Log, Stock and Two Simple Lotteries^{*}

Sergiy Verstyuk[†]

This version: 18 February 2019

Abstract

This paper studies the problem of decision-making under risk by agents whose information-processing abilities may be limited. The theoretical approach taken here relies on economic laboratory experiments, neuroscientific findings, and information-theoretic formalism. The derived mechanism for processing information can be built into the Lucas tree model, a general equilibrium macro-finance workhorse. It simplifies, though also distorts the agent's subjective perspective of the objective stochastic environment. The former converges to the latter when informationprocessing capacity is sufficiently large (inducing standard "rational expectations"); but in a more realistic case of bounded capacity, certain rational biases of perception emerge endogenously. The most non-trivial one is caterogization: it implies dropping from consideration the less important principal dimensions ("dispersion folding") and amplifying the random variables' interdependencies ("correlation inflation"). This result helps explain the existence of and variations in the practice of style investing in financial markets.

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

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

Contents

1	Introduction							
2	Theory							
	2.1	2.1 Simple lotteries, hard decisions						
	2.2	Algorithm for decision-making under risk						
	2.3 Information constraint and informational problem							
3	Model							
	3.1	Two-p	eriod preview	12				
		3.1.1	Economic setup	12				
		3.1.2	Investment problem	12				
		3.1.3	Problem's infeasibility and feasible alternative	13				
		3.1.4	Feasible investment problem	14				
	3.2	2 Infinite-horizon extension						
		3.2.1	Economic setup	14				
		3.2.2	Feasible investment problem	15				
		3.2.3	Distortion function	15				
		3.2.4	Equilibrium	20				
	3.3	Solution						
4	Discussion							
	4.1	Theore	etical results	27				
		4.1.1	Overconfidence	27				
		4.1.2	Pessimism	29				
		4.1.3	Categorization	29				
		4.1.4	Decision outcomes	31				
	4.2	Practi	ce	31				
5	Con	clusio	n	33				

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

1 Introduction

In this paper, we study a canonical economic problem of consumption-investment choice under risk when the agent's computational, or information-processing, capabilities are bounded. Finding a good decision is burdensome, hence striving for the first-best may not always be worth the efforts. Because cognitive resources are scarce, the decisionmaking process itself can be viewed as a trade-off between closeness to the optimal choice and the costs of getting ever closer to it.

Borrowing from information theory and neuroscience, we can model these informationprocessing capabilities as a capacity (or bandwidth) of a "communication channel" in the brain that passes information between its different regions involved in decision-making. There are several equivalent ways to depict how our agent evaluates a random lottery, a particularly intuitive one is to model his/her processing of the relevant information as the following three-stage process: (i) importing the given random variable's probability distribution into the brain; (ii) drawing realizations from this distribution in a Monte Carlo-type experiment; and (iii) transmitting these realizations via a communication channel to sum them up and eventually calculate the statistics of interest.

The upshot is that due to communication channel's capacity limits, some information is lost in transmission. Specifically, each random draw has to be represented by a shortened encryption code, which effectively amounts to replacing the original random variable's probability distribution with its coarse approximation. The benefit is that using an approximate probability distribution "simplifies" a given "complex" optimization problem just enough for the agent to be able to solve it. The cost is that the information loss implies that the approximating distribution exhibits less uncertainty—or, loosely speaking, dispersion—than the original probability distribution.

We formalize the communication channel capacity by a so-called "mutual information constraint" on the channel's input and output messages. While the uncertainty that characterizes a random variable is measured by the "entropy" of its probability distribution. This is standard in information theory, and is also used in the "rational inattention" theory proposed by Sims (1998, 2003, 2006). The difference is that in Sims's case the transmitted information are new signals about the agent's external environment, while here these are intermediate results of his internal computations.

As an illustration of the differences, consider the following example. There is a fruit tree that lives for one year, with known probability distribution of its future fruit crop. In our case, in Spring you need to decide whether you want to buy the whole future crop. Making this decision requires you to process information about the distribution, which is costly, so you may prefer using not the complex true distribution but some simple proxy distribution when making the decision. You choose such a proxy distribution. Then in this same Spring you buy the future crop. In Sims's case, you are also buying the crop, but you do it later. In Spring you need to decide how you will be gathering information about the fruit crop once it appears. Gathering information is costly, so in this same Spring you choose future information acquisition strategy that will imperfectly inform you about the crop. Then in Fall you gather the information and buy the crop.¹

The above-mentioned reduction in uncertainty of the agents' approximating distribution is consistent with people's choices of random lotteries that have been demonstrated in laboratory experiments like those conducted by Gabaix and Laibson (2000). However, such a "forced" distributional distortion requires from a rational agent an appropriate counter-response. For example, in our consumption-investment application within a stochastic Gaussian setting, the entropy reduction is manifested in lowered variance of the distribution, which in turn can be compensated by a decreased mean, so that the resulting decisions do not systematically deviate from the optimal ones. We refer to the former effect as "overconfidence", and to the latter as "pessimism".

The main contribution of this study is the derivation of what we call "categorization" effect. It results from the entropy-reducing simplification in a multidimensional case, and works as follows:

- (i) eigen-values of the variance-covariance matrix are decreased, so that less crucial random variables/data dimensions they represent become non-stochastic and drop out ("dispersion folding"), thus simplifying the approximating problem;
- (ii) diagonal elements of the variance-covariance matrix are decreased, so that pairwise correlation coefficients they underlie become amplified ("correlation inflation"), thus clustering covaried random variables together in the approximating problem.

As a result, different categories of stochastic objects appear, or become more prominent in agents' perception: for example, in case (i) approximate "models of the world" arise that

¹More concretely, rational inattention theory of Sims 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 Appendix §F for more details about the difference.

are distinct from the original and include a smaller set of random variables; in case (ii) distinct capital asset classes become pronounced whose constituents' small within-class variation is neglected in the face of large between-class differences. Behavior consistent with categorical perception is supported by the experimental and neuroscientific evidence, e.g. see Goldstone and Hendrickson (2010) or Fleming et al. (2013). This result is also surprisingly attuned with Herbert Simon's perspective on bounded rationality (see Simon, 1997).

Categorization, as well as overconfidence and pessimism, are popular theoretical mechanisms in economics and finance. However, instead of imposing them exogenously as has usually been the case so far, in our framework they are derived from structural foundations and so emerge endogenously.

Another contribution is an extension of the Lucas (1978) tree model, which is a microfounded general equilibirum workhorse for capital assets in macroeconomics and finance. We allow for a more general expectations formation process than in the classical formulation, dropping the "rational expectations" assumption in its strongest form and explicitly modeling the information-processing side of the problem. It is reassuring to see that recognition of the information-processing constraints does not undermine the conventially accepted results. Conveniently, the presented formulation is also as tractable as the classical model, similarly admitting analytical closed-form solutions in special cases.

Lastly, this study contributes to the literature on modeling bounded rationality, in particular on the usage of information-theoretic methods in economics. Even though our main result can be obtained in a bare-bones stylized toy model, we attempt to show how information-processing mechanisms relying on mutual information constraint could be built into a conventional structural macro model. (For instance, this task is still a work in progress for rational inattention models beyond the linear-quadratic case.)²

An intuitive theoretical prediction produced by our framework is that agents with lower information processing capacity are more susceptible to categorization. On the empirical side, an example of categorization behavior is a practice in financial markets known as style investing. It comprises analytical activity and market trading conducted in terms of aggregate asset classes, and implies an increased (that is, over and above the levels that can be attributed to certain fundamentals) comovement between investments into assets that belong to the same asset class. As predicted, available empirical findings confirm that relatively less sophisticated retail investors exhibit stronger evidence of such excess comovement than more professional institutional investors do.

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

²Finishing 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 §E); as well as an introduction of the formal distinction between effective and physical information processing capacities, which explains some confusing empirical measurements (see §E and §G).

Appendix §C). 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 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 §F). 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 our papers quite closely interconnected. 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). The application considered here is based on the exchange economy due to Lucas (1978), whose rich asset pricing implications were carefully studied further by Martin (2013). 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).

2 Theory

In this part, we present some results of the laboratory experiments that would motivate the mechanics of our information-processing framework, then we broadly formalize such information-processing mechanics in terms of information theory, and lastly we define the specific formal mechanism that can be readily incorporated into a broader theoretical economic framework.

2.1 Simple lotteries, hard decisions

Our aim here is to examine how decisions under risk³ 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.

³We use the terms "risk" and "uncertainty" interchangeably throughout.



Figure 1: Two lotteries (payoffs (circled) with their probabilities (over contigency edges)).

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 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 footnote⁴).

Formally, the problem looks as follows:

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

where the state vector \boldsymbol{x} is 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 (strictly speaking, its expected felicity) under the time constraint.

Gabaix and Laibson (2000) investigated a more complicated version of the same problem in laboratory conditions with human subjects. They found that a simple heuristic rule that ignores the less probable lottery branches most accurately matches the empirical distribution of choices, in particular, outperforming the fully rational model of behavior.⁵

Effectively, the above heuristic 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 "complexity" at the cost of some "approximation error".

 $^{^{4}}$ The lottery on the right has higher mean (5.0 vs. 4.9) and lower variance (14.5 vs. 14.6).

⁵Specifically, Gabaix and Laibson (2000) 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 compare the explanatory performance of their proposed rule against several alternative algorithms.

2.2 Algorithm for decision-making under risk

Our next task is to formalize this idea of complexity reduction that simplifies the problem of decision under risk.

Ignoring the low-probability decision branches as above implies a reduction in a certain characteristic of the probability distribution associated with the decision outcomes, namely its entropy. Entropy is a measure of uncertainty of a probability distribution; it can be related to distribution's dispersion and thought of as its complexity. Formally, in Shannon's sense entropy of a discretely distributed random variable x is defined as

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

where $|\operatorname{supp}(g)|$ is the cardinality of the support set of $g(\cdot)$.⁶ In the case of continuous space of outcomes, the so-called differential entropy just invokes integration in place of summation. 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.⁷

For example, consider a probability mass function (PMF) $g(\cdot)$ for a random variable x that constitutes some state variable such as a country's real GDP growth rate or a company's revenues (or perhaps a lottery's outcome, along the lines of §2.1). Figure 2 presents such a variable that envisions six possible realizations with probabilities $\{1/8, 1/8, 1/4, 1/4, 1/8, 1/8\}$, which implies the entropy of 2.5 bits (irrespective of numeric realization values). For a given decision-maker, such a distribution may be too complex to deal with, so he is bound to operate in terms of a simpler distribution. That is, to define only a few possible scenarios: "baseline", "positive" and "negative". In other words, to "categorize" out of six fine partitions of the probability space just three coarser ones (thus ignoring some of the contingencies in line with §2.1). In Figure 2 this is illustrated by a simplified PMF $h(\cdot)$ for a random variable \hat{x} that allows for just three different realizations with probabilities $\{1/4, 1/2, 1/4\}$, implying a reduced entropy of merely 1.5 bits.

