Uncertainty Prices when Beliefs are Tenuous*

Lars Peter Hansen† Thomas J.Sargent‡

October 10, 2016

Abstract

A decision maker expresses ambiguity about statistical models in the following ways. He has a family of benchmark parametric probability models but suspects that their parameters vary over time in unknown ways that he does not describe probabilistically. He expresses further a suspicion that all of these parametric models are misspecified by entertaining alternative nonparametric probability distributions that he represents only as positive martingales and that he restricts to be statistically close to the benchmark parametric models. Because he is averse to ambiguity, he uses a max-min criterion to evaluate alternative plans. We characterize equilibrium uncertainty prices by confronting a decision maker with a portfolio choice problem. We offer a quantitative illustration for benchmark parametric models that focus uncertainty on macroeconomic growth and its persistence. In the example there emerge nonlinearities in marginal valuations that induce time variation in the market price uncertainty. Prices of uncertainty fluctuate because the investor especially fears high persistence in bad states and low persistence in good ones.

Keywords— Risk, uncertainty, asset prices, relative entropy, Chernoff entropy, robustness, variational preferences; baseline, benchmark, parametric, and nonparametric models

---

*We thank Lloyd Han, Victor Zhorin and especially Yiran Fan for computational assistance. Hansen acknowledges support from the National Science Foundation, under grant 0951576 "DMUU: Center for Robust Decision Making on Climate and Energy Policy" and from the Policy Research Program of the MacArthur Foundation under the project "The Price of Policy Uncertainty."

†University of Chicago, E-mail: lhansen@uchicago.edu.

‡New York University, E-mail: thomas.sargent@nyu.edu.

In what circumstances is a minimax solution reasonable? I suggest that it is reasonable if and only if the least favorable initial distribution is reasonable according to your body of beliefs.   Good (1952)

1 Introduction

Applied dynamic economic models today typically rely on the rational expectations assumption that agents inside a model and nature share the same probability distribution. This paper takes a different approach by assuming that the agents inside the model confront various forms of of model uncertainty. They may not know values of parameters governing the evolution of pertinent state variables; they may suspect that these parameters vary over time; they may worry that their parametric model is incorrect. Thus, we put the agents inside our model into what they view as a complicated setting in which outcomes are sensitive to their subjective beliefs and in which learning is particularly difficult. We draw liberally from literatures on decision theory, robust control theory, and the econometrics of misspecified models to build a tractable model of how decision makers' specification concerns affect equilibrium prices and quantities.

To put our approach to work in a concrete setting, we use a consumption-based asset pricing model as a platform for studying how decision makers' specification worries influence "prices of uncertainty." These prices emerge from how the decision makers inside our dynamic economic model evaluate the utility consequences of alternative specifications of state dynamics. We show how these concerns induce variation in asset values and construct a quantitative example that assigns an important role to macroeconomic growth rate uncertainty. Because of its averse consequences to their discounted expected utilities, investors in our model fear growth rate persistence in times of weak growth. In contrast, they fear the absence of persistence when macroeconomic growth is high.

We describe procedures that simplify this apparently difficult specification challenge, for the investor and for us as outside analysts. We model decision making by blending ideas from two seemingly distinct approaches. We start by assuming that a decision maker considers a parametric family of benchmark models (with either fixed or time varying parameters) using a recursive structure suggested by Chen and Epstein (2002) for continuous time models with Brownian motion information structures. Because our decision maker distrusts all of his benchmark models, he adds nonparametric models residing within a statistical neighborhood of them. As we shall argue, the Chen and Epstein structure is too confining to include such statistical concerns about model misspecification. Instead we extend work by Hansen and Sargent (2001) and Hansen et al. (2006) that described a decision maker who expresses distrust of a probability model by surrounding it with an infinite dimensional family of difficult-to-discriminate nonparametric models. The decision maker represents alternative models by multiplying baseline probabilities with likelihood ratios whose entropies relative to the baseline model are penalized to be small. Formally, we accomplish this by applying a continuous-time counterpart of the dynamic variational preferences of Maccheroni et al. (2006b). In particular, we generalize what Hansen and Sargent (2001) and Maccheroni et al. (2006a,b) call multiplier preferences.[1]

We illustrate our approach by applying it to an environment that includes growth rate uncertainty in the macroeconomy. A representative investor who is a stand in for "the market" has specification doubts that affect prices of exposures to economic shocks. We calculate shadow prices that characterize aspects of model specifications that most concern the representative investor. These representative investor shadow prices are also uncertainty prices that clear competitive security markets. The negative of an endogenously determined vector of worst-case drift distortions equals a vector of prices that compensate the representative investor for bearing model uncertainty.[2] Time variation in uncertainty prices emerges endogenously since investor concerns about the persistence of macroeconomic growth rates differ depending on the state of the macroeconomy.

Viewed as a contribution to the consumption-based asset pricing literature, this paper extends earlier inquiries about whether responses to modest amounts of model ambiguity can substitute for implausibly large risk aversions required to explain observed market prices of risk. Viewed as a contribution to the economic literature on robust control theory and ambiguity, this paper introduces a tractable new way of formulating and quantifying a set of models against which a decision maker seeks robust decisions and evaluations.

Section 2 specifies an investor's baseline probability model and martingale perturbations to it, both cast in continuous time for analytical convenience. Section 3 describes discounted relative entropy, a statistical measure of discrepancy between martingales, and uses it to construct a convex set of probability measures that we impute to our decision maker. This martingale representation proves to be a tractable way for us to formulate robust decision problems in sections 4, 5 and 8.

[1] Applications of multiplier preferences to macroeconomic policy design and dynamic incentive problems include Karantounias (2013), Bhandari (2014) and Miao and Rivera (2016).

[2] This object also played a central role in the analysis of Hansen and Sargent (2010).

Section 6 uses Chernoff entropy, a statistical distance measure applicable to a set of martingales, to quantify difficulties in discriminating between competing specifications of probabilities. We show how to use this measure a) in the spirit of Good (1952), *ex post* to assess plausibility of worst-case models, and b) to calibrate the penalty parameter used to represent preferences. By extending estimates from Hansen et al. (2008), section 7 calculates key objects in a quantitative version of a baseline model together with worst-case probabilities with a convex set of alternative models that concern both a robust investor and a robust planner. Section 8 constructs a recursive representation of a competitive equilibrium of an economy with a representative investor who stands in for "the market". Then it links the worst-case model that emerges from a robust planning problem to equilibrium compensations that the representative investor receives in competitive markets. Section 9 offers concluding remarks.

2 Models and perturbations

This section describes nonnegative martingales that perturb a baseline probability model. Section 3.1 then describes how we use a family of parametric alternatives to a baseline model to form a convex set of martingales that in later sections we use to pose robust decision problems.

2.1 Mathematical framework

For concreteness, we use a specific *baseline* model and in section 3 a corresponding family of parametric alternatives that we call *benchmark* models. A representative investor cares about a stochastic process $X \doteq \{X_t : t \geqslant 0\}$ that he approximates with a baseline model[3]

$$dX_t = \hat{\mu}(X_t)dt + \sigma(X_t)dW_t, \tag{1}$$

where W is a multivariate Brownian motion.[4]

A decision maker cares about plans. A *plan* is a $\{C_t : t \geqslant 0\}$ that is progressively measurable process with respect to the filtration $\mathcal{F} = \{\mathcal{F}_t : t \geqslant 0\}$ associated with the

[3]We let X denote a stochastic process, X_t the process at time t, and x a realized value of the process.

[4]A Markov formulation is not essential. It could be generalized to allow other stochastic processes that can be constructed as functions of a Brownian motion information structure. Applications typically use Markov specifications.

Brownian motion W augmented by any information available at date zero. Under this restriction, the date t component C_t is measurable with respect to $\mathcal{F}_t$.

Because he does not fully trust baseline model (1), the decision maker explores the utility consequences of other probability models that he obtains by multiplying probabilities associated with (1) by likelihood ratios. Following Hansen et al. (2006), we represent a likelihood ratio by a positive martingale M^H with respect to the baseline model (1) that satisfies[5]

$$dM_t^H = M_t^H H_t \cdot dW_t \tag{2}$$

or

$$d\log M_t^H = H_t \cdot dW_t - \frac{1}{2}|H_t|^2 dt, \tag{3}$$

where H is progressively measurable with respect to the filtration $\mathcal{F} = \{\mathcal{F}_t : t \geqslant 0\}$ associated with the Brownian motion W. We adopt the convention that M_t^H is zero when $\int_0^t |H_u|^2 du$ is infinite, which happens with positive probability. In the event that

$$\int_0^t |H_u|^2 du < \infty \tag{4}$$

with probability one, the stochastic integral $\int_0^t H_u \cdot dW_u$ is an appropriate probability limit. Imposing the initial condition $M_0^H = 1$, we express the solution of stochastic differential equation (2) as the stochastic exponential

$$M_t^H = \exp\left(\int_0^t H_u \cdot dW_u - \frac{1}{2}\int_0^t |H_u|^2 du\right); \tag{5}$$

M_t^H is a local martingale, but not necessarily a martingale.[6]

Definition 2.1. *$\mathcal{M}$ denotes the set of all martingales M^H constructed as stochastic exponentials via representation* (5) *with an H that satisfies* (4) *and is progressively measurable with respect to $\mathcal{F} = \{\mathcal{F}_t : t \geqslant 0\}$.*

To construct alternative to the probability measure associated with baseline model (1), take any $\mathcal{F}_t$-measurable random variable Y_t and, before computing expectations conditioned

[5]James (1992), Chen and Epstein (2002), and Hansen et al. (2006) used this representation.

[6]Sufficient conditions for the stochastic exponential to be a martingale such as Kazamaki's or Novikov's are not convenient here. Instead we will verify that an extremum of a pertinent optimization problem does indeed result in a martingale.

on X_0, multiply it by M_t^H. Associated with H are probabilities defined by

$$E^H\left[B_t|\mathcal{F}_0\right] = E\left[M_t^H B_t|\mathcal{F}_0\right]$$

for any $t \geqslant 0$ and any bounded $\mathcal{F}_t$-measurable random variable B_t. Thus, the positive random variable M_t^H acts as a Radon-Nikodym derivative for the date t conditional expectation operator $E^H\left[\,\cdot\,|X_0\right]$. The martingale property of the process M^H ensures that conditional expectations operators satisfy a Law of Iterated Expectations.

Under baseline model (1), W is a standard Brownian motion, but under the alternative H model, it has increments

$$dW_t = H_t dt + dW_t^H, \tag{6}$$

where W^H is now a standard Brownian motion. Furthermore, under the M^H probability measure, $\int_0^t |H_u|^2 du$ is finite with probability one for each t. While (3) expresses the evolution of $\log M^H$ in terms of increment dW, the evolution in terms of dW^H is:

$$d\log M_t^H = H_t \cdot dW_t^H + \frac{1}{2}|H_t|^2 dt. \tag{7}$$

In light of (7), we can write model (1) as:

$$dX_t = \hat{\mu}_x(X_t)dt + \sigma(X_t)\cdot H_t dt + \sigma(X_t) dW_t^H.$$

3 Measuring statistical discrepancy

We use entropy relative to the baseline probability to restrict martingales that represent alternative probabilities. The process $M^H \log M^H$ evolves as an Ito process with drift (also called a local mean)

$$\nu_t = \frac{1}{2} M_t^H |H_t|^2.$$

Write the conditional mean of $M^H \log M^H$ in terms of a history of local means[7]

$$E\left[M^H \log M^H|\mathcal{F}_0\right] = E\left(\int_0^t \nu_u du|\mathcal{F}_0\right)$$

[7]For this paper, we will simply impose the first equality. There exists a variety of sufficient conditions that justify this equality.

$$= \frac{1}{2} E \left(\int_0^t M_u^H |H_u|^2 du | \mathcal{F}_0 \right).$$

To formulate a decision problem that chooses probabilities to minimize expected utility, we will use the representation after the second equality without imposing that M^H is a martingale and then verify that the solution is indeed a martingale. Hansen et al. (2006) justify this approach.[8]

To construct relative entropy with respect to a probability model affiliated with a martingale M^R defined by a drift distortion process R, we use a likelihood ratio $\log M_t^H - \log M_t^R$ with respect to the M_t^R model rather than a likelihood ratio $\log M_t^H$ with respect to the baseline model to arrive at:

$$E \left[M_t^H \left(\log M_t^H - \log M_t^R \right) | \mathcal{F}_0 \right] = \frac{1}{2} E \left(\int_0^t M_u^H |H_u - R_u|^2 du \Big| \mathcal{F}_0 \right).$$

When the following limits exist, a notion of relative entropy appropriate for stochastic processes is:

$$\begin{aligned} \lim_{t \to \infty} \frac{1}{t} E \left[M_t^H \left(\log M_t^H - \log M_t^R \right) \Big| \mathcal{F}_0 \right] &= \lim_{t \to \infty} \frac{1}{2t} E \left(\int_0^t M_u^H |H_u - R_u|^2 du \Big| \mathcal{F}_0 \right) \\ &= \lim_{\delta \downarrow 0} \frac{\delta}{2} E \left(\int_0^\infty \exp(-\delta u) M_u^H |H_u - R_u|^2 du \Big| \mathcal{F}_0 \right). \end{aligned}$$

The second line is the limit of Abel integral averages, where scaling by δ makes the weights $\delta \exp(-\delta u)$ integrate to one. We shall use Abel averages with a discount rate equaling the subjective rate that discounts expected utility flows. With that in mind, we define a discrepancy between two martingales M^H and M^R as:

$$\Delta \left(M^H ; M^R | \mathcal{F}_0 \right) = \frac{\delta}{2} \int_0^\infty \exp(-\delta t) E \left(M_t^H \mid H_t - R_t \mid^2 \Big| \mathcal{F}_0 \right) dt.$$

Hansen and Sargent (2001) and Hansen et al. (2006) set $R_t \equiv 1$ to construct relative entropy neighborhoods of a baseline model:

$$\Delta(M^H ; 1 | \mathcal{F}_0) = \frac{\delta}{2} \int_0^\infty \exp(-\delta t) E \left(M_t^H |H_t|^2 \Big| \mathcal{F}_0 \right) dt \geqslant 0, \tag{8}$$

where baseline probabilities are represented here by a degenerate martingale that is identi-

[8] See their Claims 6.1 and 6.2.

cally one. Formula (8) quantifies how a martingale M^H distorts baseline model probabilities. Following Hansen and Sargent (2001), we call $\Delta(M^H; 1|x)$ discounted entropy relative to a probability represented by the baseline martingale.

In contrast to Hansen and Sargent (2001) and Hansen et al. (2006), we start from a convex set $M^R \in \mathcal{M}^o$ of *benchmark* models represented as martingales with respect to a baseline model. We shall describe how we form $\mathcal{M}^o$ in subsection 3.1. These benchmark models are particular parametric alternatives that concern the decision maker. For scalar $\theta > 0$, define a scaled discrepancy of martingale M^H from a set of martingales $\mathcal{M}^o$ as

$$\Theta(M^H|\mathcal{F}_0) = \theta \inf_{M^R \in \mathcal{M}^o} \Delta\left(M^H; M^R|\mathcal{F}_0\right). \tag{9}$$

Scaled discrepancy $\Theta(M^H|\mathcal{F}_0)$ equals zero for M^H in $\mathcal{M}^o$ and is positive for M^H not in $\mathcal{M}^o$. We use discrepancy $\Theta(M^H|\mathcal{F}_0)$ to express the idea that a decision maker wants to investigate the utility consequences of all models that are statistically close to those in $\mathcal{M}^o$. The scaling parameter θ measures how heavily we will penalize utility-minimizing probability choices.

3.1 Constructing a family $\mathcal{M}^o$ of benchmark models

We construct a family of benchmark probabilities by forming a set of martingales M^R with respect to a baseline probability associated with model (1). Formally,

$$\mathcal{M}^o = \left\{M^R \in \mathcal{M} \text{ such that } R_t \in \Xi_t \text{ for all t}\right\} \tag{10}$$

where Ξ is a process of convex sets adapted to the filtration $\mathcal{F}$. Chen and Epstein (2002) also used an instant-by-instant constraint (10) to construct a set of probability models.[9] We use constraint (10) only as an intermediate step in constructing a larger set of statistically similar models whose utility consequences we want to study.

The set of probabilities implied by martingales in $\mathcal{M}^o$ satisfies a property called "rectangularity" by Epstein and Schneider (2003). More generally, a rectangular family of probabilities is formed by i) specifying a set of possible local (i.e., instantaneous) transitions for each t, and ii) constructing *all* possible joint probabilities having such local transitions. Because we use martingales in $\mathcal{M}$ to represent alternative probabilities, the time separability

[9]Anderson et al. (1998) also explored consequences of this type of constraint but without the state dependence in Ξ. Allowing for state dependence is important in the applications featured in this paper.

of specification (10) implies a rectangular family of probabilities.[10]

We use specification (10) because it allows us to represent time-varying parameter models. For at least two reasons, confining $\mathcal{M}^o$ to time-invariant parameter models would be too restrictive. One is that time invariance excludes learning from information that arrives with the passage of time. Another is that the passage of time alters what the decision maker cares about.

