Persistence of the Marginal Utility of Wealth”, *Econometrica*, 73(6): 1977–2016.

[3] Anderson, Anders. (2013) “Trading and Under-Diversification”, *Review of Finance*, 17: 1699–1741.

[4] Andersson, Magnus, Elizaveta Krylova and Sami Vähämaa. (2008) “Why Does the Correlation Between Stock and Bond Returns Vary Over Time?”, *Applied Financial Economics*, 18: 139–151.

[5] Backus, David, Mikhail Chernov and Stankey Zin. (2014) “Sources of Entropy in Representative Agent Models”, *Journal of Finance*, 69(1): 51–99.

[6] Baddeley, Alan. (2003) “Working Memory: Looking Back and Looking Forward.” *Nature Reviews Neuroscience*, 4: 829–839.

[7] Baddeley, R., P. Hancock and P. Földiák (eds.). (2000) *Information Theory and the Brain*, New York, NY: Cambridge University Press.

[8] Barber, Brad M., Terrance Odean, Ning Zhu. (2009) “Systematic Noise”, *Journal of Financial Markets*, 12: 547–569.

[9] Barberis, Nicholas and Andrei Shleifer. (2003) “Style investing”, *Journal of Financial Economics*, 68: 161–199.

[10] Barberis, Nicholas, Andrei Shleifer and Jeffrey Wurgler. (2005) “Comovement”, *Journal of Financial Economics*, 75: 283–317.

[11] Barlow, Horace B. (1961) “Possible Principles Underlying the Transformation of Sensory Messages.” In: W. Rosenblith (ed.), *Sensory Communication*, Cambridge, MA: MIT Press.

[12] Bayer, Hannah. (2008) “Focus on Decision Making”, *Nature Neuroscience*, 11(4): 387.

[13] Berger, Toby. (1971) *Rate-distortion Theory: A Mathematical Basis for Data Compression*, Englewood Cliffs, NJ: Prentice-Hall.

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

[15] Blume, Marshall E. and Irwin Friend (1975) “The Asset Structure of Individual Portfolios and Some Implications for Utility Functions”, *The Journal of Finance*, 30(2): 585–603.

[16] Bor, Daniel, John Duncan, Richard J. Wiseman and Adrian M. Owen. (2003) “Encoding Strategies Dissociate Prefrontal Activity from Working Memory Demand”, *Neuron* 37: 361–367.

[17] Bor, Daniel and Adrian M. Owen. (2006) “Working Memory: Linking Capacity with Selectivity”, *Current Biology*, 16(4): 136–138.

[18] Bordalo, Pedro, Nicola Gennaioli and Andrei Shleifer. (2016) “Diagnostic Expectations and Credit Cycles.” Manuscript.

[19] Boyer, Brian H. (2011) “Style-Related Comovement: Fundamentals or Labels?”, *The Journal of Finance*, 66(1): 307–332.

[20] Breeden, Douglas T. (1979) “An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities.” *Journal of Financial Economics*, 7: 265–296.

[21] Brown, Stephen J. and William N. Goetzmann. (1997) “Mutual Fund Styles”, *Journal of Financial Economics*, 43: 373–399.

[22] Brunnermeier, Markus K., Christian Gollier, and Jonathan A. Parker. (2007) "Optimal Beliefs, Asset Prices, and the Preference for Skewed Returns." *American Economic Review (Papers and Proceedings)*, 97 (2): 159–165.

[23] Brunnermeier, Markus K. and Jonathan A. Parker. (2005) "Optimal Expectations." *American Economic Review*, 95 (4): 1092–1118.

[24] Brunnermeier, Markus K., Filippos Papakonstantinou and Jonathan A. Parker. (2013) "Optimal Time-Inconsistent Beliefs: Misplanning, Procrastination, and Commitment." Manuscript.

[25] Bunge, Silvia A. (2005) "Foreword", *Cognitive Brain Research*, 23: 1.