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, 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 instance, the fact that the complex distribution $g(\cdot)$ has the entropy of 2.5 bits means that respective average codeword length is the same 2.5 bits. Indeed, such a binary alphabet would be {000, 001, 10, 11, 010, 011} for the corresponding realizations of the random variable x; with codewords "10" and "11" each used on average two times more frequently than any of the codewords "000", "001", "010" and "011". On the other hand, the simplified distribution $h(\cdot)$ with its entropy as well as the average codeword length of 1.5 bits admits a

⁶We use $\log(\cdot)$ for $\log_2(\cdot)$ and $\ln(\cdot)$ for $\log_e(\cdot)$ throughout.

⁷Shannon 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).



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

more compressed alphabet $\{00, 1, 01\}$ for the realizations of \hat{x} , thus achieving a shorter representation of the contingencies.

In turn, such a code-based perspective permits us to quantify the demands on the computational resources (or, more broadly, on the information processing capacity) and to construct an explicit mechanism underlying the decision-making process. The mechanism's main ingredient, the process of mental evaluation of a lottery, can be represented by the formal algorithm that we (very briefly) outline next.

The algorithm's mechanics is visualized in Figure 3, which provides the flowchart with pseudocode. It starts in the top left corner with a given distribution g(x) together with the necessary initializations, and finishes in the bottom left corner with a calculated lottery's expected value $E^{h}[x]$, going through the following information-processing stages:

- (i) the compression of the given information about the lottery distribution (via a communication channel transmitting the corresponding description, see an arrow marked with $\mathcal{I}_{\mathcal{A}}(g;h)$ in Figure 3);
- (ii) the loading and providing access to this information within a working memory (allocated in the memory storage, see a box marked with $\hat{n}_d \times \mathcal{M}_{\mathcal{A}}(h)$ in Figure 3);
- (iii) the calculation of lottery's value (via a communication channel transmitting intermediate iterations of such computations, see an an arrow marked with $\mathcal{I}_{\mathcal{A}}(h;h)$ in Figure 3).

At any of these three stages, depending on which of the capacity constraints is binding (description channel capacity, storage memory capacity, or computation channel capacity), an information processing "bottleneck" may arise. Such bottlenecks on the way of information flow can preclude the procedure's smooth completion that is necessary for the lottery's evaluation (and ultimately for making an optimal decision). The constraint posed by the tightest bottleneck is precisely the reason for replacing the complex distribution $g(\cdot)$ with a simpler $h(\cdot)$: the latter distribution allows for shorter codewords than those required for the former. (The functional $\mathcal{I}_{\mathcal{A}}(\cdot; \cdot)$ stands for mutual information between two probability distributions, the input and the output, and it is a standard measure of communication channels' capacity; we will return to it later.)



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

The leading illustration of such a bottleneck is probably item (ii) above, memory and its capacity, which is relatively well studied in neuroscience (see Appendix §G) as well as recognized as an important concept in the economic literature (see Appendix §C 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).

We outlined above just a primitive problem of evaluating a simple lottery.⁸ Analogous structural approach provides a recipe for the evaluation of expectations of arbitrary functions of random variables, as well as their maximization, as required by a canonical problem of decision-making under risk. Moreover, for all of these tasks, and regardless of which of the three potential bottlenecks is actually the relevant one, the corresponding restriction can be equivalently formulated in terms of a communication channel capacity, or "mutual information", constraint (elaborated in the next part). Appendix §E develops the formal algorithm in detail, and Appendix §F offers an intuitive overview of its key features.

2.3 Information constraint and informational problem

In the previous part §2.2, we have amended the standard economic optimization problem with an explicit information-processing mechanism. Now our goal is to (i) introduce and

⁸Which we formalize as ancillary procedure \mathcal{P}_{\int} .

analyze the formal constraint that a simplified distribution $h(\cdot)$ should satisfy for any given complex distribution $g(\cdot)$; and (ii) develop a disciplined way of choosing among many candidate distributions $h(\cdot)$ the one most suitable for the job at hand.

On that account, the algorithm presented earlier quantifies the computational costs of solving a stochastic optimization problem. An exogenously given bound on these costs allows to formulate an information-processing capacity or, equivalently, a mutual information constraint (in short, an information constraint):

$$\mathcal{I}(g(\boldsymbol{x}); h(\hat{\boldsymbol{x}})) := \mathcal{E}(g(\boldsymbol{x})) + \mathcal{E}(h(\hat{\boldsymbol{x}})) - \mathcal{E}(f(\boldsymbol{x}, \hat{\boldsymbol{x}})) \le \kappa$$
(3)

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

To clarify the notation, mutual information $\mathcal{I}(g(\boldsymbol{x}); h(\hat{\boldsymbol{x}}))$ is defined in terms of entropy $\mathcal{E}(\cdot)$, it is familiar from §2.2. A constant κ is a capacity bound, and it is measured in bits.⁹ The probability density $h(\hat{\boldsymbol{x}})$ 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. While $f(\boldsymbol{x}, \hat{\boldsymbol{x}})$ above is an ancillary function that captures the overall structure of stochastic interrelationships in the 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 $h(\hat{\boldsymbol{x}})$; $f(\boldsymbol{x}, \hat{\boldsymbol{x}})$ density's role is to account for all possible contingencies of \boldsymbol{x} vis-à-vis $\hat{\boldsymbol{x}}$.

Intuitively, information constraint (3) trades-off how fine [coarse] probability measure $h(\hat{x})$ is (middle term of the equation's right-hand side) versus how [in]accurate approximation of x by \hat{x} is (rightmost term), taking g(x) (leftmost term) as given. Put differently, with g(x) fixed, to satisfy the information constraint one can either (i) reduce the entropy of $h(\hat{x})$ by making it a coarser probability measure, or (ii) increase the entropy of $f(x, \hat{x})$ by making \hat{x} a less accurate approximation of 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 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).¹⁰

⁹It is important to emphasize that κ is a measure of effective capacity, in contrast to 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 the—at first sight, implausibly—low empirical measurements of implied information processing capacity (going all the way back to the classical Miller, 1956, whose very title "The Magical Number Seven, Plus or Minus Two" reflects how small the estimates of such capacity in bits often are). See Appendix §E for more details.

¹⁰As an alternative intuition, 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 entropy of the conditional distribution of \boldsymbol{x} given $\hat{\boldsymbol{x}}$: we can write $\mathcal{I}(g(\boldsymbol{x}); h(\hat{\boldsymbol{x}})) = \mathcal{E}(g(\boldsymbol{x})) - \mathcal{E}(f(\boldsymbol{x}|\hat{\boldsymbol{x}}))$; 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

To clarify the terminology, henceforth we refer to the original random variable \boldsymbol{x} and its probability distribution $g(\boldsymbol{x})$ as to "true", objective, unconstrained; while labeling the simplified random variable $\hat{\boldsymbol{x}}$ and its distribution $h(\hat{\boldsymbol{x}})$ as "approximating", subjective, constrained objects.

Next, the agent is interested in minimizing the losses due to the presence of information constraint. Thus arises the complexity-related information processing or, alternatively, the evaluation-optimization problem (in short, 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

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

as well as the necessary technical restrictions (hereafter not listed explicitly)

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

$$f(\boldsymbol{x}, \hat{\boldsymbol{x}}) \ge 0$$
 $\forall \boldsymbol{x} \in \operatorname{supp}(g), \hat{\boldsymbol{x}} \in \operatorname{supp}(h)$ $[\nu(\boldsymbol{x}, \hat{\boldsymbol{x}})]$

(Lagrange multipliers on each constraint are specified on the right in square brackets).¹¹

Here, our agent seeks to minimize the expected "distortion" from using the approximating distribution $h(\hat{x})$ instead of the true distribution g(x) subject to information constraint. The distortion function $d(x, \hat{x})$ is taken as given: it is formulated as a distance metric between the values of its two arguments, and is a kind of a "loss function" that measures the shortfalls in felicity due to using a distorted probability distribution. (The choice of appropriate distortion function is problem-specific, we will discuss it later in part §3.1.4; think of it as the utility achieved versus the utility that could hypothetically be achieved absent the information-processing costs.) The technical restrictions just ensure that the resulting $h(\hat{x})$ is a proper probability distribution.¹²

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

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 $\hat{\boldsymbol{x}}$: we can also write $\mathcal{I}(g(\boldsymbol{x}); h(\hat{\boldsymbol{x}})) = \mathcal{E}(h(\hat{\boldsymbol{x}})) - \mathcal{E}(f(\hat{\boldsymbol{x}}|\boldsymbol{x}));$ e.g., in Figure 2 knowing \boldsymbol{x} makes $\hat{\boldsymbol{x}}$ a certainty.

¹¹It may be worth emphasizing: while the information constraint has the same form as in Sims (2003, 2006), in the current study it has a different motivation (more on this in §1 and §F); hence the addition of the informational problem with its distortion minimization (introduced here in §2.3), as well as the subordinated place of the informational problem in the hierarchy of decisions (made clear later in §3.1).

¹²In terms of Appendix §E'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.

Lemma 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 $g(\boldsymbol{x})$, also let $d(\boldsymbol{x}, \hat{\boldsymbol{x}})$ 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

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

Proof. See Appendix §D.1.

This is an established result in information theory. The problem constitutes a wellposed convex minimization. The general solution involves Lagrange multipliers, but those can be identified in a more concrete setting. The form of the solution may look counterintuitive at first: $g(\boldsymbol{x})$ is known and we are looking for $h(\hat{\boldsymbol{x}})$, but the solution provides a conditional distribution of \boldsymbol{x} . However, the key is to realize that the conditional distribution of \boldsymbol{x} given $\hat{\boldsymbol{x}}$ is actually the distribution of approximation error, i.e., the deviation of the original random variable \boldsymbol{x} from its simplified counterpart $\hat{\boldsymbol{x}}$. This indeed suffices for recovering the sought solution.¹³ Later in §3.3 we will demonstrate how this can be done with an example.

Quite sensibly, whenever the information constraint does not bind, the expected distortion can be reduced to zero. That is, $f(\boldsymbol{x}|\hat{\boldsymbol{x}})$ becomes the Dirac delta function centered at $\hat{\boldsymbol{x}}$ then. Looking at it from the economics perspective, density $h(\cdot)$ coincides with $g(\cdot)$, thus inducing familiar "rational expectations" (more on this comes later).

Lastly, we claim that the following property holds.

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

-	-	-	

¹³Given the knowledge of $g(\mathbf{x})$, it is often easier to proceed by just guessing the density $h(\hat{\mathbf{x}})$, which we are chiefly interested in, and then verifying that, together with the deduced conditional distribution, the implied joint density $f(\mathbf{x}, \hat{\mathbf{x}}) := f(\mathbf{x}|\hat{\mathbf{x}})h(\hat{\mathbf{x}})$ satisfies the necessary requirements.

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

In the following part, we embed this information-processing mechanism into a concrete model economy.

3 Model

3.1 Two-period preview

For an easier exposition, first we formulate a two-period model, and an infinite-horizon extension follows later.

3.1.1 Economic setup

Consider the following economic setting, which is essentially a variation of Lucas (1978) tree model. 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 P_t and q_t , respectively. The investments in them bring stochastic dividends D_{t+1} , or "fruits", at the beginning of period (t+1). A K-sized random vector D_{t+1} is distributed, given 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 oneperiod-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. 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, assumed to have a constant relative risk-aversion functional form, that is discounted at rate β . This endowment economy comprises \hat{q} , a strictly positive Ksized constant vector, of risky trees, whose shares are initially owned by the representative agent. A risk-free tree is fictitious, the economy comprises $\hat{q}_0 = 0$ of them, i.e., it exists in zero net supply and can be thought of as a cash credit technology.

