A Quartet of Semigroups for Model Specification, Robustness, Prices of Risk, and Model Detection

Evan W. Anderson

University of North Carolina

Lars Peter Hansen

University of Chicago

Thomas J. Sargent

Stanford University and Hoover Institution

ABSTRACT

A representative agent fears that his model, a continuous time Markov process with jump and diffusion components, is misspecified and therefore uses robust control theory to make decisions. Under the decision maker's approximating model, that cautious behavior puts adjustments for model misspecification into market prices for risk factors. We use a statistical theory of detection to quantify how much model misspecification the decision maker should fear, given his historical data record. A semigroup is a collection of objects connected by something like the law of iterated expectations. The law of iterated expectations defines the semigroup for a Markov process, while similar laws define other semigroups. Related semigroups describe (1) an approximating model; (2) a model misspecification adjustment to the continuation value in the decision maker's Bellman equation; (3) asset prices; and (4) the behavior of the model detection statistics that we use to calibrate how much robustness the decision maker prefers. Semigroups 2, 3, and 4 establish a tight link between the market price of uncertainty and a bound on the error in statistically discriminating between an approximating and a worst case model.

Keywords: Approximation, misspecification, robustness, risk, uncertainty, statistical detection, pricing.

22 April 2003

1. Introduction

1.1. Rational expectations and model misspecification

A rational expectations econometrician or calibrator typically attributes no concern about specification error to agents even as he shuttles among alternative specifications.[1] Decision makers inside a rational expectations model know the model.[2] Their confidence contrasts with the attitudes of both econometricians and calibrators. Econometricians routinely use likelihood-based specification tests (information criteria or IC) to organize comparisons between models and empirical distributions. Less formally, calibrators sometimes justify their estimation procedures by saying that they regard their models as incorrect and unreliable guides to parameter selection if taken literally as likelihood functions. But the agents inside a calibrator's model do not share the model-builder's doubts about specification.

By equating agents' subjective probability distributions to the objective one implied by the model, the assumption of rational expectations precludes any concerns that agents should have about the model's specification. The empirical power of the rational expectations hypothesis comes from having decision makers' beliefs be *outcomes*, not *inputs*, of the model-building enterprise. A standard argument that justifies equating objective and subjective probability distributions is that agents would eventually detect any difference between them, and would adjust their subjective distributions accordingly. This argument implicitly gives agents an infinite history of observations, a point that is formalized by the literature on convergence of myopic learning algorithms to rational expectations equilibria of games and dynamic economies.[3]

Specification tests leave applied econometricians in doubt because they have too few observations to discriminate among alternative models. Econometricians with finite data sets thus face a model detection problem that builders of rational expectations models let agents sidestep by endowing them with infinite histories of observations "before time zero."

This paper is about models with agents whose data bases are finite, like econometricians and calibrators. Their limited data leave agents with model specification doubts that are

---

[1] For example, see the two papers about specification error in rational expectations models by Sims (1993) and Hansen and Sargent (1993).

[2] This assumption is so widely used that it rarely excites comment within macroeconomics. Kurz (1997) is an exception. The rational expectations critique of earlier dynamic models with adaptive expectations was that they implicitly contained two models, one for the econometrician and a worse one for the agents who are forecasting inside the model. See Jorgenson (1967) and Lucas (1976). Rational expectations modelling responded to this critique by attributing a common model to the econometrician and the agents within his model. Econometricians and agents can have different information sets, but they agree about the model (stochastic process).

[3] See Evans and Honkapohja (2001) and Fudenberg and Levine (1998).

quantitatively similar to those of econometricians and that make them value decision rules that perform well across a set of models. In particular, agents fear misspecifications of the state transition law that are sufficiently small that they are difficult to detect because they are obscured by random shocks that impinge on the dynamical system. Agents adjust decision rules to protect themselves against modelling errors, a precaution that puts model uncertainty premia into equilibrium security market prices. Because we work with Markov models, we can avail ourselves of a powerful tool called a semigroup.

1.2. Iterated laws and semigroups

The *law of iterated expectations* imposes consistency requirements that cause a collection of conditional expectations operators associated with a Markov process to form a mathematical object called a *semigroup*. The operators are indexed by the time that elapses between when the forecast is made and when the random variable being forecast is realized. This semigroup and its associated *generator* characterize the Markov process. Because we consider forecasting random variables that are functions of a Markov state, the current forecast depends only on the current value of the Markov state.[4]

The *law of iterated values* embodies analogous consistency requirements for a collection of economic values assigned to claims to payoffs that are functions of future values of a Markov state. The family of valuation operators indexed by the time that elapses between when the claims are valued and when their payoffs are realized forms another semigroup. Just as a Markov process is characterized by its semigroup, so prices of payoffs that are functions of a Markov state can be characterized by a semigroup. Hansen and Scheinkman (2002) exploited this insight. Here we extend their insight to other semigroups. In particular, we describe four semigroups: (1) one that describes a Markov process; (2) another that adjusts continuation values in a way that rewards decision rules that are robust to misspecification of the approximating model; (3) another that models the equilibrium pricing of securities with payoff dates in the future; and (4) another that governs statistics for discriminating between alternative Markov processes using a finite time series data record.[5] We show the close connections that bind these four semigroups.

---

[4] The semigroup formulation of Markov processes is common in the literature on applied probability. See Ethier and Kurtz (1986) for a general treatment of semigroups and Hansen and Scheinkman (1995) for their use in studying the identification of continuous-time Markov models.

[5] Here the operator is indexed by the time horizon of the available data. In effect there is a 'statistical detection operator' that measures the statistical value of information available to discriminate between two Markov processes.

1.3. Model detection errors and market prices of risk

In earlier work (Hansen, Sargent, and Tallarini (1999), henceforth denoted HST, and Hansen, Sargent, and Wang (2002), henceforth denoted HSW), we studied various discrete time asset pricing models in which decision makers' fear of model misspecification put model uncertainty premia into market prices of risk, thereby potentially helping to account for the equity premium. Transcending the detailed dynamics of our examples was a tight relationship between the market price of risk and the probability of distinguishing the representative decision maker's approximating model from a worst-case model that emerges as a byproduct of his cautious decision making procedure. Although we had offered only a heuristic explanation for that relationship, we nevertheless exploited it to help us calibrate the set of alternative models that the decision maker should plausibly seek robustness against. In the context of continuous time Markov models, this paper analytically establishes a precise link between the uncertainty component of risk prices and a *bound* on the probability of distinguishing the decision maker's approximating and worst case models. We also develop new ways of representing decision makers' concerns about model misspecification and their equilibrium consequences.

1.4. Related literature

In the context of a discrete-time, linear-quadratic permanent income model, HST considered model misspecifications measured by a single robustness parameter. HST showed how robust decision-making promotes behavior like that induced by risk aversion. They interpreted a preference for robustness as a decision maker's response to Knightian uncertainty and calculated how much concern about robustness would be required to put market prices of risk into empirically realistic regions. Our fourth semigroup, which describes model detection errors, provides a statistical method for judging whether the required concern about robustness is plausible.

HST and HSW allowed the robust decision maker to consider only a limited array of specification errors, namely, shifts in the conditional mean of shocks that are i.i.d. and normally distributed under an approximating model. In this paper, we consider more general approximating models and motivate the form of potential specification errors by using specification test statistics. We show that HST's perturbations to the approximating model emerge in linear-quadratic, Gaussian control problems as well as in a more general class of control problems in which the stochastic evolution of the state is a Markov diffusion process. However, we also show that misspecifications different from HST's must be entertained when the approximating model includes Markov jump components. As in

HST, our formulation of robustness allows us to reinterpret one of Epstein and Zin's (1989) recursions as reflecting a preference for robustness rather than aversion to risk.

As we explain in Hansen, Sargent, Turmuhambetova, and Williams (henceforth HSTW) (2002), the robust control theory described in section 5 is closely connected to the min-max expected utility or multiple priors model of Gilboa and Schmeidler (1989). A main theme of the present paper is to advocate a workable strategy for actually specifying those multiple priors in applied work. Our strategy is to use detection error probabilities to surround the single model that is typically specified in applied work with a set of empirically plausible but vaguely specified alternatives.

1.5. Robustness versus learning

A convenient feature of rational expectations models is that the model builder imputes a unique and explicit model to the decision maker. Our analysis shares this analytical convenience. While an agent distrusts his model, he still uses it to guide his decisions.[6] But the agent uses his model in a way that recognizes that it is an approximation. To quantify approximation, we measure discrepancy between the approximating model and other models with relative entropy, an expected log likelihood ratio, where the expectation is taken with respect to the distribution from the alternative model. Relative entropy is used in the theory of large deviations, a powerful mathematical theory about the rate at which uncertainty about unknown distributions is resolved as the number of observations grows.[7] An advantage of using entropy to restrain model perturbations is that we can appeal to the theory of statistical detection to provide information about how much concern about robustness is quantitatively reasonable.

Our decision maker confronts alternative models that can be discriminated among only with substantial amounts of data, so much data that, because he discounts the future, the robust decision maker simply accepts model misspecification as a permanent situation. He designs robust controls, and does not use data to improve his model specification over time. He adopts this stance because relative to his discount factor, it would take too much

---

[6] The assumption of rational expectations equates a decision maker's approximating model to the objective distribution. Empirical applications of models with robust decision makers like HST and HSW have equated those distributions too. The statement that the agent regards his model as an approximation, and therefore makes cautious decisions, leaves open the possibility that the agent's concern about model misspecification is "just in his head", meaning that the data are actually generated by the approximating model. The "just in his head" assumption justifies equating the agent's approximating model with the econometrician's model, a step that allows us to bring to bear much of the powerful empirical apparatus of rational expectations econometrics. In particular, it provides the same economical way of imputing an approximating model to the agents as rational expectations does. The difference is that we allow the agent's doubts about that model to affect his decisions.

[7] See Cho, Williams, and Sargent (2002) for a recent application of large deviation theory to a model of learning dynamics in macroeconomics.

time for enough data to accrue for him to dispose of the alternative models that concern him. In contrast, many formulations of learning have decision makers fully embrace an approximating model when making their choices.[8] Despite their different orientations, learners and robust decision makers both need a convenient way to measure the proximity of two probability distributions. This fact builds technical bridges between robust decision theory and learning theory. The same expressions from large deviation theory that govern bounds on rates of learning also provide bounds on value functions across alternative possible models in robust decision theory.[9] More importantly here, we shall show that the tight relationship between detection error probabilities and the market price of risk that was encountered by HST and HSW can be explained by formally studying the rate at which detection errors decrease as sample size grows.

1.6. Reader's guide

A reader interested only in our main results can read section 2, then jump to the empirical applications in section 9.

2. Overview

This section briefly tells how our main results apply in the special case in which the approximating model is a diffusion. Later sections provide technical details and show how things change when we allow jump components.

A representative agent's model asserts that the state of an economy x_t in a state space D follows a diffusion[10]

$$dx_t = \mu(x_t)\,dt + \Lambda(x_t)\,dB_t \tag{2.1}$$

where B_t is a Brownian vector. The agent wants decision rules that work well not just when (2.1) is true but also when the data conform to models that are statistically difficult to distinguish from (2.1). A *robust control* problem to be studied in section 5 leads to such a robust decision rule together with a value function $V(x_t)$ and a process $\gamma(x_t)$ for the marginal utility of consumption of a representative agent. As a byproduct of the robust control problem, the decision maker computes a worst-case diffusion that takes the form

$$dx_t = [\mu(x_t) + \Lambda(x_t)\,\hat{g}(x_t)]\,dt + \Lambda(x_t)\,dB_t, \tag{2.2}$$

where $\hat{g} = -\frac{1}{\theta}\Lambda'\frac{\partial V}{\partial x}$ and $\theta > 0$ is a parameter measuring the size of potential model misspecifications. Notice that (2.2) modifies the drift but not the volatility relative to

---

[8] See Bray (1982) and Kreps (1998).

[9] See Hansen and Sargent (2004) for discussions of these bounds.

[10] Diffusion (2.1) describes the 'physical probability measure'.

(2.1). The formula for $\hat{g}$ tells us that large values of θ are associated with $\hat{g}_t$'s that are small in absolute value, making model (2.2) difficult to distinguish statistically from model (2.1). The diffusion (2.6) below lets us quantify just how difficult this statistical detection problem is.

Without a preference for robustness to model misspecification, the usual approach to asset pricing is to compute the expected discounted value of payoffs with respect to the 'risk-neutral' probability measure that is associated with the following twisted version of the physical measure (diffusion (2.1)):

$$dx_t = \left[\mu\left(x_t\right) + \Lambda\left(x_t\right)\bar{g}\left(x_t\right) \right] dt + \Lambda\left(x_t\right) dB_t. \tag{2.3}$$

In using the risk-neutral measure to price assets, future expected returns are discounted at the risk-free rate $\rho(x_t)$, obtained as follows. The marginal utility of the representative household $\gamma(x_t)$ conforms to $d\gamma_t = \mu_\gamma(x_t)dt + \sigma_\gamma(x_t)dB_t$. Then the risk-free rate is $\rho(x_t) = \delta - \frac{\mu_\gamma(x_t)}{\gamma(x_t)}$, where δ is the instantaneous rate at which the household discounts future utilities; the risk-free rate thus equals the negative of the expected growth rate of the representative household's marginal utility. The price of a payoff $\phi(x_N)$ contingent on a Markov state in period N is then

$$\bar{E}\left(\exp\left[-\int_0^N \rho\left(x_u\right) du \right] \phi\left(x_N\right) | x_0 = x \right) \tag{2.4}$$

where $\bar{E}$ is the expectation evaluated with respect to the distribution generated by (2.3). This formula gives rise to a pricing operator for every horizon N. Relative to the approximating model, the diffusion (2.3) for the risk-neutral measure distorts the drift in the Brownian motion by adding the term $\Lambda(x)\bar{g}(x_t)$, where $\bar{g} = \Lambda' \frac{\partial \log \gamma(x)}{\partial x}$. Here $\bar{g}$ is a vector of 'factor risk prices' or 'market prices of risk'. The *equity premium puzzle* is the finding that with plausible quantitative specifications for the marginal utility $\gamma(x)$, factor risk prices $\bar{g}$ are too small relative to their empirically estimated counterparts.

In section 7, we show that when the planner and a representative consumer want robustness, the diffusion associated with the risk-neutral measure appropriate for pricing becomes

$$dx_t = \left(\mu\left(x_t\right) + \Lambda\left(x_t\right)\left[\bar{g}\left(x_t\right) + \hat{g}\left(x_t\right)\right] \right) dt + \Lambda\left(x_t\right) dB_t, \tag{2.5}$$

where $\hat{g}$ is the *same* process that appears in (2.2). With robustness sought over a set of alternative models that is indexed by θ, factor risk prices become augmented to $\bar{g} + \hat{g}$. The representative agent's concerns about model misspecification contribute the $\hat{g}$ component of the factor risk prices. To evaluate the quantitative potential for attributing parts of the market prices of risk to agents' concerns about model misspecification, we need to calibrate θ and therefore $|\hat{g}|$.