[26] Camerer, Colin F. (1995) "Individual Decision Making." In: John H. Kagel and Alvin E. Roth (eds.), *Handbook of Experimental Economics*, chapter 8, pages 587–703, Princeton, NJ: Princeton University Press.

[27] Campbell, John Y. and John Ammer. (1993) "What Moves the Stock and Bond Markets? A Variance Decomposition for Long-Term Asset Returns", *The Journal of Finance*, 48(1): 3–37.

[28] 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.

[29] Campbell, John Y. and Luis M. Viceira. (2002a) *Appendix for "Strategic Asset Allocation: Portfolio Choice for Long-Term Investors"*, Oxford University Press.

[30] Campbell, John Y. and Luis M. Viceira. (2002b) *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors*, Oxford University Press.

[31] Campbell, John Y., Yeung Lewis Chan and Luis M. Viceira. (2003) "A Multivariate Model of Strategic Asset Allocation", *Journal of Financial Economics*, 67: 41–80.

[32] Canner, Niko, N. Gregory Mankiw and David N. Weil. (1997) "An Asset Allocation Puzzle", *American Economic Review*, 87(1): 181–191.

[33] Caplin, Andrew and Mark Dean. (2013) "The Behavioral Implications of Rational Inattention with Shannon Entropy." Manuscript.

[34] Caplin, Andrew and Mark Dean. (forthcoming) "Revealed Preference, Rational Inattention, and Costly Information Acquisition." *American Economic Review.*

[35] Carroll, Christopher D. (2003) "Macroeconomic Expectations of Households and Professional Forecasters." *Quarterly Journal of Economics*, 118 (1): 269–298.

Too Good to Be True?" *American Economic Review*, 90 (4): 787–805.

[36] Chan, Louis K.C., Jason Karceski, and Josef Lakonishok. (2000) "New Paradigm or Same Old Hype in Equity Investing?", *Financial Analysts Journal*, 56(4): 23–36.

[37] Chen, Hui, Winston Wei Dou and Leonid Kogan. (2015) "Measuring the "Dark Matter" in Asset Pricing Models." Manuscript.

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

[39] Choi, Nicole and Richard W. Sias. (2009) "Institutional Industry Herding", *Journal of Financial Economics*, 94: 469–491.

[40] Connolly, Robert, Chris Stivers and Licheng Sun. (2005) "Stock Market Uncertainty and the Stock-Bond Return Relation", *Journal of Financial and Quantitative Analysis*, 40(1): 161–194.

[41] Cooper, Michael J., Huseyin Gulen and P. Raghavendra Rau. (2005) "Changing Names with Style: Mutual Fund Name Changes and Their Effects on Fund Flows", *The Journal of Finance*, 60(6): 2825–2858.

[42] Cover, Thomas M. and Joy A. Thomas. (2006) *Elements of Information Theory*, 2nd ed., New York: Wiley-Interscience.

[43] Cowan, Nelson. (2000) "The Magical Number 4 in Short-term Memory: A Reconsideration of Mental Storage Capacity", *Behavioral and Brain Sciences*, 24(1): 87–185.

[44] Damper, Robert I. and Stevan Harnad. (2000) "Neural Network Modeling of Categorical Perception", *Perception & Psychophysics*, 62(4): 843–867.

[45] Dayan, Peter and Laurence F. Abbott. (2001) *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*, The MIT Press.

[46] Doeswijk, Ronald Q., Trevin W. Lam and Laurens Swinkels. (2014) "The Global Multi-Asset Market Portfolio 1959–2012." Manuscript.

[47] Doya, Kenji, Shin Ishii, Alexandre Pouget and Rajesh P.N. Rao (eds.). (2007) *Bayesian Brain: Probabilistic Approaches to Neural Coding*, The MIT Press.

[48] Eliasmith,Chris, Terrence C. Stewart, Xuan Choo, Trevor Bekolay, Travis DeWolf, Yichuan Tang, Daniel Rasmussen. (2012) "A Large-Scale Model of the Functioning Brain." *Science*, 338(6111): 1202–1205.