3.1.2 Investment problem

The consumer-investor is interested in solving the following consumption and portfolio choice problem, \mathcal{P}_q :¹⁴

$$\max_{C_{t},\{q_{0,t},q_{t}\}} \{u(C_{s}) + \beta \mathsf{E}_{t}^{g} \left[u(C_{t+1})\right]\} = \{u(C_{s}) + \beta \int_{\mathbb{R}_{+}^{K}} u(C_{t+1}) g(\boldsymbol{D}_{t+1}|\boldsymbol{D}_{t}) d\boldsymbol{D}_{t+1}\} \quad \{\mathcal{P}_{q}\}$$

¹⁴Henceforth, we omit the probability densities' subscripts when there is no room for an ambiguity in understanding.

subject to budget constraints (for convenience, defining cum-dividend wealth W_t here)

$$C_{t} + P_{0,t}q_{0,t} + \boldsymbol{P}_{t}^{\mathsf{T}}\boldsymbol{q}_{t} = q_{0,t-1} + (\boldsymbol{P}_{t} + \boldsymbol{D}_{t})^{\mathsf{T}}\boldsymbol{q}_{t-1} =: W_{t},$$
$$C_{t+1} = q_{0,t} + \boldsymbol{D}_{t+1}^{\mathsf{T}}\boldsymbol{q}_{t},$$

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

$$g(\boldsymbol{D}_{t+1}|\boldsymbol{D}_t)$$
 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).

3.1.3 Problem's infeasibility and feasible alternative

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

$$\mathbf{E}_{t}^{g}\left[\varphi^{\sharp}(\boldsymbol{x}|\boldsymbol{\theta})\right] := u(W_{t} - \{P_{0,t}, \boldsymbol{P}_{t}\}^{\mathsf{T}}\boldsymbol{\theta}) + \beta \mathbf{E}_{t}^{g}\left[u([1 \ \boldsymbol{x}^{\mathsf{T}}]\boldsymbol{\theta})\right],$$

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

Our consumer-investor is assumed to know¹⁵ 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 distribution $g(\boldsymbol{x})$ that drives the random variable \boldsymbol{x} and is used in the conditional expectation operator $\mathbf{E}_t^g[\cdot]$. Nevertheless, in general the problem \mathcal{P}_q may be infeasible to solve — potentially it violates the information-processing capacity constraint. Because, even though the agent obtains $g(\cdot)$ as an input into the algorithm of part §2.2, he may be unable to execute the full procedure.

Therefore, the agent focuses instead on maximizing

$$\mathbf{E}_t^h[\varphi(\hat{\boldsymbol{x}}|\boldsymbol{\theta})] := u(W_t - \{P_{0,t}, \boldsymbol{P}_t\}^{\mathsf{T}}\boldsymbol{\theta}) + \beta \mathbf{E}_t^h[u([1 \; \hat{\boldsymbol{x}}^{\mathsf{T}}]\boldsymbol{\theta})],$$

where

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

This more general formulation makes it clear that in solving the stochastic optimization problem, we are using the subjective probability density $h(\hat{x})$ in place of the objective density g(x), with the discrepancy between the two densities depending on the distortion function $d(x, \hat{x})$ and the available information processing capacity κ . The distortion

¹⁵Say, to have learned by time t.

function $d(\boldsymbol{x}, \hat{\boldsymbol{x}})$ between the original and simplified random variables is chosen to be just some reasonable measure of distance in terms of felicities $\varphi^{\sharp}(\boldsymbol{x}|\boldsymbol{\theta})$ and $\varphi(\hat{\boldsymbol{x}}|\boldsymbol{\theta})$. Of course, a high enough capacity (denote it as κ^{\sharp}) allows density $h(\cdot)$ not to diverge from $g(\cdot)$, reducing the distortion to zero. 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 $\varphi^{\sharp}(\boldsymbol{x}|\boldsymbol{\theta})$. More explicit formulations can be found in Appendix §H.¹⁶

3.1.4 Feasible investment problem

A feasible version of our consumption and investment problem would recognize that the information constraint may in fact be binding, and combines the consumption and portfolio choice problem \mathcal{P}_q with the informational problem $\mathcal{P}_{\mathcal{I}}$. This allows to relax the standard problem's idealistic assumptions; but, importantly, is done without loss of generality.

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

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

subject to budget constraints

$$C_{t} + P_{0,t}q_{0,t} + \boldsymbol{P}_{t}^{\mathsf{T}}\boldsymbol{q}_{t} = q_{0,t-1} + (\boldsymbol{P}_{t} + \hat{\boldsymbol{D}}_{t})^{\mathsf{T}}\boldsymbol{q}_{t-1} =: \hat{W}_{t},$$
$$C_{t+1} = q_{0,t} + \hat{\boldsymbol{D}}_{t+1}^{\mathsf{T}}\boldsymbol{q}_{t},$$

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 $\hat{\boldsymbol{D}}_t$ given, with $u(C_t) = C_t^{1-\gamma}/(1-\gamma)$, and where

$$h(\hat{\boldsymbol{D}}_{t+1}|\hat{\boldsymbol{D}}_t)$$
 solves $\mathcal{P}_{\mathcal{I}}$ given $d(\boldsymbol{D}_{t+1}, \hat{\boldsymbol{D}}_{t+1})$ and κ ,
 $g(\boldsymbol{D}_{t+1}|\boldsymbol{D}_t)$ 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 (with respect to the distortion metric used) solution to the auxiliary informational problem. Also, note that in time t, the subjective random variables \hat{D}_t and \hat{W}_t coincide with their objective counterparts (almost surely), but we signify them with hats nevertheless to preserve notational succession across the time periods.

3.2 Infinite-horizon extension

3.2.1 Economic setup

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

¹⁶Our understanding is that the solution to the informational problem $\mathcal{P}_{\mathcal{I}}$ has already been learned by time t or is just computationally easy relative to the solution of the unconstrained problem \mathcal{P}_q .

 D_{s+1} at the beginning of period (s + 1). A K-sized random vector D_s follows a timehomogeneous (stationary) Markov chain defined by transition probability density function $g(D_{s+1}|D_s) = g(D_{t+1}|D_t)$ for all periods s > t. The risk-free asset is still composed of a one-period-lived tree. Its deterministic dividends $D_{0,s+1}$ are normalized to 1 for all periods s > t.

3.2.2 Feasible investment problem

The Bellman equation¹⁷ corresponding to the infinite-horizon feasible problem, \mathcal{P}_{QI} , is

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

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

subject to

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

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

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

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

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

Above we denote the maximum value function as $v^{\sharp}(\{q_{0,t-1}, q_{t-1}\}, D_t)$ for the infeasible unconstrained case, and as $v(\{q_{0,t-1}, q_{t-1}\}, \hat{D}_t)$ for the feasible constrained one. In terms of the notation introduced earlier in §2, now $\varphi^{\sharp}(\boldsymbol{x}|\boldsymbol{\theta}) := v^{\sharp}(\boldsymbol{\theta}, \boldsymbol{x})$ and $\varphi(\hat{\boldsymbol{x}}|\boldsymbol{\theta}) := v(\boldsymbol{\theta}, \hat{\boldsymbol{x}})$.

The Bellman equation's formulation is standard except that the probability density function $h_D(\cdot)$ with respect to which it is defined has been endogenized: now it stems from (PQI-4), the solution to auxiliary sub-problem $\mathcal{P}_{\mathcal{I}}$.¹⁸

3.2.3 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 a larger

¹⁷The 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.

¹⁸Lastly, note that the environment in terms of dynamics and stochasticity is as primitive as possible, the agent solves the same problem again and again. The task of embedding into our framework non-trivial inference and updating with regard to $g(\cdot)$ is outside the scope of this paper, and it is pursued elsewhere.

problem \mathcal{P}_{QI} , and convenient to work with analytically, is a laborous task that we attend to next.

In the interest of clarity, a few points deserve calling attention to. To define a distortion function that is internally consistent, we rely on the structure of unconstrained investment problem's solution (and on that of its constrained counterpart, using the "guess and verify" approach).

Moreover, for analytical convenience we make a number of assumptions (further details on them are introduced in due course):

- (D1) independent identical distributions of asset returns, and
- (D2) log-Normality of asset returns;
- (D3) the norm used in a distortion function is L^2 (squared),
- (D4) the transformation used there is logarithmic, and
- (D5) the (endogenous) wealth share parameter used is determined by the requirement to minimize maximum loss.

To start with, define the new variables, gross returns $R_{0,t+1}$ and \mathbf{R}_{t+1} , as well as, preventing any possible confusion, nominal investments $y_{0,t}$ and y_t :

$$W_{t+1} = q_{0,t} + (\boldsymbol{P}_{t+1} + \boldsymbol{D}_{t+1})^{\mathsf{T}} \boldsymbol{q}_{t} =:$$

=: $P_{0,t} R_{0,t+1} q_{0,t} + (\operatorname{diag}(\boldsymbol{P}_{t}) \boldsymbol{R}_{t+1})^{\mathsf{T}} \boldsymbol{q}_{t} =:$ (4)

$$=: R_{0,t+1}y_{0,t} + \boldsymbol{R}_{t+1}^{\mathsf{T}}\boldsymbol{y}_t; \tag{5}$$

(and similarly, \hat{R}_{t+1}). Also, define the total and tree-specific shares of wealth invested in risky assets, respectively, $\hat{\omega}_t$ and $\boldsymbol{\omega}_t$:

$$\widehat{\omega}_t := \mathbf{1}^{\mathsf{T}} \boldsymbol{\omega}_t := \mathbf{1}^{\mathsf{T}} \operatorname{diag}(\boldsymbol{P}_t) \boldsymbol{q}_t / W_t.$$

(Here, the function $\operatorname{diag}(\cdot)$ takes as an argument a vector and returns a diagonal matrix with a given vector's elements on the main diagonal. Later, the function $\operatorname{diag}^{-1}(\cdot)$ will be the corresponding inverse function that takes a diagonal (or just square) matrix and returns a column vector with the main diagonal's elements of a given matrix.)

First, realize that the value function has the following form:

$$v^{\sharp}(\{q_{0,t}, \boldsymbol{q}_t\}, \boldsymbol{D}_{t+1}) = A(W_{t+1})^{1-\gamma},$$
(6)

where $A := (1 - \beta)^{-\gamma} / (1 - \gamma)$; with analogous expression held for $v(\{q_{0,t}, \boldsymbol{q}_t\}, \hat{\boldsymbol{D}}_{t+1})$.¹⁹

¹⁹We 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.