To calibrate θ and $\hat{g}$, we turn to a closely related fourth diffusion that governs the probability distribution of errors from using likelihood ratio tests to detect which of two models generated a continuous record of length N of observations on x_t. Here the key idea is that we can represent the average error in using a likelihood ratio test to detect the difference between the two models (2.1) and (2.3) from a continuous record of data of length N as $.5E\left(\min\{\exp(\ell^N), 1\}|x_0 = x\right)$ where E is evaluated with respect to model (2.1) and ℓ^N is a likelihood ratio of the data record of model (2.2) with respect to model (2.1). For each $\alpha \in (0, 1)$, we can use the inequality $E\left(\min\{\exp(\ell^N), 1\}|x_0 = x\right) \leq E\left(\{\exp(\alpha\ell^N)\}|x_0 = x\right)$ to attain a bound on the detection error probability. For each α, we show that the bound can be calculated by forming a new diffusion that uses (2.1) and (2.2) as ingredients, and in which the drift distortion $\hat{g}$ from (2.2) plays a key role. In particular, for $\alpha \in (0, 1)$, define

$$dx_t^\alpha = \left[\mu\left(x_t\right) + \alpha\Lambda\left(x_t\right)\hat{g}\left(x_t\right)\right]dt + \Lambda\left(x_t\right)dB_t, \tag{2.6}$$

and define the local rate function $\rho^\alpha(x) = \frac{(1-\alpha)\alpha}{2}\hat{g}(x)'\hat{g}(x)$. Then the bound on the average error in using a likelihood ratio test to discriminate between the approximating model (2.1) and the worst case model (2.2) from a continuous data record of length N is

$$\text{av error} \leq .5E^\alpha\left[\exp\left(-\int_0^N \rho^\alpha\left(x_t\right)\right)dt\Big|x_0 = x\right], \tag{2.7}$$

where E^α is the mathematical expectation evaluated with respect to the diffusion (2.6). The error rate $\rho^\alpha(x)$ is maximized by setting $\alpha = .5$. Notice that the right side of (2.7) is one half the price of pure discount bond that pays off one unit of consumption for sure N periods in the future, treating ρ^α as the risk-free rate and the measure induced by (2.6) as the risk-neutral probability measure.

It is remarkable that the three diffusions (2.2), (2.5), and (2.6) that describe the worst case model, asset pricing under a preference for robustness, and the local behavior of a bound on model detection errors, respectively, are all obtained by perturbing the drift in the approximating model (2.1) with functions of the *same* drift distortion $\hat{g}(x)$ that emerges from the robust control problem. To the extent that the bound on detection probabilities is informative about the detection probabilities themselves, our theoretical results thus neatly explain the pattern that was observed in the empirical applications of HST and HSW, namely, that there is a tight link between calculated detection error probabilities and the market price of risk. That link transcends all details of the model specification.[11] In section 9, we shall encounter this tight link again when we calibrate the contribution to market prices of risk that can plausibly be attributed to a preference for robustness in the context of three continuous time asset pricing models.

---

[11] See figure 8 of HSW.

Subsequent sections of this paper substantiate these and other results in a more general Markov setting that permits x to have jump components, so that jump distortions also appear in the Markov processes for the worst case model, asset pricing, and model detection error. We shall exploit and extend the asset-pricing structure of formulas like (2.4) and (2.7) by recognizing that they reflect that collections of expectations, values, and bounds on detection error rates can all be described with semigroups.

3. Mathematical Preliminaries

The remainder of this paper studies continuous-time Markov formulations of model specification, robust decision-making, pricing, and statistical model detection. We use Feller semigroups indexed by time for all four purposes. This section develops the semigroup theory needed for our paper.

3.1. Semigroups and their generators

Let D be a Markov state space that is a locally compact and separable subset of $\mathbf{R}^m$. We distinguish two cases. First, when D is compact, we let C denote the space of continuous functions mapping D into $\mathbf{R}$. Second, when we want to study cases in which the state space is unbounded so that D is not compact, we shall use a one-point compactification that enlarges the state space by adding a point at ∞. In this case we let C be the space of continuous functions that vanish at ∞. We can think of such functions as having domain D or domain $D \cup \infty$. The compactification is used to limit the behavior of functions in the tails when the state space is unbounded. We use the sup-norm to measure the magnitude of functions on C and to define a notion of convergence.

We are interested in a strongly continuous semigroup of operators $\{\mathcal{S}_t : t \geq 0\}$ with an *infinitesimal generator* $\mathcal{G}$. For $\{\mathcal{S}_t : t \geq 0\}$ to be a semigroup we require that $\mathcal{S}_0 = \mathcal{I}$ and $\mathcal{S}_{t+\tau} = \mathcal{S}_t \mathcal{S}_\tau$ for all $\tau, t \geq 0$. A semigroup is strongly continuous if

$$\lim_{\tau \downarrow 0} \mathcal{S}_\tau \phi = \phi$$

where the convergence is uniform for each ϕ in C. Continuity allows us to compute a time derivative and to define a generator

$$\mathcal{G}\phi = \lim_{\tau \downarrow 0} \frac{\mathcal{S}_\tau \phi - \phi}{\tau}. \tag{3.1}$$

This is again a uniform limit and it is well defined on a dense subset of C. A generator describes the instantaneous evolution of a semigroup. A semigroup can be constructed

from a generator by solving a differential equation. Thus applying the semigroup property gives

$$\lim_{\tau \downarrow 0} \frac{\mathcal{S}_{t+\tau}\phi - \mathcal{S}_t\phi}{\tau} = \mathcal{G}\mathcal{S}_t\phi, \tag{3.2}$$

a differential equation for a semigroup that is subject to the initial condition that $\mathcal{S}_0$ is the identity operator. The solution to differential equation (3.2) is depicted heuristically as:

$$\mathcal{S}_t = \exp(t\mathcal{G})$$

and thus satisfies the semigroup requirements. The exponential formula can be justified rigorously using a Yosida approximation, which formally constructs a semigroup from its generator.

In what follows, we will use semigroups to model Markov processes, intertemporal prices, and statistical discrimination. Using a formulation of Hansen and Scheinkman (2002), we first examine semigroups that are designed to model Markov processes,

3.2. Representation of a generator

We describe a convenient representation result for a strongly continuous, positive, contraction semigroup. Positivity requires that $\mathcal{S}_t$ maps nonnegative functions ϕ into nonnegative functions ϕ for each t. When the semigroup is a contraction, it is referred to as a *Feller semigroup*. The contraction property restricts the norm of $\mathcal{S}_t$ to be less than or equal to one for each t and is satisfied for semigroups associated with Markov processes. Generators of Feller semigroups have a convenient characterization:

$$\mathcal{G}\phi = \mu \cdot \left(\frac{\partial\phi}{\partial x}\right) + \frac{1}{2}\text{trace}\left(\Sigma\frac{\partial^2\phi}{\partial x\partial x'}\right) + \mathcal{N}\phi - \rho\phi \tag{3.3}$$

where $\mathcal{N}$ has the product form

$$\mathcal{N}\phi(x) = \int [\phi(y) - \phi(x)]\,\eta(dy|x) \tag{3.4}$$

where ρ is a nonnegative continuous function, μ is an m-dimensional vector of continuous functions, Σ is a matrix of continuous functions that is positive semidefinite on the state space, and $\eta(\cdot|x)$ is a finite measure for each x and continuous in x for Borel subset of D. We require that $\mathcal{N}$ map C_K^2 into C where C_K^2 is the subspace of functions that are twice continuously differentiable functions with compact support in D. Formula (3.4) is valid at least on C_K^2.[12]

---

[12] See Theorem 1.13 in Chapter VII of Revuz and Yor (1994). Revuz and Yor give a more general representation that is valid provided that the functions in C_K^∞ are in the domain of the generator. Their representation does not require that $\eta(\cdot|x)$ be a finite measure for each x but imposes a weaker restriction on this measure. As we will see, when $\eta(\cdot|x)$ is finite, we can define a jump intensity. Weaker restrictions permit there to be an infinite number of expected jumps in finite intervals that are arbitrarily small in magnitude. As a consequence, this extra generality involves more cumbersome notation and contributes nothing essential to our analysis.

To depict equilibrium prices we will sometimes go beyond Feller semigroups. Pricing semigroups are not necessarily contraction semigroups unless the instantaneous yield on a real discount bond is nonnegative. When we use this approach for pricing, we will allow ρ to be negative. While this puts us out of the realm of Feller semigroups, as argued by Hansen and Scheinkman (2002), known results for Feller semigroups can often be extended to pricing semigroups.

We can think of the generator (3.3) as being composed of three parts. The first two components are associated with well known continuous-time Markov process models, namely, diffusion and jump processes. The third part discounts. The next three subsections will interpret these components of equation (3.3).

3.2.1. Diffusion processes

The generator of a Markov diffusion process is a second-order differential operator:

$$\mathcal{G}_d \phi = \mu \cdot \left(\frac{\partial \phi}{\partial x} \right) + \frac{1}{2} \text{trace} \left(\Sigma \frac{\partial^2 \phi}{\partial x \partial x'} \right)$$

where the coefficient vector μ is the drift or local mean of the process and the coefficient matrix Σ is the diffusion or local covariance matrix. The corresponding stochastic differential equation is:

$$dx_t = \mu\left(x_t\right) dt + \Lambda\left(x_t\right) dB_t$$

where $\{B_t\}$ is a multivariate standard Brownian motion and $\Lambda\Lambda' = \Sigma$. Sometimes the resulting process will have attainable boundaries, in which case we either stop the process at the boundary or impose other boundary protocols.

3.2.2. Jump processes

The generator for a Markov jump process is:

$$\mathcal{G}_n \phi = \mathcal{N} \phi = \lambda \left[\mathcal{Q}\phi - \phi \right] \tag{3.5}$$

where the coefficient $\lambda \doteq \int \eta(dy|x)$ is a possibly state-dependent Poisson intensity parameter that sets the jump probabilities and $\mathcal{Q}$ is a conditional expectation operator that encodes the transition probabilities conditioned on a jump taking place. Without loss of generality, we can assume that the transition distribution associated with the operator $\mathcal{Q}$ assigns probability zero to the event $y = x$ provided that $x \neq \infty$, where x is the current Markov state and y the state after a jump takes place. That is, conditioned on a jump taking place, the process cannot stay put with positive probability unless it reaches a boundary.

The jump and diffusion components can be combined in a model of a Markov process. That is,

$$\mathcal{G}_d\phi + \mathcal{G}_n\phi = \mu \cdot \left(\frac{\partial\phi}{\partial x}\right) + \frac{1}{2}\text{trace}\left(\Sigma\frac{\partial^2\phi}{\partial x\partial x'}\right) + \mathcal{N}\phi \tag{3.6}$$

is the generator of a family (semigroup) of conditional expectation operators of a Markov process $\{x_t\}$, say $\mathcal{S}_t(\phi)(x) = E[\phi(x_t)|x_0 = x]$.

3.2.3. Discounting

The third part of (3.3) accounts for discounting. Thus, consider a Markov process $\{x_t\}$ with generator $\mathcal{G}_d + \mathcal{G}_n$. Construct the semigroup:

$$\mathcal{S}_t\phi = E\left(\exp\left[-\int_0^t \rho\left(x_\tau\right)d\tau\right]\phi\left(x_t\right)|x_0 = x\right)$$

on C. We can think of this semigroup as *discounting* the future state at the stochastic rate $\rho(x)$. Discount rates will play essential roles in representing shadow prices from a robust resource allocation problem and in measuring statistical discrimination between competing models.[13]

3.3. Extending the domain to bounded functions

While it is mathematically convenient to construct the semigroup on C, sometimes it is necessary for us to extend the domain to a larger class of functions. For instance, indicator functions $\mathbf{1}_D$ of nondegenerate subsets D are omitted from C. Moreover, $\mathbf{1}_D$ is not in C when D is not compact; nor can this function be approximated uniformly. Thus to extend the semigroup to bounded, Borel measurable functions, we need a weaker notion of convergence. Let $\{\phi_j : j = 1, 2, ...\}$ be a sequence of uniformly bounded functions that converges pointwise to a bounded function ϕ_o. We can then extend the $\mathcal{S}_\tau$ semigroup to ϕ_o using the formula:

$$\mathcal{S}_\tau\phi_o = \lim_{j\to\infty}\mathcal{S}_\tau\phi_j$$

---

[13] When $\rho \geq 0$, the semigroup is a contraction. In this case, we can use $\mathcal{G}$ as a generator of a Markov process in which the process is curtailed at rate ρ. Formally, we can let ∞ be a terminal state at which the process stays put. Starting the process at state $x \neq \infty$, $E\left(\exp\left[-\int_0^t \rho(x_\tau)d\tau\right]|x_0 = x\right)$ is the probability that the process is not curtailed after t units of time. See Revuz and Yor (1994, page 280) for a discussion. As in Hansen and Scheinkman (2002), we will use the discounting interpretation of the semigroup and not use ρ as a curtailment rate. Discounting will play an important role in our discussion of detection and pricing. In pricing problems, ρ can be negative in some states as might occur in a real economy, an economy with a consumption numeraire.

where the limit notion is now pointwise. The choice of approximating sequence does not matter and the extension is unique.[14]

With this construction, we define the instantaneous discount or interest rate as the pointwise derivative

$$-\lim_{\tau \downarrow 0} \frac{1}{\tau} \log \mathcal{S}_\tau \mathbf{1}_D = \rho$$

when the derivative exists.

3.4. Extending the generator to unbounded functions

Value functions for control problems on noncompact state spaces are often not bounded. Thus for our study of robust counterparts to optimization, we must extend the semigroup and its generator to unbounded functions. We adopt an approach that is specific to a Markov process and hence we study this extension only for a semigroup generated by $\mathcal{G} = \mathcal{G}_d + \mathcal{G}_n$.

We extend the generator using martingales. To understand this approach, we first remark that for a given ϕ in the domain of the generator,

$$M_t = \phi\left(x_t\right) - \phi\left(x_0\right) - \int_0^t \mathcal{G}\phi\left(x_\tau\right) d\tau$$

is a martingale. In effect, we produce a martingale by subtracting the integral of the local means from the process $\{\phi(x_t)\}$. This martingale construction suggests a way to build the *extended generator*. Given ϕ we find a function ψ such that

$$M_t = \phi\left(x_t\right) - \phi\left(x_0\right) - \int_0^t \psi\left(x_\tau\right) d\tau \tag{3.7}$$

is a local martingale (a martingale under all members of a sequence of stopping times that increases to ∞). We then define $\mathcal{G}\phi = \psi$. This construction extends the operator $\mathcal{G}$ to a larger class of functions than those for which the operator differentiation (3.1) is well defined. For every ϕ in the domain of the generator, $\psi = \mathcal{G}\phi$ in (3.7) produces a martingale. However, there are ϕ's not in the domain of the generator for which (3.7) also produces a martingale.[15] In the case of a Feller process defined on a state-space D that is an open subset of $\mathbf{R}^m$, this extended domain contains at least functions in $\tilde{C}^2$, functions that are twice continuously differentiable on D. Such functions can be unbounded when the original state space D is not compact.

---

[14] This extension was demonstrated by Dynkin (1956). Specifically, Dynkin defines a weak (in the sense of functionals) counterpart to this semigroup and shows that there is a weak extension of this semigroup to bounded, Borel measurable functions.

[15] There are other closely related notions of an extended generator in the probability literature. Sometimes calendar time dependence is introduced into the function ϕ, or martingales are used in place of local martingales.