[49] Fama, Eugene F. and Kenneth R. French. (1993) "Common Risk Factors in the Returns on Stocks and Bonds", *Journal of Financial Economics*, 33: 3–56.

[50] Fleming, Jeff, Chris Kirby and Barbara Ostdiek. (1998) "Information and Volatility Linkages in the Stock, Bond, and Money Markets", *Journal of Financial Economics*, 49: 111–137.

[51] Fleming, Stephen M., Laurence T. Maloney and Nathaniel D. Daw. (2013) "The Irrationality of Categorical Perception", *Journal of Neuroscience*, 33(49): 19060–19070.

[52] French, Kenneth R. and James M. Poterba. (1991) "Investor Diversification and International Equity Markets", *American Economic Review*, 81(2): 222–226.

[53] Friston, Karl. (2009) "The Free-energy Principle: A Rough Guide to the Brain?" *Trends in Cognitive Sciences*, 13(7): 293–301.

[54] Friston, Karl. (2010) "The Free-energy Principle: A Unified Brain Theory?" *Nature Reviews Neuroscience*, 11: 127–138.

[55] Froot, Kenneth and Melvyn Teo. (2008) "Style Investing and Institutional Investors", *Journal of Financial and Quantitative Analysis*, 43(4): 883–906.

[56] Fuster, Andreas, Benjamin Hebert, and David Laibson. (2012) "Natural Expectations, Macroeconomic Dynamics, and Asset Pricing." *NBER Macroeconomics Annual*, 26 (1): 1–48.

[57] Gabaix, Xavier. (2014a) "A Sparsity-Based Model of Bounded Rationality." *Quarterly Journal of Economics*, 129 (4): 1661–1710.

[58] Gabaix, Xavier. (2014b) "Sparse Dynamic Programming and Aggregate Fluctuations." Manuscript.

[59] Gabaix, Xavier and David Laibson. (2000) "A Boundedly Rational Decision Algorithm." *American Economic Review (Papers and Proceedings)*, 90 (2): 433–438.

[60] Gabaix, Xavier and David Laibson. (2005) "Bounded Rationality and Directed Cognition." Manuscript.

[61] Gabaix, Xavier, David Laibson, Guillermo Moloche, and Stephen Weinberg. (2006) "Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model." *American Economic Review*, 96 (4): 1043–1068.

[62] Ghosh, Anisha, Christian Julliard, Alex P. Taylor. (2013) "What is the Consumption-CAPM missing? An Information-Theoretic Framework for the Analysis of Asset Pricing Models." Manuscript.

[63] Gigerenzer, Gerd and Reinhard Selten (Eds.). (2001) *Bounded Rationality: The Adaptive Toolbox*, Cambridge, MA: MIT Press.

[64] Gigerenzer, Gerd, Peter M. Todd, and the ABC Research Group. (2000) *Simple Heuristics that Make Us Smart*, New York, NY: Oxford University Press.

[65] Gilboa, Itzhak and David Schmeidler. (1995) "Case-Based Decision Theory." *Quarterly Journal of Economics*, 110 (3): 605–639.

[66] Gilovich, Thomas, Dale W. Griffin, and Daniel Kahneman (Eds.). (2002) *Heuristics and Biases: The Psychology of Intuitive Judgment*, New York, NY: Cambridge University Press.

[67] Glimcher, Paul W., Colin F. Camerer, Ernst Fehr, Russell A. Poldrack (eds.). (2008) *Neuroeconomics: Decision Making and the Brain*, Academic Press.

[68] Goetzmann, William N. and Alok Kumar. (2008) "Equity Portfolio Diversification", *Review of Finance*, 12: 433–463.

[69] Goetzmann, William N., Lingfeng Li, and K. Geert Rouwenhorst. (2005) "Long-Term Global Market Correlations", *The Journal of Business*, 78(1): 1–38.