Second, we assume asset returns are i.i.d. rather than just Markovian (D1 above): $g(\mathbf{R}_{t+1}|\mathbf{R}_t) := g(\mathbf{R}_{t+1})$; as well as log-Normal (D2): $g(\mathbf{R}_{t+1})$ is $\log \mathcal{N}(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r)$.²⁰ For log-returns defined as $\mathbf{r}_{t+1} := \ln \mathbf{R}_{t+1}$ (also reserving $r_{0,t+1} := \ln R_{0,t+1}$) this means

$$g(\boldsymbol{r}_{t+1}) ext{ is } \mathcal{N}(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r)$$

$$\tag{7}$$

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

Next, given the value function's form (6), a reasonable choice for the distortion function in the sense of squared L^2 norm (D3) would be $d(\mathbf{r}_{t+1}, \hat{\mathbf{r}}_{t+1}) := ||v^{\sharp}(\{q_{0,t}, \mathbf{q}_t\}, \mathbf{D}_{t+1}) - v(\{q_{0,t}, \mathbf{q}_t\}, \hat{\mathbf{D}}_{t+1})||_2^2$. However, given the CRRA functional form, we modify it further and take logarithms (D4):

$$d(\boldsymbol{r}_{t+1}, \hat{\boldsymbol{r}}_{t+1}) := \frac{1}{(1-\gamma)^2} ||\ln|v^{\sharp}(\{q_{0,t}, \boldsymbol{q}_t\}, \boldsymbol{D}_{t+1})| - \ln|v(\{q_{0,t}, \boldsymbol{q}_t\}, \hat{\boldsymbol{D}}_{t+1})|||_2^2 = = (\ln W_{t+1} - \ln \hat{W}_{t+1})^2.$$
(8)

Lastly, an additional assumption, required only for K > 1 cases, restricts the dependence of $d(\mathbf{r}_{t+1}, \hat{\mathbf{r}}_{t+1})$ on $\boldsymbol{\omega}_t$ by replacing the dependence on exact $\boldsymbol{\omega}_t$ with agnostic motive of minimization of maximum loss for any possible $\boldsymbol{\omega}_t$ (per D5 above). (See Appendix §D.2 for more details.)

Now, we adapt the distortion function's formulation (8) further, making it even more convenient for our use.

Proposition 1 (Distortion Function). The above distortion function, in the context of problem \mathcal{P}_{QI} and given the distributional assumptions, can be reformulated as follows:

$$d(\mathbf{r}_{t+1}, \hat{\mathbf{r}}_{t+1}) = (\ln W_{t+1} - \ln \hat{W}_{t+1})^2 \approx \\ \approx (\boldsymbol{\omega}_t^{\mathsf{T}} (\mathbf{r}_{t+1} - \hat{\mathbf{r}}_{t+1} + \check{\boldsymbol{\mu}}_r))^2, \qquad (P1-1)$$

where

$$\hat{\boldsymbol{\mu}}_r := \boldsymbol{\mu}_r + \check{\boldsymbol{\mu}}_r, \tag{P1-2}$$

is the mean of the simplified random variable \hat{r}_{t+1} , and where

$$\check{\boldsymbol{\mu}}_r := \frac{1}{2} \operatorname{diag}^{-1} (\boldsymbol{\Sigma}_r - \hat{\boldsymbol{\Sigma}}_r) - \frac{1}{2} (\boldsymbol{\Sigma}_r - \hat{\boldsymbol{\Sigma}}_r) \boldsymbol{\omega}_t, \qquad (P1-3)$$

is a bias term, with $\hat{\Sigma}_r$ denoting the variance-covariance matrix for \hat{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:

$$(\boldsymbol{\omega}_{t}^{\mathsf{T}}(\boldsymbol{r}_{t+1} - \hat{\boldsymbol{r}}_{t+1} + \boldsymbol{\check{\mu}}_{r}))^{2} \propto (\boldsymbol{r}_{t+1} - \hat{\boldsymbol{r}}_{t+1} + \boldsymbol{\check{\mu}}_{r}(\widehat{\omega}_{t}))^{\mathsf{T}}(\boldsymbol{r}_{t+1} - \hat{\boldsymbol{r}}_{t+1} + \boldsymbol{\check{\mu}}_{r}(\widehat{\omega}_{t})) =:$$

=: $d(\boldsymbol{r}_{t+1}, \hat{\boldsymbol{r}}_{t+1}),$ (P1-4)

 $^{^{20}}$ 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 a practically relevant problem.

where

$$\check{\boldsymbol{\mu}}_{r}(\widehat{\omega}_{t}) := \frac{1}{2} \operatorname{diag}^{-1} (\boldsymbol{\Sigma}_{r} - \widehat{\boldsymbol{\Sigma}}_{r})(1 - \widehat{\omega}_{t}) = \\
= \frac{1}{2} \operatorname{diag}(\sigma_{r,1}^{2} - \widehat{\sigma}_{r,1}^{2}, \cdots, \sigma_{r,K}^{2} - \widehat{\sigma}_{r,K}^{2}) \mathbf{1}(1 - \widehat{\omega}_{t})$$
(P1-5)

is another bias term.

Proof. See Appendix §D.2.

(From now on, the elements of a matrix are denoted with the same letter as the matrix, but in lowercase.)

The main aim of the Proposition above is the appropriate form of the distortion function that we are going to use for the informational sub-problem of problem \mathcal{P}_{QI} . Specifically, the refinement given by equation (P1-4) produces a convenient sum-of-squares formulation for the (bias-corrected) differences between the true and approximated logreturns, which will eventually allow us to claim Gaussianity of these differences.²¹

The Proposition also contains an intermediate result, which stems from the continuoustime approximation of the wealth dynamics based on the (geometric) Brownian motion. We would like to cancel out the effect of having replaced the original variance Σ_r with the simplified variance $\hat{\Sigma}_r$ so as to ensure the expected growth rate of the simplified log-wealth $\ln \hat{W}_{t+1}$ equals that of the original log-wealth $\ln W_{t+1}$. This can be achieved by adjusting the mean of the simplified random variable $\hat{\mu}_r$ by a bias term $\check{\mu}_r$, as shown in equation (P1-2). For positive risky investments and a positive discrepancy between the original and simplified variances²², the adjustment entails a downward shift (second term on the RHS of equation P1-3) from the origin point²³ (first term). Therefore, we are dealing with a simplified distribution of log-returns $h_r(\hat{r}_{t+1})$ that is biased, but biased in an optimal, expected distortion-minimizing way (see Appendix §D.2.1).

The following example can illustrate what is going on. In the case of one risky asset that is not held in the investment portfolio (i.e., $\omega_t = 0$), we have $\hat{\mu}_r = \mu_r + 0.5(\Sigma_r - \hat{\Sigma}_r)$, which leads to matching of the expected values of simplified and true returns, $\mathbf{E}_t^h[\hat{W}_{t+1}/W_t] = \mathbf{E}_t^h[\hat{R}_{t+1}] = \exp(\hat{\mu}_r + 0.5\hat{\Sigma}_r) = \exp(\mu_r + 0.5\Sigma_r) = \mathbf{E}_t^g[R_{t+1}] = \mathbf{E}_t^g[W_{t+1}/W_t]$). However, in the case of one risky asset that is the sole constituent of the investment portfolio ($\omega_t = 1$), we have a matching of the means, $\hat{\mu}_r = \mu_r$, which leads to undershooting of the expected return on the simplified portfolio, $\mathbf{E}_t^h[\hat{W}_{t+1}/W_t] =$

²¹Note that the refined distortion function depends on the scalar $\hat{\omega}_t$ only through vector $\check{\boldsymbol{\mu}}_r(\hat{\omega}_t)$; in other words, given the bias term $\check{\boldsymbol{\mu}}_r(\hat{\omega}_t)$, the solution to the informational problem is invariant to the chosen value of the bound $\hat{\omega}_t$. This fact will be put to work later.

 $^{^{22}}$ The latter is shown to be the case later in §3.3.

 $^{^{23}}$ This origin point effectively presumes that the simplified wealth indeed follows the dynamics determined by the simplified rather than the true variance.

 $E_t^h[\hat{R}_{t+1}] = \exp(\hat{\mu}_r + 0.5\hat{\Sigma}_r) < \exp(\mu_r + 0.5\Sigma_r) = E_t^g[R_{t+1}] = E_t^g[W_{t+1}/W_t]$. The undershooting can be interpreted as a "pessimistic" view at the potential investment opportunities, and such "pessimism" is nothing more than an implication of the simplified variance being dominated by the original one.²⁴

Remark 1 (Decision-Rule Adjustment vs. Subjective-Perception Adjustment). A more direct way of correcting for the difference between the original and simplified variances is to adjust the decision rules (i.e., policy functions dictating the choice of control variables $C_t, \{q_{0,t}, q_t\}$) rather than the subjective perception (i.e., the mean of the simplified distribution $\hat{\mu}_r$). Adjusting the decision rules shifts control variables $C_t, \{q_{0,t}, q_t\}$ directly; which takes extra (K + 1) adjustment parameters. Adjusting the mean shifts control variables $C_t, \{q_{0,t}, q_t\}$ indirectly, by affecting the chosen subjective probability density $h(\hat{D}_{t+1}|\hat{D}_t)$, via $h(\hat{r}_{t+1})$, that control variables are functionally dependent on; which takes extra K adjustment parameters. Judging in terms of degrees of freedom, the latter option is more restrictive (lower number of parameters that can adjust), but less computationally greedy (lower number of parameters that have to be determined).²⁵

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 Σ_r , $\hat{\Sigma}_r$ and ω_t) the bias term $\check{\mu}_r$ 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 all 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 nontrivial 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.]

As a last technicality, part of our solution method relies on random variables being uncorrelated. Decorrelation is achieved using eigenvalue decomposition of the variancecovariance matrix Σ_r into a diagonal matrix of eigenvalues Σ and a square matrix of eigenvectors Ξ ; which allows to transform the raw objects \mathbf{r}_{t+1} , $\hat{\mathbf{r}}_{t+1}$ and $\check{\boldsymbol{\mu}}_r(\hat{\omega}_t)$ into, respectively, \mathbf{x} , $\hat{\mathbf{x}}$ and $\check{\boldsymbol{\mu}}(\hat{\omega}_t)$. For details, see Appendix §I; Proposition I.1 there states

²⁴The above pessimism-entailing 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 (8), immunizing the distortion function from the effect of risk attitude.