4. A tour of four semigroups

In the remainder of the paper we will study four semigroups. Before describing each in detail, it is useful to tabulate the four semigroups and their uses. We have already introduced the first semigroup, which describes the evolution of a state vector process $\{x_t\}$. This semigroup portrays a decision maker's approximating model. It has the generator displayed in (3.3) with $\rho = 0$, which we repeat here for convenience:

$$\mathcal{G}\phi = \mu \cdot \left(\frac{\partial \phi}{\partial x}\right) + \frac{1}{2}\text{trace}\left(\Sigma \frac{\partial^2 \phi}{\partial x \partial x'}\right) + \mathcal{N}\phi. \tag{4.1}$$

While up to now we used $\mathcal{G}$ to denote a generic semigroup, from this point forward we will reserve it for the approximating model. We can think of the decision maker as using the semigroup generated by $\mathcal{G}$ to forecast functions $\phi(x_t)$. This semigroup for the approximating model can have both jump and Brownian components, but the discount rate ρ is zero. In some settings, the semigroup associated with the approximating model includes a description of endogenous state variables and therefore embeds robust decision rules of one or more decision makers, as for example when the approximating model emerges from a robust resource allocation problem of the kind to be described in section 5.

With our first semigroup as a point of reference, we will consider three additional semigroups. The second semigroup represents an endogenous worst-case model that a decision maker uses to promote robustness to possible misspecification of his approximating model (4.1). For reasons that we discuss in section 8, we shall focus the decision maker's attention on worst-case models that are absolutely continuous with respect to his approximating model. Following Kunita (1969), we shall assume that the decision maker believes that the data are actually generated by a member of a class of models that are obtained as Markov perturbations of the approximating model (4.1). We parameterize this class of models by a pair of functions (g, h), where g is a continuous function of the Markov state x that has the same number of coordinates as the underlying Brownian motion, and h is a nonnegative function of (y, x) that distorts the jump intensities. For the worst-case model, we have the particular settings $g = \hat{g}$ and $h = \hat{h}$. Then we can represent the worst-case generator $\hat{\mathcal{G}}$ as

$$\hat{\mathcal{G}}\phi = \hat{\mu} \cdot \left(\frac{\partial \phi}{\partial x}\right) + \frac{1}{2}\text{trace}\left(\hat{\Sigma} \frac{\partial^2 \phi}{\partial x \partial x'}\right) + \hat{\mathcal{N}}\phi, \tag{4.2}$$

where

$$\hat{\mu} = \mu + \Lambda \hat{g}$$
$$\hat{\Sigma} = \Sigma$$
$$\hat{\eta}\left(dy|x\right) = \hat{h}\left(y, x\right)\eta\left(dy|x\right).$$

The distortion $\hat{g}$ to the diffusion and the distortion $\hat{h}$ to the jump component in the worst case model will also play essential roles both in asset pricing and in the detection probabilities formulas. From (3.5), it follows that the jump intensity under this parameterization is given by $\hat{\lambda}(x) = \int \hat{h}(y,x)\eta(dy|x)$ and the jump distribution conditioned on x is $\frac{\hat{h}(y,x)}{\hat{\lambda}(x)}\eta(dy|x)$. A generator of the form (4.2) emerges from a robust decision problem, the perturbation pair $(\hat{g}, \hat{h})$ being chosen by a malevolent player, as we discuss below. Our third semigroup modifies one that Hansen and Scheinkman (2002) developed for computing the time zero price of a state contingent claim that pays off $\phi(x_t)$ at time t. Hansen and Scheinkman showed that the time zero price can be computed with a risk-free rate $\bar{\rho}$ and a *risk-neutral* probability measure embedded in a semigroup with generator:

$$\bar{\mathcal{G}}\phi = -\bar{\rho}\phi + \bar{\mu} \cdot \left(\frac{\partial \phi}{\partial x}\right) + \frac{1}{2}\text{trace}\left(\bar{\Sigma}\frac{\partial^2 \phi}{\partial x \partial x'}\right) + \bar{\mathcal{N}}\phi. \tag{4.3a}$$

Here

$$\bar{\mu} = \mu + \Lambda\bar{\pi}$$
$$\bar{\Sigma} = \Sigma \tag{4.3b}$$
$$\bar{\eta}(dy|x) = \bar{\Pi}(y,x)\eta(dy|x).$$

In the absence of a concern about robustness, $\bar{\pi} = \bar{g}$ is a vector of prices for the Brownian motion *factors* and $\bar{\Pi} = \bar{h}$ encodes the jump risk prices. In Markov settings without a concern for robustness, (4.3b) represents the connection between the physical probability and the so-called risk-neutral probability that is widely used for asset pricing along with the interest rate adjustment.

We alter generator (4.3) to incorporate a representative consumer's concern about robustness to model misspecification. Specifically a preference for robustness changes the ordinary formulas for $\bar{\pi}$ and $\bar{\Pi}$ that are based solely on pricing risks under the assumption that the approximating model is true. A concern about robustness alters the relationship between the semigroups for representing the underlying Markov processes and pricing. With a concern for robustness, we represent factor risk prices by relating $\bar{\mu}$ to the worst-case drift $\hat{\mu}$: $\bar{\mu} = \hat{\mu} + \Lambda\bar{g}$ and risk-based jump prices by relating $\bar{\eta}$ to the worst-case jump measure $\hat{\eta}$: $\bar{\eta}(dy|x) = \bar{h}(y,x)\hat{\eta}(dy|x)$. Combining this decomposition with the relation between the worst-case and the approximating models gives the new vectors of pricing functions

$$\bar{\pi} = \bar{g} + \hat{g}$$
$$\bar{\Pi} = \bar{h}\hat{h}$$

where the pair $(\hat{g}, \hat{h})$ is used to portray the (constrained) worst-case model in (4.2). Later we will supply formulas for $(\bar{\rho}, \bar{g}, \bar{h})$.

semigroup	generator	rate	drift distortion	jump dist. density
approximating model	$\mathcal{G}$	0	0	1
worst-case model	$\hat{\mathcal{G}}$	0	$\hat{g}(x)$	$\hat{h}(y,x)$
pricing	$\bar{\mathcal{G}}$	$\bar{\rho}(x)$	$\bar{\pi}(x) = \bar{g}(x) + \hat{g}(x)$	$\bar{\Pi}(x) = \bar{h}(y,x)\hat{h}(y,x)$
detection	$\mathcal{G}^{\alpha}$	$\rho^{\alpha}(x)$	$g^{\alpha}(x)$	$h^{\alpha}(y,x)$

Table 4.1: Parameterizations of the generators of four semigroups. The rate modifies the generator associated with the approximating model by adding $-\rho\phi$ to the generator for a test function ϕ. The drift distortion adds a term $\Lambda g \cdot \frac{\partial \phi}{\partial x}$ to the generator associated with the approximating model. The jump distortion density is $h(y,x)\eta(dy|x)$ instead of the jump distribution $\eta(dy|x)$ in the generator for the approximating model.

A fourth semigroup statistically quantifies the discrepancy between two competing models as a function of the time interval of available data. We are particularly interested in measuring the discrepancy between the approximating and worst case models. For each $\alpha \in (0,1)$, we develop a bound on a detection error probability in terms of a semigroup and what looks like an associated 'risk-free interest rate'. The counterpart to the risk-free rate serves as an instantaneous discrimination rate. For each α, the generator for the bound on the detection error probability can be represented as:

$$\mathcal{G}^{\alpha}\phi = -\rho^{\alpha}\phi + \mu^{\alpha} \cdot \left(\frac{\partial \phi}{\partial x}\right) + \frac{1}{2}\text{trace}\left(\Sigma^{\alpha}\frac{\partial^2 \phi}{\partial x \partial x'}\right) + \mathcal{N}^{\alpha}\phi,$$

where

$$\mu^{\alpha} = \mu + \Lambda g^{\alpha}$$

$$\Sigma^{\alpha} = \Sigma$$

$$\eta^{\alpha}\left(dy|x\right) = h^{\alpha}\left(y,x\right)\eta\left(dy|x\right).$$

The semigroup generated by $\mathcal{G}^{\alpha}$ governs the behavior as sample size grows of a bound on the fraction of errors made when distinguishing two Markov models using likelihood ratios or posterior odds ratios. The α associated with the best bound is determined on a case by case basis and is especially easy to find in the special case that the Markov process is a pure diffusion.

Table 4.1 summarizes our parameterization of these four semigroups. Subsequent sections supply formulas for the entries in this table.

5.   Model misspecification and robust control

We now study the continuous-time robust resource allocation problem. In addition to an approximating model, this analysis will produce a constrained worst case model that by helping the decision maker to assess the fragility of any given decision rule can be used as a device to choose a robust decision rule.

5.1. Lyapunov equation under Markov approximating model and a fixed decision rule

Under a Markov approximating model with generator $\mathcal{G}$ and a fixed policy function $i(x)$, the decision maker's value function is

$$V\left(x\right) = \int_0^\infty \exp\left(-\delta t\right) E\left[U\left[x_t, i\left(x_t\right)\right]|x_0 = x\right] dt.$$

The value function V satisfies the continuous-time Lyapunov equation:

$$\delta V\left(x\right) = U\left[x, i\left(x\right)\right] + \mathcal{G}V\left(x\right). \tag{5.1}$$

Since V may not be bounded, we interpret $\mathcal{G}$ as the weak extension of the generator (3.6) defined using local martingales. The local martingale associated with this equation is:

$$M_t = V\left(x_t\right) - V\left(x_0\right) - \int_0^t \left(\delta V\left(x_s\right) - U\left[x_s, i\left(x_s\right)\right]\right) ds.$$

As in (3.6), this generator can include diffusion and jump contributions.

We will eventually be interested in optimizing over a control i, in which case the generator $\mathcal{G}$ will depend explicitly on the control. For now we suppress that dependence. We refer to $\mathcal{G}$ as the approximating model; $\mathcal{G}$ can be modelled using the triple (μ, Σ, η) as in (3.6). The pair (μ, Σ) consists of the drift and diffusion coefficients while the conditional measure η encodes both the jump intensity and the jump distribution.

We want to modify the Lyapunov equation (5.1) to incorporate a concern about model misspecification. We shall accomplish this by replacing $\mathcal{G}$ with another generator that expresses the decision maker's precaution about the specification of $\mathcal{G}$.

5.2. Entropy penalties

We now introduce perturbations to the decision maker's approximating model that are designed to make finite horizon transition densities of the perturbed model be absolutely continuous with respect to those of the approximating model. We use a notion of absolute continuity that pertains only to finite intervals of time. In particular, imagine a Markov process evolving for a finite length of time. Our notion of absolute continuity restricts probabilities induced by the path $\{x_\tau : 0 \leq \tau \leq t\}$ for all finite t. See HSTW (2002), who discuss this notion as well as an infinite history version of absolute continuity. Kunita (1969) shows how to preserve both the Markov structure and absolute continuity.

Following Kunita (1969), we shall consider a Markov perturbation that can be parameterized by a pair (g, h), where g is a continuous function of the Markov state x and has the same number of coordinates as the underlying Brownian motion, and h is a nonnegative function of (y, x) used to model the jump intensities. In section 8, we will have more to say about these perturbations including a discussion of why we do not perturb Σ. For the pair (g, h), the perturbed generator is portrayed using a drift $\mu + \Lambda g$, a diffusion matrix Σ, and a jump measure $h(y, x)\eta(dy|x)$. Thus the perturbed generator is

$$\mathcal{G}(g, h)\phi(x) = \mathcal{G}\phi(x) + [\Lambda(x) g(x)] \cdot \frac{\partial \phi(x)}{\partial x} + \int [h(y, x) - 1][\phi(y) - \phi(x)]\eta(dy|x).$$

For this perturbed generator to be a Feller process would require that we impose additional restrictions on h. For analytical tractability we will only limit the perturbations to have finite entropy. We will be compelled to show, however, that the perturbation used to implement robustness does indeed generate a Markov process. This perturbation will be constructed formally as the solution to a constrained minimization problem. In what follows, we continue to use the notation $\mathcal{G}$ to be the approximating model in place of the more tedious $\mathcal{G}(0, 1)$.

5.3. Conditional relative entropy

At this point, it is useful to have a local measure of conditional relative entropy.[16] Conditional relative entropy plays a prominent role in large deviation theory and in classical statistical discrimination where it is sometimes used to study the decay in the so called type II error probabilities, holding fixed type I errors (Stein's Lemma). For the purposes of this section, we will use relative entropy as a discrepancy measure. In section 8 we will elaborate on its connection to the theory of statistical discrimination. As a measure of discrepancy, it has been axiomatized by Csiszar (1991) although his defense shall not concern us here.

---

[16] This will turn out to be a limiting version of a local Chernoff measure ρ^α to be defined in section 8.

By ℓ_t we denote the log of the ratio of the likelihood of model one to the likelihood of model zero, given a data record of length t. For now, let the data be either a continuous or a discrete time sample. The relative entropy conditioned on x_0 is defined to be:

$$
\begin{aligned}
E\left(\ell_t \middle| x_0, \ \text{model } 1\right) &= E\left[\ell_t \exp\left(\ell_t\right) \middle| x_0, \ \text{model } 0\right] \\
&= \frac{d}{d\alpha} E\left[\exp\left(\alpha \ell_t\right) \middle| x_0, \ \text{model } 0\right]\bigg|_{\alpha=1},
\end{aligned}
\tag{5.2}
$$

where we have assumed that the model zero probability distribution is absolutely continuous with respect to the model one probability distribution. To evaluate entropy, the second relation differentiates the moment-generating function for the log-likelihood ratio. The same information inequality that justifies maximum likelihood estimation implies that relative entropy is nonnegative.

When the model zero transition distribution is absolutely continuous with respect to the model one transition distribution, entropy collapses to zero as the length of the data record $t \to 0$. Therefore, with a continuous data record, we shall use a concept of conditional relative entropy as a *rate*, specifically the time derivative of (5.2). Thus, as a local counterpart to (5.2), we have the following measure:

$$
\epsilon\left(g, h\right)\left(x\right) = \frac{g\left(x\right)' g\left(x\right)}{2} + \int \left[1 - h\left(y, x\right) + h\left(y, x\right) \log h\left(y, x\right)\right] \eta\left(dy | x\right)
\tag{5.3}
$$

where model zero is parameterized by $(0, 1)$ and model one is parameterized by (g, h). The quadratic form $g'g/2$ comes from the diffusion contribution, and the term

$$
\int \left[1 - h\left(y, x\right) + h\left(y, x\right) \log h\left(y, x\right)\right] \eta\left(dy | x\right)
$$

measures the discrepancy in the jump intensities and distributions. It is positive by the convexity of $h \log h$ in h.

Let Δ denote the space of all such perturbation pairs (g, h). Conditional relative entropy ϵ is convex in (g, h). It will be finite only when

$$
0 < \int h\left(y, x\right) \eta\left(dy | x\right) < \infty.
$$

When we introduce adjustments for model misspecification, we modify Lyapunov equation (5.1) in the following way to penalize entropy

$$
\delta V\left(x\right) = \min_{(g,h) \in \Delta} U\left[x, i\left(x\right)\right] + \theta \epsilon\left(g, h\right) + \mathcal{G}\left(g, h\right) V\left(x\right),
$$

where $\theta > 0$ is a penalty parameter. We are led to the following entropy penalty problem.