[70] Goldstone, Robert L. and Andrew T. Hendrickson. (2010) "Categorical perception", *WIREs Cognitive Science*, 1(1): 69–78.

[71] Guidolin, Massimo and Allan Timmermann. (2007) "Asset Allocation under Multivariate Regime Switching", *Journal of Economic Dynamics and Control*, 31: 3503–3544.

[72] Hansen, Lars P. (2007) "Beliefs, Doubts and Learning: Valuing Macroeconomic Risk." *American Economic Review*, 97 (2): 1–30.

[73] Hansen, Lars P. (2014) "Nobel Lecture: Uncertainty Outside and Inside Economic Models", *Journal of Political Economy*, 122(5): 945–987.

[74] Hansen, Lars P. and Thomas J. Sargent. (2007a) *Robustness*, Princeton, NJ: Princeton University Press.

[75] Hansen, Lars P. and Thomas J. Sargent. (2007b) "Recursive Robust Estimation and Control Without Commitment." *Journal of Economic Theory*, 136 (1): 1–27.

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

[77] Jame, Russell and Qing Tong. (2014) "Industry-based Style Investing", *Journal of Financial Markets*, 19: 110–130.

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

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

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

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

[82] Jehiel, Philippe. (2005) "Analogy-Based Expectation Equilibrium." *Journal of Economic Theory*, 123 (2): 81–104.

[83] Kelly, Morgan. (1995) "All Their Eggs in One Basket: Portfolio Diversification of US Households", *Journal of Economic Behavior and Oraanization*, 27: 87–96.

[84] Kleinberg, Joel and Herbert Kaufman. (1971) "Constancy in Short-term Memory: Bits and Chunks", *Journal of Experimental Psychology*, 90(2): 326–333.

[85] Kumar, Alok. (2009) "Dynamic Style Preferences of Individual Investors and Stock Returns", *Journal of Financial and Quantitative Analysis*, 44(3): 607–640.

[86] Kumar, Alok and Charles M. C. Lee. (2006) "Retail Investor Sentiment and Return Comovements", *The Journal of Finance*, 61(5): 2451–2486.

[87] Ledoit, Olivier and Michael Wolf (2003) "Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portofolio Selection." *Journal of Empirical Finance*, 10(5): 603–621.

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

[89] Ledoit, Olivier and Michael Wolf (2004b) "Honey, I Shrunk the Sample Covariance Matrix." *Journal of Portfolio Management*, 30(4): 110–119.

[90] Li, Lingfeng. (2002) "Macroeconomic Factors and the Correlation of Stock and Bond Returns". *Yale International Center for Finance Working Paper*, No. 02-46.

[91] Li, Lingfeng. (2003) "An Economic Measure of Diversification Benefits". *Yale International Center for Finance Working Paper*, No. 03-11.

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

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

[94] Manski, Charles F. (2017) "Survey Measurement of Probabilistic Macroeconomics Expectations: Progress and Promise." Manuscript.

[95] Marschak, Jacob. (1959) "Remarks on the Economics of Information." *Cowles Foundation Discussion Paper*, No. 70.

[96] Marschak, Jacob. (1968) "Economics of Inquiring, Communicating, Deciding." *American Economic Review*, 58(2): 1–18.

[97] Marschak, Jacob. (1971) "Economics of Information Systems." *Journal of the American Statistical Association*, 66(333): 192–219.

[98] Matějka, Filip. (2014a) "Rigid Pricing and Rationally Inattentive Consumer." Manuscript.

[99] Matějka, Filip. (2014b) "Rationally Inattentive Seller: Sales and Discrete Pricing." Manuscript.

[100] Matějka, Filip and Alisdair McKay. (2014) "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model." Manuscript.

[101] Matějka, Filip and Christopher A. Sims. (2010) "Discrete Actions in Information-Constrained Tracking Problems." Manuscript.

[102] Merton, Robert C. (1969) "Lifetime Portfolio Selection under Uncertainty: The Continuous-Time Case." *The Review of Economics and Statistics*, 51(3): 247–257.