The following example illustrates how learning breaks time invariance:

Example 3.1. *Apply Bayes' rule to a finite collection of models characterized by R^j where M^{R^j} is in $\mathcal{M}^o$ for $j = 0, 1, ..., n$. Let $\pi_0^j \geqslant 0$ be a prior probability of model R^j where $\sum_{i=1}^n \pi_0^j = 1$. A martingale*

$$M = \sum_{j=1}^n \pi_0^j M^{R^j}$$

corresponds to a mixture of R^j models. The mathematical expectation of M conditioned on date zero information equals unity. The law of motion for M is

$$\begin{aligned} dM_t =& \sum_{j=1}^n \pi_0^j dM_t^{R^j} \\ =& \sum_{j=1}^n \pi_0^j M_t^{R^j} R_t^j \cdot dW_t \\ =& M_t \left(\pi_t^j R_t^j \right) \cdot dW_t \end{aligned}$$

where π_t^j is the date t posterior

$$\pi_t^j = \frac{\pi_0^j M_t^{R^j}}{M_t}.$$

The drift distortion is

$$R_t = \sum_{i=1}^n \pi_t^j R_t^j.$$

The example illustrates how Bayes' rule leads naturally to a particular form of history-dependent weights on the R_t^j's that characterize alternative models.

Another reason for history dependence is that decision maker's perspective changes over time. A decision maker with a set of priors (i.e., a robust Bayesian) would want to evaluate the utility consequences of the set of posteriors implied by Bayes' law from different perspectives over time. With an aversion to ambiguity, this robust Bayesian would

[10]Rectangularity, per se, does not require Ξ_t to be convex, a property that we impose for other reasons.

rank alternative plans by minimizing expected continuation utilities over the set of priors. Epstein and Schneider (2003) note that for many possible sets of models and priors this approach induces a form of dynamic inconsistency. Thus, consider a given plan. A decision maker has more information at $t > 0$ than at $t = 0$ and he cares only about the continuation of the plan for dates $s \geqslant t$. To evaluate a plan under ambiguity aversion at $t > 0$, the decision maker would minimize over the set of date zero priors. His change in perspective would in general lead the decision maker to choose different worst-case date zero priors as time passes. As a result, a date t conditional preference order could conflict with a date zero preference order. This leads Epstein and Schneider to examine the implications of a dynamic consistency axiom. To make preferences satisfy that axiom, they argue that the decision maker's set of probabilities should satisfy their rectangularity property. Epstein and Schneider make

> ... an important conceptual distinction between the set of probability laws that the decision maker views as possible, such as $Prob$, and the set of priors P that is part of the representation of preference.

Regardless of whether they are subjectively or statistically plausible, Epstein and Schneider recommend augmenting a decision maker's original set of "possible" probabilities (i.e., their $Prob$) with enough additional probabilities to make the enlarged (i.e., their P) set satisfy a rectangularity property that suffices to render the conditional preferences orderings dynamically consistent as required by their axioms.

We illustrate the mechanics in terms of the simplified setting of Example 3.1 with $n = 2$. Suppose that we have a set of priors $\underline{\pi}_0^1 \leqslant \pi_0^1 \leqslant \overline{\pi}_0^1$. For each possible prior, we use Bayes' rule to construct a posterior residing in an interval $[\underline{\pi}_t, \overline{\pi}_t]$, an associated set of drift processes $\{R_t : t \geqslant 0\}$, and implied probability measures over the filtration $\{\mathcal{F}_t : t \geqslant 0\}$. This family of probabilities will typically not be rectangular. We can construct the smallest rectangular family of probabilities by considering the larger space $\{R_t : t \geqslant 0\}$ with $R_t \in \Xi_t$, where

$$\Xi_t = \left\{\pi_t R_t^1 + (1 - \pi_t^1) R_t^2,\ \underline{\pi}_t^1 \leqslant \pi_t \leqslant \overline{\pi}_t^1,\ \pi_t \text{ is } \mathcal{F}_t \text{ measurable}\right\} \tag{11}$$

Augmenting the set $\{R_t : t \geqslant 0\}$ in this way succeeds in making conditional preference orderings over plans remain the same as time passes. This set of probabilities includes elements that can emerge from *no* single date zero prior when updated using Bayes' rule. Instead, we construct the set $\{R_t : t \geqslant 0\}$ by allowing a different date zero prior at each calendar date t. The freedom to do that makes the set of alternative probability models substantially larger and intertemporally disconnects restrictions on local transition probabilities.

There is a conflict between augmenting a family of probabilities to be rectangular and embracing the concept of admissibility that is widely used in statistical decision theory. An admissible decision rule cannot be dominated under all possible probability specifications entertained by the decision maker. Verifying optimality against a worst-case model is a common way to establish that a statistical decision rule is admissible. Epstein and Schneider's proposal to achieve dynamic consistency by adding probabilities to those that the decision maker thinks are possible can render the resulting decision rule inadmissible. Good (1952)'s recommendation for assessing max-min decision making is then unworkable.[11]

We expand the set of probability models for a purpose other than Epstein and Schneider's goal of dynamic consistency and we do it in a way that protects Good's (1952) recommendation. We start with a finite-dimensional parameterization of transition dynamics and note that if R_t^j is a time invariant function of the Markov state X_t for each j, then so is a convex combination of R_t^j's. We can use a convex combination of R_t^j's to represent time-varying parameters whose evolution we do not characterize probabilistically. We illustrate this by modifying Example 3.1 to capture a finite dimensional parameter vector that varies over time.

Example 3.2. *Suppose that R_t^j is a time invariant function of the Markov state, X_t for each $j = 1, \ldots, n$. Linear combinations of R_t^j's can generate the following set of time-invariant parameter models:*

$$\left\{ M^R \in \mathcal{M} : R_t = \sum_{j=1}^{n} \pi^j R_t^j, \pi \in \Pi \textit{ for all } t \geqslant 0 \right\}. \tag{12}$$

Here the unknown parameter vector $\pi = \begin{bmatrix} \pi^1 & \pi^2 & \ldots & \pi^n \end{bmatrix}' \in \Pi$, a convex subset of $\mathbb{R}^n$. We

[11]Presumably, an advocate of Epstein and Schneider's dynamic consistency axiom could respond that admissibility is too limiting in a dynamic context because it takes what amounts to committing to a time 0 perspective and not allowing a decision maker to revaluate later. Nevertheless, it is common in the control theory literature to maintain the date zero perspective and in effect solve a commitment problem under ambiguity aversion.

can enlarge this set to include time-varying parameter models:

$$\left\{ M^R \in \mathcal{M} : R_t = \sum_{j=1}^{n} \widetilde{\pi}_t^j R_t^j, \widetilde{\pi}_t \in \Pi \textit{ for all } t \geqslant 0 \right\}, \tag{13}$$

where the unknown time-varying parameter vector $\widetilde{\pi}_t = \begin{bmatrix} \widetilde{\pi}_t^1 & \widetilde{\pi}_t^2 & \ldots & \widetilde{\pi}_t^n \end{bmatrix}'$ *has realizations confined to* Π*, the same convex subset of* $\mathbb{R}^n$ *that appears in* (12)*. The decision maker has an incentive to compute the mathematical expectation of* $\widetilde{\pi}_t$ *conditional on date* t *information, which we denote* π_t*. Since the realizations of* $\widetilde{\pi}_t$ *are restricted to be in* Π*, this same restriction applies to their conditional expectations, and thus*

$$\Xi_t = \left\{ \sum_{j=1}^{n} \pi_t^j R_t^j,\ \pi_t \in \Pi,\ \pi_t \textit{ is } \mathcal{F}_t \textit{ measurable} \right\}. \tag{14}$$

The family of probabilities induced by martingales $M^R \in \mathcal{M}$ *with drift distortion processes* $R_t \in \Xi_t$ *restricted by* (14) *is rectangular.*

Remark 3.3. *We can maintain convexity of a set of* R_t^j *while also imposing restrictions motivated by statistical plausibility.*

As the quantitative example in section 7 demonstrates, even though the benchmark models are linear in a Markov state, max-min expressions of ambiguity aversion discover worst-case models with nonlinearities in the underlying dynamics. In this case, an *ex post* assessment of empirical plausibility of the type envisioned by Good (1952) asks whether such nonlinear outcomes are plausible.

Our applications employ a special case of constraint (14), but the approach is more generally applicable.

3.2 Misspecification of benchmark models

Our decision maker wants to evaluate the utility consequences not just of the benchmark models in $\mathcal{M}^o$ but also of less structured models that statistically are difficult to distinguish from them. For that purpose, we employ the scaled statistical discrepancy measure $\Theta(M^H|\mathcal{F}_0)$ defined in (9).[12] The decision maker uses the scaling parameter $\theta < \infty$ and

[12] See Watson and Holmes (2016) and Hansen and Marinacci (2016) for recent discussions of longstanding misspecification challenges confronted by statisticians and economists.

the relative entropy that it implies to calibrate a set of additional unstructured models. We pose a minimization problem in which θ serves as a penalty parameter that prohibits exploring alternative probabilities that statistically deviate too far from the benchmark models. This minimization problem induces a preference ordering within a broader class of dynamic variational preference that Maccheroni et al. (2006b) have shown to be dynamically consistent.

To understand how our formulation relates to dynamic variational preferences, notice that benchmark models represented in terms of their drift distortion processes R_t enter separately on the right side of the statistical discrepancy measure

$$\Delta\left(M^H; M^R|\mathcal{F}_0\right) = \frac{\delta}{2}\int_0^\infty \exp(-\delta t) E\left(M_t^H \mid H_t - R_t \mid^2 \Big| \mathcal{F}_0\right) dt.$$

Specification (10) leads to a conditional discrepancy

$$\xi_t(H_t) = \inf_{R_t \in \Xi_t} |H_t - R_t|^2$$

and an associated integrated discounted discrepancy

$$\Theta\left(M^H|\mathcal{F}_0\right) = \frac{\theta}{2}\int_0^\infty \exp(-\delta t) E\left[M_t^H \xi_t(H_t) \Big| \mathcal{F}_0\right] dt.$$

We want a decision maker to care also about the utility consequences of statistically close models that we describe in terms of the discrepancy measure

$$\Theta\left(M^H|\mathcal{F}_0\right). \tag{15}$$

For any hypothetical state- and date-contingent plan – consumption in the example of section 5 – we follow Hansen and Sargent (2001) in minimizing a discounted expected utility function plus the θ-scaled relative entropy penalty $\Theta(M^H|\mathcal{F}_0)$ over the set of models. As the following remark verifies, this procedure induces a dynamically consistent preference ordering over decision processes.

Remark 3.4. *Define what Maccheroni et al. (2006b) call an ambiguity index process:*

$$\Theta_t\left(M^H\right) = \frac{\theta}{2}\int_0^\infty \exp(-\delta u) E\left[\left(\frac{M_{t+u}^H}{M_t^H}\right)\xi_{t+u}(H_{t+u}) \Big| \mathcal{F}_t\right] du.$$

This process solves the following continuous-time counterpart to equations (11) and (12) of Maccheroni et al.:

$$0 = -\delta\Theta_t\left(M^H\right) + \frac{\theta}{2}\xi_t(H_t),$$

where the counterpart to our γ_t in their analysis is $\frac{\theta}{2}\xi_t(H_t)$.[13]

3.3 An unworkable alternative

We have formulated our family of benchmark models $\mathcal{M}^o$ to be rectangular and therefore consistent with a formulation of Chen and Epstein (2002). But our decision maker's concern that all benchmark models might be misspecified leads him to want explore the utility consequences of unstructured probability models that are not rectangular even though they are statistically close, as measured by relative entropy.

An alternative approach would be first to construct a set that includes relative entropy neighborhoods of all martingales in $\mathcal{M}^o$. For instance, we could start with a set

$$\left\{M^H \in \mathcal{M} : \Theta(M^H|\mathcal{F}_0) < \epsilon\right\} \tag{16}$$

that yields a set of implied probabilities that are not rectangular. Why not at this point follow Epstein and Schneider's (2003) recommendation to add enough martingales to attain a rectangular set of probability measures? The answer is that doing so would include all martingales in $\mathcal{M}$ – a set too large for a max-min robustness analysis.

To show this, it suffices to look at relative entropy neighborhood of the baseline model.[14] To construct a rectangular set of models that include the baseline model, for a fixed date τ, consider a random vector $\overline{H}_\tau$ that is observable at that date and that satisfies

$$E\left(|\overline{H}_\tau|^2 \mid X_0 = x\right) < \infty. \tag{17}$$

Form a stochastic process

$$H_t^u = \begin{cases} 0 & 0 \leqslant t < \tau \\ \overline{H}_\tau & \tau \leqslant t < \tau + u \\ 0 & t \geqslant \tau + u. \end{cases} \tag{18}$$

[13]In Hansen and Sargent (2001) and Hansen et al. (2006), $\gamma_t = \frac{\theta}{2}|H_t|^2$. This contrasts with how equation (17) of Maccheroni et al. (2006b) describes Hansen and Sargent and Hansen et al.'s "multiplier preferences". We regard the disparity as a minor blemish in Maccheroni et al. (2006b). It is pertinent to point this out here only because the analysis in this paper generalizes our earlier work.

[14]Including additional benchmark models would only make the set of martingales larger.

The martingale M^{H^u} associated with H^u equals one both before time τ and after time $\tau + u$. Compute relative entropy:

$$\begin{aligned}\Delta(M^{H^u}|x) &= \left(\frac{1}{2}\right)\int_{\tau}^{\tau+u} \exp(-\delta t) E\left[M_t^{H^u}|\overline{H}_\tau|^2 dt \Big| X_0 = x\right] dt \\ &= \left[\frac{1-\exp(-\delta u)}{2\delta}\right] \exp(-\delta\tau) E\left(|\overline{H}_\tau|^2 \mid X_0 = x\right).\end{aligned}$$

Evidently, relative entropy $\Delta(M^{H^u}|x)$ can be made arbitrarily small by shrinking u to zero. This means that any rectangular set that contains $\overline{\mathcal{M}}$ must allow for a drift distortion $\bar{H}_\tau$ at date τ. We summarize this argument in the following proposition:

Proposition 3.5. *Any rectangular set of probabilities that contains the probabilities induced by martingales in* (16) *must also contain the probabilities induced by any martingale in* $\mathcal{M}$.

This rectangular set of martingales allows us too much freedom in setting date τ and random vector $\bar{H}_\tau$: all martingales in the set $\mathcal{M}$ identified in definition 2.1 are included in the smallest rectangular set that embeds the set described by (16). That set is too big to pose a meaningful robust decision problem.

4 Robust planning problem

To illustrate how a robust planner would evaluate the utility consequences of alternative models that are difficult to distinguish according to our relative entropy measure, we deliberately consider a simple setup with an exogenous consumption process. We use a continuous time planning problem to deduce shadow prices of uncertainty from the martingales that generate worst-case probabilities.[15]

To construct a set of models, the planner follows the recipe:

1) Begin with a baseline model.

2) Create a set $\mathcal{M}^o$ of *benchmark* models by naming a sequence of closed, convex sets $\{\Xi_t\}$ and associated drift distortion processes $\{R_t\}$ that satisfy "rectangularity constraint" (14).

[15]Richer models would include production, capital accumulation, and distinct classes of decision makers with differential access to financial markets. Before adding such features, we want to understand uncertainty in our simple environment. In doing this, we follow a tradition extending back to Lucas (1978) and Mehra and Prescott (1985).

3) Guided by a discrepancy measure (9), augment $\mathcal{M}^o$ with additional models that, although they violate (14), are statistically close to models that do satisfy it.

For step 1), we use the diffusion (1) as a baseline model. For step 2), we begin with Markov alternatives to (1) of the form

$$dX_t = \mu^j(X_t) + \sigma(X_t)dW_t^{R^j},$$

where W^{R^j} is a Brownian motion and (6) continues to describe the relationship between the processes W and W^{R^j}. The vector of drifts μ^j differs from $\hat{\mu}$ in baseline model (1), but the volatility vector σ is common to both models. These initial benchmark models have drift distortions that are time invariant functions of the Markov state, namely, linear combinations of $R_t^j = \eta^j(X_t)$, where

$$\eta^j(x) = \sigma(x)^{-1}\left[\mu^j(x) - \hat{\mu}(x)\right].$$

As in Example 3.2, we add benchmark models of the form (14), so we construct an initial set of time invariant parameter models by

$$r(x) = \sum_{j=1}^{n} \pi^j \eta^j(x), \;\; \pi \in \Pi \tag{19}$$

where a π in the convex set Π characterizes an alternative model. In forming the set $\mathcal{M}^o$ of benchmark models, we allow the π's to be adapted to filtration $\mathcal{F}$ in order to accommodate time variation in the underlying parameters. Step 3) includes statistically similar models.

We depict preferences with an instantaneous utility function $U(x)$ and a subjective discount rate δ. Where m is a realized value of a martingale, a value function $mV(x)$ that satisfies the following HJB equation determines a worst-case model:

$$\begin{aligned} 0 = \min_{h,r} & -\delta mV(x) + mU(x) + m\hat{\mu}(x)\cdot\frac{\partial V}{\partial x}(x) + m[\sigma(x)h]\cdot\frac{\partial V}{\partial x}(x) \\ & + \frac{m}{2}\text{trace}\left[\sigma(x)'\frac{\partial^2 V}{\partial x \partial x'}(x)\sigma(x)\right] + \frac{\theta m}{2}|h-r|^2 \end{aligned} \tag{20}$$

subject to (19). Here r (for "rectangular") represents benchmark models in $\mathcal{M}^o$ and h represents nonparametric models that are statistically similar to models in $\mathcal{M}^o$. Because m multiplies all terms on the right side, it can be omitted.