Problem A

$$J(V) = \inf_{(g,h)\in\Delta} \theta\epsilon(g,h) + \mathcal{G}(g,h)V. \tag{5.4}$$

Theorem 5.1. *Suppose that (i) V is in $\tilde{C}^2$ and (ii) $\int \exp[-V(y)/\theta]\eta(dy|x) < \infty$ for all x. The minimizer of Problem A is*

$$\hat{g}(x) = -\frac{1}{\theta}\Lambda(x)'\frac{\partial V(x)}{\partial x}$$
$$\hat{h}(y,x) = \exp\left[\frac{V(x)-V(y)}{\theta}\right]. \tag{5.5a}$$

The optimized value of the criterion is:

$$J(V) = -\theta\frac{\mathcal{G}\left[\exp\left(-\frac{V}{\theta}\right)\right]}{\exp\left(-\frac{V}{\theta}\right)}. \tag{5.5b}$$

Finally, the implied measure of conditional relative entropy is:

$$\epsilon^* = \frac{V\mathcal{G}\left[\exp\left(-V/\theta\right)\right] - \mathcal{G}\left[V\exp\left(-V/\theta\right)\right] - \theta\mathcal{G}\left[\exp\left(-V/\theta\right)\right]}{\theta\exp\left(-V/\theta\right)}. \tag{5.5c}$$

Proof. The proof is in Appendix A. ∎

The formulas ($5.5a$) for the distortions will play a key role in our applications to asset pricing and statistical detection.

5.4. Risk-sensitivity as an alternative interpretation

In light of Theorem 5.1, our modified version of Lyapunov equation (5.1) is

$$\delta V(x) = \min_{(g,h)\in\Delta} U[x,i(x)] + \theta\epsilon(g,h) + \mathcal{G}(g,h)V(x)$$
$$= U[x,i(x)] - \theta\frac{\mathcal{G}\left[\exp\left(-\frac{V}{\theta}\right)\right](x)}{\exp\left[-\frac{V(x)}{\theta}\right]}. \tag{5.6}$$

If we ignore the minimization prompted by fear of model misspecification and instead simply start with that modified Lyapunov equation as a description of preferences, then replacing $\mathcal{G}V$ in the Lyapunov equation (5.1) by $-\theta\frac{\mathcal{G}[\exp(-\frac{V}{\theta})]}{\exp(-\frac{V}{\theta})}$ can be interpreted as adjusting the continuation value for risk. For undiscounted problems, the connection between risk-sensitivity and robustness is developed in a literature on risk-sensitive control (*e.g.* see James (1992) and Runolfsson (1994)). Hansen and Sargent's (1995) recursive formulation of risk sensitivity accommodates discounting.

The connection between the robustness and the risk-sensitivity interpretations is most evident when $\mathcal{G} = \mathcal{G}_d$ so that x is a diffusion. Then

$$-\theta \frac{\mathcal{G}_d \left[\exp\left(-\frac{V}{\theta}\right)\right]}{\exp\left(-\frac{V}{\theta}\right)} = \mathcal{G}_d\left(V\right) - \frac{1}{2\theta}\left(\frac{\partial V}{\partial x}\right)' \Sigma \left(\frac{\partial V}{\partial x}\right).$$

In this case, (5.6) is a partial differential equation. Notice that $-\frac{1}{2\theta}$ scales $\left(\frac{\partial V}{\partial x}\right)' \Sigma \left(\frac{\partial V}{\partial x}\right)$, the local variance of the value function process $\{V(x_t)\}$. The interpretation of (5.6) under risk sensitive preferences would be that the decision maker is concerned not about robustness but about both the local mean and the local variance of the continuation value process. The parameter θ is inversely related to the size of the risk adjustment. Larger values of θ assign a smaller concern about risk. The term $\frac{1}{\theta}$ is the so-called risk sensitivity parameter.

Runolfsson (1994) deduced the $\delta = 0$ (ergodic control) counterpart to (5.6) to obtain a robust interpretation of risk sensitivity. Partial differential equation (5.6) is also a special case of the equation system that Duffie and Epstein (1992), Duffie and Lions (1992), and Schroder and Skiadas (1999) have analyzed for stochastic differential utility. They showed that for diffusion models, the recursive utility generalization introduces a variance multiplier that can be state dependent. The counterpart to this multiplier in our setup is state independent and equal to the risk sensitivity parameter $\frac{1}{\theta}$. For a robust decision maker, this variance multiplier restrains entropy between the approximating and alternative models. The mathematical connections between robustness, on the one hand, and risk sensitivity and recursive utility, on the other, let us draw on a set of analytical results from those literatures.[17]

5.5. The θ-constrained worst case model

Given a value function, Theorem 5.1 reports the formulas for the distortions $(\hat{g}, \hat{h})$ for a worst-case model used to enforce robustness. This worst case model is Markov and depicted in terms of the value function. This theorem thus gives us a generator $\hat{\mathcal{G}}$ and shows us how to fill out the second row in Table 4.1. In fact, a separate argument is needed to show formally that $\hat{\mathcal{G}}$ does in fact generate a Feller process or more generally a Markov process. There is a host of alternative sufficient conditions in the probability theory literature. Kunita (1969) gives one of the more general treatments of this problem and goes outside the realm of Feller semigroups. Also, Ethier and Kurtz (1985, Chapter 8) give some sufficient conditions for operators to generate Feller semigroups, including restrictions on the jump component $\hat{\mathcal{G}}_n$ of the operator.

---

[17] See section 9.2 for alternative interpretations of a particular empirical application in terms of risk-sensitivity and robustness. For that example, we show how the robustness interpretation helps us to restrict θ.

Using the Theorem 5.1 characterization of $\hat{\mathcal{G}}$, we can apply Theorem 8.1 to obtain the generator of a detection semigroup that measures the statistical discrepancy between the approximating model and the worst-case model.

5.6. An alternative entropy constraint

We briefly consider an alternative but closely related way to compute worst-case models and to enforce robustness. In particular, we consider:

Problem B

$$J^* \left(V \right) = \inf_{(g,h) \in \Delta, \epsilon(g,h) \leq \varepsilon} \mathcal{G} \left(g, h \right) V. \tag{5.7}$$

This problem has the same solution as that given by Problem A except that θ must now be chosen so that the relative entropy constraint is satisfied. That is, θ should be chosen so that $\epsilon(g, h)$ satisfies the constraint. The resulting θ will typically depend on x. The optimized objective must now be adjusted to remove the penalty:

$$J^* \left(V \right) = J \left(V \right) - \theta \epsilon^* = \frac{V \mathcal{G} \left[\exp \left(-V/\theta \right) \right] - \mathcal{G} \left[V \exp \left(-V/\theta \right) \right]}{\exp \left(-V/\theta \right)},$$

which follows from (5.5c).

These formulas simplify greatly when the approximating model is a diffusion. Then θ satisfies

$$\theta^2 = \frac{1}{2\varepsilon} \left(\frac{\partial V \left(x \right)}{\partial x} \right)' \Sigma \left(\frac{\partial V \left(x \right)}{\partial x} \right).$$

This formulation embeds a version of the continuous-time preference order that Chen and Epstein (2001) proposed to capture uncertainty aversion. We had also suggested the diffusion version of this robust adjustment in our earlier paper (Anderson, Hansen and Sargent (1998)).

5.7. Enlarging the class of perturbations

In this paper we focus on misspecifications or perturbations to an approximating Markov model that themselves are Markov models. But in HSTW, we took a more general approach and began with a family of absolutely continuous perturbations to an approximating model that is a Markov diffusion. Absolute continuity over finite intervals puts a precise structure on the perturbations, even when the Markov specification is not imposed on these perturbations. As a consequence, HSTW follow James (1992) by considering path dependent specifications of the drift of the Brownian motion $\int_0^t g_s ds$, where g_s is constructed as a general function of past x's. Given the Markov structure of this control problem,

its solution can be represented as a time-invariant function of the state vector x_t that we denote $\hat{g}_t = \hat{g}(x_t)$.

5.8. Adding controls to the original state equation

We now allow the generator to depend on a control vector. Consider an approximating Markov control law of the form $i(x)$ and let the generator associated with an approximating model be $\mathcal{G}(i)$. For this generator, we introduce perturbation (g, h) as before. We write the corresponding generator as $\mathcal{G}(g, h, i)$. To attain a robust decision rule, we use the Bellman equation for a two-player zero-sum Markov *multiplier game*:

$$\delta V = \max_{i} \min_{(g,h) \in \Delta} U(x, i) + \theta \epsilon(g, h) + \mathcal{G}(g, h, i) V. \tag{5.8}$$

The Bellman equation for a corresponding *constraint game* is:

$$\delta V = \max_{i} \min_{(g,h) \in \Delta(i), \epsilon(g,h) \leq \varepsilon} U(x, i) + \mathcal{G}(g, h, i) V.$$

Sometimes infinite-horizon counterparts to terminal conditions must be imposed on the solutions to these Bellman equations. Moreover, application of a Verification Theorem will be needed to guarantee that the implied control laws actually solve the game. Finally, these Bellman equations presume that the value function is twice continuously differentiable. It is well known that this differentiability is not always present in problems in which the diffusion matrix can be singular. In these circumstances there is typically a *viscosity* generalization to each of these Bellman equations with very similar structures. (See Fleming and Soner (1991) for a development of the viscosity approach to controlled Markov processes.)

6. Portfolio allocation

To put some of the results of section 5 to work, we now consider a robust portfolio problem. In section 7 we will use this problem to exhibit how asset prices can be deduced from the shadow prices of a robust resource allocation problem. We depart somewhat from our previous notation and let $\{x_t : t \geq 0\}$ denote a state vector that is exogenous to the individual investor. The investor influences the evolution of his wealth, which we denote by w_t. Thus the investor's composite state at date t is (w_t, x_t). We first consider the case in which the exogenous component of the state vector evolves as a diffusion process. Later we let it be a jump process. Combining the diffusion and jump pieces is straightforward. We focus on the formulation with the entropy penalty used in Problem (5.4), but the constraint counterpart is similar.

6.1. Diffusion

An investor confronts asset markets that are driven by a Brownian motion. Under an approximating model, the Brownian increment factors have date t prices given by $\pi(x_t)$ and x_t evolves according to a diffusion:

$$dx_t = \mu(x_t) \, dt + \Lambda(x_t) \, dB_t. \tag{6.1}$$

Equivalently, the x process has a generator $\mathcal{G}_d$ that is a second-order differential operator with drift μ and diffusion matrix $\Sigma = \Lambda\Lambda'$. A control vector b_t entitles the investor to an instantaneous payoff $b_t \cdot dB_t$ with a price $\pi(x_t) \cdot b_t$ in terms of the consumption numeraire. This cost can be positive or negative. Adjusting for cost, the investment has payoff $-\pi(x_t) \cdot b_t dt + b_t \cdot dB_t$. There is also a market in a riskless security with an instantaneous risk free rate $\rho(x)$. The wealth dynamics are therefore

$$dw_t = [w_t \rho(x_t) - \pi(x_t) \cdot b_t - c_t] \, dt + b_t \cdot dB_t, \tag{6.2}$$

where c_t is date t consumption. The control vector is $i' = (b', c)$. Only consumption enters the instantaneous utility function. By combining (6.1) and (6.2), we form the evolution for a composite Markov process.

But the investor has doubts about this approximating model and wants a robust decision rule. Therefore he solves a version of game (5.8) with (6.1), (6.2) governing the dynamics of his composite state vector w, x. With only the diffusion component, the investor's Bellman equation is

$$\delta V(w, x) = \max_{(c,b)} \min_g U(c) + \theta \epsilon(g) + \mathcal{G}(g, b, c) V$$

where $\mathcal{G}(g, b, c)$ is constructed using drift vector

$$\begin{bmatrix} \mu(x) + \Lambda(x) g \\ w\rho(x) - \pi(x) \cdot b - c + b \cdot g \end{bmatrix}$$

and diffusion matrix

$$\begin{bmatrix} \Lambda \\ b' \end{bmatrix} \begin{bmatrix} \Lambda' & b \end{bmatrix}.$$

The choice of the worst case shock g satisfies the first-order condition:

$$\theta g + V_w b + \Lambda' V_x = 0 \tag{6.3}$$

where $V_w \doteq \frac{\partial V}{\partial w}$ and similarly for V_x. Solving (6.3) for g gives a special case of the formula in (5.5a). The resulting worst-case shock would depend on the control vector b. In what follows we seek a solution that does not depend on b.

The first-order condition for consumption is

$$V_w\left(w, x\right) = U_c\left(c\right),$$

and the first-order condition for the risk allocation vector b is

$$-V_w \pi + V_{ww} b + \Lambda' V_{xw} + V_w g = 0. \tag{6.4}$$

In the limiting case in which the robustness penalty parameter is set to ∞, we obtain the familiar result that

$$b = \frac{\pi V_w - \Lambda' V_{xw}}{V_{ww}},$$

in which the portfolio allocation rule has a contribution from risk aversion measured by $-V_w / w V_{ww}$ and a hedging demand contributed by the dynamics of the exogenous forcing process x.

Take the Markov perfect equilibrium of the relevant version of game (5.8). Provided that V_{ww} is negative, the same equilibrium decision rules prevail no matter whether one player or the other chooses first, or whether they choose simultaneously. The first-order conditions (6.3) and (6.4) are linear in b and g. Solving these two linear equations gives the control laws for b and g as a function of the composite state (w, x):

$$
\begin{aligned}
\hat{b} &= \frac{\theta \pi V_w - \theta \Lambda' V_{xw} + V_w \Lambda' V_x}{\theta V_{ww} - \left(V_w\right)^2} \\
\hat{g} &= \frac{V_w \Lambda' V_{xw} - \left(V_w\right)^2 \pi - V_{ww} \Lambda' V_x}{\theta V_{ww} - \left(V_w\right)^2}.
\end{aligned}
\tag{6.5}
$$

Notice how the robustness penalty adds terms to the numerator and denominator of the portfolio allocation rule. Of course, the value function V also changes when we introduce θ. Notice also that (6.5) gives decision rules of the form

$$
\begin{aligned}
\hat{b} &= \hat{b}\left(w, x\right) \\
\hat{g} &= \hat{g}\left(w, x\right),
\end{aligned}
\tag{6.6}
$$

and in particular how the worst case shock g feeds back on the consumer's endogenous state variable w. Permitting g to depend on w expands the kinds of misspecifications that the consumer considers.

6.1.1. Related formulations

So far we have studied portfolio choice in the case of a constant robustness parameter θ. Maenhout (2001) considers portfolio problems in which the robustness penalty depends on the continuation value. In his case, the preference for robustness is designed so that asset demands are not sensitive to wealth levels as is typical in constant θ formulations. Lei (2000) uses the instantaneous constraint formulation of robustness described in section 5.6 to investigate portfolio choice. His formulation also makes θ state dependent, since θ now formally plays the role of a Lagrange multiplier that restricts conditional entropy at every instant. Lei specifically considers the case of incomplete asset markets in which the counterpart to b has a lower dimension than the Brownian motion.

6.1.2. Ex post Bayesian interpretation

While the dependence of g on the endogenous state w seems reasonable as a way to enforce robustness, it can be unattractive if we wish to interpret the implied worst case model as one with misspecified exogenous dynamics. It is sometimes asked whether a prescribed decision rule can be rationalized as being optimal for *some* set of beliefs, and then to find what those beliefs must be. The dependence of the shock distributions on an endogenous state variable such as wealth w might be regarded as a peculiar set of beliefs because it is egotistical to let an adverse nature feedback on personal state variables.