²⁵The author thanks Michael Woodford for raising this issue.

that this decorrelating transformation is innocuous.²⁶

3.2.4 Equilibrium

An equilibrium of the economy formed by the exogenously given economic setting introduced in §3.2.1 and the endogenously chosen optimal solutions to the feasible problem $\mathcal{P}_{Q\mathcal{I}}$ in §3.2.2 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 $F(\boldsymbol{D}_{t+1}, \hat{\boldsymbol{D}}_{t+1} | \boldsymbol{D}_t, \hat{\boldsymbol{D}}_t) :$ $\mathbb{R}_+^K \times \mathbb{R}_+^K \mapsto [0, 1]$ such that:

- (i) [consumption and investment optimality] Bellman equation (PQI-1) subject to the budget constraint (PQI-2), control variable's domain restriction and with given utility function specification is satisfied;
- (ii) [consumption and investment coherence] goods and asset markets clear, i.e.,

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

- (iii) [informational optimality] joint probability density function $f_r(\mathbf{r}_{t+1}, \hat{\mathbf{r}}_{t+1})$ solves (utilizing the decorrelating transformation) the informational problem $\mathcal{P}_{\mathcal{I}}$ with $g_r(\mathbf{r}_{t+1})$ as the true density and κ as the information-processing capacity;
- (iv) [informational coherence] probability density functions for dividends $f(\boldsymbol{D}_{t+1}, \hat{\boldsymbol{D}}_{t+1} | \boldsymbol{D}_t, \hat{\boldsymbol{D}}_t)$ (referred to in equation PQI-4), $g(\boldsymbol{D}_{t+1} | \boldsymbol{D}_t)$ (referred to in PQI-5) and $h(\hat{\boldsymbol{D}}_{t+1} | \hat{\boldsymbol{D}}_t)$ (referred to in PQI-3) are consistent with the densities for returns $f(\boldsymbol{r}_{t+1}, \hat{\boldsymbol{r}}_{t+1})$, $g(\boldsymbol{r}_{t+1})$ and $h(\hat{\boldsymbol{r}}_{t+1})$, also the correlated random variables' densities $f(\boldsymbol{r}_{t+1}, \hat{\boldsymbol{r}}_{t+1})$, $g(\boldsymbol{r}_{t+1})$ and $h(\hat{\boldsymbol{r}}_{t+1})$ are consistent with the decorrelated variables' densities $f(\boldsymbol{x}, \hat{\boldsymbol{x}})$, $g(\boldsymbol{x})$ and $h(\hat{\boldsymbol{x}})$, i.e., $\forall \boldsymbol{D}_{t+1}, \hat{\boldsymbol{D}}_{t+1} \in \mathbb{R}_+^K$:

$$\begin{split} f(\boldsymbol{D}_{t+1}, \hat{\boldsymbol{D}}_{t+1} | \boldsymbol{D}_t, \hat{\boldsymbol{D}}_t) &= f(\boldsymbol{r}_{t+1}, \hat{\boldsymbol{r}}_{t+1}) = f(\boldsymbol{\Xi} \boldsymbol{x}, \boldsymbol{\Xi} \hat{\boldsymbol{x}}) = f(\boldsymbol{x}, \hat{\boldsymbol{x}}), \\ g(\boldsymbol{D}_{t+1} | \boldsymbol{D}_t) &= g(\boldsymbol{r}_{t+1}) = g(\boldsymbol{\Xi} \boldsymbol{x}) = g(\boldsymbol{x}), \\ h(\hat{\boldsymbol{D}}_{t+1} | \hat{\boldsymbol{D}}_t) &= h(\hat{\boldsymbol{r}}_{t+1}) = h(\boldsymbol{\Xi} \hat{\boldsymbol{x}}) = h(\hat{\boldsymbol{x}}). \end{split}$$

A policy function determining the optimal investment in tree shares $\boldsymbol{q}(\{q_{0,t-1}, \boldsymbol{q}_{t-1}\}, \boldsymbol{D}_t)$ could be added to the list of equilibrium objects, but it is a constant function that is identically equal to $\hat{\boldsymbol{q}}$ because the considered economy is an autarky.

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, which is done next.

²⁶In terms of Appendix §E's algorithm, our understanding is that this transformation is performed when solving the informational problem before the Generating codebook step of the algorithm (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).

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 distinctly segregated formulations of these two sub-problems. (The full solution to the feasible consumption and portfolio choice problem \mathcal{P}_{QI} is available in Appendix §B.)

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

$$f(\boldsymbol{x}|\hat{\boldsymbol{x}}) = \exp\left(\frac{1}{\lambda}\nu(\boldsymbol{x},\hat{\boldsymbol{x}}) - \frac{1}{\lambda}\mu(\boldsymbol{x}) - \frac{1}{\lambda}(\boldsymbol{x}-\hat{\boldsymbol{x}}+\check{\boldsymbol{\mu}}(\widehat{\omega}_t))^{\mathsf{T}}(\boldsymbol{x}-\hat{\boldsymbol{x}}+\check{\boldsymbol{\mu}}(\widehat{\omega}_t))\right), \quad \forall \hat{\boldsymbol{x}} \in \operatorname{supp}(h).$$
(9)

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 Theorem. But an attentive reader has already spotted the kernel of a Gaussian probability density function in the last equation, which suggests the subsequent direction.

Theorem 1 (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 (9), where the random vector $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

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

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

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

where

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}), \qquad \boldsymbol{\Psi} = \frac{\lambda}{2} \boldsymbol{I}_{K},$$
$$\hat{\boldsymbol{x}} \sim \mathcal{N}(\hat{\boldsymbol{\mu}}(\hat{\omega}_{t}), \hat{\boldsymbol{\Sigma}}), \qquad \hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} - \boldsymbol{\Psi};$$
$$\lambda = 2 \left(e^{-2\kappa} |\boldsymbol{\Sigma}| \right)^{\frac{1}{K}}.$$

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).

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

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

where

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

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

$$\begin{split} \{\sigma_k^2\}_1^K &:= \operatorname{sortdescending}(\{\sigma_k^2\}_1^K), \\ k^* &:= \arg\min_{k \in \{1, \dots, K\}} \{\sigma_k^2 \mid \sigma_k^2 > \frac{\lambda}{2}\}, \\ \lambda &= 2\left(e^{-2\kappa}\sigma_{k^*+1}^{-2}\cdots \sigma_K^{-2}|\boldsymbol{\Sigma}|\right)^{\frac{1}{k^*}}. \end{split}$$

(The last $(K - k^*)$ elements of vector \hat{x} are to be understood as deterministic scalars, or alternatively as Dirac delta functions centered at $\{\hat{\mu}_{k^*+1}(\hat{\omega}_t), \cdots, \hat{\mu}_K(\hat{\omega}_t)\}$.) 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 §D.3.

The qualifiers 'interior' and 'boundary' in the formulation of the above Theorem should be understood in relation to the Cartesian product set $X_1^K[0,\sigma_k^2]$. Figure 4 illustrates the "reverse water-filling" logic of the (boundary) solution of Theorem 1. From the information-processing perspective, σ_k^2 (diagonal elements of the original variancecovariance matrix Σ) represent the total information available for processing, $\hat{\sigma}_k^2$ (diagonal elements of the simplified variance-covariance matrix $\hat{\Sigma}$) represent the information that is actually processed, while ψ_k^2 (elements on the diagonal of the variance-covariance matrix for approximation errors Ψ) represent the information that is omitted and constitutes the approximation errors; and the solution logic implies "filling" σ^2 -s with $\hat{\sigma}^2$ -s, proceeding in "reverse", from the top to the bottom.

Consider an extreme situation when information-processing capacity is unavailable, $\kappa \to -\infty$ (the information constraint is binding with the Lagrange multiplier $\lambda \to +\infty$): then no information can be processed, $\hat{\sigma}_k^2 = 0$, $\forall k \in \{1, \ldots, K\}$, while approximation errors are at their maxima, $\psi_k^2 = \sigma_k^2$, $\forall k \in \{1, \ldots, K\}$. If κ rises somewhat (the information constraint is binding, but now the Lagrange multiplier takes a finite value $\lambda < \infty$), we appear in the boundary solution case depicted in Figure 4: then some information can be processed, $\hat{\sigma}_k^2 > 0$ for certain k, with the corresponding approximation errors subsiding, $\psi_k^2 < \sigma_k^2$ for these same k. If κ rises a lot more, but with $\kappa < \kappa^{\sharp}$ (the information



Figure 4: "Reverse water-filling".

constraint is still biniding and λ falls even further), a qualitative change in the picture occurs and we move to the the interior solution case: $\hat{\sigma}_k^2 = \sigma_k^2 - \psi_k^2 > 0$ for all k, with the approximation errors subsiding even more, $\psi_k^2 < \sigma_k^2$ for all k, and now also equalizing, $\psi_k^2 = \psi_l^2$ for all k, l. At the other extreme, $\kappa = \kappa^{\sharp}$ (the information constraint does not bind any more and $\lambda = 0$): then all the information is processed, $\hat{\sigma}_k^2 = \sigma_k^2$ for all k, while $\psi_k^2 = 0$ for all k.²⁷

A non-trivial result here is the following: when information processing capacity decreases, the information omissions disproportionally affect the data dimensions with low informational content (small eigenvalues σ_k^2). This 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 Theorem 1, is characterized by the folded dispersions of the less volatile components of random vector \hat{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 Theorem 1 (take the lower-right elements of $\hat{\Sigma}$ in the boundary solution case).

Considering the more revealing here case 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

²⁷Lastly, note that in our case we benefit from the Normality of distribution $g(\cdot)$, among other assumptions: 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 related analysis of Matějka and Sims (2010).

the effects of entropy reduction culminating in its complete "folding",²⁸ 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, according to Corollary 1, in the case of two uncorrelated random variables [alternatively, two correlated random variables' dimensions] x_1 and x_2 , 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 grapevines, with wild raspberries growing on the sidelines. These two plants' "payoffs" depend on different subsets of natural as well as market conditions, and are thus uncorrelated. Moreover, market prices on wild raspberries are not subject to quality-related fluctuations or large demand swings, hence they are not too volatile. As a result, a binding capacity constraint may lead to total disregard of the presence of these relatively less prominent fruits.

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 (e.g., 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 3 develops this theme a little further.)

The results of Theorem 1 in economically interesting terms such as returns (i.e., after the inversion of the decorrelating transformation), are presented in the subsequent Theorem 2.

Theorem 2 (Specific Solution to Informational Problem: Representation in Economic Terms). The specific solution to the informational problem given in the statement of Theorem 1 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

 $\Sigma_r = \hat{\Sigma}_r + \Psi_r, \quad \forall \Sigma_r \text{ that is } K \times K \text{ positive semi-definite,}$

where

$$\hat{m{r}}_{t+1} \sim \mathcal{N}(\hat{m{\mu}}_r(\widehat{\omega}_t), m{\Sigma}_r),
onumber \ m{\epsilon}_{r,t+1} \sim \mathcal{N}(m{0}, m{\Psi}_r),$$

with $\hat{\boldsymbol{\mu}}_r(\hat{\omega}_t)$ and $\check{\boldsymbol{\mu}}_r(\hat{\omega}_t)$ given by Proposition 1, as well as with $\hat{\boldsymbol{\Sigma}}_r$ and $\boldsymbol{\Psi}_r$ defined as

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

basing on the results of Theorem 1.

²⁸Terms "dispersion folding", "randomness collapse" or "distribution contraction" all seem fitting for this phenomenon.

Moreover, we can 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 Theorem 2 (as well as in Theorem 1) by appealing to the arguments from Appendix §B.2.^{29,30}

The main economic results of Theorem 2 boil down to the following.

Corollary 2 (Specific Solution to Informational Problem: Overconfidence). The specific solution to the informational problem represented in economic terms, as given in the statement of Theorem 2, is characterized by "overconfidence". I.e., the subjective variance-covariance matrix is dominated by its objective counterpart:

$$(\Sigma_r - \hat{\Sigma}_r)$$
 is $K \times K$ positive semi-definite.

Proof. Immediate from Theorem 2 (take the variance decomposition equation).

Thus, $\hat{\Sigma}_r$, the variance-covariance matrix of the simplified log-returns \hat{r}_{t+1} , is smaller than Σ_r , its counterpart for the original log-returns r_{t+1} . This is a direct consequence of the entropy-reducing simplification prompted by the information-processing capacity constraint: recall that the entropy of a Gaussian random vector is a one-to-one map with the determinant of its variance-covariance matrix. (Moreover, as compensation for the simplification above, $\hat{\mu}_r$, the mean of \hat{r}_{t+1} , is biased toward pessimism in comparison to μ_r , its counterpart for r_{t+1} ; this follows right from Proposition 1.)