The problem on the right side of HJB equation (20) can be simplified by first minimizing with respect to h given r, or equivalently, by minimizing with respect to $h - r$ given r. First-order conditions for this simpler problem lead to

$$h - r = -\frac{1}{\theta}\sigma(x)'\frac{\partial V}{\partial x}(x). \tag{21}$$

Substituting from (21) into HJB equation (20) gives the reduced HJB equation in:

Problem 4.1.

$$\begin{aligned} 0 = \min_r \; & -\delta V(x) + U(x) + \hat{\mu}(x) \cdot \frac{\partial V}{\partial x}(x) + [\sigma(x) r] \cdot \frac{\partial V}{\partial x}(x) \\ & + \frac{1}{2}\text{trace}\left[\sigma(x)'\frac{\partial^2 V}{\partial x \partial x'}(x)\sigma(x)\right] - \frac{1}{2\theta}\left[\frac{\partial V}{\partial x}(x)\right]'\sigma(x)\sigma(x)'\left[\frac{\partial V}{\partial x}(x)\right] \end{aligned} \tag{22}$$

where minimization is subject to (19).

After problem 4.1 has been solved for an optimal $r^*(x)$, we can recover the optimal h from $h^*(x) = r^*(x) - \frac{1}{\theta}\sigma(x)'\frac{\partial V^*}{\partial x}(x)$ where V^* solves HJB equation (22).

This minimization problem generates two minimizers, namely, r and h. The minimizing r is a benchmark drift taking the form $r^*(x) = \sum_{j=1}^{n} \pi^{j*}(x) r_j(x)$, which evidently depends on the state x. The associated minimizing h is a worst-case drift distortion $h^*(x)$ relative to the worst-case benchmark model. The worst-case drift distortion $h^*(x)$ adjusts for the decision maker's suspicion that the data are generated by a model not in $\mathcal{M}^o$.

5 Uncertainty about Macroeconomic Growth

To prepare the way for the quantitative illustration in section 7, this section applies our setup within a particular macro-finance setting. We start with a baseline parametric model, then form a family of parametric benchmark probability models for a representative investor's consumption process. We deduce the pertinent version of HJB equation (20) that describes the value function attained by worst-case drift distortions r and h. The baseline model is

$$\begin{aligned} dY_t &= .01\left(\hat{\alpha}_y + \hat{\beta} Z_t\right) dt + .01\sigma_y \cdot dW_t \\ dZ_t &= (\hat{\alpha}_z - \hat{\kappa} Z_t)\, dt + \sigma_z \cdot dW_t. \end{aligned} \tag{23}$$

We scale by .01 because Y is typically expressed in logarithms and we want to work with growth rates. Let

$$X = \begin{bmatrix} Y \\ Z \end{bmatrix}.$$

We focus on the following collection of benchmark parametric models that nests that the baseline model (23):

$$\begin{aligned} dY_t &= .01\left(\alpha_y + \beta Z_t\right) dt + .01\sigma_y \cdot dW_t^R \\ dZ_t &= \left(\alpha_z - \kappa Z_t\right) dt + \sigma_z \cdot dW_t^R, \end{aligned} \tag{24}$$

where W^R is a Brownian motion and (6) continues to describe the relationship between the processes W and W^R. Here $(\alpha_y, \alpha_z, \beta, \kappa)$ are parameters distinguishing the benchmark models (24) from the baseline model, and (σ_y, σ_z) are parameters common to models (23) and (24).

We can represent members of a parametric class defined by (24) in terms of our section 2.1 structure with drift distortions R of the form

$$R_t = \eta(Z_t) \equiv \eta_0 + \eta_1 Z_t,$$

then use (1), (6), and (24) to deduce the following restrictions on η_0 and η_1:

$$\begin{aligned} \sigma\eta_0 &= \begin{bmatrix} \alpha_y - \hat{\alpha}_y \\ \alpha_z - \hat{\alpha}_z \end{bmatrix} \\ \sigma\eta_1 &= \begin{bmatrix} \beta - \hat{\beta} \\ \hat{\kappa} - \kappa \end{bmatrix}. \end{aligned} \tag{25}$$

where

$$\sigma = \begin{bmatrix} (\sigma_y)' \\ (\sigma_z)' \end{bmatrix}.$$

Suppose that $Y = \log C$, where C is consumption, δ is a subjective rate of discount and instantaneous utility $U(x) = y$. The formulation in Example 3.2 is equivalent in this case to imposing restrictions directly on $(\alpha_y, \alpha_z, \beta, \kappa)$. The HJB equation corresponding to (22) that determines a worst-case benchmark model – i.e., a worst-case $r(z)$ – is:

$$
\begin{aligned}
0 = \min_{(\alpha_y, \alpha_z, \beta, \kappa) \in \Pi} & -\delta \upsilon(z) + .01(\alpha_y + \beta z) + \frac{d\upsilon}{dz}(z)\left(\alpha_z - \kappa z\right) \\
& - \frac{1}{2\theta} \begin{bmatrix} .01 & \frac{d\upsilon}{dz}(z) \end{bmatrix} \sigma\sigma' \begin{bmatrix} .01 \\ \frac{d\upsilon}{dz}(z) \end{bmatrix} + \frac{1}{2}|\sigma_z|^2 \frac{d^2\upsilon}{dz^2}(z)
\end{aligned} \tag{26}
$$

where $V(x) = y + \upsilon(z)$ and Π is a convex set of parameter vectors characterizing a set of benchmark models. A worst-case benchmark model induces a worst-case nonparametric model via equation (21). (In the portfolio problem of section 8, we will also maximize over portfolio weights and the consumption process C.)

6 Entropy discrepancy

Good (1952) suggests that evaluating a max-min expected utility approach involves verifying that the associated worst-case model is plausible.[16] We implement that suggestion by using entropy to measure how far a worst-case model is from a set of benchmark models, then applying Good's idea to help us calibrate the penalty parameter θ in HJB equation (20). We describe two approaches.

6.1 Relative entropy revisited

Recall from section 3 that relative entropy for a stochastic process conditioned on date 0 information is:

$$
\begin{aligned}
\varepsilon(M^H) &= \lim_{t\to\infty} \frac{1}{t} E\left(M_t^H \log M_t^H \middle| \mathcal{F}_0\right) \\
&= \lim_{t\to\infty} \frac{1}{2t} \int_0^t E\left(M_t^H |H_t|^2 \middle| \mathcal{F}_0\right) dt \\
&= \lim_{\delta\downarrow 0} \frac{\delta}{2} \int_0^\infty \exp(-\delta t) E\left(M_t^H |H_t|^2 \middle| \mathcal{F}_0\right) dt.
\end{aligned}
$$

If the stochastic process under the probability implied by M^H is stationary and ergodic, then the limit on the right side is the unconditional expectation of $\frac{1}{2}|H_t|^2$ under the M^H distribution. The mathematical expectation of discounted relative entropy under the sta-

[16]See Berger (1994) and Chamberlain (2000) for related discussions.

tionary distribution implied by M^H equals $\varepsilon(M^H)$ and does not depend on δ . Hence, relative entropy is simply one half times the expectation of $|H_t|^2$ under this measure.

Under the parametric family posited in the example of section 5, the stationary distribution of M^H is normal with mean $\frac{\alpha_z}{\kappa}$ and variance $\frac{|\sigma_z|^2}{2\kappa}$. For

$$H_t = \sigma^{-1} \begin{bmatrix} \alpha_y - \hat{\alpha}_y \\ \alpha_z - \hat{\alpha}_z \end{bmatrix} + \sigma^{-1} \begin{bmatrix} \beta - \hat{\beta} \\ \hat{\kappa} - \kappa \end{bmatrix} Z_t,$$

relative entropy is

$$\begin{aligned} \varepsilon\left(M^H\right) =& \frac{1}{2} \left| \sigma^{-1} \begin{bmatrix} \alpha_y - \hat{\alpha}_y \\ \alpha_z - \hat{\alpha}_z \end{bmatrix} \right|^2 + \begin{bmatrix} \beta - \hat{\beta} & \hat{\kappa} - \kappa \end{bmatrix} \left(\sigma^{-1}\right)' \sigma^{-1} \begin{bmatrix} \alpha_y - \hat{\alpha}_y \\ \alpha_z - \hat{\alpha}_z \end{bmatrix} \left(\frac{\alpha_z}{\kappa} \right) \\ & + \frac{1}{2} \left| \sigma^{-1} \begin{bmatrix} \beta - \hat{\beta} \\ \hat{\kappa} - \kappa \end{bmatrix} \right|^2 \left[\frac{|\sigma_z|^2}{2\kappa} + \left(\frac{\alpha_z}{\kappa} \right)^2 \right] \end{aligned} \tag{27}$$

6.2 Chernoff entropy

Chernoff entropy emerges from studying how, by disguising probability distortions of a baseline model, Brownian motions make it challenging to distinguish models statistically. Chernoff entropy's connection to a statistical decision problem makes it interesting, but it is less tractable than relative entropy. In this section, we characterize Chernoff entropy by extending a construction of Chernoff (1952). In the spirit of Anderson et al. (2003), we use Chernoff (1952) entropy to measure a distortion Z to a baseline model. Anderson et al. (2003) use Chernoff entropy measured as a local rate to draw direct connections between magnitudes of market prices of uncertainty and statistical discrimination. This local rate is state dependent and for diffusion models proportional to the local drift in relative entropy. Important distinctions arise when we measure statistical discrepancy globally as did Newman and Stuck (1979). In this section, we characterize the global version Chernoff entropy and show how to compute it.

6.2.1 Bounding mistake probabilities

Think of a pairwise model selection problem that statistically compares the baseline model (1) with a model generated by a martingale M^H whose logarithm evolves according to

$$d \log M_t^H = -\frac{1}{2} |H_t|^2 dt + H_t \cdot dW_t.$$

Consider a statistical model selection rule based on a data history of length t that takes the form $\log M_t^H \geqslant \log \tau$, where M_t^H is the likelihood ratio associated with the alternative model for a sample size t. This selection rule might incorrectly choose the alternative model when the baseline model governs the data. We can bound the probability of this outcome by using an argument from large deviations theory that starts from

$$\mathbf{1}_{\left\{\log M_t^H \geqslant \tau\right\}} = \mathbf{1}_{\left\{-\mathsf{s}\tau + \mathsf{s}\log M_t^H \geqslant 0\right\}} = \mathbf{1}_{\left\{\exp(-\mathsf{s}\tau)(M_t^H)^{\mathsf{s}} \geqslant 1\right\}} \leqslant \exp(-\mathsf{s}\tau)(M_t^H)^{\mathsf{s}}.$$

This inequality holds for $0 \leqslant \mathsf{s} \leqslant 1$. Under the baseline model, the expectation of the term on the left side equals the probability of mistakenly selecting the alternative model when data are a sample of size t generated by the baseline model. We bound this mistake probability for large t by following Donsker and Varadhan (1976) and Newman and Stuck (1979) and studying

$$\limsup_{t\to\infty} \frac{1}{t} \log E\left[\exp(-\mathsf{s}\tau)\left(M_t^H\right)^{\mathsf{s}} | X_0 = x\right] = \limsup_{t\to\infty} \frac{1}{t} \log E\left[\left(M_t^H\right)^{\mathsf{s}} | \mathcal{F}_0\right]$$

for alternative choices of s. The threshold τ does not affect this limit. Furthermore, the limit is often independent of the initial conditioning information. To get the best bound, we compute

$$\inf_{0 \leqslant \mathsf{s} \leqslant 1} \limsup_{t\to\infty} \frac{1}{t} \log E\left[\left(M_t^H\right)^{\mathsf{s}} | X_0 = x\right],$$

a limit supremum that is typically negative because mistake probabilities decay with sample size. Chernoff entropy is then

$$\chi(M^H) = -\inf_{0 \leqslant \mathsf{s} \leqslant 1} \limsup_{t\to\infty} \frac{1}{t} \log E\left[\left(M_t^H\right)^{\mathsf{s}} | \mathcal{F}_0\right]. \tag{28}$$

Setting $\chi(M^H) = 0$ would mean including only alternative models that cannot be distinguished on the basis of histories of infinite length. In effect, that is what is done in papers that extend the rational expectations equilibrium concept to self-confirming equilibria associated with probability models that are wrong off equilibrium paths, i.e., for events that do not occur infinitely often.[17] Because we want to include alternative parametric probability models, we entertain positive values of $\chi(M^H)$. Our decision theory differs from that typically used for self confirming equilibria because our decision maker acknowledges model uncertainty and wants to adjust decisions accordingly.

[17] See Sargent (1999) and Fudenberg and Levine (2009).

To interpret $\chi(M^H)$, consider the following argument. If the decay rate of mistake probabilities were constant, then mistake probabilities for two sample sizes $T_i, i = 1, 2$, would be

$$\text{mistake probability}_i = \frac{1}{2}\exp\left(-T_i\bar{\rho}\right)$$

for $\bar{\rho} = \chi(M^H)$. We define a 'half-life' as an increase in sample size $T_2 - T_1 > 0$ that multiplies a mistake probability by a factor of one half:

$$\frac{1}{2} = \frac{\text{mistake probability}_2}{\text{mistake probability}_1} = \frac{\exp\left(-T_2\bar{\rho}\right)}{\exp\left(-T_1\bar{\rho}\right)},$$

so the half-life is approximately

$$T_2 - T_1 = \frac{\log 2}{\bar{\rho}}. \tag{29}$$

The preceding back-of-the-envelope calculation justifies the detection error bound computed by Anderson et al. (2003). The bound on the decay rate should be interpreted cautiously because it is constant although the actual decay rate is not. Furthermore, the pairwise comparison oversimplifies the true challenge, which is statistically to discriminate among *multiple* models.

We can make a symmetrical calculation that reverses the roles of the two models and instead conditions on the perturbed model implied by martingale Z^H. It is straightforward to show that the limiting rate remains the same. Thus, when we select a model by comparing a log likelihood ratio to a constant threshold, the two types of mistakes share the same asymptotic decay rate.

6.2.2 Using Chernoff entropy

To implement Chernoff entropy, we follow an approach suggested by Newman and Stuck (1979). Because our worst case models are Markovian, we can use Perron-Frobenius theory to characterize

$$\lim_{t\to\infty}\frac{1}{t}\log E\left[\left(M_t^H\right)^{\mathsf{s}}|X_0 = x\right]$$

for a given $\mathsf{s} \in (0,1)$ as a dominant eigenvalue for a semigroup of linear operators. When this approach is appropriate, the limit does not depend on the initial state x and is characterized as a dominant eigenvalue associated with an eigenfunction that is strictly positive. Given the restrictions on s, since M^H is a martingale, $\left(M_t^H\right)^{\mathsf{s}}$ is a super martingale and its expectation typically decays to zero at an asymptotically exponential rate. Because we expect the dominant eigenvalue to be negative, we represent it as $-\rho$ where the positive number ρ informs us about the rate at which mistake probabilities converge to zero as we increase the observation interval.

Let a positive eigenfunction e and an associated eigenvalue $-\rho$ solve:

$$\exp(-\rho t)e(x) = E\left[\left(M^H\right)^{\mathsf{s}} e(X_t)|X_0 = x\right]$$

where we assume

$$H_t = h^*(X_t)$$

and h^* is the solution to Problem 4.1. "Differentiating with respect to t" gives:

$$\begin{aligned}-\rho e(x) = & \frac{\mathsf{s}(\mathsf{s}-1)}{2} h^*(x) \cdot h^*(x) e(x) + [\mu(x) + \mathsf{s}\sigma(x)h^*(x)] \cdot \left[\frac{\partial e}{\partial x}(x)\right] \\ & + \frac{1}{2}\text{trace}\left[\sigma(x)' \frac{\partial^2 e}{\partial x \partial x'}(x)\sigma(x)\right].\end{aligned}$$

Notice that this equation restricts the function e only up to scale. Implicitly, ρ depends on the parameter s. We maximize $\rho(\mathsf{s})$ by choice of $\mathsf{s} \in [0,1]$.

See Appendix A for some quasi-analytic formulas when H is among the parametric family of baseline models in the example of section 5. Section 7 outcomes will also impel us to evaluate Chernoff entropy for nonlinear Markov specifications, in which case the calculations will use numerical methods without resort to quasi-analytic formulas. It will be handy to move between Chernoff entropy and half-lives of decay in mistake probabilities:

$$\tilde{\rho} = \frac{\log 2}{\text{half life}}.$$

7 Quantitative example

Our quantitative example builds on the setup of section 5 and features a representative investor who wants to explore utility consequences of alternative models portrayed by $\{M_t^H\}$ and $\{M_t^R\}$ processes, some of which contribute difficult to detect and troublesome predictable components of consumption growth.[18] Chernoff entropy shapes and quantifies the doubts that we impute to investors.

[18]While we appreciate the value of a more comprehensive empirical investigation with multiple macroeconomic time series, here our aim is to illustrate a mechanism within the context of relatively simple time series models of predictable consumption growth.

7.1 Parametric benchmark models

Our example blends parts of Bansal and Yaron (2004) and Hansen et al. (2008). We use a vector autoregression (VAR) to construct a quantitative version of a baseline model like (23) that approximates responses of consumption to permanent shocks. In contrast to Bansal and Yaron (2004), we introduce no stochastic volatility because we want to focus exclusively on fluctuations in uncertainty prices that are induced by the representative investor's specification concerns.