But there is a way to make this feature more acceptable. It requires using a dynamic counterpart to an argument of Blackwell and Girshick (1954). We can produce a different representation of the solution to the decision problem by forming an exogenous state vector W that conforms to the Markov perfect equilibrium of the game. We can confront a decision maker with this law of motion for the exogenous state vector, have him not be concerned with robustness against misspecification of this law by setting $\theta = \infty$, and pose an ordinary decision problem in which the decision maker has a unique model. We initialize the exogenous state at $W_0 = w_0$. The optimal decision processes for $\{(b_t, c_t)\}$ (but not the control laws) will be identical for this decision problem and for game (5.8) (see HSWT). It can be said that this alternative problem gives a Bayesian rationale for the robust decision procedure.

6.2. Jumps

Suppose now that the exogenous state vector $\{x_t\}$ evolves according to a Markov jump process with jump measure η. To accommodate portfolio allocation, introduce the choice of a function a that specifies how wealth changes when a jump takes place. Consider an investor who faces asset markets with date-state Arrow security prices given by $\Pi(y, x_t)$ where $\{x_t\}$ is an exogenous state vector with jump dynamics. In particular, a choice a with instantaneous payoff $a(y)$ if the state jumps to y has a price $\int \Pi(y, x_t) a(y) \eta(dy|x)$ in terms of the consumption numeraire. This cost can be positive or negative. When a jump does not take place, wealth evolves according to

$$dw_t = \left[\rho(x_{t-}) w_{t-} - \int \Pi(y, x_{t-}) a(y) \eta(dy|x_{t-}) - c_{t-} \right] dt$$

where $\rho(x)$ is the riskfree rate given state x and for any variable z, $z_{t-} = \lim_{\tau \uparrow t} z_\tau$. If the state x jumps to y at date t, the new wealth is $a(y)$. The Bellman equation for this problem is

$$\delta V(w, x) = \max_{c,a} \min_{h \in \Delta} U(c) + V_w(w, x) \left[\rho(x) w_t - \int \Pi(y, x) a(y) \eta(dy|x) - c \right]$$
$$+ \theta \int [1 - h(y, x) + h(y, x) \log h(y, x)] \eta(dy|x)$$
$$+ \int h(y, x) (V[a(y), y] - V(w, x)) \eta(dy|x)$$

The first-order condition for c is the same as for the diffusion case and equates V_w to the marginal utility of consumption. The first-order condition for a requires

$$\hat{h}(y, x) V_w[\hat{a}(y), y] = V_w(w, x) \Pi(y, x),$$

and the first-order condition for h requires

$$-\theta \log \hat{h}(y, x) = V[\hat{a}(y), y] - V(w, x).$$

Solving this second condition for $\hat{h}$ gives the jump counterpart to the solution asserted in Theorem 5.1. Thus the robust $\hat{a}$ satisfies:

$$\frac{V_w[\hat{a}(y), y]}{V_w(w, x)} = \frac{\Pi(y, x)}{\exp\left(\frac{-V[\hat{a}(y), y] + V(x)}{\theta}\right)}.$$

In the limiting no-concern-about-robustness case $\theta = \infty$, $\hat{h}$ is set to one. Since V_w is equated to the marginal utility for consumption, the first-order condition for a equates the marginal rate of substitution of consumption before and after the jump to the price $\Pi(y, x)$. Introducing robustness scales the price by the jump distribution distortion.

In this portrayal, the worst case h depends on the endogenous state w, but it is again possible to obtain an alternative representation of the probability distortion that would give an ex post Bayesian justification for the decision process of a.

7.   Pricing risky claims

By building on findings of Hansen and Scheinkman (2002), we now consider a third semigroup that is to be used to price risky claims. We denote this semigroup by $\{\mathcal{P}_t : t \geq 0\}$ where $\mathcal{P}_t \phi$ assigns a price at date zero to a date t payoff $\phi(x_t)$. That pricing can be described by a semigroup follows from the Law of Iterated Values: a date 0 state-date claim $\phi(x_t)$ can be replicated by first buying a claim $\mathcal{P}_\tau \phi(x_{t-\tau})$ and then at time $t - \tau$ buying a claim $\phi(x_t)$. Like our other semigroups, this one has a generator, say $\bar{\mathcal{G}}$, that we write as in (3.3):

$$\bar{\mathcal{G}} \phi = -\bar{\rho}\phi + \bar{\mu} \cdot \left(\frac{\partial \phi}{\partial x}\right) + \frac{1}{2}\text{trace}\left(\bar{\Sigma}\frac{\partial^2 \phi}{\partial x \partial x'}\right) + \bar{\mathcal{N}}\phi$$

where

$$\bar{\mathcal{N}}\phi = \int [\phi(y) - \phi(x)]\,\bar{\eta}\,(dy|x)\,.$$

The coefficient on the level term $\bar{\rho}$ is the instantaneous riskless yield to be given in formula (7.3). It is used to price locally riskless claims. Taken together, the remaining terms

$$\bar{\mu} \cdot \left(\frac{\partial \phi}{\partial x}\right) + \frac{1}{2}\text{trace}\left(\bar{\Sigma}\frac{\partial^2 \phi}{\partial x \partial x'}\right) + \bar{\mathcal{N}}\phi$$

comprise the generator of the so called *risk neutral* probabilities. The risk neutral evolution is Markov.

As discussed by Hansen and Scheinkman (2002), we should expect there to be a connection between the semigroup underlying the Markov process and the semigroup that underlies pricing. Like the semigroup for Markov processes, a pricing semigroup is positive: it assigns nonnegative prices to nonnegative functions of the Markov state. We can thus relate the semigroups by importing the measure-theoretic notion of *equivalence*. Prices of contingent claims that pay off only in probability measure zero events should be zero. Conversely, when the price of a contingent claim is zero, the event associated with that claim should occur only with measure zero, which states the principle of no-arbitrage. We can capture these properties by specifying that the generator $\bar{\mathcal{G}}$ of the pricing semigroup satisfies:

$$\bar{\mu}(x) = \mu(x) + \Lambda(x)\bar{\pi}(x)$$
$$\bar{\Sigma}(x) = \Sigma(x) \tag{7.1}$$
$$\bar{\eta}(x) = \bar{\Pi}(y,x)\eta(dy|x)$$

where $\bar{\Pi}$ is strictly positive. Thus we construct equilibrium prices by producing a triple $(\bar{\rho}, \bar{\pi}, \bar{\Pi})$. We now show how to construct this triple both with and without a preference for robustness.

7.1. Marginal rate of substitution pricing

To compute prices, we follow Lucas (1978) and focus on the consumption side of the market. While Lucas used an endowment economy, Brock (1982) showed that the essential thing in Lucas's analysis was not the pure endowment feature. Instead it was the idea of pricing assets from marginal utilities that are evaluated at a candidate equilibrium consumption process that can be computed prior to computing prices. In contrast to Brock, we use a robust planning problem to generate a candidate equilibrium allocation. As in Breeden (1979), we use a continuous-time formulation that provides simplicity along some dimensions.[18]

7.2. Pricing without a concern for robustness

First consider the case in which the consumer has no concern about model misspecification. Proceeding in the spirit of Lucas (1978) and Brock (1982), we can construct market prices of risk from the shadow prices of a planning problem. Following Lucas and Prescott (1971) and Prescott and Mehra (1980), we solve a representative agent planning problem to get a state process $\{x_t\}$, an associated control process $\{i_t\}$, and a marginal utility of consumption process $\{\gamma_t\}$, respectively. We let $\mathcal{G}^*$ denote the generator for the state vector process that emerges when the optimal controls from the resource allocation problem with no concern for robustness are imposed. In effect, $\mathcal{G}^*$ is the generator for the $\theta = \infty$ robust control problem.

We construct a stochastic discount factor process by evaluating the marginal rate of substitution at the proposed equilibrium consumption process:

$$\mathrm{mrs}_t = \exp\left(-\delta t\right) \frac{\gamma\left(x_t\right)}{\gamma\left(x_0\right)}$$

where $\gamma(x)$ denotes the marginal utility process for consumption as a function of the state x. Without a preference for robustness, the pricing semigroup satisfies

$$\mathcal{P}_t \phi\left(x\right) = E^* \left[\mathrm{mrs}_t \phi\left(x_t\right) | x_0 = x\right] \tag{7.2}$$

where the expectation operator E^* is the one implied by $\mathcal{G}^*$.

---

[18] This analysis differs from that of Breeden (1979) by its inclusion of jumps.

Individuals solve a version of the portfolio problem described in section 6 without a concern for robustness. This supports the following representation of the generator for the equilibrium pricing semigroup $\mathcal{P}_t$:

$$\bar{\rho} = -\frac{\mathcal{G}^*\gamma}{\gamma} + \delta$$

$$\bar{\mu} = \mu^* + \Lambda^*\bar{\pi} = \mu^* + \Lambda^*\Lambda^{*\prime}\frac{\partial \log \gamma}{\partial x} \tag{7.3}$$

$$\bar{\eta}\left(dy|x\right) = \bar{\Pi}\left(y,x\right)\eta^*\left(dy|x\right) = \left[\frac{\gamma\left(y\right)}{\gamma\left(x\right)}\right]\eta^*\left(dy|x\right).$$

These are the usual rational expectations risk prices. The risk-free rate is the subjective rate of discount reduced by the local mean of the equilibrium marginal utility process scaled by the marginal utility. The vector $\bar{\pi}$ of Brownian motion risk prices are weights on the Brownian increment in the evolution of the marginal utility of consumption, again scaled by the marginal utility. Finally the jump risk prices $\bar{\Pi}$ are given by the equilibrium marginal rate of substitution between consumption before and after a jump.

7.3. Pricing with a concern for robustness under the worst case model

As in our previous analysis, let $\mathcal{G}$ denote the approximating model. This is the model that emerges after imposing the robust control law $\hat{i}$ while assuming that there is no model misspecification ($g = 0$ and $h = 1$). It differs from $\mathcal{G}^*$, which also assumes no model misspecification but instead imposes a rule derived *without* any preference for robustness. But simply attributing the beliefs $\mathcal{G}$ to private agents in (7.3) will not give us the correct equilibrium prices when there is a preference for robustness. Let $\hat{\mathcal{G}}$ denote the worst case model that emerges as part of the Markov perfect equilibrium of the two-player, zero-sum game. However, formula (7.3) *will* yield the correct equilibrium prices if we in effect impute to the individual agents the worst-case generator $\hat{\mathcal{G}}$ instead of $\mathcal{G}^*$ as their model of state evolution when making their decisions *without* any concerns about its possible misspecification.

To substantiate this claim, we consider individual decision-makers who, when choosing their portfolios, use the worst-case model $\hat{\mathcal{G}}$ as if it were correct (i.e., they have no concern about the misspecification of *that* model, so that rather than entertaining a family of models, the individuals commit to the worst-case $\hat{\mathcal{G}}$ as a model of the state vector $\{x_t : t \geq 0\}$). The pricing semigroup then becomes

$$\mathcal{P}_t\phi\left(x\right) = \hat{E}\left[\mathrm{mrs}_t\phi\left(x_t\right)|x_0 = x\right] \tag{7.4}$$

where $\hat{E}$ denotes the mathematical expectation with respect to the *distorted* measure described by the generator $\hat{\mathcal{G}}$. The generator for this pricing semigroup is parameterized

by

$$\bar{\rho} = -\frac{\hat{\mathcal{G}}\gamma}{\gamma} + \delta$$

$$\bar{\mu} = \hat{\mu} + \Lambda\bar{g} = \hat{\mu} + \Lambda\Lambda'\frac{\partial \log \gamma}{\partial x} \tag{7.5}$$

$$\bar{\eta}\left(dy|x\right) = \bar{h}\left(y, x\right)\hat{\eta}\left(dy|x\right) = \left[\frac{\gamma\left(y\right)}{\gamma\left(x\right)}\right]\hat{\eta}\left(dy|x\right).$$

As in subsection 7.2, $\gamma(x)$ is the log of the marginal utility of consumption except it is evaluated at the solution of the robust planning problem. Individuals solve the portfolio problem described in section 6 using the worst-case model of the state $\{x_t\}$ with pricing functions $\bar{\pi} = \bar{g}$ and $\bar{\Pi} = \bar{h}$ specified relative to the worst-case model. We refer to $\bar{g}$ and $\bar{h}$ as *risk prices* because they are equilibrium prices that emerge from an economy in which individual agents use the worst-case model as if it were the correct model to assess risk. The vector $\bar{g}$ contains the so-called factor risk prices associated with the vector of Brownian motion increments. Similarly, $\bar{h}$ prices jump risk.

Comparison of (7.3) and (7.5) shows that the formulas for factor risk prices and the risk free rate are identical except that we have used the distorted generator $\hat{\mathcal{G}}$ in place of $\mathcal{G}^*$. This comparison shows that we can use standard characterizations of asset pricing formulas if we simply replace the generator for the approximating model $\mathcal{G}$ with the distorted generator $\hat{\mathcal{G}}$.[19]

7.4. Pricing under the approximating model

There is another portrayal of prices that uses the approximating model $\mathcal{G}$ as a reference point and that provides a vehicle for defining model uncertainty prices and for distinguishing between the contributions of risk and model uncertainty. The $\bar{g}$ and $\bar{h}$ from subsection 7.3 give the risk components. We now use the discrepancy between $\mathcal{G}$ and $\hat{\mathcal{G}}$ to produce the model uncertainty prices.

To formulate model uncertainty prices, we consider how prices can be represented under the approximating model when the consumer has a preference for robustness. We want to represent the pricing semigroup as

$$\mathcal{P}_t\phi\left(x\right) = E\left[\left(\text{mrs}_t\right)\left(mpu_t\right)\phi\left(x_t\right)|x_0 = x\right] \tag{7.6}$$

where mpu is a multiplicative adjustment to the marginal rate of substitution that allows us to evaluate the conditional expectation with respect to the approximating model rather

---

[19] In the applications in HST, HSW, and section 9, we often take the *actual* data generating model to be the approximating model to study implications. In that sense, the approximating model supplies the same kinds of empirical restrictions that a rational expectations econometric model does.

than the distorted model. Instead of (7.3), to attain (7.6), we portray the drift and jump distortion in the generator for the pricing semigroup as

$$\bar{\mu} = \hat{\mu} + \Lambda \bar{g} = \mu + \Lambda \left(\bar{g} + \hat{g} \right)$$
$$\bar{\eta} \left(dy | x \right) = \bar{h} \left(y, x \right) \hat{\eta} \left(dy | x \right) = \bar{h} \left(y, x \right) \hat{h} \left(y, x \right) \eta \left(dy | x \right).$$

Changing expectation operators in depicting the pricing semigroup will not change the instantaneous risk-free yield. Thus from Theorem 5.1 we have:

Theorem 7.1. *Let V^p be the value function for the robust resource allocation problem. Suppose that (i) V^p is in $\tilde{C}^2$ and (ii) $\int \exp[-V^p(y)/\theta]\eta(dy|x) < \infty$ for all x. Moreover, γ is assumed to be in the domain of the extended generator $\hat{\mathcal{G}}$. Then the equilibrium prices can be represented by:*

$$\bar{\rho} = -\frac{\hat{\mathcal{G}}\gamma}{\gamma} + \delta$$
$$\bar{\pi} \left(x \right) = -\frac{1}{\theta} \Lambda \left(x \right)' V_x^p \left(x \right) + \Lambda \left(x \right)' \left[\frac{\gamma_x \left(x \right)}{\gamma \left(x \right)} \right] = \hat{g} \left(x \right) + \bar{g} \left(x \right)$$
$$\log \bar{\Pi} \left(y, x \right) = -\frac{1}{\theta} \left[V^p \left(y \right) - V^p \left(x \right) \right] + \log \gamma \left(y \right) - \log \gamma \left(x \right) = \log \hat{h} \left(y, x \right) + \log \bar{h} \left(y, x \right).$$

This theorem follows directly from the relation between $\mathcal{G}$ and $\hat{\mathcal{G}}$ given in Theorem 5.1 and from the risk prices of subsection 7.3. It supplies the third row of Table 4.1.