The claim that Σ_r is smaller than Σ_r has somewhat surprising economic ramifications.

Corollary 3 (Specific Solution to Informational Problem: Correlation Inflation). The specific solution to the informational problem represented in economic terms, as given in the statement of Theorem 2, is characterized by:

(a) The inflated correlations between the elements of $\hat{\mathbf{r}}_{t+1}$ relative to those for the elements of \mathbf{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]:

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

²⁹To sum up, Proposition 1 demonstrates the appropriateness of our approximation procedure and allows to derive the normality of \hat{r}_{t+1} with the subsequent decomposition equations of Theorems 1–2, thus providing us with analytical convenience; while Appendix §B.2 ensures that approximation accuracy result from Proposition 1 has not been lost in the process.

³⁰There 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: $\Psi_r = \Psi$. Another revealing result in the interior solution case is that the optimal mean bias term defined in Theorem 1.1, $\check{\mu}_r$, takes the form conformable with the bias term from Theorem 1.3, $\check{\mu}_r(\widehat{\omega}_t)$: $\check{\mu}_r = \frac{1}{2} \text{diag}^{-1}(\Sigma_r - \hat{\Sigma}_r) \odot (1 - \omega_t)$, where \odot denotes the Hadamard product.

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

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

(b) Either inflated or shrinking correlations between the elements of r

 *î*_{t+1} relative to those for the elements of r

 *i*_{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]:

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

Proof. See Appendix §D.6.

The "correlation inflation" outcome is essentially robust, in spite of its reversal in some special instances of the boundary solution case. The reason is that globally, both the shrinking diagonal variance terms and the inflating off-diagonal covariance terms contribute toward variance-covariance matrix $\hat{\Sigma}_r$ being smaller than Σ_r . Since the effect of shrinking covariances works in the opposite direction, the reversal may happen only locally.³¹

Intuitively, Corollary 3 shows how the inflation of the correlations between the elements of \hat{r}_{t+1} as compared to the correlations between the elements of r_{t+1} emerges subjectively. Effectively, this leads to further attraction of the positively correlated elements and the repulsion of the negatively correlated ones, which ultimately results in the pooling of the random vector's components into relatively detached "categories".³²

Returning to our vineyard 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, with the two categories of plants being pushed apart as very distinct kinds of capital goods (say, "red" vs. "white") that are characterized by different combinations of attributes such as acidity or tannin levels.

Lastly, we note the following result.

 $^{^{31}}$ By the way, notice in the Corollary 1 and, especially, Corollary 2 the interplay between reduced variances and biased correlations, which stems from the trade-off stipulated by the information-processing capacity constraint (3).

³²In 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).

Corollary 4 (Satisficing). A binding information-processing capacity bound κ , such that $\kappa < \kappa^{\sharp}$, implies a positive expected value-function shortfall:

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

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". The constrained, i.e. a second-best, value function $v(\cdot)$ serves the role of the (endogenous) "aspiration level".

Proof. Take $\lambda > 0$ in Theorems 1 and 2, recognizing their premises from Lemma 1 and Proposition 1.

This is a straight-forward instance of the concept of "satisficing"—as opposed to "maximizing"—introduced by Simon (1997). Here, it literally corresponds to optimality in a constrained sense that generally does not achieve the first-best level.³³

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 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}_{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 (left panel) and the subjective perspective on this stochastic landscape (right panel) for a case with two risky assets.

4.1.1 Overconfidence

One 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

³³Also 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).



Figure 5: Objective and subjective probability densities, two risky assets (parameterizations used are $\mathbf{R}_{t+1} \sim \log \mathcal{N}(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r), \, \mathbf{r}_{t+1} \sim \mathcal{N}(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r),$ $\boldsymbol{\mu}_r = [0.10; 0.20], \, \boldsymbol{\Sigma}_r = [0.10, 0.08; 0.08, 0.16], \, \mathbf{E}_t^g[\mathbf{R}_{t+1}] = [1.16; 1.32] \text{ (left panel)};$ and, for $\kappa = 1$ nat ≈ 1.44 bits with $\boldsymbol{\omega}_t = [0.5; 0.5],$ $\hat{\mathbf{R}}_{t+1} \sim \log \mathcal{N}(\hat{\boldsymbol{\mu}}_r, \hat{\boldsymbol{\Sigma}}_r), \, \hat{\mathbf{r}}_{t+1} \sim \mathcal{N}(\hat{\boldsymbol{\mu}}_r, \hat{\boldsymbol{\Sigma}}_r),$ $\hat{\boldsymbol{\mu}}_r = [0.11; 0.21], \, \hat{\boldsymbol{\Sigma}}_r = [0.06, 0.08; 0.08, 0.12], \, \mathbf{E}_t^h[\hat{\mathbf{R}}_{t+1}] = [1.15; 1.31] \text{ (right panel)}).$ is a coarser version of the latter. In our case with log-normal payoffs, the entropy reduction is achieved solely by reducing the variance of the relevant random variables.

Specifically, Σ_r , the variance-covariance matrix of \hat{r}_{t+1} , is smaller than Σ_r , its counterpart for r_{t+1} , as stated in Corollary 2. This is reflected in Figure 5 by the relatively more peaked probability densities in the right panel.

4.1.2 Pessimism

As a by product of the previous result, we obtain what can be viewed as necessary "pessimism". In order to compensate for the entropy-reducing simplification discussed above, the means of approximating random variables have to be adjusted. Effectively, the required adjustment amounts to adopting a subjectively pessimistic view at the future state of the world.

The exact bias term $\check{\boldsymbol{\mu}}_r$, as given by Proposition 1, is stronger for a larger share of wealth invested in risky assets $\boldsymbol{\omega}_t$ and a larger discrepency between objective and subjective variances $(\boldsymbol{\Sigma}_r - \hat{\boldsymbol{\Sigma}}_r)$. For a positive size of risky investments, the bias shifts the subjectively expected value $\mathbf{E}_t^h[\hat{\boldsymbol{R}}_{t+1}]$ downward. (Otherwise, a risk-averse agent would be relatively over-invested in risky assets.³⁴) Figure 5 demonstrates the optimal result with the expected value of the top right panel's probability distribution moved slightly to the west, so that $\mathbf{E}_t^h[\hat{\boldsymbol{R}}_{t+1}] < \mathbf{E}_t^g[\boldsymbol{R}_{t+1}].^{35}$

4.1.3 Categorization

Our key result, concerning only the multivariate settings, is emerging "categorization". Conceptually, it is worthwhile to distinguish two kinds of categorization:

- (i) 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);³⁶
- (ii) pooling of the random variables themselves, i.e. pooling of types (e.g., conditionally on the value of one random variable the other converges toward a non-stochastic Dirac delta function).

Technically, entropy reduction for uncorrelated random variables leads, at its extreme, to "dispersion folding": that is, folding and dropping out low-volatility categories (think of

 $^{^{34}}$ 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.

³⁵Although for the parameterization used here, the bottom right panel's distribution is actually shifted slightly to the east.

³⁶In 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.

it as "categorization in" vs. "categorization out"). This formally corresponds to bundling of states, and is dealt with by Corrollary 1.

However, for correlated random variables the above pertains to random variables' dimensions, and leads to "correlation inflation": that is, clustering into similar categories (think of it as "categorization together" vs. "categorization apart"). This formally corresponds to pooling of types, and is taken care of by Corrollary 3

While originally belonging to the domain of cognitive psychology and neuroscience, these ideas were introduced to economics a long time ago. For instance, Herbert Simon (1947, quoted from 1997 edition) noted:

"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."

Indeed, 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 to variance reduction, or "overconfidence", more generally).

In this paper we examine only one of the two categorization mechanisms: while ignoring categorization as folding and dropping out the less important random variables, here we focus on categorization as the inflation of correlations between random variables.³⁷

In our case, the positively correlated elements of \mathbf{r}_{t+1} become even more correlated in $\hat{\mathbf{r}}_{t+1}$, 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 give rise to subjective clustering of different assets into relatively distinct categories, "asset classes". The mechanics behind it is especially transparent in the leading case of interior solution: the covariance terms in $\hat{\boldsymbol{\Sigma}}_r$, the variance-covariance matrix of simplified returns $\hat{\boldsymbol{r}}_{t+1}$, are unchanged, but the variance terms are reduced, consequently leading to the correlation coefficients exceeding in absolute magnitude those in $\boldsymbol{\Sigma}_r$, the variance-covariance matrix of original returns \boldsymbol{r}_{t+1} . 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 $(\hat{\rho}_{r,12} = 0.90 > 0.63 = \rho_{r,12}).^{38}$

 $^{^{37}}$ We discuss the economic role of the other mechanism in more detail elsewhere, as a part of a separate line of investigation.

³⁸A useful way to look at this categorization result is through the lens of principal component analysis. Theorems 1 and 2 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 also increases (in absolute terms) the correlation between the two variables.

In a real-world practice, 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.

4.1.4 Decision outcomes

Nevertheless, when agents do follow the above approximation procedures, the decision outcomes (such as prices $P_{0,t}^*$ and P_t^*) achieve constrained optimality by construction, as reflected in Corollary 4. That is, they 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 $E_t^f[d(\mathbf{r}_{t+1}, \hat{\mathbf{r}}_{t+1})] = \Psi_r \ge \mathbf{0}$. The approximation is very accurate in practice though (see the comments on Proposition 1.1 in §D.2). Also note that decision errors are symmetric around the fully optimal levels (see Theorem 2), hence they at least partically cancel out on aggregate. Moreover, the errors are relatively smaller for contingencies that impact welfare the most (see §H).

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 (with a caveat that agent multiplicity is not modeled explicitly here).³⁹

4.2 Practice

Our main result regarding the subjective ampification of the correlations between different risky assets and their consequent categorization has a number of practical implications.

In the case of positively correlated assets, endogenously emerges their subjective clustering into asset classes. 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 entail the (self-reinforcing) popularity of operating in terms of aggregated asset classes instead of disaggregated assets among the investors (as well as among econometricians). This fuels the interest in "asset allocation", "asset comovement" and "style investing".⁴⁰

³⁹We 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 for trade.

 $^{^{40}}$ 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 straight-forward theoretical prediction of our analysis is that agents with lower information-processing capacity will be relatively more predisposed to such simplifying categorizations and clustering. The style investing phenomenon is well studied in the literature and provides enough evidence to verify the above prediction. First of all, signs of style investing are indeed present in the data: investor trading demands exhibit notable correlations; correlations within the same category of assets are stronger than between different categories; and measured correlations exceed the levels warranted by the assets' undelying fundamental characteristics (e.g., see Pindyck and Rotemberg, 1990; Chan et al., 2000; Teo and Woo, 2004; Froot and Teo, 2008; Choi and Sias, 2009). Secondly, the variation in the strength of correlations is consistent with the presumed differences in capacity: the clustering is stronger among less sophisticated retail investors than it is among institutional investors, who tend to be professional market participants that are supposed to be relatively less capacity-constrained; Kumar and Lee (2006) already found some evidence of this, while an explicit comparison of Jame and Tong (2014) reports the values for one popular measure of correlations in investment decisions ("herding") in two market constituencies we are concerned about at 4.01% and 2.09%, respectively. This is consistent with our theoretical prediction.