In constucting a VAR, we follow Hansen et al. (2008) by using additional macroeconomic time series to infer information about long-term consumption growth. We report a calibration of our baseline model (23) deduced from a trivariate VAR for the first difference of log consumption, the difference between logs of business income and consumption, and the difference between logs of personal dividend income and consumption. This specification makes consumption, business income, and personal dividend income cointegrated. Business income is measured as proprietor's income plus corporate profits per capita. Dividends are personal dividend income per capita.[19]

We fit a trivariate vector autoregression that imposes cointegration among the three series. Since we presume that all three time series grow, the coefficients in the cointegrating relation are known. In Appendix B we tell how we used the discrete time VAR estimates to deduce the following parameters for the baseline model (23):

$$\begin{aligned} \hat{\alpha}_y &= .386 & \hat{\beta} &= 1 \\ \hat{\alpha}_z &= 0 & \hat{\kappa} &= .019 \end{aligned}$$

$$\begin{aligned} \sigma_y &= \begin{bmatrix} .488 \\ 0 \end{bmatrix} \\ \sigma_z &= \begin{bmatrix} .013 \\ .028 \end{bmatrix} \end{aligned} \tag{30}$$

We suppose that $\delta = .002$ and $U(x) = y$, where y is the logarithm of consumption. The

[19]The time series are quarterly data from 1948 Q1 to 2015 Q1. Our consumption measure is nondurables plus services consumption per capita. The business income data are from NIPA Table 1.12 and the dividend income from NIPA Table 7.10. By including proprietors' income in addition to corporate profits, we use a broader measure of business income than Hansen et al. (2008) who used only corporate profits. Hansen et al. (2008) did not include personal dividends in their VAR analysis.

standard deviation of the Z process is .158 in the implied stationary distribution under this model.

Having described the representative investor's baseline model, we now describe his benchmark parametric models. We include benchmark models associated with alternative values of $(\alpha_z, \beta, \kappa)$, but for simplicity suppose that $\alpha_y = \hat{\alpha}_y$. When $\alpha_y = \hat{\alpha}_y$, the formula for relative entropy reported in (27) simplifies to:

$$\begin{aligned}\varepsilon\left(\alpha_z, \beta, \kappa\right) =& \frac{1}{2}\left|\sigma^{-1}\begin{bmatrix} 0 \\ \alpha_z - \hat{\alpha}_z \end{bmatrix}\right|^2 + \begin{bmatrix} \beta - \hat{\beta} & \hat{\kappa} - \kappa \end{bmatrix}\left(\sigma^{-1}\right)'\sigma^{-1}\begin{bmatrix} 0 \\ \alpha_z - \hat{\alpha}_z \end{bmatrix}\left(\frac{\alpha_z}{\kappa}\right) \\ &+ \frac{1}{2}\left|\sigma^{-1}\begin{bmatrix} \beta - \hat{\beta} \\ \hat{\kappa} - \kappa \end{bmatrix}\right|^2\left[\frac{|\sigma_z|^2}{2\kappa} + \left(\frac{\alpha_z}{\kappa}\right)^2\right]\end{aligned}$$

Evidently $\varepsilon\left(\alpha_z, \beta, \kappa\right)$ is one half the second moment of H_t under the altered probability. We want to study a set of benchmark models constrained by

$$\left\{\left(\alpha_z, \beta, \kappa\right) : \varepsilon\left(\alpha_z, \beta, \kappa\right) \leqslant \frac{1}{2}\mathsf{q}^2\right\} \tag{31}$$

for some choice of q.

7.2 Uncertain macroeconomic growth

To locate the boundary of the set defined by (31), we solve for $\alpha_z = f(\beta, \kappa)$. This entails solving a quadratic equation for α_z. We take the smaller of the two solutions for α_z and pose:

$$\min_{\beta,\kappa} \frac{d\nu}{dz}(z)\left[f(\beta,\kappa) - \kappa z\right] + .01\beta z,$$

which is the portion of HJB equation (26) pertinent for minimization. Given z, first-order conditions with respect to β and κ are:

$$\begin{aligned}\frac{d\nu}{dz}(z)\frac{\partial f}{\partial \beta}(\beta,\kappa) + .01z &= 0 \\ \frac{\partial f}{\partial \kappa}(\beta,\kappa) - z &= 0.\end{aligned} \tag{32}$$

Substitute solutions of (32) along with the implied value of α_z into HJB equation (26). Both β and κ depend on z. The value function $\nu(z)$ that solves this HJB equation is approximately linear in z for large values of $|z|$, a fact that in Appendix C we use to construct Neumann boundary conditions for the HJB equation (26) that we shall use to compute a numerical approximation.

7.2.1 Dependence on z

Had we also restricted $\beta = 1$, the minimizers for (α_z, κ) would not depend on the value function, provided that the derivative $\frac{d\nu}{dz}(z)$ is always positive. When $\beta = 1$, only the second equation of (32) comes into play. This very special outcome also emerges from other models of ambiguity aversion. For instance, worst-case models of Ilut and Schneider (2014) can be deduced *ex ante* before solving an HJB equation. While computationally convenient, such an outcome looses the important aspect of robustness analysis that unearths features of probabilities that depend on utility functions.

7.2.2 Attitudes toward persistence

The robust planner dislikes persistence of bad growth states and likes persistence of good growth states. Figures 1, 2, and 3 quantify these effects. We consider two choices of q, .1 and .05. We refer to these as *ex ante* choices for reasons that we explain later in this subsection.

Figure 1 depicts sets of $(\alpha_z, \beta, \kappa)$ for the case in which $\mathsf{q} = .1$. It depicts the portion of the region implied by the lower of the two solutions of the quadratic equation for α_z as a function of (β, κ), the part pertinent for the minimization problem. Two curves on the boundary of this region depict minimizing values of the parameters as functions of the state z, one for $z > 0$, the other for $z < 0$. Since β and κ drop out of the objective for the minimization when $z = 0$, the solution there is located at the bottom of this region and corresponds to the smallest possible value of α_z.

Although the contours reported at the surface of Figure 1 compute relative entropy presuming constant coefficients, the domain of the minimization problem includes state-dependent "parameters" $\alpha_z(z), \beta(z), \kappa(z)$. As a consequence, the minimization problem on the right side of the HJB equation allows for many more models than the time invariant parameter models associated with the contours at the boundary. The resulting average relative entropy at the minimizing $(\alpha_z, \beta, \kappa)$ is substantially more than that implied by the *ex ante* choices of the q's. Therefore, we also report what we call *ex post* q's, by which we mean relative entropies for the worst-case models. These are .139 and .068, which as anticipated, are larger than the *ex ante* value .1 and .05, respectively.

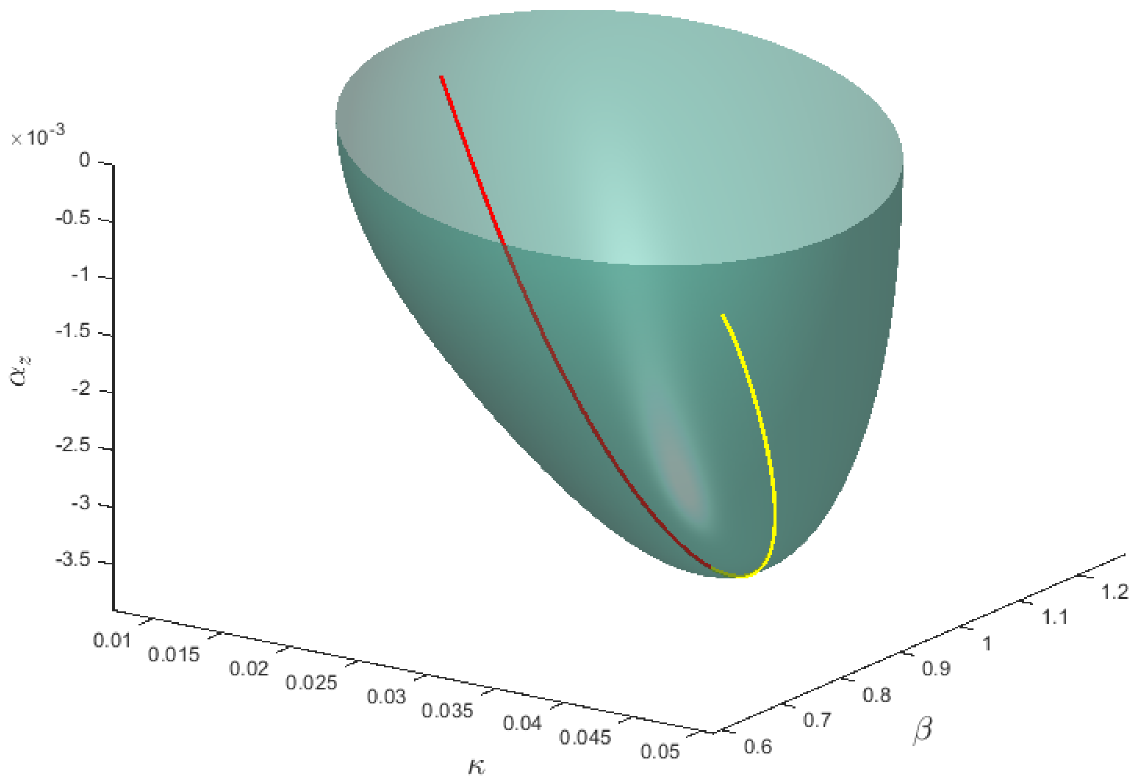

Figure 1: Set of alternative parameter values $(\alpha_z, \beta, \kappa)$ constrained by relative entropy. To generate the parameter region, we set $\mathsf{q} = .1$ and plot only the lower portion of the set as captured by α_z. The red curve plots the chosen parameter configuration for $z < 0$ and the yellow curve for $z > 0$. The $z = 0$ solution is at the bottom of the region.

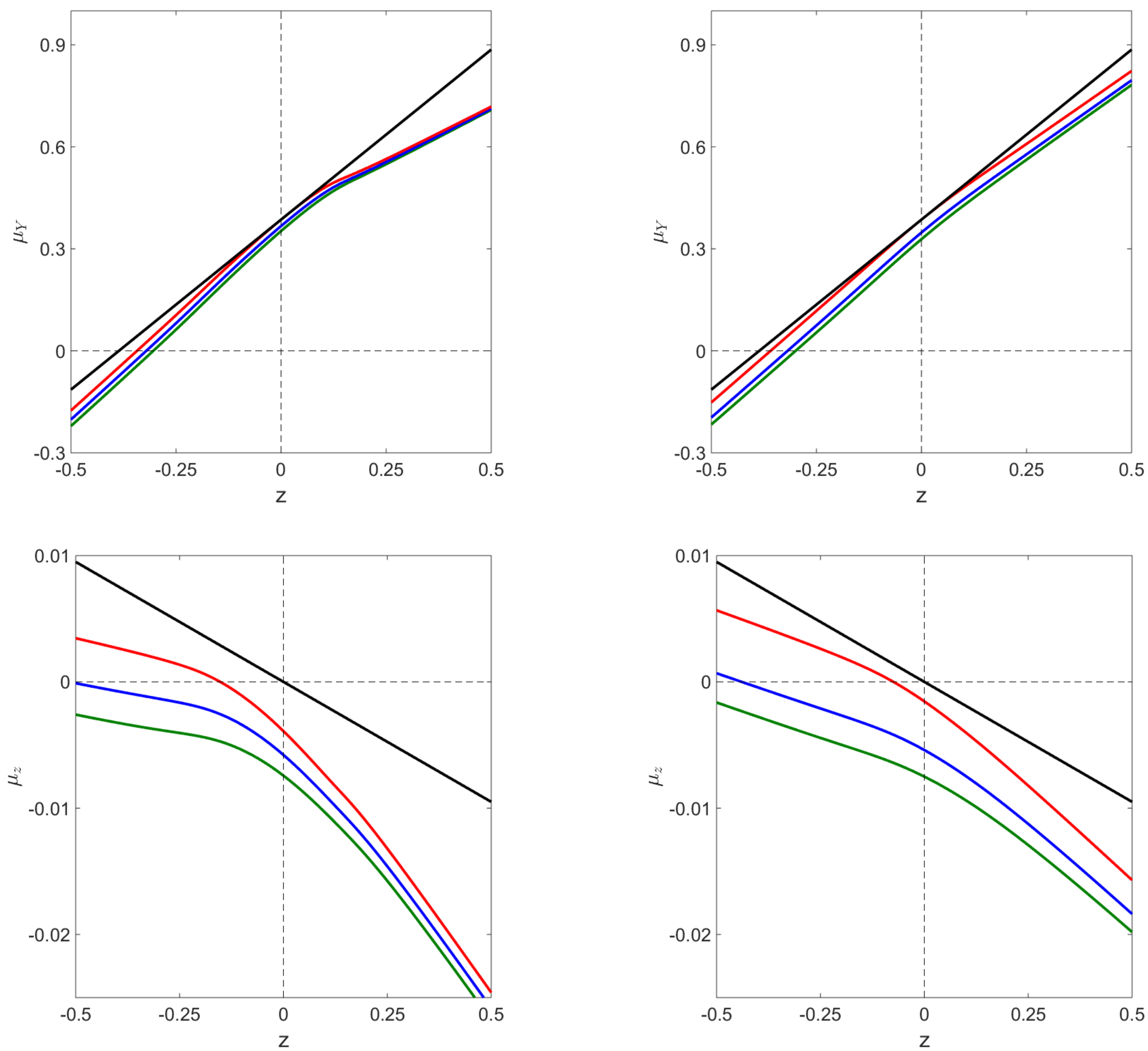

Figure 2: Distorted drifts. Left panels: larger benchmark entropy (*ex ante* $\mathsf{q} = .1$, *ex post* q =.139) . Right panels: smaller benchmark entropy (*ex ante* $\mathsf{q} = .05$, *ex post* $\mathsf{q} = .068$). Upper panels: μ_y's. Lower panels: μ_z's. **Black**: baseline model; **red**: worst-case benchmark model; **blue**: Chernoff half life 120; and **green**: Chernoff half life 60.

Figure 2 reports drift adjustments for ambiguity aversion and concerns about misspecification. Impacts on the drift for the logarithm of consumption are modest compared to bigger drift distortions for the expected growth rate in consumption. Positive values of z are good macroeconomic growth states, so the robust planner worries that these will not persist. For sufficiently negative values of z, the reverse is true, so the robust planner fears persistence of bad growth states, concerns about alternatives within the set of benchmark models. Adding concerns about misspecification of the set of benchmark models shifts the drift of the growth state down, more in bad growth states than in good ones.

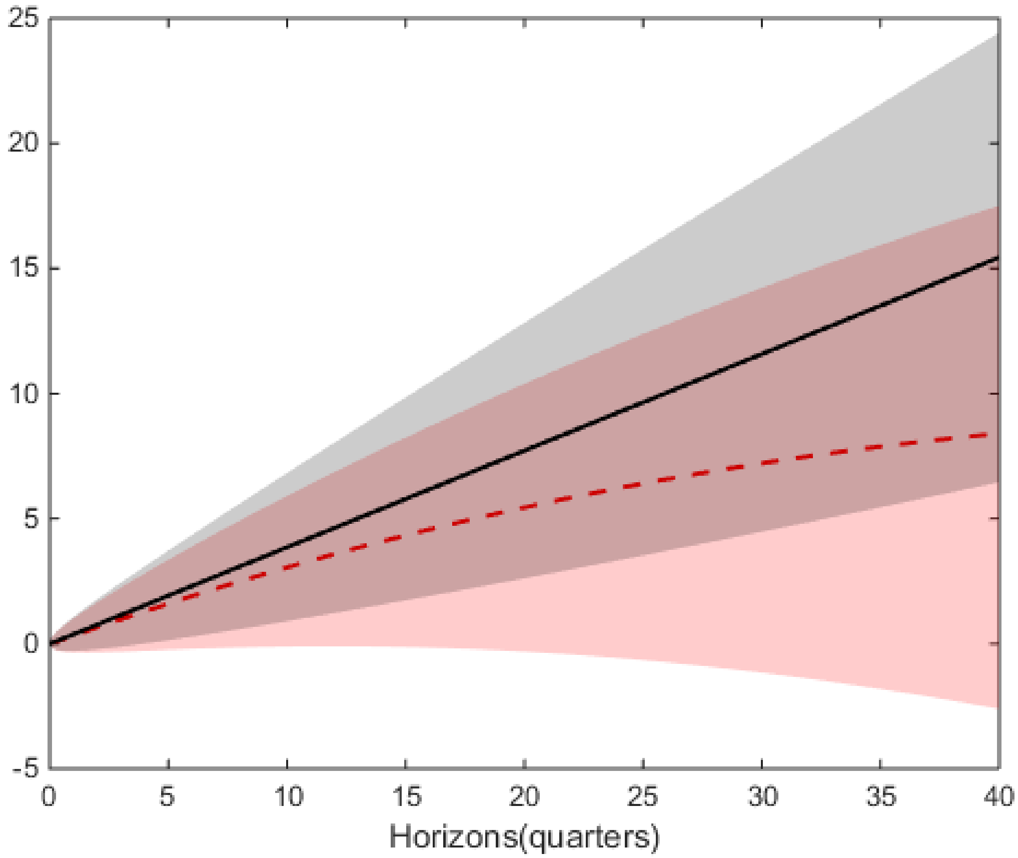

Figure 3: Distribution of $Y_t - Y_0$ under the baseline model and worst-case model for $\mathsf{q} = .1$ and a Chernoff half life of 60 quarters. The **black** solid line depicts the median under the baseline model and the associated gray shaded region gives the region within the .1 and .9 deciles. The red dashed line is the median under the worst-case model and the red shaded region gives the region within the .1 and .9 deciles.

Figure 3 extrapolates impacts of the drift distortion on distributions of future consumption growth over alternative horizons. It shows how the consumption growth distribution adjusted for ambiguity aversion and misspecification tilts down relative to the baseline distribution.