7.5. Model uncertainty prices: diffusion and jump components

We have already interpreted $\bar{g}$ and $\bar{h}$ as risk prices. Thus we view $\hat{g} = -\frac{1}{\theta}\Lambda' V_x^p$ as the contribution to the Brownian exposure prices that comes from model uncertainty. Similarly, we think of $\hat{h}(y, x) = -\frac{1}{\theta} \exp[V^p(y) - V^p(x)]$ as the model uncertainty contribution to the jump exposure prices. HST obtained the additive decomposition for the Brownian motion exposure asserted in Theorem 7.1 as an approximation for linear-quadratic, Gaussian resource allocation problems. By studying continuous time diffusion models we have been able to sharpen their results and relax the linear-quadratic specification of constraints and preferences.

7.6. Subtleties about decentralization

In Hansen and Sargent (2003), we confirm that the solution of a robust planning problem can be decentralized with households who also solve robust decision problems while facing the state-date prices that we derived above. We confront the household with a recursive representation of state-date prices, give the household the same robustness parameter θ as the planner, and allow the household to choose a new worst-case model. The recursive representation of the state-date prices is portrayed in terms of the state vector X for the planning problem. As in the portfolio problems of section 6, among the households' state variables is their endogenously determined financial wealth, w. In equilibrium, the household's wealth can be expressed as a function of the state vector X of the planner. However, in posing the household's problem, it is necessary to include both wealth w and the state vector X that propels the state-date prices as distinct state components of the household's state. More generally, it is necessary to include both economy-wide and individual versions of household capital stocks and physical capital stocks in the household's state vector, where the economy-wide components are used to provide a recursive representation of the date-state prices.

Thus the controls and the worst case shocks chosen by both the planner, on the one hand, and the households in the decentralized economy, on the other hand, will depend on different state vectors. However, in a competitive equilibrium, the decisions that emerge from these distinct rules will be perfectly aligned. That is, if we take the decision rules of the household in the decentralized economy and impose the equilibrium conditions requiring that 'the representative agent be representative', then the decisions and the motion of the state will match. The worst-case models will also match. In addition, although the worst-case models depend on different state variables, they coincide along an equilibrium path.

7.7. Ex post Bayesian equilibrium interpretation of robustness

In a decentralized economy, Hansen and Sargent (2003) also confirm that it is possible to compute robust decision rules for both the planner and the household by a) endowing each such decision maker with his own worst-case model, and b) having each solve his decision problem *without* a preference for robustness, while treating those worst-case models as if they were true. *Ex post* it is possible to interpret the decisions made by a robust decision maker who has a concern about the misspecification of his model as also being made by an equivalent decision maker who has no concern about the misspecification of a *different* model that can be constructed from the worst case model that is computed by the robust decision maker. Hansen and Sargent's (2003) results thus extend results of

HSTW, discussed in section 6.1.2, to a setting where both a planner and a representative household choose worst case models, and where their worst case models turn out to be aligned.

8.   Statistical discrimination

A weakness in what we have achieved up to now is that we have provided the practitioner with no guidance on how to calibrate our model uncertainty premia of Theorem 7.1, or what formulas $(5.5a)$ tell us is virtually the same thing, the decision maker's robustness parameter θ. It is at this critical point that our fourth semigroup enters the picture.[20]

Our fourth semigroup governs bounds on detection statistics that we can use to guide our thinking about how to calibrate a concern about robustness. We shall synthesize this semigroup from the objects in two other semigroups that represent alternative models that we want to choose between given a finite data record. We apply the bounds associated with distinguishing between the decision maker's approximating and worst-case models. In designing a robust decision rule, we assume that our decision maker worries about alternative models that available time series data cannot readily dispose of. Therefore, we study a stylized model selection problem. Suppose that a decision-maker chooses between two models that we will refer to as zero and one. Both are continuous-time Markov process models. We construct a measure of how much time series data are needed to distinguish these models and then use it to calibrate our robustness parameter θ. Our statistical discrepancy measure is the same one that in section 5 we used to adjust continuation values in a dynamic programming problem that is designed to acknowledge concern about model misspecification.

8.1. Measurement and prior probabilities

We assume that there are direct measurements of the state vector $\{x_t : 0 \leq t \leq N\}$ and aim to discriminate between two Markov models: model zero and model one. We assign prior probabilities of one-half to each model. If we choose the model with the maximum posterior probability, two types of errors are possible, choosing model zero when model one is correct and choosing model one when model zero is correct. We weight these errors by the prior probabilities and, following Chernoff (1952), study the error probabilities as the sample interval becomes large.

---

[20] As we shall see in section 9, our approach to disciplining the choice of θ depends critically on our adopting a robustness and not a risk-sensitivity interpretation.

8.2. A semigroup formulation of bounds on error probabilities

We evade the difficult problem of precisely calculating error probabilities for nonlinear Markov processes and instead seek bounds on those error probabilities. To compute those bounds, we adapt Chernoff's (1952) large deviation bounds to discriminate between Markov processes. Large deviation tools apply here because the two types of error both get small as the sample size increases. Let $\mathcal{G}^0$ denote the generator for Markov model zero and $\mathcal{G}^1$ the generator for Markov model one. Both can be represented as in (3.6).

8.2.1. Discrimination in discrete time

Before developing results in continuous time, we discuss discrimination between two Markov models in discrete time. Associated with each Markov process is a family of transition probabilities. For any interval τ, these transition probabilities are mutually absolutely continuous when restricted to some event that has positive probability under both probability measures. If no such event existed, then the probability distributions would be orthogonal, making statistical discrimination easy. Let $p_\tau(y|x)$ denote the ratio of the transition density over a time interval τ of model one relative to that for model zero. We include the possibility that $p_\tau(y|x)$ integrates to a magnitude less than one using the model zero transition probability distribution. This would occur if the model one transition distribution assigned positive probability to an event that has measure zero under model zero. We also allow the density p_τ to be zero with positive model zero transition probability.

If discrete time data were available, say $x_0, x_\tau, x_{2\tau}, ..., x_{T\tau}$ where $N = T\tau$, then we could form the log likelihood ratio:

$$\ell_\tau^N = \sum_{j=1}^{T} \log p_\tau\left(x_{j\tau}, x_{(j-1)\tau}\right).$$

Model one is selected when

$$\ell_\tau^N > 0, \tag{8.1}$$

and model zero is selected otherwise. The probability of making a classification error at date zero conditioned on model zero is

$$\Pr\{\ell_\tau^N > 0 | x_0 = x, \ \text{model } 0\} = E\left(\mathbf{1}_{\{\ell_\tau^N > 0\}} | x_0 = x, \ \text{model } 0\right).$$
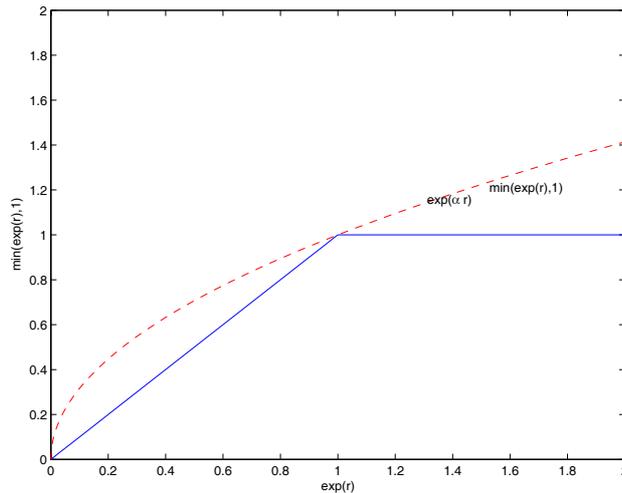
It is convenient that the probability of making a classification error conditioned on model one can also be computed as an expectation of a transformed random variable conditioned on model zero. Thus,

$$\begin{aligned} \Pr\{\ell_\tau^N < 0 | x_0 = x, \ \text{model } 1\} &= E\left[\mathbf{1}_{\{\ell_\tau^N < 0\}} | x_0 = x, \ \text{model } 1\right] \\ &= E\left[\exp\left(\ell_\tau^N\right)\mathbf{1}_{\{\ell_\tau^N < 0\}} | x_0 = x, \ \text{model } 0\right]. \end{aligned}$$

The second equality follows because multiplication of the indicator function by the likelihood ratio $\exp\left(\ell_\tau^N\right)$ converts the conditioning model from one to zero. Combining these two expressions, the average error is:

$$\text{av error} = \frac{1}{2}E\left(\min\{\exp\left(\ell_\tau^N\right),1\}\Big|x_0 = x, \quad \text{model} \quad 0\right). \tag{8.2}$$

Because we compute expectations only under the model zero probability measure, from now on we leave implicit the conditioning on model zero.

Figure 8.1: Graph of $\min\{\exp(r),1\}$ and the dominating function $\exp(r\alpha)$ for $\alpha = .5$.

Instead of using formula (8.2) to compute the probability of making an error, we will use a convenient upper bound originally suggested by Chernoff (1952) and refined by Hellman and Raviv (1970). To motivate the bound, note that for any $0 < \alpha < 1$ the piecewise linear function $\min\{s,1\}$ is dominated by the concave function s^α and that the two functions agree at the kink point $s = 1$. The smooth dominating function gives rise to more tractable computations as we alter the amount of available data. Thus, setting $\log s = r = \ell_\tau^N$ and using (8.2) gives the bound:

$$\text{av error} \leq \frac{1}{2}E\left[\exp\left(\alpha\ell_\tau^N\right)\Big|x_0 = x\right] \tag{8.3}$$

where the right side is the moment-generating function for the log-likelihood ratio ℓ_τ^N (see Fig. 8.1). (Later we shall discuss how to choose $\alpha \in (0,1)$ in order to maximize error detection rates.) Define an operator:

$$\mathcal{K}_\tau^\alpha \phi\left(x\right) = E\left[\exp\left(\alpha\ell_\tau^\tau\right)\phi\left(x_\tau\right)\Big|x_0 = x\right].$$

Then inequality (8.3) can be portrayed as:

$$\text{av error} \leq \frac{1}{2} \left(\mathcal{K}_\tau^\alpha\right)^T \mathbf{1}_D (x) \tag{8.4}$$

where $\mathbf{1}_D$ is again the indictor function of the state space D for the Markov process, and where superscript T on the right side denotes sequential application of an operator T times. This bound applies for any integer choice of T and any choice of α between zero and one.[21]

When restricted to a function space C, we have the inequalities

$$\begin{aligned}
|\mathcal{K}_\tau^\alpha \phi (x)| &\leq E\left(\left[\exp\left(\ell_\tau^\tau\right)\right]^\alpha |\phi|\Big|x_0 = x\right) \\
&\leq E\left[\exp\left(\ell_\tau^\tau\right)|\phi|^{\frac{1}{\alpha}}\Big|x_0 = x\right]^\alpha \\
&\leq \|\phi\|
\end{aligned}$$

where the second inequality is an application of Jensen's inequality. Thus $\mathcal{K}_\tau^\alpha$ is a contraction on C.

8.3. Rates for measuring discrepancies between models locally

Classification errors become less frequent as more data become available. One common way to study the large sample behavior of classification error probabilities is to investigate the limiting behavior of the operator $\left(\mathcal{K}_\tau^\alpha\right)^T$ as T gets large. This amounts to studying how fast $\left(\mathcal{K}_\tau^\alpha\right)^T$ contracts for large T and results in a large deviation characterization. Chernoff (1952) proposed such a characterization for i.i.d. data that later researchers extended to Markov processes. Formally, a large-deviation analysis can give rise to an *asymptotic rate* for discriminating between models.

Given the state dependence in the Markov formulation, there are two different possible notions of discrimination rates that are based on Chernoff entropy. One notion that we shall dub 'long run' is state independent; to construct it requires additional assumptions. This long run rate is computed by studying the semigroup $\{\mathcal{K}_t : t \geq 0\}$ as t gets large. This semigroup can have a positive eigenfunction, that is, a function ϕ that solves

$$\mathcal{K}_t\phi = \exp\left(-\tilde{\delta}t\right)\phi \tag{8.5}$$

for some positive $\tilde{\delta}$. When it exists, this eigenfunction dominates the remaining ones as the time horizon t gets large. As a consequence, for large t this semigroup decays approximately at an exponential rate $\tilde{\delta}$. Therefore, $\tilde{\delta}$ is a long run measure of the rate at which information for discriminating between two models accrues. By construction and as part of its 'long run' nature, the rate $\tilde{\delta}$ is independent of the state.

---

[21] This bound covers the case in which the model one density omits probability, and so equivalence between the two measures is not needed for this bound to be informative.

In this paper, we use another approximation that results in a state-dependent or 'short-run' discrimination rate. It is this state-dependent rate that is closely linked to our robust decision rule, in the sense that it is governed by the same objects that emerge from the worst-case analysis for our robust control problem. The semigroup $\{\mathcal{K}_t : t \geq 0\}$ has the same properties as a pricing semigroup, and furthermore it contracts. We can define a discrimination rate in the same way that we define an instantaneous interest rate from a pricing semigroup. This leads us to use Chernoff entropy as $\rho^\alpha(x)$. It differs from the decay rate $\tilde{\delta}$ defined by (8.5). For a given state x, it measures the statistical ability to discriminate between models when a small interval of data becomes available. When the rate $\rho^\alpha(x)$ is large, the time series data contain more information for discriminating between models.

Before characterizing a *local* discrimination rate ρ^α that is applicable to continuous-time processes, we consider the following example.

8.3.1. Constant drift

Consider sampling a continuous multivariate Brownian motion with a constant drift. Let μ^0, Σ^0 and μ^1, Σ^1 be the drift vectors and constant diffusion matrices for models zero and one, respectively. Thus under model zero, $x_{j\tau} - x_{(j-1)\tau}$ is normally distributed with mean $\tau\mu^0$ and covariance matrix $\tau\Sigma^0$. Under an alternative model one, $x_{j\tau} - x_{(j-1)\tau}$ is normally distributed with mean $\tau\mu^1$ and covariance matrix $\tau\Sigma^1$.