In the case of negatively correlated assets, from a subjective point of view they may seem like effective hedging instruments. For instance, portfolios of government bonds and portfolios of stocks may have a slightly negative true correlation of returns in some economic 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). Different hedging motives are the main focus of the "strategic asset allocation" literature.⁴¹

Lastly, correlation inflation may have real consequences for investors that maximize portfolio return subject to a constraint on the accepted level of portoflio variance (i.e., "mean-variance investors"), in case they rely on the simplified variance $\hat{\Sigma}_r$ but still use the matched expected return $\mathbf{E}^h[\hat{\mathbf{R}}_{t+1}] = \mathbf{E}^g[\mathbf{R}_{t+1}]$. In general, such investors 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 $\hat{\rho}_{r,12}$ reduces the portfolio variance by $0.5\hat{\sigma}_r(1-\hat{\rho}_{r,12})$. Here, the magnitude of variance reduction decreases as $\hat{\sigma}_r$ falls and $\hat{\rho}_{r,12}$ rises further, a typical situation since the simplified variance $\hat{\Sigma}_r$ is necessarily lower than its original counterpart Σ_r .

 $^{^{41}}$ E.g., see Brennan et al. (1997), Campbell and Viceira (2002b), also refer to Wachter (2010) for a more recent review paper; additionally, a modern practitioners' view can be found in Asl and Etula (2012).

5 Conclusion

The goal of the current paper is to improve our understanding of how in a stochastic 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, descriptive theoretical framework of decision-making under risk. It presumes rational optimizing behavior of the agents, 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. The ensuing structural mechanism allows to supplement a classical general equilibrium Lucas tree model with costly information processing.

As a result, the constrained-optimal behavior of our agent exhibits categorization. The latter occurs due to simplification and pruning of the perceived environment (what we called dispersion folding), which in turn lead to subjective clustering of covarying random variables together (correlation inflation). Empirically, categorization induced by information-processing constraints seems to be responsible for the financial markets regularity known as style investing.

The present paper develops an approach to evaluating expectations of stochastic objects that explicitly accounts for constraints imposed by the bounded information-processing capacity. Our information-theoretic approach carries through without any loss of generality, as traditional rational expectations are nested within and emerge as a special case when the information constraint is not binding. This generality allows to examine the information-processing capacity demands of the rational expectation formation. From an operational standpoint, as a process of computing an integral with respect to some probability measure, the demands it poses are not prohibitive, because various adjustments can reduce the computational costs dramatically without substantial efficiency losses. However, from a more conceptual standpoint, as an equilibrium notion, its demands may be too restrictive in practice (for instance, forcing the true objective and the approximate subjective distributions to coincide is a strong restriction that is rarely innocuous).⁴²

⁴²Lastly, note 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 practical parallels going beyond toy examples. Appendix §K fleshes out this point in more detail.

References

- 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] Asl, Farshid M. and Erkko Etula. (2012) "Advancing Strategic Asset Allocation in a Multi-Factor World", The Journal of Portfolio Management, 39(1): 59–66.
- [6] Backus, David, Mikhail Chernov and Stankey Zin. (2014) "Sources of Entropy in Representative Agent Models", *Journal of Finance*, 69(1): 51–99.
- [7] Baddeley, Alan. (2003) "Working Memory: Looking Back and Looking Forward." Nature Reviews Neuroscience, 4: 829–839.
- [8] Baddeley, R., P. Hancock and P. Földiák (eds.). (2000) Information Theory and the Brain, New York, NY: Cambridge University Press.
- Barber, Brad M., Terrance Odean, Ning Zhu. (2009) "Systematic Noise", Journal of Financial Markets, 12: 547–569.
- Barberis, Nicholas and Andrei Shleifer. (2003) "Style investing", Journal of Financial Economics, 68: 161–199.
- [11] Barberis, Nicholas, Andrei Shleifer and Jeffrey Wurgler. (2005) "Comovement", Journal of Financial Economics, 75: 283–317.
- [12] Barlow, Horace B. (1961) "Possible Principles Underlying the Transformation of Sensory Messages." In: W. Rosenblith (ed.), *Sensory Communication*, Cambridge, MA: MIT Press.
- [13] Bayer, Hannah. (2008) "Focus on Decision Making", Nature Neuroscience, 11(4): 387.
- [14] Berger, Toby. (1971) Rate-distortion Theory: A Mathematical Basis for Data Compression, Englewood Cliffs, NJ: Prentice-Hall.
- [15] Bickel, Peter J. and Bo Li (2006) "Regularization in Statistics." Test, 15(2): 271-344.
- [16] 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.
- [17] 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.
- [18] Bor, Daniel and Adrian M. Owen. (2006) "Working Memory: Linking Capacity with Selectivity", *Current Biology*, 16(4): 136–138.
- [19] Bordalo, Pedro, Nicola Gennaioli and Andrei Shleifer. (2016) "Diagnostic Expectations and Credit Cycles." Manuscript.
- [20] Boyer, Brian H. (2011) "Style-Related Comovement: Fundamentals or Labels?", The Journal of Finance, 66(1): 307–332.
- [21] Brennan, Michael J., Eduardo S. Schwartz, and Ronald Lagnado. (1997) "Strategic Asset Allocation", Journal of Economic Dynamics and Control, 21: 1377–1403.

- [22] Brown, Stephen J. and William N. Goetzmann. (1997) "Mutual Fund Styles", Journal of Financial Economics, 43: 373–399.
- [23] 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.
- [24] Brunnermeier, Markus K. and Jonathan A. Parker. (2005) "Optimal Expectations." American Economic Review, 95 (4): 1092–1118.
- [25] Brunnermeier, Markus K., Filippos Papakonstantinou and Jonathan A. Parker. (2013) "Optimal Time-Inconsistent Beliefs: Misplanning, Procrastination, and Commitment." Manuscript.
- [26] Bunge, Silvia A. (2005) "Foreword", Cognitive Brain Research, 23: 1.
- [27] 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.
- [28] 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.
- [29] 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.
- [30] Campbell, John Y. and Luis M. Viceira. (2002a) Appendix for "Strategic Asset Allocation: Portfolio Choice for Long-Term Investors", Oxford University Press.
- [31] Campbell, John Y. and Luis M. Viceira. (2002b) Strategic Asset Allocation: Portfolio Choice for Long-Term Investors, Oxford University Press.
- [32] 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.
- [33] Canner, Niko, N. Gregory Mankiw and David N. Weil. (1997) "An Asset Allocation Puzzle", American Economic Review, 87(1): 181–191.
- [34] Caplin, Andrew and Mark Dean. (2013) "The Behavioral Implications of Rational Inattention with Shannon Entropy." Manuscript.
- [35] Caplin, Andrew and Mark Dean. (forthcoming) "Revealed Preference, Rational Inattention, and Costly Information Acquisition." American Economic Review.
- [36] 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.

- [37] 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.
- [38] Chen, Hui, Winston Wei Dou and Leonid Kogan. (2015) "Measuring the "Dark Matter" in Asset Pricing Models." Manuscript.
- [39] Chen, Zhe and Simon Haykin. (2002). "On Different Facets of Regularization Theory." Neural Computation, 14: 2791–2846.
- [40] Choi, Nicole and Richard W. Sias. (2009) "Institutional Industry Herding", Journal of Financial Economics, 94: 469–491.
- [41] 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.

- [42] 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.
- [43] Cover, Thomas M. and Joy A. Thomas. (2006) *Elements of Information Theory*, 2nd ed., New York: Wiley-Interscience.
- [44] 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.
- [45] Damper, Robert I. and Stevan Harnad. (2000) "Neural Network Modeling of Categorical Perception", Perception & Psychophysics, 62(4): 843–867.
- [46] Dayan, Peter and Laurence F. Abbott. (2001) Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems, The MIT Press.
- [47] Doeswijk, Ronald Q., Trevin W. Lam and Laurens Swinkels. (2014) "The Global Multi-Asset Market Portfolio 1959–2012." Manuscript.
- [48] Doya, Kenji, Shin Ishii, Alexandre Pouget and Rajesh P.N. Rao (eds.). (2007) Bayesian Brain: Probabilistic Approaches to Neural Coding, The MIT Press.
- [49] 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.
- [50] 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.
- [51] 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.
- [52] Fleming, Stephen M., Laurence T. Maloney and Nathaniel D. Daw. (2013) "The Irrationality of Categorical Perception", *Journal of Neuroscience*, 33(49): 19060–19070.
- [53] French, Kenneth R. and James M. Poterba. (1991) "Investor Diversification and International Equity Markets", American Economic Review, 81(2): 222–226.
- [54] Friston, Karl. (2009) "The Free-energy Principle: A Rough Guide to the Brain?" Trends in Cognitive Sciences, 13(7): 293–301.
- [55] Friston, Karl. (2010) "The Free-energy Principle: A Unified Brain Theory?" Nature Reviews Neuroscience, 11: 127–138.
- [56] Froot, Kenneth and Melvyn Teo. (2008) "Style Investing and Institutional Investors", Journal of Financial and Quantitative Analysis, 43(4): 883–906.
- [57] Fuster, Andreas, Benjamin Hebert, and David Laibson. (2012) "Natural Expectations, Macroeconomic Dynamics, and Asset Pricing." NBER Macroeconomics Annual, 26 (1): 1–48.
- [58] Gabaix, Xavier. (2014a) "A Sparsity-Based Model of Bounded Rationality." Quarterly Journal of Economics, 129 (4): 1661–1710.
- [59] Gabaix, Xavier. (2014b) "Sparse Dynamic Programming and Aggregate Fluctuations." Manuscript.
- [60] Gabaix, Xavier and David Laibson. (2000) "A Boundedly Rational Decision Algorithm." American Economic Review (Papers and Proceedings), 90 (2): 433–438.
- [61] Gabaix, Xavier and David Laibson. (2005) "Bounded Rationality and Directed Cognition." Manuscript.