7.3 Uncertain dependence on the growth state

The growth state z affects the evolution of both consumption and the growth rate. We now focus on the uncertainty on these terms only by restricting the constant terms to satisfy: $\alpha_z = \hat{\alpha}_z = 0$ and $\alpha_y = \hat{\alpha}_y$. Under these restrictions, average relative entropy simplifies to:

$$\varepsilon\left(M^R\right) = \frac{1}{2}\left|\sigma^{-1}\begin{bmatrix}\beta - \hat{\beta} \\ \hat{\kappa} - \kappa\end{bmatrix}\right|^2 \left(\frac{|\sigma_z|^2}{2\kappa}\right)$$

Consider an entropy set

$$\left\{(\beta, \kappa) : \varepsilon\left(M^R\right) \leqslant \frac{1}{2}\mathsf{q}^2\right\}. \tag{33}$$

For a given κ, to find the boundary of the entropy set (33) we solve a quadratic equation for β and take the high β solution when $\kappa < \hat{\kappa}$ and the low β solution when $\kappa > \hat{\kappa}$. We obtain two functions $\beta = f^-(\kappa)$ for $\kappa < \hat{\kappa}$ and $\beta = f^+(\kappa)$ for $\kappa > \hat{\kappa}$. The minimizing choice of κ then solves:

$$.01\frac{df^-}{d\kappa}(\kappa) - \frac{dv}{dz}(z) = 0$$

when $z < 0$ and

$$.01\frac{df^+}{d\kappa}(\kappa) - \frac{dv}{dz}(z) = 0$$

when $z > 0$. Evidently, the worst-case κ depends on the value function.

The value function solves two coupled HJB equations, one for $z < 0$ and another for $z > 0$. For $z < 0$, the HJB equation is:

$$\begin{aligned} 0 = \min_{\kappa} & -\delta v(z) + .01[\hat{\alpha}_y + f^-(\kappa)z] - z\kappa\frac{dv}{dz}(z) \ + \frac{1}{2}|\sigma_z|^2\frac{d^2v}{dz^2}(z) \\ & - \frac{1}{2\theta}\begin{bmatrix}.01 & \frac{dv}{dz}(z)\end{bmatrix}\sigma\sigma'\begin{bmatrix}.01 \\ \frac{dv}{dz}(z)\end{bmatrix}. \end{aligned}$$

There is an analogous equation for $z > 0$. The discontinuity in worst-case benchmark models at zero means that at $z = 0$ the value function does not have a second derivative at zero. We obtain two second-order differential equations in value functions and their derivatives; these value functions coincide at $z = 0$, as do their first-derivatives.

Figures 4 and 5 offer quantitative explorations of ambiguity aversion that focus exclusively on the slope coefficients (β, κ) of the benchmark models.

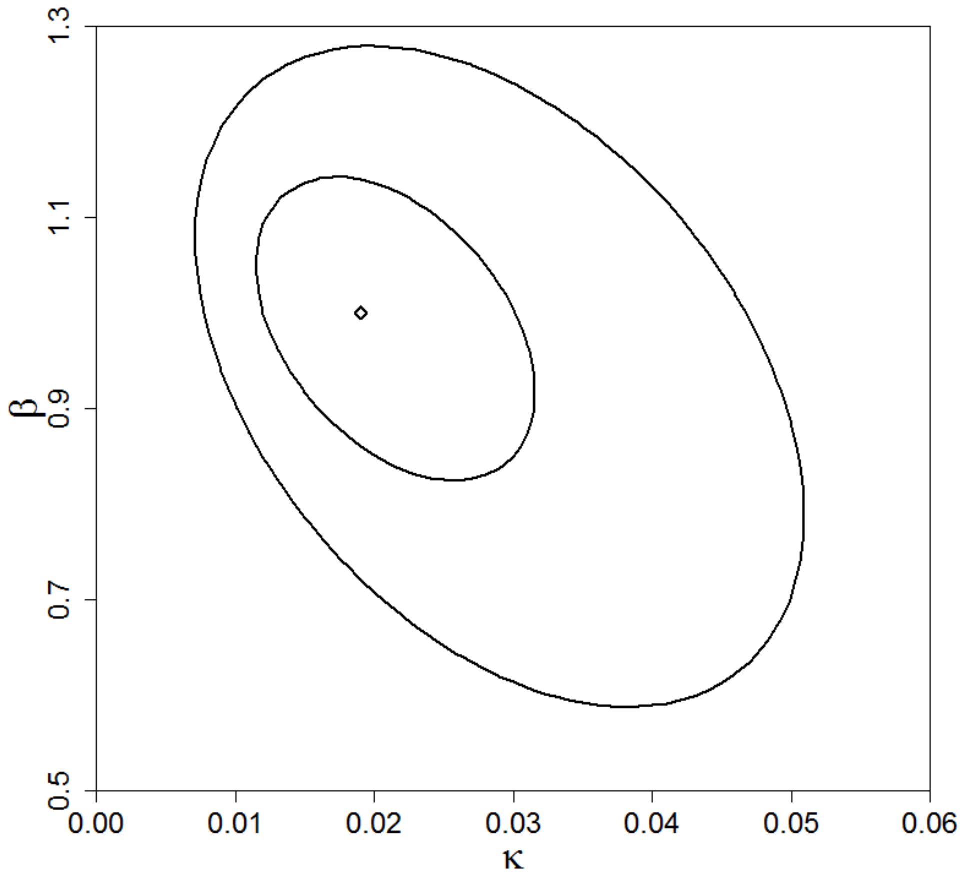

Figure 4: Parameter contours for (β, κ) holding relative entropy fixed. The outer curve $\mathsf{q} = .1$ and the inner curve $\mathsf{q} = .05$. The small diamond in the model depicts the baseline model.

Figure 4 shows the boundaries for the $\mathsf{q} = .1$ and $\mathsf{q} = .05$ restrictions on relative entropy. It is the lower segment of these contours that is targeted in the decision problem.

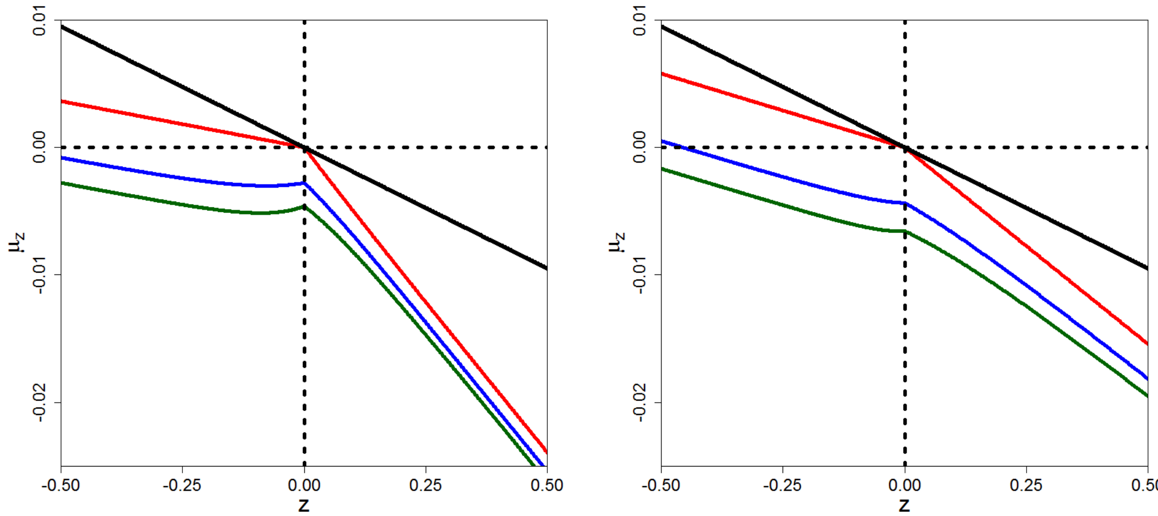

Figure 5: Distorted growth rate drifts. Left panel: larger benchmark entropy ($\mathsf{q} = .1$). Right panel: smaller benchmark entropy ($\mathsf{q} = .05$). **red**: worst-case benchmark model; **blue**: half life 120; and **green**: half life 60.

Figure 5 shows adjustments of the drifts due to ambiguity aversion and concerns about misspecification of the benchmark models. Setting $\theta = \infty$ silences concerns about about misspecification of the benchmark models, all of which must be expressed through minimization over h. When we set $\theta = +\infty$, the implied worst-case benchmark model has state dynamics that take the form of a threshold autoregression with a kink at zero. The distorted drifts again show less persistence than does the baseline model for negative values of z and more persistence for larger values of z. Activating a concern for misspecification of the benchmark models by letting θ be finite shifts the drift as a function of the state downwards, even more so for negative values of z than positive ones.

7.4 Commitment to a worst-case benchmark model

Partly to make contact with an alternative formulation proposed by Hansen and Sargent (2016), we now alter timing protocols. Instead of (H_t, R_t) being chosen simultaneously each instant as depicted in HJB equation (20), a decision maker now confronts a single benchmark model that has been chosen by a statistician in charge of choosing our R process. The decision maker chooses an H process because he does not trust that benchmark model.

Until now, in choosing (H, R) our decision maker has used discounted relative entropy.

We now assume that in choosing the R process, the statistician uses an undiscounted counterpart of relative entropy. Recall that

$$\lim_{t\to\infty} \frac{1}{2t} E\left(\int_0^t M_u^R |R_u|^2 du \Big| \mathcal{F}_0 \right) = \lim_{\delta\downarrow 0} \frac{\delta}{2} E\left(\int_0^\infty \exp(-\delta u) M_u^R |R_u|^2 du \Big| \mathcal{F}_0 \right)$$

In choosing a baseline model, the statistician uses the $\delta \downarrow 0$ measure and imposes the restriction:

$$\lim_{\delta\downarrow 0} \frac{\delta}{2} E\left[\int_0^\infty \exp(-\delta u) M_u^R \left(|R_u|^2 - \xi_u \right) du \Big| \mathcal{F}_0 \right] \leqslant 0. \tag{34}$$

This constraint is satisfied whenever $|R_t|^2 \leqslant \xi_t$ but also for other models as well. The process $\{\xi_t\}$ is specified exogenously. Relative entropy neighborhoods of interior martingales are included in this set. Taking the $\delta \downarrow 0$ limit eliminates the dependence on conditioning information for a convex set of martingales M^R.

To make these results comparable to earlier ones, we have the statistician construct a set of models as follows

- Start with $\underline{\kappa} < \hat{\kappa}$ that is the smallest κ on one of the two curves in Figure 4.
- Let $\mathsf{u} = \hat{\kappa} - \underline{\kappa}$. Then compute $\tilde{\mathsf{s}}(\mathsf{u})$ by solving:

$$\mathsf{b} = \min_{\mathsf{s}} \left| \sigma^{-1} \begin{bmatrix} \mathsf{s} \\ \mathsf{u} \end{bmatrix} \right|$$

This construction assures that

$$\left\{ (\beta, \kappa) : \left| \sigma^{-1} \begin{bmatrix} \beta - \hat{\beta} \\ \hat{\kappa} - \kappa \end{bmatrix} \right|^2 \leqslant \mathsf{b}^2 \right\}$$

contains the region inside the corresponding contour set in Figure 4. This leads us to specify:

$$\xi_t = \mathsf{b}^2 |Z_t|^2$$

in our computations.

We start with a date-zero perspective. The statistician uses the same instantaneous utility function as the decision maker and takes a process of instantaneous utilities as given. The statistician then uses a martingale relative to the baseline model to construct a benchmark model by solving a continuous-time analogue of a control problem posed by Petersen et al. (2000).

The decision maker discounts with $\delta > 0$ and accepts the statistician's model as a benchmark. Because he doubts it he makes a robustness adjustment of the type suggested by Hansen and Sargent (2001). This problem is reminiscent of an optimal expectations game formulated by Brunnermeier and Parker (2005). They formulate two-agent decision problems in which one agent chooses beliefs using an undiscounted utility function while the other agent takes those beliefs as fixed when evaluating alternative plans. Despite formal connections between their work and ours, their work does not originate in concerns about robustness.

An equilibrium for our robust planner game is particularly easy to compute because the instantaneous utility is specified *a priori*. This allows us to first to solve the statistician problem and then the decision maker's problem.[20]

7.4.1 Statistician Problem

We solve the statistician problem first for a given δ and Lagrange multiplier ℓ. The statistician's value function ς is quadratic in z and solves:

$$
\begin{aligned}
0 = \min_r & -\delta\varsigma(z,\delta,\ell) + .01(\hat{\alpha}_y + \hat{\beta}z + \sigma_y \cdot r) + \frac{d\varsigma}{dz}(z,\delta,\ell)\left(\hat{\alpha}_z - \hat{\kappa}z + \sigma_z \cdot r\right) \\
& + \frac{1}{2}|\sigma_z|^2 \frac{d^2\varsigma}{dz^2}(z,\delta,\ell) + \frac{\ell}{2}\left(r \cdot r - \mathsf{b}^2 z^2\right) \\
= & -\delta\varsigma(z,\delta,\ell) + .01(\hat{\alpha}_y + \hat{\beta}z) + \frac{d\varsigma}{dz}(z,\delta,\ell)\left(\hat{\alpha}_z - \hat{\kappa}z\right) + \frac{1}{2}|\sigma_z|^2 \frac{d^2\varsigma}{dz^2}(z,\delta,\ell) \\
& - \frac{1}{2\ell}\begin{bmatrix} .01 & \frac{d\varsigma}{dz}(z,\delta,\ell) \end{bmatrix} \sigma\sigma' \begin{bmatrix} .01 \\ \frac{d\varsigma}{dz}(z,\delta,\ell) \end{bmatrix} - \frac{\ell \mathsf{b}^2 z^2}{2}
\end{aligned}
$$

We take limits as δ gets small and set the multiplier ℓ to satisfy the relative entropy constraint (34) by maximizing:

$$\max_{\ell} \lim_{\delta \downarrow 0} \delta\varsigma(\cdot,\delta,\ell).$$

Since in the small δ limit we do not discount the future, the maximizing multiplier ℓ^* does not depend on z. See Appendix D for a characterization of the solution. The implied drift

[20] In a model with production, this two-step approach would no long apply.

adjustment used to represent the statisician's benchmark model is

$$r^* = \iota^*(z) = -\frac{1}{\ell}^* \sigma' \begin{bmatrix} .01 \\ \frac{d\varsigma}{dz}(z, 0, \ell^*) \end{bmatrix},$$

where we evaluate $\frac{d\varsigma}{dz}(z, 0, \ell^*)$ by computing a small δ limit of $\frac{d\varsigma}{dz}(z, \delta, \ell^*)$. Since ς is quadratic in z, the statistician's worst-case benchmark model alters the probability distribution for W to have a drift that is linear in z. This leads us to express the local dynamics for the benchmark model as:

$$\begin{aligned} \hat{\alpha}_y + \hat{\beta} z + \sigma_y \cdot \iota^*(z) &= \alpha_y^* + \beta^* z \\ \hat{\alpha}_z - \hat{\kappa} z + \sigma_z \cdot \iota^*(z) &= \alpha_z^* - \kappa^* z. \end{aligned}$$

Evidently, here the worst-case benchmark model remains within our parametric class. Note that this benchmark model is not a time-varying or state-dependent coefficient model, in contrast to the situation found under the distinct section 7.3 setting.

7.4.2 Robust Control Problem

At date zero, the decision maker accepts the statistician's model as a benchmark but because he doubts it, he makes a robustness adjustment of the type suggested by Hansen and Sargent (2001). Where the value function is $m[y + \psi(z)]$, write the decision maker's HJB equation as[21]

$$\begin{aligned} 0 = \min_h & -\delta\psi(z) + .01\sigma_y \cdot h + \sigma_z \cdot h \frac{d\psi}{dz}(z) + \frac{\theta}{2} h \cdot h \\ & + .01\left(\alpha_y^* + \beta^* z\right) + \frac{d\psi}{dz}(z)\left(\alpha_z^* - \kappa^* z\right) + \frac{1}{2}|\sigma_z|^2 \frac{d^2\psi}{dz^2}(z) \\ = & -\delta\psi(z) + .01\left(\alpha_y^* + \beta^* z\right) + \frac{d\psi}{dz}(z)\left(\alpha_z^* - \kappa^* z\right) + \frac{1}{2}|\sigma_z|^2 \frac{d^2\psi}{dz^2}(z) \\ & - \frac{1}{2\theta}\begin{bmatrix} .01 & \frac{d\psi}{dz}(z) \end{bmatrix} \sigma\sigma' \begin{bmatrix} .01 \\ \frac{d\psi}{dz}(z) \end{bmatrix}. \end{aligned}$$

In this case we can say more. It is straightforward to show that $\psi(z) = \psi_0 + \psi_1 z$, where in particular ψ_1 solves:

$$0 = -\delta\psi_1 + .01\beta^* - \psi_1\kappa^*$$

[21] As in earlier decision problems, we can omit m from the HJB equation because it multiplies all terms.

which implies

$$\psi_1 = \frac{.01\beta^*}{\delta + \kappa^*}.$$

7.4.3 Results

The following composite[22] worst-case model emerges from our statistician-decision maker game:

$$h^* = \eta^*(z) = \iota^*(z) - \frac{1}{\theta}\sigma' \begin{bmatrix} .01 \\ \frac{d\psi}{dz}(z) \end{bmatrix} = \iota^*(z) - \frac{1}{\theta}\sigma' \begin{bmatrix} .01 \\ \psi_1 \end{bmatrix}$$

In light of how we have formulated things, it is not surprising that the second term takes a form found by Hansen et al. (2006). While this second term is constant in the present example, the first term ι^* is state dependent; η^* remains within the parametric class of benchmark models. Dynamic consistency prevails in the sense that if we ask the players to re-assess their choices at some date $t > 0$, each player would remain content with its original choice.