[103] Migliore, Michele, Gaspare Novara and Domenico Tegolo. (2008) "Single Neuron Binding Properties and the Magical Number 7", *Hippocampus*, 18: 1122–1130.

[104] Miller, George A. (1956). "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information." *Psychological Review*, 63(2): 81–97.

[105] Mullainathan, Sendhil. (2002a) "A Memory-Based Model of Bounded Rationality." *Quarterly Journal of Economics*, 117 (3): 735–774.

[106] Mullainathan, Sendhil. (2002b) "Thinking Through Categories." Manuscript.

[107] Ortega, Pedro A. and Daniel A. Braun. (2013) "Thermodynamics as a Theory of Decision-Making with Information-Processing Costs", *Proceedings of the Royal Society A*, 469: 20120683.

[108] Owen, Adrian M. (2004) "Working Memory: Imaging the Magic Number Four", *Current Biology*, 14: 573–574.

[109] Padoa-Schioppa, Camillo, Aldo Rustichini. (2014) "Rational Attention and Adaptive Coding: A Puzzle and a Solution." *American Economic Review*, 104(5): 507–513.

[110] Pammi, V.S. Chandrasekhar and Narayanan Srinivasan. (2013) "Preface", *Progress in Brain Research*, 202: xi.

[111] Peng, Lin and Wei Xiong. (2006) "Investor Attention, Overconfidence and Category Learning." *Journal of Financial Economics*, 80 (3): 563–602.

[112] Pindyck, Robert S. and Julio J. Rotemberg. (1990) "The Excess Co-Movement of Commodity Prices", *The Economic Journal*, 100(403): 1173–1189.

[113] Polkovnichenko, Valery. (2005) "Household Portfolio Diversification: A Case for Rank-Dependent Preferences", *Review of Financial Studies*, 18(4): 1467–1502.

[114] Rao, Rajesh P. N. (2010) "Decision Making Under Uncertainty: a Neural Model Based on Partially Observable Markov Decision Processes." *Frontiers in Computational Neuroscience*, 4(146): 1–18.

[115] Ravid, Doron. (2016) "Bargaining with Rational Inattention." Manuscript.

[116] Rustichini, Aldo and Camillo Padoa-Schioppa. (2015) "A Neuro-Computational Model of Economic Decisions." *Journal of Neurophysiology*, 114(5): 1382–1398.

[117] Samuelson, Paul A. (1969) "Lifetime Portfolio Selection By Dynamic Stochastic Programming." *The Review of Economics and Statistics*, 51(3): 239–246.

[118] Scheinkman, José A. and Wei Xiong. (2003) "Overconfidence and Speculative Bubbles." *Journal of Political Economy*, 111 (6): 1183–1219.

[119] Schultz, Wolfram. (2008) "Introduction. Neuroeconomics: the Promise and the Profit", *Philosophical Transactions of the Royal Society B*, 363: 3767–3769.

[120] Shannon, Claude E. (1948) "A Mathematical Theory of Communication." *Bell System Technical Journal*, 27: 379–423, 623–656.

[121] Sharpe, William F. (1992) "Asset Allocation: Management Style and Performance Measurement", *Journal of Portfolio Management*, 18(2): 7–19.

[122] Shiffrin, Richard M. and Robert M. Nosofsky. (1994) "Seven Plus or Minus Two: A Commentary On Capacity Limitations", *Psychological Review*, 101(2): 357–361.

[123] Shiller, Robert J. (1989) "Comovements in Stock Prices and Comovements in Dividends", *The Journal of Finance*, 44(3): 719–729.

[124] Shiller, Robert J. and Andrea E. Beltratti. (1992) "Stock Prices and Bond Yields: Can their comovements be explained in terms of present value models?", *Journal of Monetary Economics*, 30: 25–46.

[125] Simon, Herbert A. (1997) *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organizations*, 4th edn, New York: Free Press.