- [62] 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.
- [63] 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.
- [64] Gigerenzer, Gerd and Reinhard Selten (Eds.). (2001) Bounded Rationality: The Adaptive Toolbox, Cambridge, MA: MIT Press.
- [65] Gigerenzer, Gerd, Peter M. Todd, and the ABC Research Group. (2000) Simple Heuristics that Make Us Smart, New York, NY: Oxford University Press.
- [66] Gilboa, Itzhak and David Schmeidler. (1995) "Case-Based Decision Theory." Quarterly Journal of Economics, 110 (3): 605–639.
- [67] Gilovich, Thomas, Dale W. Griffin, and Daniel Kahneman (Eds.). (2002) Heuristics and Biases: The Psychology of Intuitive Judgment, New York, NY: Cambridge University Press.
- [68] Glimcher, Paul W., Colin F. Camerer, Ernst Fehr, Russell A. Poldrack (eds.). (2008) Neuroeconomics: Decision Making and the Brain, Academic Press.
- [69] Goetzmann, William N. and Alok Kumar. (2008) "Equity Portfolio Diversification", Review of Finance, 12: 433–463.
- [70] Goetzmann, William N., Lingfeng Li, and K. Geert Rouwenhorst. (2005) "Long-Term Global Market Correlations", *The Journal of Business*, 78(1): 1–38.
- [71] Goldstone, Robert L. and Andrew T. Hendrickson. (2010) "Categorical perception", WIREs Cognitive Science, 1(1): 69–78.
- [72] Guidolin, Massimo and Allan Timmermann. (2007) "Asset Allocation under Multivariate Regime Switching", Journal of Economic Dynamics and Control, 31: 3503–3544.
- [73] Hansen, Lars P. (2007) "Beliefs, Doubts and Learning: Valuing Macroeconomic Risk." American Economic Review, 97 (2): 1–30.
- [74] Hansen, Lars P. (2014) "Nobel Lecture: Uncertainty Outside and Inside Economic Models", Journal of Political Economy, 122(5): 945–987.
- [75] Hansen, Lars P. and Thomas J. Sargent. (2007a) Robustness, Princeton, NJ: Princeton University Press.
- [76] Hansen, Lars P. and Thomas J. Sargent. (2007b) "Recursive Robust Estimation and Control Without Commitment." *Journal of Economic Theory*, 136 (1): 1–27.
- [77] Jagannathan, Ravi and Tongshu Ma (2003). "Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps." *Journal of Finance*, 58(4): 1651–1684.
- [78] Jame, Russell and Qing Tong. (2014) "Industry-based Style Investing", Journal of Financial Markets, 19: 110–130.
- [79] James, W. and Charles M. Stein (1961), "Estimation with Quadratic Loss", Proc. Fourth Berkeley Symp. Math. Statist. Prob. 1: 361–379.
- [80] Jaynes, Edwin T. (2003) Probability Theory: The Logic of Science, Cambridge University Press.
- [81] Jorion, Philippe. (1985). "International Portfolio Diversification with Estimation Risk." Journal of Business, 58(3): 259–278.
- [82] Jorion, Philippe. (1986). "Bayes-Stein Estimation for Portfolio Analysis." Journal of Financial and Quantitative Analysis, 21(3): 279–292.

- [83] Jehiel, Philippe. (2005) "Analogy-Based Expectation Equilibrium." Journal of Economic Theory, 123 (2): 81–104.
- [84] Kelly, Morgan. (1995) "All Their Eggs in One Basket: Portfolio Diversification of US Households", Journal of Economic Behavior and Oraanization, 27: 87–96.
- [85] Kleinberg, Joel and Herbert Kaufman. (1971) "Constancy in Short-term Memory: Bits and Chunks", Journal of Experimental Psychology, 90(2): 326–333.
- [86] Kumar, Alok. (2009) "Dynamic Style Preferences of Individual Investors and Stock Returns", Journal of Financial and Quantitative Analysis, 44(3): 607–640.
- [87] Kumar, Alok and Charles M. C. Lee. (2006) "Retail Investor Sentiment and Return Comovements", *The Journal of Finance*, 61(5): 2451–2486.
- [88] 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.
- [89] Ledoit, Olivier and Michael Wolf (2004a) "A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices." Journal of Multivariate Analysis, 88(2): 365–411.
- [90] Ledoit, Olivier and Michael Wolf (2004b) "Honey, I Shrunk the Sample Covariance Matrix." Journal of Portfolio Management, 30(4): 110–119.
- [91] Li, Lingfeng. (2002) "Macroeconomic Factors and the Correlation of Stock and Bond Returns". Yale International Center for Finance Working Paper, No. 02-46.
- [92] Li, Lingfeng. (2003) "An Economic Measure of Diversification Benefits". Yale International Center for Finance Working Paper, No. 03-11.
- [93] Lucas, Robert E., Jr. (1978) "Asset Prices in an Exchange Economy." Econometrica, 46(6): 1429– 1445.
- [94] MacKay, David J. C. (2003) Information Theory, Inference, and Learning Algorithms, Cambridge University Press.
- [95] Manski, Charles F. (2017) "Survey Measurement of Probabilistic Macroeconomics Expectations: Progress and Promise." Manuscript.
- [96] Marschak, Jacob. (1959) "Remarks on the Economics of Information." Cowles Foundation Discussion Paper, No. 70.
- [97] Marschak, Jacob. (1968) "Economics of Inquiring, Communicating, Deciding." American Economic Review, 58(2): 1–18.
- [98] Marschak, Jacob. (1971) "Economics of Information Systems." Journal of the American Statistical Association, 66(333): 192–219.
- [99] Martin, Ian. (2013) "The Lucas Orchard." *Econometrica*, 81(1): 55–111.
- [100] Matějka, Filip. (2014a) "Rigid Pricing and Rationally Inattentive Consumer." Manuscript.
- [101] Matějka, Filip. (2014b) "Rationally Inattentive Seller: Sales and Discrete Pricing." Manuscript.
- [102] Matějka, Filip and Alisdair McKay. (2014) "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model." Manuscript.
- [103] Matějka, Filip and Christopher A. Sims. (2010) "Discrete Actions in Information-Constrained Tracking Problems." Manuscript.
- [104] Merton, Robert C. (1969) "Lifetime Portfolio Selection under Uncertainty: The Continuous-Time Case." The Review of Economics and Statistics, 51(3): 247–257.

- [105] Migliore, Michele, Gaspare Novara and Domenico Tegolo. (2008) "Single Neuron Binding Properties and the Magical Number 7", *Hippocampus*, 18: 1122–1130.
- [106] 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.
- [107] Mullainathan, Sendhil. (2002a) "A Memory-Based Model of Bounded Rationality." Quarterly Journal of Economics, 117 (3): 735–774.
- [108] Mullainathan, Sendhil. (2002b) "Thinking Through Categories." Manuscript.
- [109] 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.
- [110] Owen, Adrian M. (2004) "Working Memory: Imaging the Magic Number Four", Current Biology, 14: 573–574.
- [111] Padoa-Schioppa, Camillo, Aldo Rustichini. (2014) "Rational Attention and Adaptive Coding: A Puzzle and a Solution." American Economic Review, 104(5): 507–513.
- [112] Pammi, V.S. Chandrasekhar and Narayanan Srinivasan. (2013) "Preface", Progress in Brain Research, 202: xi.
- [113] Peng, Lin and Wei Xiong. (2006) "Investor Attention, Overconfidence and Category Learning." Journal of Financial Economics, 80 (3): 563–602.
- [114] Pindyck, Robert S. and Julio J. Rotemberg. (1990) "The Excess Co-Movement of Commodity Prices", *The Economic Journal*, 100(403): 1173–1189.
- [115] Polkovnichenko, Valery. (2005) "Household Portfolio Diversification: A Case for Rank-Dependent Preferences", *Review of Financial Studies*, 18(4): 1467–1502.
- [116] 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.
- [117] Ravid, Doron. (2016) "Bargaining with Rational Inattention." Manuscript.
- [118] Rustichini, Aldo and Camillo Padoa-Schioppa. (2015) "A Neuro-Computational Model of Economic Decisions." Journal of Neurophysiology, 114(5): 1382–1398.
- [119] Samuelson, Paul A. (1969) "Lifetime Portfolio Selection By Dynamic Stochastic Programming." The Review of Economics and Statistics, 51(3): 239–246.
- [120] Scheinkman, José A. and Wei Xiong. (2003) "Overconfidence and Speculative Bubbles." Journal of Political Economy, 111 (6): 1183–1219.
- [121] Schultz, Wolfram. (2008) "Introduction. Neuroeconomics: the Promise and the Profit", Philosophical Transactions of the Royal Society B, 363: 3767–3769.
- [122] Shannon, Claude E. (1948) "A Mathematical Theory of Communication." Bell System Technical Journal, 27: 379–423, 623–656.
- [123] Sharpe, William F. (1992) "Asset Allocation: Management Style and Performance Measurement", Journal of Portfolio Management, 18(2): 7–19.
- [124] Shiffrin, Richard M. and Robert M. Nosofsky. (1994) "Seven Plus or Minus Two: A Commentary On Capacity Limitations", *Psychological Review*, 101(2): 357–361.
- [125] Shiller, Robert J. (1989) "Comovements in Stock Prices and Comovements in Dividends", The Journal of Finance, 44(3): 719–729.

- [126] 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.
- [127] Simon, Herbert A. (1997) Administrative Behavior: A Study of Decision-Making Processes in Administrative Organizations, 4th edn, New York: Free Press.
- [128] Simon, Herbert A. (1957) Models of Man: Social and Rational, New York: John Wiley and Sons.
- [129] Sims, Christopher A. (1998) "Stickiness." Carnegie-Rochester Conference Series on Public Policy, 49(1): 317–356.
- [130] Sims, Christopher A. (2003) "Implications of Rational Inattention." Journal of Monetary Economics, 50: 665–690.
- [131] Sims, Christopher A. (2006) "Rational Inattention: A Research Agenda." Manuscript.
- [132] 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.
- [133] Squire, Larry R., Floyd Bloom, Nicholas C. Spitzer, Sascha du Lac, Anirvan Ghosh, Darwin Berg (eds.). (2008) Fundamental Neuroscience, 3rd ed., Academic Press.
- [134] Statman, Meir. (1987) "How Many Stocks Make a Diversified Portfolio?", The Journal of Financial and Quantitative Analysis, 22(3): 353–363.
- [135] 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.
- [136] Still, Susanne, David A. Sivak, Anthony J. Bell and Gavin E. Crooks. (2012) "Thermodynamics of Prediction", *Physical Review Letters*, 109(120604).
- [137] Stutzer, Michael. (1995) "A Bayesian Approach to Diagnosis of Asset Pricing Models", Journal of Econometrics, 68: 367–397.
- [138] Stutzer, Michael. (1996) "A Simple Nonparametric Approach to Derivative Security Valuation", Journal of Finance, 51(5): 1633–1652.
- [139] Teo, Melvyn, Sung-Jun Woo. (2004) "Style Effects in the Cross-Section of Stock Returns", Journal of Financial Economics, 74: 367–398.
- [140] 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.
- [141] Tibshirani, Robert. (1996) "Regression Shrinkage and Selection via the Lasso", Journal of the Royal Statistical Society. Series B, 58(1): 267–288.
- [142] Van Nieuwerburgh, Stijn and Laura Veldkamp. (2009) "Information Immobility and the Home Bias Puzzle", Journal of Finance, 64(3): 1187–1215.
- [143] Van Nieuwerburgh, Stijn and Laura Veldkamp. (2010) "Information Acquisition and Under-Diversification", *Review of Economic Studies*, 77: 779–805.
- [144] Veldkamp, Laura L. (2006) "Information Markets and the Comovement of Asset Prices", Review of Economic Studies, 73: 823–845.
- [145] Wachter, Jessica A. (2010) "Asset Allocation", Annual Review of Financial Economics, 2: 175–206.

- [146] Wahal, Sunil and M. Deniz Yavuz. (2013) "Style Investing, Comovement and Return Predictability", Journal of Financial Economics, 107: 136–154.
- [147] Wilson, Andrea. (2014) "Bounded Memory and Biases in Information Processing." *Econometrica*, 82 (6): 2257–2294.
- [148] Won, Joong-Ho, Johan Lim, Seung-Jean Kim and Bala Rajaratnam. (2012) "Condition Number Regularized Covariance Estimation." Manuscript.
- [149] Woodford, Michael. (2014) "An Optimizing Neuroeconomic Model of Discrete Choice." Manuscript.
- [150] Woodford, Michael. (2012) "Inattentive Valuation and Reference-Dependent Choice." Manuscript.
- [151] 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.