We study values of θ chosen to match Chernoff half lives of 60 quarters and 120 quarters, respectively. The statistician's worst-case benchmark model has a linear drift for the state variable but its slope is flatter than in the baseline model. It is also shifted downwards. The decision maker's concerns about misspecification of the statistician's benchmark model lead him to shift the worst-case drift as a function of the state variable downwards, while keeping it parallel to that function in the benchmark model. Figure 6 illustrates these effects.

[22] It is composed of worst-case R and H processes.

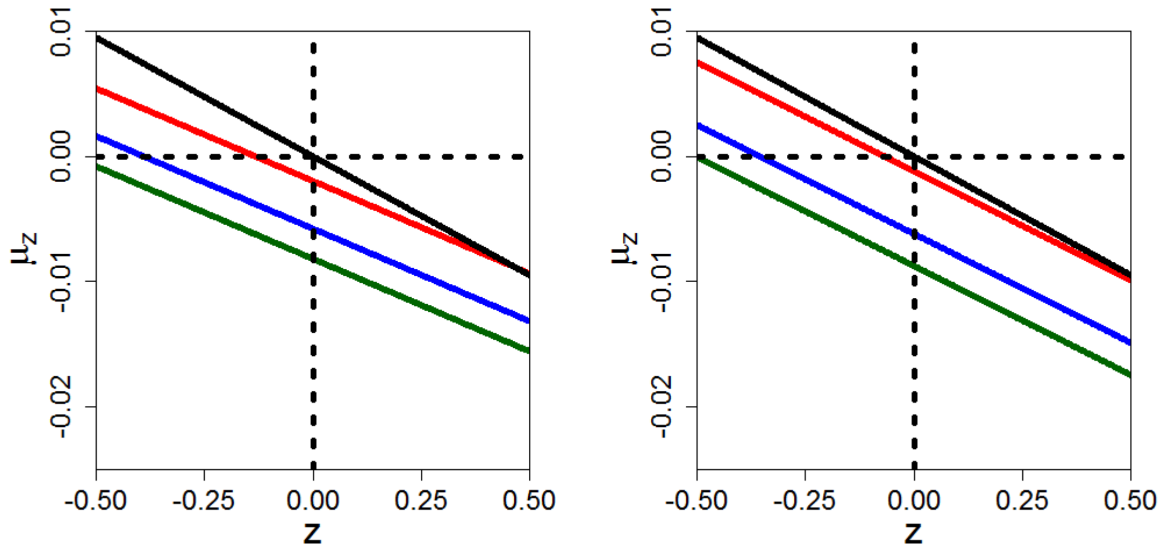

Figure 6: Distorted growth rate drifts. Left panel: larger constraint set ($\underline{\kappa} = .0071$). Right panel: smaller constraint set ($\underline{\kappa} = .0114$). **Black**: baseline model; **red**: worst-case benchmark model; **blue**: half life 120; and **green**: half life 60.

7.4.4 Alternative formulations

For readers uncomfortable with this game against a statistician formulation, we note that there are other formulations with a very similar quantitative implications. For example, we could impose a common discount rate on the statistician and the decision maker and change the timing protocol so that the h and r processes are both chosen once and for all at $t = 0$. Some work in control theory (e.g., Petersen et al. (2000)) formulates things in this way. In that work, the Lagrange multiplier ℓ typically depends on the initial state. In Hansen and Sargent (2016), we suggested robust averaging over this initial state, but this still requires a form of *ex ante* commitment.

As an alternative to imposing constraint (34), we could have formed the set of all martingales M^R for which:

$$\lim_{\delta \downarrow 0} \frac{\delta}{2} E\left(\int_0^\infty \exp(-\delta u) M_u^R |R_u|^2 du \Big| \mathcal{F}_0 \right) \leqslant \mathsf{b}^*$$

$$\mathsf{b}^* = \sup_R \lim_{\delta \downarrow 0} \frac{\delta}{2} E\left(\int_0^\infty \exp(-\delta u) M_u^R |R_u|^2 du \middle| \mathcal{F}_0 \right)$$

subject to $|R_t|^2 \leqslant \xi_t$. This modeling choice would have two consequences. First it would result in an even larger set of martingales. Second, if applied to the present quantitative example, the implied worst-case benchmark and worst-case models would entail adding constant (and therefore state-independent) drifts to the Brownian motions.

8 Robust portfolio choice and pricing

In this section, we describe equilibrium prices that make a representative investor willing to bear risks accurately described by baseline model (1) inspite of his concerns about model misspecification. We construct equilibrium prices by appropriately extracting shadow prices from the robust planning problem of section 4. We decompose equilibrium risk prices into distinct compensations for bearing risk and for bearing model uncertainty. We begin by posing the representative investor's portfolio choice problem.

8.1 Robust investor portfolio problem

A representative investor faces a continuous-time Merton portfolio problem in which individual wealth K evolves as

$$dK_t = -C_t dt + K_t \iota(Z_t) dt + K_t A_t \cdot dW_t + K_t \pi(Z_t) \cdot A_t dt, \tag{35}$$

where $A_t = a$ is a vector of chosen risk exposures, $\iota(x)$ is the instantaneous risk free rate expressed, and $\pi(z)$ is the vector of risk prices evaluated at state $Z_t = z$. Initial wealth is K_0. The investor discounts the logarithm of consumption and distrusts his probability model.

Key inputs to a representative investor's robust portfolio problem are the baseline model (1), the wealth evolution equation (35), the vector of risk prices $\pi(z)$, and the quadratic function ξ that defines the alternative explicit models that concern the representative investor.

Under a guess that the value function takes the form $m\tilde{v}(z) + m \log k + m \log \delta$, the

HJB equation for the robust portfolio allocation problem is

$$
\begin{aligned}
0 = \max_{a,c} \min_{h,r} & -\delta m \tilde{\upsilon}(z) - \delta m \log k - \delta m \log \delta + \delta m \log c - \frac{mc}{k} + m\iota(z) \\
& + m\pi(z) \cdot a + ma \cdot h - \frac{m|a|^2}{2} + m\frac{d\tilde{\upsilon}}{dz}(z)\left(\hat{\alpha}_z - \hat{\kappa} z + \sigma_z \cdot h\right) \\
& + \frac{m}{2}|\sigma_z|^2 \frac{d^2\tilde{\upsilon}}{dz^2}(z) + \left(\frac{m\theta}{2}\right)|h - r|^2
\end{aligned}
\tag{36}
$$

subject to

$$
r = \sum_{j=1}^{n} \pi^j \eta^j(z), \ \ \pi \in \Pi \tag{37}
$$

First-order conditions for consumption are

$$
\frac{\delta}{c^*} = \frac{1}{k},
$$

which imply that $c^* = \delta k$, an implication that follows from the unitary elasticity of intertemporal substitution. First-order conditions for a and h are

$$
\pi(z) + h^* - a^* = 0 \tag{38a}
$$

$$
a^* + \theta(h^* - r^*) + \frac{d\tilde{\upsilon}}{dz}(z)\sigma_z = 0. \tag{38b}
$$

These two equations determine a^* and $h^* - r^*$ as a function of $\pi(z)$ and the value function $\tilde{\upsilon}$. We determine r^* as a function of h^* by solving:

$$
\min_r \frac{\theta}{2}|h - r|^2.
$$

subject to (37). Taken together, these determine (a^*, h^*, r^*). We can appeal to arguments like those of Hansen and Sargent (2008, ch. 7) to justify stacking first-order conditions as a way to collect equilibrium conditions for the pertinent two-person zero-sum game.[23]

[23]If we were to use a timing protocol that allows the maximizing player to take account of the impact of its decisions on the minimizing agent, we would obtain the same equilibrium decision rules described in the text.

8.2 Competitive equilibrium prices

We now impose $\log C = Y$ as an equilibrium condition. We show here that the drift distortion η^* that emerges from the robust planner's problem of section 5 determines prices that a competitive equilibrium awards for bearing model uncertainty. To compute a vector $\pi(x)$ of competitive equilibrium risk prices, we find a robust planner's marginal valuations of exposures to the W shocks. We decompose that price vector into separate compensations for bearing *risk* and for accepting *model uncertainty*.

Noting from the robust planning problem that the shock exposure vectors for $\log K$ and Y must coincide implies

$$a^* = (.01)\sigma_y.$$

From (38b) and the solution for r^*,

$$h^* = \eta^*(z),$$

where η^* is the worst-case drift from the robust planning problem provided that we show that $\tilde{\upsilon} = \upsilon$, where υ is the value function for the robust planning problem. Thus, from (38a), $\pi = \pi^*$, where

$$\pi^*(z) = (.01)\sigma_y - \eta^*(z). \tag{39}$$

Similarly, in the problem for a representative investor within a competitive equilibrium, the drifts for $\log K$ and Y coincide:

$$-\delta + \iota(z) + [(.01)\sigma_y - \eta^*(z)] \cdot a^* - \frac{.0001}{2}\sigma_y \cdot \sigma_y = (.01)(\hat{\alpha}_y + \hat{\beta} z),$$

so that $\iota = \iota^*$, where

$$\iota^*(z) = \delta + .01(\hat{\alpha}_y + \hat{\beta} z) + .01\sigma_y \cdot \eta^*(z) - \frac{.0001}{2}\sigma_y \cdot \sigma_y. \tag{40}$$

We use these formulas for equilibrium prices to construct a solution to the HJB equation of a representative investor in a competitive equilibrium by letting $\tilde{\upsilon} = \upsilon$.

8.3 Local uncertainty prices

The equilibrium stochastic discount factor process for our robust representative investor economy is

$$d\log S_t = -\delta dt - .01\left(\hat{\alpha}_y + \hat{\beta} Z_t\right) dt - .01\sigma_y \cdot dW_t + H_t^* \cdot dW_t - \frac{1}{2}|H_t^*|^2 dt. \qquad (41)$$

The components of the vector $\pi^*(Z_t)$ given by (39) equal minus the local exposures to the Brownian shocks. These are usually interpreted as local "risk prices," but we shall reinterpret them. Motivated by the decomposition

$$\textbf{minus stochastic discount factor exposure} = \underbrace{.01\sigma_y}_{\textbf{risk price}} \quad \underbrace{-H_t^*}_{\textbf{uncertainty price}},$$

we prefer to think of $.01\sigma_y$ as risk prices induced by the curvature of log utility and $-H_t^*$ as "uncertainty" prices induced by a representative investor's doubts about the baseline model. Here H_t^* is state dependent.

8.4 Uncertainty prices over alternative investment horizons

In the context of our quantitative models, we now report the shock-price elasticities that Borovička et al. (2014) showed are horizon-dependent uncertainty prices of risk exposures. Shock price elasticities describe the dependence of logarithms of expected returns on an investment horizon. The logarithm of the expected return from a consumption payoff at date t consists of two terms:

$$\log E\left(\frac{C_t}{C_0}\middle| X_0 = x\right) - \log E\left[S_t\left(\frac{Y_t}{Y_0}\right)\middle| X_0 = x\right]. \qquad (42)$$

where $\log C_t = Y_t$. The first term is the expected payoff and the second is the cost of purchasing that payoff. Notice that in our example, since imposed a unitary elasticity of substitution:

$$S_t\left(\frac{C_t}{C_0}\right) = M_t^{H^*}$$

so the second term features the martingale computed to implement robustness.

To compute an elasiticity, we change locally the exposure of consumption to the underlying Brownian motion compute the consequences for the expected return. From a mathematical perspective, an important inputs into this calculation are Malliavin derivatives. These derivatives measure how a shock at given date effects the consumption and the stochastic discount factor processes. Both S_t and C_t depend on Brownian motion between dates zero and t. We are particularly interested in the impact of date t shock on S_t and C_t. Computing the derivative of the logarithm of the expected return given in (42) results in

$$\frac{E\left[\mathcal{D}_t C_t | \mathcal{F}_0\right]}{E\left[C_t | \mathcal{F}_0\right]} - E\left[\mathcal{D}_t M_t^{H^*} | \mathcal{F}_0\right]$$

where $\mathcal{D}_t C_t$ and $M_t^{H^*}$ denote the two dimensional vectors of Mallianvin derivatives (with respect to the two dimensional Brownian increment at date t for consumption and the worst-case martingale.

Using the convenient formula familiar from other forms of differentiation

$$\mathcal{D}_t C_t = C_t \left(\mathcal{D}_t \log C_t\right).$$

The Mallivian derivative of $\log C_t = Y_t$ is the vector $.01\sigma_y$ or the exposure vector $\log C_t$ to the Brownian increment dW_t:

$$\mathcal{D}_t C_t = .01 C_t \sigma_y,$$

and thus

$$\frac{E\left[\mathcal{D}_t C_t | \mathcal{F}_0\right]}{E\left[C_t | \mathcal{F}_0\right]} = .01\sigma_y$$

Similarly,

$$\mathcal{D}_t M_t^H = H_t^*.$$

Therefore the term structure of prices that interest us are given by

$$.01\sigma_y - E\left[M_t^{H^*} H_t^* | \mathcal{F}_0\right]$$

The first term is the familiar risk price for comsumption-based asset pricing. It is state independent and contributes a (small) term that is independent of the horizon. In contrast, for the second term the equilibrium drift distortion provides state dependent component and its expectation under the distorted probability measure gives a time and state dependent contribution to the term structure of uncertainty prices.[24]

[24] There are other horizon dependent elasticities that we could compute. For instance, we might look at the impact of a shock at date zero on C_t and $M_t^{H^*}$ and trace out the impact of changing the horizon but keeping the date of the shock fixed.

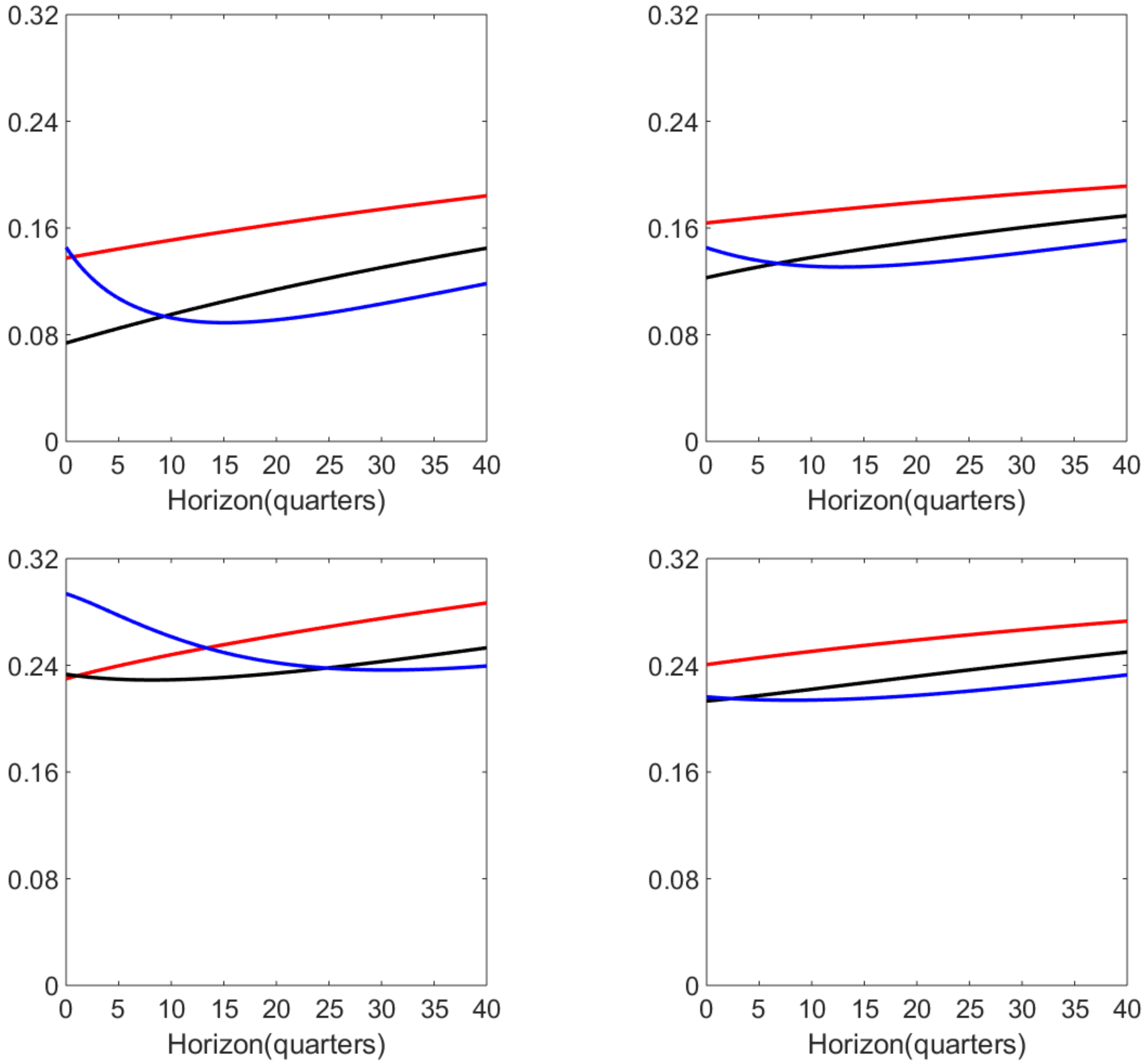

Figure 7: Shock price elasticities for alternative horizons and deciles for the specification with $(\alpha_z, \beta, \kappa)$ uncertainty. Left panels: larger benchmark entropy (*ex ante* $\mathsf{q} = .1$, *ex post* $\mathsf{q} = .1394$). Right panels: smaller benchmark entropy (*ex ante* $\mathsf{q} = .05$, *ex post* $\mathsf{q} = .0681$). Top panels: first shock. Bottom panels: second shock. **Black**: median of the Z stationary distribution; **red**: .1 decile; and **blue**: .9 decile.