[126] Simon, Herbert A. (1957) *Models of Man: Social and Rational*, New York: John Wiley and Sons.

[127] Sims, Christopher A. (1998) "Stickiness." *Carnegie-Rochester Conference Series on Public Policy*, 49(1): 317–356.

[128] Sims, Christopher A. (2003) "Implications of Rational Inattention." *Journal of Monetary Economics*, 50: 665–690.

[129] Sims, Christopher A. (2006) "Rational Inattention: A Research Agenda." Manuscript.

[130] Sims, Christopher A. (2010) "Rational Inattention and Monetary Economics." In: Benjamin M. Friedman and Michael Woodford (eds.), *Handbook of Monetary Economics*, edition 1, volume 3, chapter 4, pages 155–181, Elsevier.

[131] Squire, Larry R., Floyd Bloom, Nicholas C. Spitzer, Sascha du Lac, Anirvan Ghosh, Darwin Berg (eds.). (2008) *Fundamental Neuroscience*, 3rd ed., Academic Press.

[132] Statman, Meir. (1987) "How Many Stocks Make a Diversified Portfolio?", *The Journal of Financial and Quantitative Analysis*, 22(3): 353–363.

[133] 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.

[134] Still, Susanne, David A. Sivak, Anthony J. Bell and Gavin E. Crooks. (2012) "Thermodynamics of Prediction", *Physical Review Letters*, 109(120604).

[135] Stutzer, Michael. (1995) "A Bayesian Approach to Diagnosis of Asset Pricing Models", *Journal of Econometrics*, 68: 367–397.

[136] Stutzer, Michael. (1996) "A Simple Nonparametric Approach to Derivative Security Valuation", *Journal of Finance*, 51(5): 1633–1652.

[137] Teo, Melvyn, Sung-Jun Woo. (2004) "Style Effects in the Cross-Section of Stock Returns", *Journal of Financial Economics*, 74: 367–398.

[138] Tijsseling, Adriaan and Harnad, Stevan. (1997) "Warping Similarity Space in Category Learning by Backprop Nets." In: Ramscar, M., Hahn, U., Cambouropolos, E. & Pain, H. (eds.) Proceedings of SimCat 1997: Interdisciplinary Workshop on Similarity and Categorization. Department of Artificial Intelligence, Edinburgh University: 263–269.

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

[140] Van Nieuwerburgh, Stijn and Laura Veldkamp. (2009) "Information Immobility and the Home Bias Puzzle", *Journal of Finance*, 64(3): 1187–1215.

[141] Van Nieuwerburgh, Stijn and Laura Veldkamp. (2010) "Information Acquisition and Under-Diversification", *Review of Economic Studies*, 77: 779–805.

[142] Veldkamp, Laura L. (2006) "Information Markets and the Comovement of Asset Prices", *Review of Economic Studies*, 73: 823–845.

[143] Wachter, Jessica A. (2010) "Asset Allocation", *Annual Review of Financial Economics*, 2: 175–206.

[144] Wahal, Sunil and M. Deniz Yavuz. (2013) "Style Investing, Comovement and Return Predictability", *Journal of Financial Economics*, 107: 136–154.

[145] Wilson, Andrea. (2014) "Bounded Memory and Biases in Information Processing." *Econometrica*, 82 (6): 2257–2294.

[146] Won, Joong-Ho, Johan Lim, Seung-Jean Kim and Bala Rajaratnam. (2012) “Condition Number Regularized Covariance Estimation.” Manuscript.

[147] Woodford, Michael. (2014) “An Optimizing Neuroeconomic Model of Discrete Choice.” Manuscript.

[148] Woodford, Michael. (2012) “Inattentive Valuation and Reference-Dependent Choice.” Manuscript.

[149] Yang, Jian, Yinggang Zhou, Zijun Wang. (2009) “The Stock-Bond Correlation and Macroeconomic Conditions: One and a Half Centuries of Evidence”, *Journal of Banking and Finance*, 33: 670–680.