Suppose that $\Sigma^0 \neq \Sigma^1$ and that the probability distributions implied by the two models are *equivalent* (i.e., mutually absolutely continuous). Equivalence will always be satisfied when Σ^0 and Σ^1 are nonsingular but will also be satisfied when the degeneracy implied by the covariance matrices coincides. It can be shown that

$$\lim_{\tau \downarrow 0} \mathcal{K}_\tau^\alpha 1_D < 1$$

suggesting that a continuous-time limit will not result in a semigroup. Recall that a semigroup of operators must collapse to the identity when the elapsed interval becomes arbitrarily small. When the covariance matrices Σ^0 and Σ^1 differ, the detection-error bound remains positive even when the data interval becomes small. This reflects the fact that while absolute continuity is preserved for each positive τ, it is known from Cameron and Martin (1947) that the probability distributions implied by the two limiting continuous-time Brownian motions will *not* be mutually absolutely continuous when the covariance matrices differ. Since diffusion matrices can be inferred from high frequency data, differences in these matrices are easy to detect.[22]

---

[22] The continuous time diffusion specification carries with it the implication that diffusion matrices can be inferred from high frequency data without knowledge of the drift. That data are discrete in practice tempers

Suppose that $\Sigma^0 = \Sigma^1 = \Sigma$. If $\mu^0 - \mu^1$ is not in the range of Σ, then the discrete-time transition probabilities for the two models over an interval τ are not equivalent, making the two models easy to distinguish using data. If, however, $\mu^0 - \mu^1$ is in the range of Σ, then the probability distributions are equivalent for any transition interval τ. Using a complete-the-square argument, it can be shown that

$$\mathcal{K}_\tau^\alpha \phi(x) = \exp(-\tau\rho_\alpha) \int \phi(y) P_\tau^\alpha(y-x)\, dy$$

where P_α^τ is a normal distribution with mean $\tau(1-\alpha)\mu^0 + \tau\alpha\mu^1$ and covariance matrix $\tau\Sigma$,

$$\rho^\alpha = \frac{\alpha(1-\alpha)}{2}(\mu^0 - \mu^1)' \Sigma^{-1}(\mu^0 - \mu^1) \tag{8.6}$$

and Σ^{-1} is a generalized inverse when Σ is singular. It is now the case that

$$\lim_{\tau\downarrow 0} \mathcal{K}_\tau^\alpha \mathbf{1}_D = 1.$$

The parameter ρ^α acts as a discount rate, since

$$\mathcal{K}_\tau^\alpha \mathbf{1}_D = \exp(-\tau\rho^\alpha).$$

The best probability bound (the largest ρ^α) is obtained by setting $\alpha = 1/2$, and the resulting discount rate is referred to as Chernoff entropy. The continuous-time limit of this example is known to produce probability distributions that are absolutely continuous over any finite horizon (see Cameron and Martin, 1947).

For this example, the long-run discrimination rate $\tilde{\delta}$ and the short-run rate ρ^α coincide because ρ^α is state independent. This equivalence emerges because the underlying processes have independent increments. For more general Markov processes, this will not be true and the counterpart to the short-run rate will depend on the Markov state. The existence of a well defined long-run rate requires special assumptions.

---

the notion that the diffusion matrix can be inferred exactly. Nevertheless, estimating conditional means is much more difficult than estimating conditional covariances. Our continuous time formulation simplifies our analysis by focusing on the more challenging drift inference problem.

8.3.2. Continuous time

There is a semigroup probability bound analogous to (8.3) that helps to understand how data are informative in discriminating between Markov process models. Suppose that we model two Markov processes as Feller semigroups. The generator of semigroup zero is

$$\mathcal{G}^0\phi = \mu^0 \cdot \left(\frac{\partial \phi}{\partial x}\right) + \frac{1}{2}\text{trace}\left(\Sigma\frac{\partial^2 \phi}{\partial x\partial x'}\right) + \mathcal{N}^0\phi$$

and the generator of semigroup one is

$$\mathcal{G}^1\phi = \mu^1 \cdot \left(\frac{\partial \phi}{\partial x}\right) + \frac{1}{2}\text{trace}\left(\Sigma\frac{\partial^2 \phi}{\partial x\partial x'}\right) + \mathcal{N}^1\phi$$

In specifying these two semigroups, we assume identical Σ's. As in the example, this assumption is needed to preserve absolute continuity. Moreover, we require that μ^1 can be represented as:

$$\mu^1 = \Lambda g + \mu^0.$$

for some continuous function g of the Markov state, where we assume that the rank of Σ is constant on the state space and can be factored as $\Sigma = \Lambda\Lambda'$ where Λ has full rank. This is equivalent to requiring that $\mu^0 - \mu^1$ is in the range of Σ.[23]

In contrast to the example, however, both of the μ's and Σ can depend on the Markov state. Jump components are allowed for both processes. These two operators are restricted to imply jump probabilities that are mutually absolutely continuous for at least some nondegenerate event. We let $h(\cdot, x)$ denote the density function of the jump distribution of $\mathcal{N}^1$ with respect to the distribution of $\mathcal{N}^0$. We assume that $h(y, x)d\eta^0(y|x)$ is finite for all x. Under absolute continuity we write:

$$\mathcal{N}^1\phi(x) = \int h(y, x)\left[\phi(y) - \phi(x)\right]\eta^0(dy|x).$$

Associated with these two Markov processes is a positive, contraction semigroup $\{\mathcal{K}_t^\alpha : t \geq 0\}$ for each $\alpha \in (0, 1)$ that can be used to bound the probability of classification errors:

$$\text{av error} \leq \frac{1}{2}\left(\mathcal{K}_N^\alpha\right)\mathbf{1}_D(x).$$

This semigroup has a generator $\mathcal{G}^\alpha$ with the Feller form:

$$\mathcal{G}^\alpha\phi = -\rho^\alpha\phi + \mu^\alpha \cdot \left(\frac{\partial \phi}{\partial x}\right) + \frac{1}{2}\text{trace}\left(\Sigma\frac{\partial^2 \phi}{\partial x\partial x'}\right) + \mathcal{N}^\alpha\phi. \tag{8.7}$$

The drift μ^α is formed by taking convex combinations of the drifts for the two models

$$\mu^\alpha = (1 - \alpha)\mu^0 + \alpha\mu^1 = \mu^0 + \alpha\Lambda g;$$

---

[23] This can be seen by writing $\Lambda g = \Sigma\Lambda'(\Lambda'\Lambda)^{-1}g$.

the diffusion matrix Σ is the common diffusion matrix for the two models, and the jump operator $\mathcal{N}^\alpha$ is given by:

$$\mathcal{N}^\alpha \phi(x) = \int [h(y,x)]^\alpha [\phi(y) - \phi(x)] \, \eta^0(dy|x)$$

Finally, the rate ρ^α is nonnegative and state dependent and is the sum of contributions from the diffusion and jump components:

$$\rho^\alpha(x) = \rho_d^\alpha(x) + \rho_n^\alpha(x). \tag{8.8}$$

The diffusion contribution

$$\rho_d^\alpha(x) = \frac{(1-\alpha)\,\alpha}{2} g(x)' \, g(x) \tag{8.9}$$

is a positive semi-definite quadratic form and the jump contribution

$$\rho_n^\alpha(x) = \int \left((1-\alpha) + \alpha h(y,x) - [h(y,x)]^\alpha \right) \eta^0(dy|x) \tag{8.10}$$

is positive because the tangent line to the concave function $(h)^\alpha$ at $h = 1$ must lie above the function.

Theorem 8.1. *(Newman, 1973 and Newman and Stuck, 1979). The generator of the positive contraction semigroup $\{\mathcal{K}_t^\alpha : t \geq 0\}$ on C is given by (8.7).*

Thus we can interpret the generator $\mathcal{G}^\alpha$ that bounds the detection error probabilities as follows. Take model zero to be the approximating model, and model one to be some other competing model. We use $0 < \alpha < 1$ to build a mixed diffusion-jump process from the pair (g^α, h^α) where $g^\alpha \doteq \alpha g$ and $h^\alpha \doteq (h)^\alpha$.

Use the notation E^α to depict the associated expectation operator. Then

$$\text{av error} \leq \frac{1}{2} E^\alpha \left[\exp\left(-\int_0^N \rho^\alpha(x_t) \, dt \right) \Big| x_0 = x \right]. \tag{8.11}$$

Of particular interest to us is formula (8.8) for ρ^α, which can be interpreted as a local statistical discrimination rate between models. In the case of two diffusion processes, this measure is a state-dependent counterpart to formula (8.6) in the example presented in section 8.3.1. The diffusion component of the rate is maximized by setting $\alpha = 1/2$. But when jump components are also present, $\alpha = 1/2$ will not necessarily give the maximal rate. Theorem 8.1 completes the fourth row of Table 4.1.

8.4. Detection statistics and robustness

Formulas (8.9) and (8.10) show how the local Chernoff entropy rate is closely related to the conditional relative entropy measure that we used to formulate robust control problems. In particular, the conditional relative entropy rate ϵ in continuous time satisfies

$$\epsilon = -\frac{d\rho^\alpha}{d\alpha}\bigg|_{\alpha=1}.$$

In the case of diffusion process, the Chernoff rate equals the conditional relative entropy rate without the proportionality factor $\alpha(1-\alpha)$.[24]

8.5. Further discussion

In some ways, the statistical decision problem posed above is too simple. It entails a pairwise comparison of *ex ante* equally likely models and gives rise to a statistical measure of distance. That the contenders are both Markov models greatly simplifies the bounds on the probabilities of making a mistake when choosing between models. The implicit loss function that justifies model choice based on the maximal posterior probabilities is symmetric (*e.g.* see Chow, 1957). Finally, the detection problem compels the decision-maker to select a specific model after a fixed amount of data have been gathered.

Bounds like Chernoff's can be obtained when there are more than two models and also when the decision problem is extended to allow waiting for more data before making a decision (*e.g.* see Hellman and Raviv, 1970 and Moscarini and Smith, 2002).[25] Like our problem, these generalizations can be posed as Bayesian problems with explicit loss functions and prior probabilities.

While the statistical decision problem posed here is by design too simple, we nevertheless find it useful in bounding a reasonable taste for robustness. The hypothesis of rational expectations instructs agents and the model builder to eliminate as misspecified those models that are detectable from *infinite* histories of data. Chernoff entropy gives us one way to extend rational expectations by asking agents to exclude specifications rejected by *finite* histories of data but to contemplate alternative models that are difficult to detect from *finite* histories of data. When Chernoff entropy is small, it is challenging to choose between competing models on purely statistical grounds.

---

[24] The distributions associated with these rates differ, however. Bound (8.11) also uses a Markov evolution indexed by α, whereas we used the $\alpha = 1$ model in evaluating the robust control objective.

[25] In particular, Moscarini and Smith (2002) consider Bayesian decision problems with a more general but finite set of models and actions. Although they restrict their analysis to i.i.d. data, they obtain a more refined characterization of the large sample consequences of accumulating information. Chernoff entropy is a key ingredient in their analysis too.

8.6. Detection and plausibility

In section 6.1.2, we reinterpreted the equilibrium allocation under a preference for robustness as one that instead would be chosen by a Bayesian social planner who holds fast to a particular model that differs from the approximating model. If the approximating model is actually true, then this artificial Bayesian planner has a false model of forcing processes, one that enough data should disabuse him of. However, our detection probability tools let us keep the Bayesian planner's model sufficiently close to the approximating model that more data than are available would be needed to detect that the approximating model is really better. That means that a rational expectations econometrician would have a difficult time distinguishing the forcing process under the approximating model from the Bayesian planner's model. Nevertheless, that the Bayesian planner uses such a nearby model can have important quantitative implications for decisions and/or asset prices. We demonstrate such effects on asset prices in section 9.[26]

9. Entropy and the market price of uncertainty

In comparing different discrete time representative agent economies with robust decision rules, HSW computationally uncovered a connection between the market price of uncertainty and the detection error probabilities for distinguishing between a planner's approximating and worst case models. The connection was so tight that the market price of uncertainty could be expressed as nearly the *same* linear function of the detection error probability, regardless of the details of the model specification.[27] HSW used this fact to calibrate θ's, which differed across different models because the relationship between detection error probabilities and θ *did* depend on the detailed model dynamics.

As emphasized in section 2, the tight link that we have formally established between the semigroup for pricing under robustness and the semigroup for detection error probability bounds provides the key to understanding HSW's empirical finding, *provided* that the detection error probability bounds do a good enough job of describing the actual detection error probabilities. Subject to that proviso, our formal results thus provide a way of moving directly from a specification of a sample size and a detection error probability to a prediction of the market price of uncertainty that transcends details of the model.

---

[26] For heterogeneous agent economies, worst case models can differ across agents because their preferences differ. But an argument like the one in the text could still be used to keep each agent's worst case model close to the approximating model as measured by detection probabilities.

[27] See Figure 8 in HSW.

Partly to explore the quality of the information in our detection error probability bounds, this section takes three distinct example economies and shows within each of them that the probability bound is quite informative and that consequently the links between the detection error probability and the market price of uncertainty are very close across the three different models. All three of our examples use diffusion models, so that the formulas summarized in section 2 apply. Recall from formula (8.9) that for the case of a diffusion the local Chernoff rate for discriminating between the approximating model and the worst-case model is

$$\alpha \left(1 - \alpha\right) \frac{|\hat{g}|^2}{2}, \tag{9.1}$$

which is maximized by setting $\alpha = .5$. Small values of the rate suggest that the competing models would be difficult to detect using time series statistical methods.

For a diffusion, we have seen how the price vector for the Brownian increments can be decomposed as $\bar{g} + \hat{g}$. In the standard model without robustness, the conditional slope of a mean-standard deviation frontier is the absolute value of the factor risk price vector, $|\bar{g}|$, but with robustness it is $|\bar{g} + \hat{g}|$, where $\hat{g}$ is the part attributable to aversion to model uncertainty. One possible statement of the equity-premium puzzle is that standard models imply a much smaller slope than is found in the data because plausible risk aversion coefficients imply a small value for $\bar{g}$. This conclusion extends beyond the comparison of stocks and bonds and is also apparent in equity return heterogeneity. See Hansen and Jagannathan (1991), Cochrane and Hansen (1992), and Cochrane (1997) for discussions.

In this section we explore the potential contribution from model uncertainty. In particular, for three models we compute $|\hat{g}|$, the associated bounds on detection error probabilities, and the detection error probabilities themselves. The three models are: (1) a generic one where the worst-case drift distortion $\hat{g}$ is independent of x; (2) our robust interpretation of a model of Bansal and Yaron (2000) in which $\hat{g}$ is again independent of x but where its magnitude depends on whether a low frequency component of growth is present; and (3) a continuous time version of HST's equilibrium permanent income model, in which $\hat{g}$ depends on x. Very similar relations between detection error probabilities and market prices of risk emerge across these three models, though they are associated with different values of θ.

mpu (= $\lvert \hat{g} \rvert$)	Chernoff rate ($\alpha = .5$)	probability bound	probability
.02	.0001	.495	.444
.04	.0002	.480	.389
.06	.0004	.457	.336
.08	.0008	.426	.286
.10	.0013	.389	.240
.12	.0018	.349	.198
.14	.0015	.306	.161
.16	.0032	.264	.129
.18	.0040	.222	.102
.20	.0050	.184	.079
.30	.0113	.053	.017
.40	.0200	.009	.002

Table 9.1: Prices of Model Uncertainty and Detection-Error Probabilities when $\hat{g}$ is independent of x, $N = 200$; mpu denotes the market price of model uncertainty, measured by $\lvert \hat{g} \rvert$. The Chernoff rate is given by (9.1).