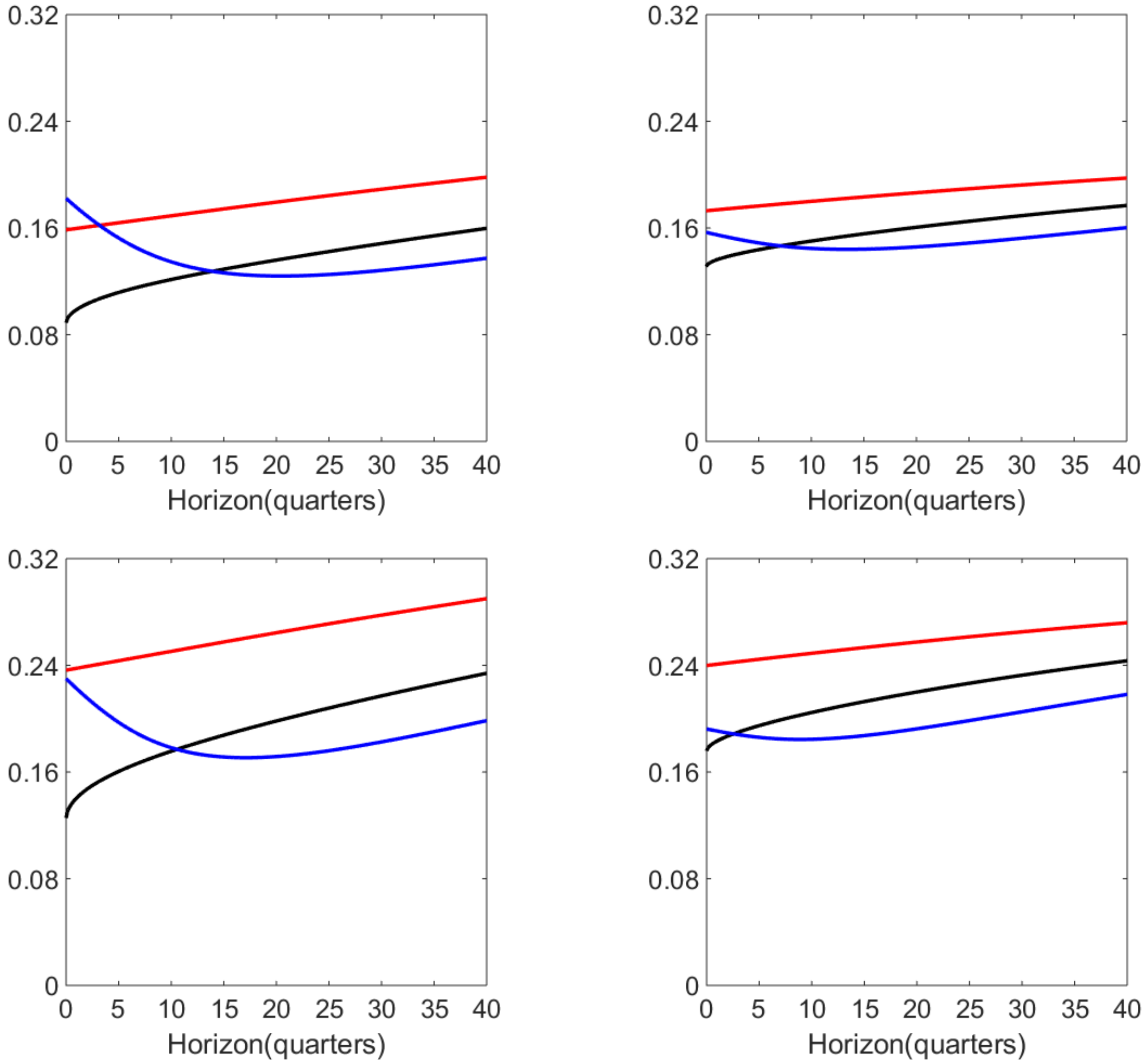

Figure 8: Shock price elasticities for alternative horizons and deciles for the specification with (β, κ) uncertainty. Left panels: larger baseline entropy ($\mathsf{q} = .1$). Right panels: smaller benchmark entropy ($\mathsf{q} = .05$). Top panels: first shock. Bottom panels: second shock. **Black**: median of the Z stationary distribution **red**: .1 decile; and **blue**: .9 decile.

Notice that although the price elasticity is initially smaller for the median specification of z than for the .9 quantile, this inequality is eventually reversed as the horizon is increased. (The blue curve and black curve cross.) The uncertainty price for positive z initially diminishes because the probability measure implied by the martingale has reduced persistence for the positive states. Under this probability, the growth rate state variable is expected to spends less time positive region. This is reflected in the smaller prices for the .9 quantile than for the median over longer investment horizons. The endogenous nonlinearity in valuation has the uncertainty prices larger for negative values of z than positive for longer investment horizons, but not necessarily for very short ones.

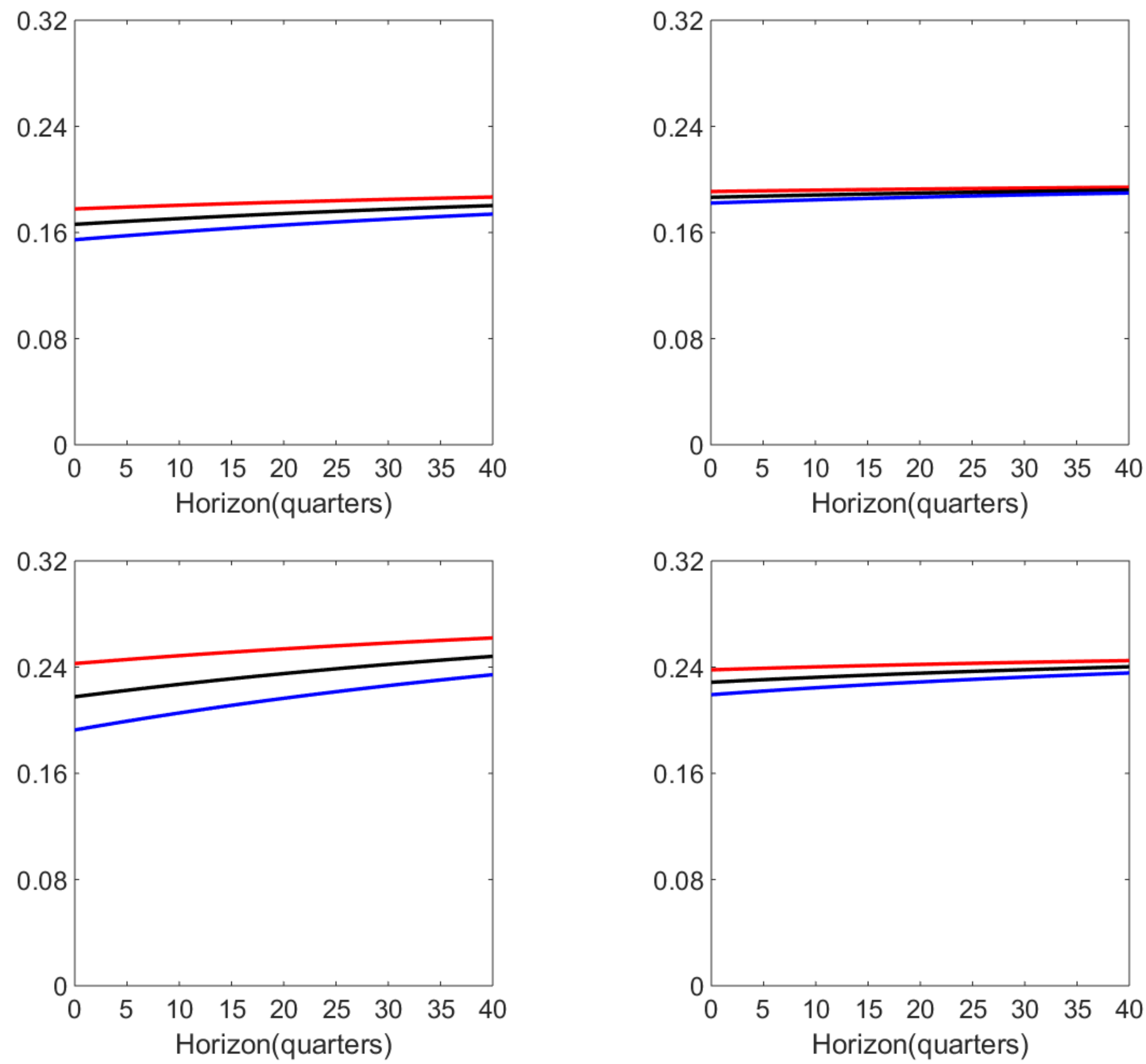

Figure 9: Shock price elasticities for alternative horizons and deciles for the specification with (β, κ) uncertainty. Left panels: larger constraint set ($\underline{\kappa} = .0071$). Right panels: smaller constraint set ($\underline{\kappa} = .0114$). Top panels: first shock. Bottom panels: second shock. **Black**: median of the Z stationary distribution **red**: .1 decile; and **blue**: .9 decile.

We have structured our quantitative examples to investigate a particular mechanism for generating fluctuations in uncertainty prices from statistically plausible amounts of uncertainty. We infer parameters of the baseline model for these examples solely from time series of macroeconomic quantities, thus completely ignoring asset prices during calibration. As a consequence, we do not expect to track closely the high frequency movements in financial markets. By limiting our empirical inputs, we respect concerns that Hansen (2007) and Chen et al. (2015) expressed about using asset market data to calibrate macro-finance models that assign a special role to investors' beliefs about the future asset prices.[25]

9 Concluding remarks

This paper formulates and applies a tractable model of the effects on equilibrium prices of exposures to macroeconomic uncertainties. Our analysis uses models' consequences for discounted expected utilities to quantify investors' concerns about model misspecification. We characterize the effects of concerns about misspecification of a baseline stochastic process for individual consumption as shadow prices for a planner's problem that supports competitive equilibrium prices.

To illustrate our approach, we have focused on the growth rate uncertainty featured in the "long-run risk" literature initiated by Bansal and Yaron (2004). Other applications seem natural. For example, the tools developed here could shed light on a recent public debate between two groups of macroeconomists, one prophesizing secular stagnation because of technology growth slowdowns, the other dismissing those pessimistic forecasts. The tools that we describe can be used, first, to quantify how challenging it is to infer persistent changes in growth rates, and, second, to guide macroeconomic policy design in light of available empirical evidence.

Specifically, we have produced a model of the log stochastic discount factor whose uncertainty prices reflect a robust planner's worst-case drift distortions η^*. We have argued that these drift distortions should be interpreted as prices of model uncertainty. The dependency of these uncertainty prices η^* on the growth state z is shaped partly by the

[25] Hansen (2007) and Chen et al. (2015) describe situations in which it is the behavior of rates of return on assets that, through the cross-equation restrictions, lead an econometrician to make inferences about the behavior of macroeconomic quantities like consumption that are much more confident than can be made from the quantity data alone. That opens questions about how the investors who are supposedly putting those cross-equation restrictions into returns came to know those quantity processes before *they* observed returns.

alternative parametric models that the decision maker entertains. In this way, our theory of state dependence in uncertainty prices is all about how our robust investor responds to the presence of the alternative parametric models among a huge set of unspecified alternative models that also concern him.

It is worthwhile comparing this paper's way of inducing time varying prices of risk with three other macro/finance models that also get them. Campbell and Cochrane (1999) proceed in the standard rational expectations single-known-probability-model tradition and so exclude any fears of model misspecification from the mind of their representative investor. They construct a history-dependent utility function in which the history of consumption expresses an externality. This history dependence makes the investor's local risk aversion depend in a countercyclical way on the economy's growth state. Ang and Piazzesi (2003) use an exponential quadratic stochastic discount factor in a no-arbitrage statistical model and explore links between the term structure of interest rates and other macroeconomic variables. Their approach allows movements in risk prices to be consistent with historical evidence without specifying an explicit general equilibrium model. A third approach introduces stochastic volatility into the macroeconomy by positing that the volatilities of shocks driving consumption growth are themselves stochastic processes. A stochastic volatility model induces time variation in risk prices via exogenous movements in the conditional volatilities impinging on macroeconomic variables.

In Hansen and Sargent (2010), countercyclical uncertainty prices are driven by a representative investor's robust model averaging. The investor carries along two difficult-to-distinguish models of consumption growth, one asserting i.i.d. log consumption growth, the other asserting that the growth in log consumption is a process with a slowly moving conditional mean. The investor uses observations on consumption growth to update a Bayesian prior over these two models, starting from an initial prior probability of .5. The prior wanders over a post WWII sample period, but ends where it started. Each period, the Hansen and Sargent representative investor expresses his specification distrust by pessimistically exponentially twisting a posterior over the two baseline models. That leads the investor to interpret good news as temporary and bad news as persistent, causing him to put countercyclical uncertainty components into equilibrium "risk" prices. Learning occurs in their analysis because the parameterized set of baseline models are time invariant and hence learnable. The decision makers in their analysis confront dynamic consistency by playing a game against future versions of themselves.

In this paper, we propose a different way to make uncertainty prices vary in a qualitatively similar way. We exclude learning and instead consider alternative models with parameters whose future variations cannot be inferred from historical data. These time-varying parameter models differ from the decision maker's baseline model, a fixed parameter model whose parameters can be well estimated from historical data. The alternative models include ones that allow parameters persistently to deviate from those of the baseline model in statistically subtle and time-varying ways. In addition to this particular class of alternative models, the decision maker also includes other statistical specifications in the set of models that concern him. The robust planner's worst-case model responds to these forms of model ambiguity partly by having more persistence in bad states and less persistence in good states. Adverse shifts in the shock distribution that drive up the absolute magnitudes of uncertainty prices were also present in some of our earlier work (for example, see Hansen et al. (1999) and Anderson et al. (2003)). In this paper, we induce state dependence in uncertainty prices in a different way, namely, by specifying a set of alternative models to capture concerns about the baseline model's specification of persistence in consumption growth.

Relative to rational expectations models, models of robustness and ambiguity aversion bring new parameters. In this paper, we extend work by Anderson et al. (2003) that calibrated those additional parameters by exploiting connections between models of statistical model discrimination and our way of formulating robustness. We build on mathematical formulations of Newman and Stuck (1979), Petersen et al. (2000), and Hansen et al. (2006). We pose an *ex ante* robustness problem that pins down a robustness penalty parameter θ by linking it to an asymptotic measure of statistical discrimination between models. This asymptotic measure allows us to quantify a half-life for reducing the mistakes in selecting between competing models based on historical evidence. A large statistical discrimination rate implies a short half-life for reducing discrimination mistake probabilities. Anderson et al. (2003) and Hansen (2007) had studied the connection between conditional discrimination rates and uncertainty prices that clear security markets. By following Newman and Stuck (1979) and studying asymptotic discrimination rates, we link statistical discrimination half-lives to calibrated equilibrium uncertainty prices.

Appendices

A Operationalizing Chernoff entropy

In this appendix we show how to compute entropies for parametric models of the form (24). Because the H's associated with them take the form

$$H_t = \eta(Z_t),$$

these alternative models are Markovian.

Given the Markov structure of both models, we compute Chernoff entropy by using an eigenvalue approach of Donsker and Varadhan (1976) and Newman and Stuck (1979). We start by computing the drift of $\left(M_t^H\right)^{\mathsf{s}} g(Z_t)$ for $0 \leqslant \mathsf{s} \leqslant 1$ at $t = 0$:

$$\begin{aligned}[\mathbb{G}(\mathsf{s})g](z) \doteq & \frac{(-\mathsf{s} + \mathsf{s}^2)}{2}|\eta(z)|^2 g(z) + \mathsf{s}g(z)'\sigma \cdot \eta(z) \\ & - g'(z)\kappa z + \frac{g''(z)}{2}|\sigma|^2,\end{aligned}$$

where $[\mathbb{G}(\mathsf{s})g](x)$ is the drift given that $Z_0 = z$. Next we solve the eigenvalue problem

$$[\mathbb{G}(\mathsf{s})]e(z,\mathsf{s}) = -\rho(\mathsf{s})e(z,\mathsf{s}),$$

whose eigenfunction $e(z,\mathsf{s})$ is the exponential of a quadratic function of z. We compute Chernoff entropy numerically by solving:

$$\chi(M^H) = \max_{\mathsf{s}\in[0,1]} \rho(\mathsf{s}).$$

To deduce a corresponding equation for $\log e$, notice that

$$(\log e)'(z) = \frac{e'(z)}{e(z)}$$

and

$$(\log e)''(z) = \frac{e''(z)}{e(z)} - \left[\frac{e'(z)}{e(z)}\right]^2.$$

For a positive g

$$\frac{[\mathbb{G}(\mathsf{s})g](z)}{g(z)} \doteq \frac{(-\mathsf{s}+\mathsf{s}^2)}{2}|\eta(z)|^2 + \mathsf{s}(\log g)'(z)\sigma \cdot \eta(z) - (\log g)'(z)\hat{\kappa} z$$
$$+ \frac{\log g''(z)}{2}|\sigma|^2 + \frac{[\log g'(z)]^2}{2}|\sigma|^2. \tag{43}$$

Using formula (43), define

$$\left[\overline{\mathbb{G}}(\mathsf{s}) \log g\right](z) = \frac{[\mathbb{G}(\mathsf{s})g](z)}{g(z)}.$$

Then we can solve

$$\left[\overline{\mathbb{G}}(\mathsf{s}) \log e\right](z, \mathsf{s}) = -\rho(\mathsf{s})$$

for $\log e$ to compute the positive eigenfunction e.

These calculations allow us numerically to compute Chernoff entropies:

$$d \log Y_t = (.01)\left(\alpha_y + \beta Z_t\right) dt + (.01)\sigma_y \cdot d\widetilde{W}_t$$
$$dZ_t = \alpha_z dt - \kappa Z_t dt + \sigma_z \cdot d\widetilde{W}_t.$$

1. Input values of κ, β, α_y, and α_z. Construct the implied $h(z) = \eta_1 z + \eta_0$ by solving;

$$\begin{bmatrix} \alpha_y - \hat{\alpha}_y \\ \alpha_z - \hat{\alpha}_z \end{bmatrix} = \sigma \eta_0$$
$$\begin{bmatrix} \beta - \hat{\beta} \\ \hat{\kappa} - \kappa \end{bmatrix} = \sigma \eta_1$$

   for η_1 and η_0.

2. For a given s, construct ζ_0, ζ_1, ζ_2, $\check{\kappa}$, $\check{\beta}$, $\check{\phi}$, and $\check{\mu}$ from:

$$(-\mathsf{s}+\mathsf{s}^2)|\eta_0 + \eta_1 z|^2 = -\left(\zeta_0 + 2\zeta_1 z + \zeta_2 z^2\right)$$

$$\check{\kappa} = (1-\mathsf{s})\hat{\kappa} + \mathsf{s}\kappa$$
$$\check{\beta} = (1-\mathsf{s})\hat{\beta} + \mathsf{s}\beta$$

$$\check{\mu}_y = (1 - \mathsf{s})\hat{\mu}_y + \mathsf{s}\mu_y$$
$$\check{\mu}_z = (1 - \mathsf{s})\hat{\mu}_z + \mathsf{s}\mu_z$$

3. Solve

$$-\rho(\mathsf{s}) = -\frac{1}{2}\left(\zeta_0 + 2\zeta_1 x + \zeta_2 z^2\right) + (\check{\mu}_z - z\check{\kappa})(\log e)'(z) + \frac{(\log e)''(z)}{2}|\sigma_z|^2 + \frac{[(\log e)'(z)]^2}{2}|\sigma_z|^2$$

where $\log e(z, \mathsf{s}) = \lambda_1 z + \frac{1}{2}\lambda_2 z^2$. Thus,

$$\lambda_2 = \frac{\check{\kappa} - \sqrt{(\check{\kappa})^2 + \zeta_2|\sigma_z|^2}}{|\sigma_z|^2}.$$

Given λ_2, λ_1 solves

$$-\zeta_1 - \check{\kappa}\lambda_1 + \lambda_1\lambda_2|\sigma_z|^2 + \check{\mu}_z\lambda_2 = 0$$

or

$$\lambda_1 = \frac{\zeta_1 - \check{\mu}_z\lambda_2}{\lambda_2|\sigma_z|^2 - \check{\kappa}} = -\frac{\zeta_1 - \check{\mu}_z\lambda_2}{\sqrt{(\check{\kappa})^2 + \zeta_2|\sigma_z|^2}}.$$

Finally,

$$\rho(\mathsf{s}) = \frac{1}{2}\zeta_0 - \frac{1}{2}|\sigma_z|^2\lambda_2 - \frac{1}{2}|\sigma_z|^2(\lambda_1)^2 - \check{\mu}_z\lambda_1.$$

4. Repeat for alternative s's and maximize $\rho(\mathsf{s})$ as a function of s.

B Statistical calibration

We fit a trivariate VAR with the following variables:

$$\log Y_{t+1} - \log Y_t$$
$$\log G_{t+1} - \log Y_{t+1}$$
$$\log D_{t+1} - \log Y_{t+1}$$

where G_t is the sum of corporate profits and proprietors' income and D_t is personal income.

Provided that the VAR has stable coefficients, this is a co-integrated system. All three time series have stationary increments, but there one common martingale process. The shock to this process is identified as the only one with long-term consequences. We set $\hat{\alpha}_z = 0$ and $\hat{\beta}_y = 1$. For the remaining parameters we:

i) fit a VAR with a constant and four lags of the first variable and five of the other two;

ii) compute the implied mean for for $\log Y_{t+1} - \log Y_t$ and set this to $\hat{\alpha}_y$;

iii) compute the state dependent component of the expected long-term growth rate by calculating:

$$\log Y_t^p = \lim_{j\to\infty} E\left(\log Y_{t+j} - \log Y_t - j\hat{\alpha}_y | \mathcal{F}_t\right)$$

implied by the VAR estimates, to compare to the counterpart calculation in the continuous-time model:

$$Z_t^p = \lim_{j\to\infty} E\left(\log Y_{t+j} - \log Y_t - j\hat{\alpha}|Z_t\right) = \frac{1}{\hat{\kappa}} Z_t.$$

iv) compute the implied autoregressive coefficient for $\{\log Y_t^p\}$ in the discrete-time specification using the VAR parameter estimates and equate this coefficient to $1 - \hat{\kappa}$.

v) compute the VAR implied covariance matrix for the one-step-ahead forecast error for $\{\log Y^p\}$, the direct shock to consumption and equate this to

$$\begin{bmatrix} (\sigma_y)' \\ \frac{1}{\hat{\kappa}}(\sigma_z)' \end{bmatrix} \begin{bmatrix} (\sigma_y) & \frac{1}{\hat{\kappa}}(\sigma_z) \end{bmatrix}$$

where we achieve identification of σ_z and σ_y by imposing a zero restriction on the second entry of σ_y and positive signs on the first coefficient of σ_y and on the second coefficient of σ_z.

C Solving the ODE's

The value function is approximately linear in the state variable for large $|z|$. This gives a good boundary Neumann boundary condition to use in an approximation in which z is restricted to a compact interval that includes $z = 0$. To deduce this boundary restriction solve:

$$
\begin{aligned}
(-\delta - \kappa)\underline{\nu} + .01\beta &= 0 \\
\underline{\nu}\frac{\partial \varepsilon}{\partial \beta}(\alpha_z, \beta, \kappa) + .01\frac{\partial \varepsilon}{\partial \kappa}(\alpha_z, \beta, \kappa) &= 0 \\
\frac{\partial \varepsilon}{\partial \alpha_z}(\alpha_z, \beta, \kappa) &= 0 \\
\varepsilon(\alpha_z, \beta, \kappa) - \frac{1}{2}\mathsf{q}^2 &= 0
\end{aligned}
$$

for $(\underline{\nu}, \alpha_z, \beta, \kappa)$. The first equation is the derivative of the value function for constant coefficients, putting aside the minimization. The next two equations are large z approximations to the first-order conditions (32). The last equation targets the boundary of the feasible region for the parameter values $(\alpha_z, \beta, \kappa)$. Notice that the last two equations imply that the α_z minimizes the quadratic function and is a zero of it. Thus the discriminate of this associated quadratic equation is zero. This gives a function of κ and β alone along with a formula for $\alpha_z = f(\beta, \kappa)$. There will be multiple ways to make the discriminate zero, and we need to pick the one corresponding to low κ and high β for $z << 0$ in the final solution. We proceed analogously for $z >> 0$ by selecting the high κ and low β solution.

Consider now the special case in which $\alpha_z = \hat{\alpha}_z$. We proceed as before except that we drop the third equation.

D Statistician value function

Recall that we studied a two-agent game: a statistician that chooses a baseline model, and a decision maker who doesn't fully trust the baseline model. In this appendix we provide some formulas for the statistician value function.

The HJB equation for the statistician is:

$$
\begin{aligned}
0 = &-\delta\varsigma(z, \delta, \ell) + .01(\hat{\alpha}_y + \hat{\beta} z) + \frac{d\varsigma}{dz}(z, \delta, \ell)\left(\hat{\alpha}_z - \hat{\kappa} z\right) + \frac{1}{2}|\sigma_z|^2 \frac{d^2\varsigma}{dz^2}(z, \delta, \ell) \\
&- \frac{1}{2\ell}\begin{bmatrix} .01 & \frac{d\varsigma}{dz}(z, \delta, \ell) \end{bmatrix} \sigma\sigma' \begin{bmatrix} .01 \\ \frac{d\varsigma}{dz}(z, \delta, \ell) \end{bmatrix} - \frac{\ell \mathsf{b}^2 z^2}{2}
\end{aligned}
$$

Guess that that the value function is quadratic

$$\varsigma(z, \delta, \ell) = -\frac{1}{2}\left[\varsigma_2(\delta, \ell)z^2 + 2\varsigma_1(\delta, \ell)z + \varsigma_0(\delta, \ell)\right]$$

which implicitly depends on ℓ. The implied worst-case benchmark distortion is

$$\iota^*(z) = -\frac{1}{\ell}\left[.01\sigma_y - \sigma_z\varsigma_2(\delta, \ell)z - \sigma_z\varsigma_1(\delta, \ell)\right].$$

We solve for ς_2, ς_1, and ς_0 by matching the coefficients for z^2, z and the constant terms, respectively. Solving first for ς_2 collection the z^2 terms. This results in the quadratic equation of the form $A(\varsigma_2)^2 + B\varsigma_2 + C = 0$ where

$$\begin{aligned} A &= -\frac{1}{2\ell}\left|\sigma_z\right|^2 \\ B &= \hat{\kappa} + \frac{\delta}{2} \\ C &= -\frac{\ell}{2}\mathsf{b}^2. \end{aligned}$$

Then

$$\begin{aligned} \varsigma_2(\delta, \ell) &= \frac{-B + \sqrt{B^2 - 4AC}}{2A} \\ &= \ell\left[\frac{\delta + 2\hat{\kappa} - \sqrt{(\delta + 2\hat{\kappa})^2 - 4\left|\sigma_z\right|^2\mathsf{b}^2}}{2\left|\sigma_z\right|^2}\right] \end{aligned}$$

$$\begin{aligned} \varsigma_1(\delta, \ell) =& 2\left[\frac{-.01\hat{\beta} + \hat{\alpha}_z\varsigma_2(\delta, \ell) - \frac{.01}{\ell}\left(\sigma_y \cdot \sigma_z\right)\varsigma_2(\delta, \ell)}{\delta + \sqrt{(\delta + 2\hat{\kappa})^2 - 4\left|\sigma_z\right|^2\mathsf{b}^2}}\right] \\ \varsigma_0(\delta, \ell) =& \frac{1}{\delta}\left[-.02\hat{\alpha}_y + 2\hat{\alpha}_z\varsigma_1(\delta, \ell) + \left|\sigma_z\right|^2\varsigma_2(\delta, \ell) + \frac{1}{\ell}\left|.01\sigma_y - \sigma_z\varsigma_1(\delta, \ell)\right|^2\right] \end{aligned}$$

To solve the undiscounted problem, we now compute three limits:

$$\varsigma_2(0, \ell) = \lim_{\delta \downarrow 0}\varsigma_2(\delta, \ell) = \ell\left[\frac{\hat{\kappa} - \sqrt{(\hat{\kappa})^2 - \left|\sigma_z\right|^2\mathsf{b}^2}}{\left|\sigma_z\right|^2}\right]$$

$$\varsigma_1(0,\ell) = \lim_{\delta\downarrow 0} \varsigma_1(\delta) = \left[\frac{-.01\hat{\beta} + \hat{\alpha}_z \varsigma_2(0,\ell) - \frac{.01}{\ell}\left(\sigma_y \cdot \sigma_z\right) \varsigma_2(0,\ell)}{\sqrt{\left(\hat{\kappa}\right)^2 - \left|\sigma_z\right|^2 \mathsf{b}^2}} \right]$$

$$\lim_{\delta\downarrow 0} \delta \varsigma_0(\delta,\ell) = \left[-.02\hat{\alpha}_y + 2\hat{\alpha}_z \varsigma_1(0,\ell) + \left|\sigma_z\right|^2 \varsigma_2(0,\ell) + \frac{1}{\ell} \left|.01\sigma_y - \sigma_z \varsigma_1(0,\ell)\right|^2 \right].$$

We compute ℓ by solving:

$$\max_{\ell} \lim_{\delta\downarrow 0} -\delta\varsigma_0(\delta,\ell).$$

References

Anderson, Evan W., Lars Peter Hansen, and Thomas J. Sargent. 1998. Risk and Robustness in Equilibrium. Available on webpages.

———. 2003. A Quartet of Semigroups for Model Specification, Robustness, Prices of Risk, and Model Detection. *Journal of the European Economic Association* 1 (1):68–123.

Ang, Andrew and Monika Piazzesi. 2003. A No-Arbitrage Vector Autoregression of the Term Structure Dynamics with Macroeconomic and Latent Variables. *Journal of Monetary Economics* 50:745–787.

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

Berger, James O. 1994. An Overview of Robust Bayesian Analysis (with discussion). *Test* 3 (1):5–124.

Bhandari, Anmol. 2014. Doubts, Asymmetries, and Insurance. University of Minnesota.

Borovička, Jaroslav, Lars Peter Hansen, and José A. Scheinkman. 2014. Shock Elasticities and Impulse Response Functions. *Mathematics and Financial Economics* 8 (4):333–354.

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

Campbell, John Y. and John Cochrane. 1999. Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior. *Journal of Political Economy* 107 (2):205–251.

Chamberlain, Gary. 2000. Econometric Applications of Maxmin Expected Utility. *Journal of Applied Econometrics* 15 (6):625–644.

Chen, Hui, Winston Wei Dou, and Leonid Kogan. 2015. Measuring the 'Dark Matter' in Asset Pricing Models. MIT Sloan School of Management.

Chen, Zengjing and Larry Epstein. 2002. Ambiguity, Risk, and Asset Returns in Continuous Time. *Econometrica* 70:1403–1443.

Chernoff, Herman. 1952. A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on the Sum of Observations. *Annals of Mathematical Statistics* 23 (4):pp. 493–507.

Donsker, Monroe E. and S. R. Srinivasa Varadhan. 1976. On the Principal Eigenvalue of Second-Order Elliptic Differential Equations. *Communications in Pure and Applied Mathematics* 29:595–621.

Epstein, Larry G. and Martin Schneider. 2003. Recursive multiple-priors. *Journal of Economic Theory* 113 (1):1–31.

Fudenberg, Drew and David K. Levine. 2009. Self-confirming Equilibrium and the Lucas Critique. *Journal of Economic Theory* 144 (6):2354–2371.

Good, I. J. 1952. Rational Decisions. *Journal of the Royal Statistical Society. Series B (Methodological)* 14 (1):pp. 107–114.

Hansen, L. P., J. C. Heaton, and N. Li. 2008. Consumption Strikes Back?: Measuring Long Run Risk. *Journal of Political Economy* .

Hansen, Lars P. and Thomas J. Sargent. 2016. Sets of Models and Prices of Uncertainty. NBER Working Papers 22000, National Bureau of Economic Research, Inc.

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

Hansen, Lars Peter and Massimo Marinacci. 2016. Ambiguity Aversion and Model Misspecification: An Economic Perspective. *Statistical Science* Forthcoming.

Hansen, Lars Peter and Thomas Sargent. 2010. Fragile beliefs and the price of uncertainty. *Quantitative Economics* 1 (1):129–162.

Hansen, Lars Peter and Thomas J. Sargent. 2001. Robust Control and Model Uncertainty. *American Economic Review* 91 (2):60–66.

———. 2008. *Robustness.* Princeton, New Jersey: Princeton University Press.

Hansen, Lars Peter, Thomas J. Sargent, and Jr. Tallarini, Thomas D. 1999. Robust Permanent Income and Pricing. *The Review of Economic Studies* 66 (4):873–907.

Hansen, Lars Peter, Thomas J. Sargent, Gauhar A. Turmuhambetova, and Noah Williams. 2006. Robust Control and Model Misspecification. *Journal of Economic Theory* 128 (1):45–90.

Ilut, Cosmin L. and Martin Schneider. 2014. Ambiguous Business Cycles. *American Economic Review* 104 (8):2368–99.

James, Matthew R. 1992. Asymptotic analysis of nonlinear stochastic risk-sensitive control and differential games. *Mathematics of Control, Signals and Systems* 5 (4):401–417.

Karantounias, Anastasios G. 2013. Managing pessimistic expectations and fiscal policy. *Theoretical Economics* 8 (1).

Lucas, Jr, Robert E. 1978. Asset Prices in an Exchange Economy. *Econometrica* 46 (6):1429–45.

Maccheroni, Fabio, Massimo Marinacci, and Aldo Rustichini. 2006a. Ambiguity Aversion, Robustness, and the Variational Representation of Preferences. *Econometrica* 74 (6):1447–1498.

———. 2006b. Dynamic Variational Preferences. *Journal of Economic Theory* 128:4–44.

Mehra, Rajnish and Edward C. Prescott. 1985. The equity premium: A puzzle. *Journal of Monetary Economics* 15 (2):145–161.

Miao, Jianjun and Alejandro Rivera. 2016. Robust Contracts in Continuous Time. *Econometrica* 84 (4):1405–1440.

Newman, C. M. and B. W. Stuck. 1979. Chernoff Bounds for Discriminating between Two Markov Processes. *Stochastics* 2 (1-4):139–153.

Petersen, I.R., M.R. James, and P. Dupuis. 2000. Minimax optimal control of stochastic uncertain systems with relative entropy constraints. *Automatic Control, IEEE Transactions on* 45 (3):398–412.

Sargent, Thomas J. 1999. *The Conquest of American Inflation.* Princeton, New Jersey: Princeton University Press.

Watson, James and Chris Holmes. 2016. Approximate Models and Robust Decisons. *Statistical Science* Forthcoming.