9.1. Links among sample size, detection probabilities, and mpu when $\hat{g}$ is independent of x

In this section we assume that the approximating model is a diffusion and that the worst case model is such that $\hat{g}$ is independent of the Markov state x. Without knowing anything more about the model, we can quantitatively explore the links among the market price of model uncertainty (mpu $\equiv \lvert \hat{g} \rvert$) and the detection error probabilities.

For $N = 200$, Table 9.1 reports values of mpu $= \lvert \hat{g} \rvert$ together with Chernoff entropy for $\alpha = .5$, the associated probability-error bound (8.11), and the actual probability of detection on the left side of (8.11) (which we can calculate analytically in this case). The probability bounds and the probabilities are computed under the simplifying assumption that the drift and diffusion coefficients are constant, as in the example in section 8.3.1. With constant drift and diffusion coefficients, the log-likelihood ratio is normally distributed, which allows us easily to compute the actual detection-error probabilities.[28]

---

[28] Thinking of a quarter as the unit of time, we took the sample interval to be 200. Alternatively, we might have used a sample interval of 600 to link to monthly postwar data. The market prices of risk and model uncertainty are associated with specific time unit normalizations. Since, at least locally, drift coefficients and diffusion matrices scale linearly with the time unit, the market prices of risk and model uncertainty scale with the square root of the time unit.

The numbers in Table 9.1 indicate that market prices of uncertainty somewhat less than .2 are associated with misspecified models that are difficult to detect. However, market prices of uncertainty of .40 are associated with easily detectable alternative models. The table also reveals that although the probability bounds are weak, they display patterns similar to those of the actual probabilities.

Empirical estimates of the slope of the mean-standard deviation frontier are about .25 for quarterly data. Given the absence of volatility in aggregate consumption, risk considerations only explain a small component of the measured risk-return tradeoff using aggregate data (Hansen-Jagannathan, 1991). In contrast, our calculations suggest that concerns about statistically small amounts of model misspecification could account for a substantial component of the empirical estimates. The following subsections confirm that this quantitative conclusion transcends details of the specification of the approximating model.

9.2. Low frequency growth

Using a recursive utility specification, Bansal and Yaron (2000) study how low frequency components in consumption and/or dividends become encoded in risk premia. Here we take up Bansal and Yaron's theme that the frequency decomposition of risks matters but reinterpret their risk premia as reflecting model uncertainty.

Consider a pure endowment economy where the state x_{1t} driving the consumption endowment $\exp(x_{1t})$ is governed by the following process:

$$
\begin{aligned}
dx_{1t} &= (.0020 + .0177x_{2t})\, dt + .0048 dB_t \\
dx_{2t} &= -.0263x_{2t}dt + .0312 dB_t.
\end{aligned}
\tag{9.2}
$$

The logarithm of the consumption endowment has a constant component of the growth rate of .002 and a time varying component of the growth rate of x_{2t}; x_{2t} has mean zero and is stationary but is highly temporally dependent. Relative to the i.i.d. specification that would be obtained by attaching a coefficient of zero on x_{2t} in the first equation of (9.2), the inclusion of the x_{2t} component alters the long run properties of consumption growth.

We calibrated the state evolution equation (9.2) by taking a discrete-time consumption process that was fit by Bansal and Yaron and embedding it in a continuous state-space model.[29] We accomplished this by sequentially applying the conversions `tf`, `d2c`, and `ss` in the MATLAB control toolbox.

---

[29] We base our calculations on an ARMA model for consumption growth, namely, $\log c_t - \log c_{t-1} = .002 + [(1 - .860L)/(1 - .974L)](.0048)\nu_t$ reported by Bansal and Yaron (2000) where $\{\nu_t : t \geq 0\}$ is a serially uncorrelated shock with mean zero and unit variance.
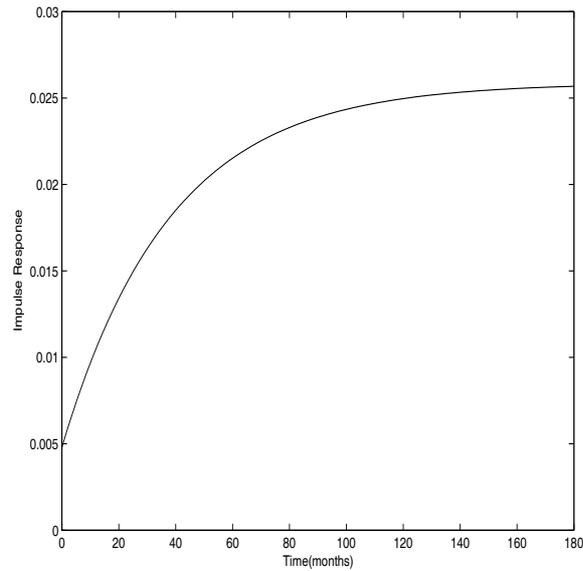
Figure 9.1: Impulse response function for the logarithm of the consumption endowment x_{1t} to a Brownian increment.
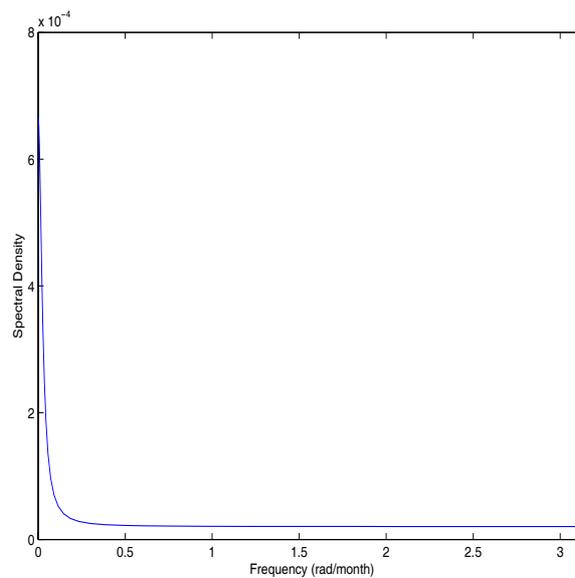
Figure 9.2: Spectral density function for the consumption growth rate dx_{1t}.

The impulse response of log consumption to a Brownian motion shock, which is portrayed in Fig. 9.1, and the spectral density function, which is shown in Fig. 9.2, both show the persistence in consumption growth. The low frequency component is a key feature in the analysis of Bansal and Yaron (2000). The impulse response function converges to its supremum from below. It takes about ten years before the impulse response function is

close to its supremum. Corresponding to this behavior of the impulse response function, there is a narrow peak in the spectral density of consumption growth at frequency zero, with the spectral density being much greater than its infimum only for frequencies with periods of more than ten years.

In what follows, we will also compute model uncertainty premia for an alternative economy in which the coefficient on x_{2t} in the first equation of (9.2) is set to zero. We calibrate this economy so that the resulting spectral density for consumption growth is flat at the same level as the infimum depicted in Fig. 9.2 and so that the corresponding impulse response function is also flat at the same initial response reported in Fig. 9.1. Because we have reduced the long run variation by eliminating x_{2t} from the first equation, we should expect to see smaller risk-premia in this economy.

Suppose that the instantaneous utility function is logarithmic. Then the value function implied by Theorem 5.1 is linear in x, so we can write V as $V(x) = v_0 + v_1 x_1 + v_2 x_2$. The distortion in the Brownian motion is given by the following special case of equation (6.3)

$$\hat{g} = -\frac{1}{\theta} \left(.0048 v_1 + .0312 v_2\right)$$

and is independent of the state x. The coefficients on the v_i's are the volatility coefficients attached to dB in equation (9.2).[30]

Since $\hat{g}$ is constant, the worst-case model implies the same impulse response functions but different mean growth rates in consumption. Larger values of $1/\theta$ lead to more negative values of the drift $\hat{g}$. We have to compute this mapping numerically because the value function itself depends on θ. Fig. 9.3 reports $\hat{g}$ as a function of $1/\theta$ (the $\hat{g}$'s are negative). Larger values of $|\hat{g}|$ imply larger values of the rate of Chernoff entropy. As in the previous example this rate is constant, and the probability bounds in Table 9.1 continue to apply to this economy.[31]

The instantaneous risk free rate for this economy is:

$$r_t^f = \delta + .0020 + .1154 x_{2t} + \sigma_1 \hat{g}_t - \frac{(\sigma_1)^2}{2}$$

where $\sigma_1 = .0048$ is the coefficient on the Brownian motion in the evolution equation for x_{1t}. Our calculations hold the risk free rate fixed as we change θ. This requires that we adjust the subjective discount rate. The predictability of consumption growth emanating from the state variable x_{2t} makes the risk free rate vary over time.

---

[30] The state independence implies that we can also interpret these calculations as coming from a decision problem in which $|\hat{g}|$ is constrained to be less than or equal to the same constant for each date and state. This specification satisfies the dynamic consistency postulate advocated by Epstein and Schneider (2001).

[31] However, the exact probability calculations reported in Table 9.1 will not apply to the present economy.
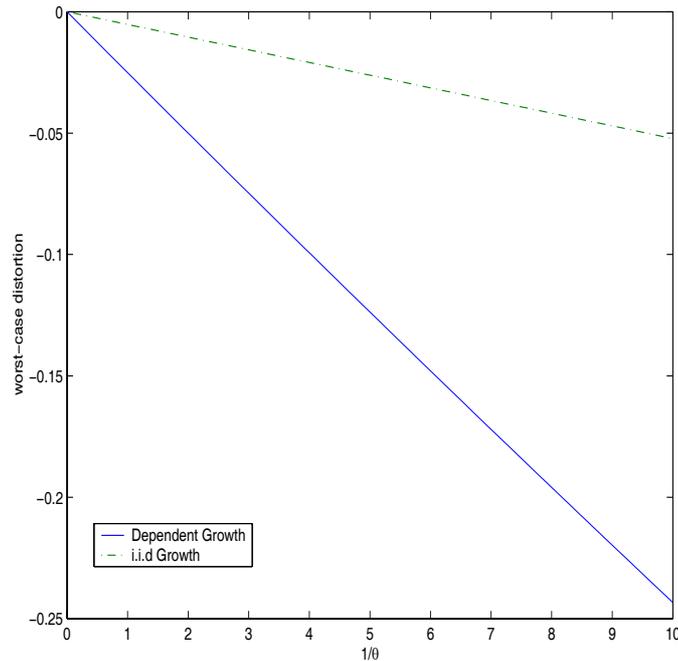
Figure 9.3: Drift distortion $\hat{g}$ for the Brownian motion. This distortion is plotted as a function of $1/\theta$. For comparison, the drift distortions are also given for the economy in which the first-difference in the logarithm of consumption is i.i.d. In generating this figure, we set the discount parameter δ so that the mean risk-free rate remained at 3.5% when we changed θ. In the model with temporally dependent growth the volatility of the risk-free rate is .38%. The risk-free rate is constant in the model with i.i.d. growth rates.

9.3. Risk-sensitivity and calibration of θ

This economy has a risk-sensitive interpretation along the lines asserted by Tallarini (2000) in his analysis of business cycle models, one that makes $1/\theta$ a risk-sensitivity parameter that imputes risk aversion to the continuation utility and makes $\hat{g}$ become the incremental contribution to risk aversion contributed by this risk adjustment. Under Tallarini's risk-sensitivity interpretation, $1/\theta$ is a risk aversion parameter, and as such is presumably fixed across environments.

Fig. 9.3 shows that holding θ fixed but changing the consumption endowment process changes the detection error rates. For a given θ, the implied worst case model could be more easily detected when there is positive persistence in the endowment growth rate. On the robustness interpretation, such detection error calculations suggest that θ should not be taken to be invariant across environments. However, on the risk sensitivity interpretation, θ should presumably be held fixed across environments. Thus while concerns about risk and robustness have similar predictions within a given environment, calibrations of θ can

depend on whether it is interpreted as a risk-sensitivity parameter that is fixed across environments, or a robustness parameter to be adjusted across environments depending on how detection error probabilities differ across environments.

9.4. Permanent income economy
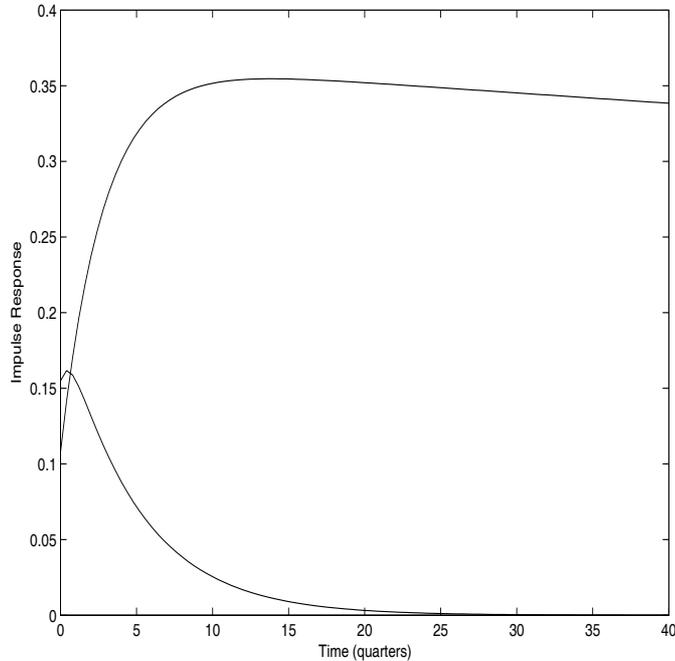
Figure 9.4: This figure gives the impulse response functions for the two income processes to two independent Brownian motion shocks.

Our previous two examples make Chernoff entropy be independent of the Markov state. In our third example, we computed detection-error bounds and detection-error probabilities for the version of HST's robust permanent income model that includes habits.[32] The Chernoff entropies are state dependent for this model, but the probability bounds can still be computed numerically. We used the parameter values from HST's discrete-time model to form an approximate continuous-time robust permanent income model, again using conversions in the MATLAB control toolbox. HST allowed for two independent income components when estimating their model. The impulse responses for the continuous-time version of the model are depicted in Fig. 9.4. The responses are to independent Brownian motion shocks. One of the processes is much more persistent than the other one. That persistent process is the one that challenges a permanent-income-style saver.

---

[32] HST estimated versions of this model with and without habits.

When we change the robustness parameter θ, we alter the subjective discount rate in a way that completely offsets the precautionary motive for saving in HST's economy and its continuous-time counterpart, so that consumption and investment profiles and real interest rates remain fixed.[33] It happens that the worst case $\hat{g}$ vector is proportional to the marginal utility of consumption and therefore is highly persistent. This outcome reflects that the decision rule for the permanent income model is well designed to protect the consumer against transient fluctuations, but that it is vulnerable to model misspecifications that are highly persistent under the approximating model. Under the approximating model, the marginal utility process is a martingale, but the (constrained) worst case model makes this process become an explosive scalar autoregression. The choice of θ determines the magnitude of the explosive autoregressive coefficient for the marginal utility process. The distortion is concentrated primarily in the persistent income component. Fig. 9.5 compares the impulse response of the distorted income process to that of the income process under the approximating model. Under the distorted model there is considerably more long-run variation. Decreasing θ increases this variation.
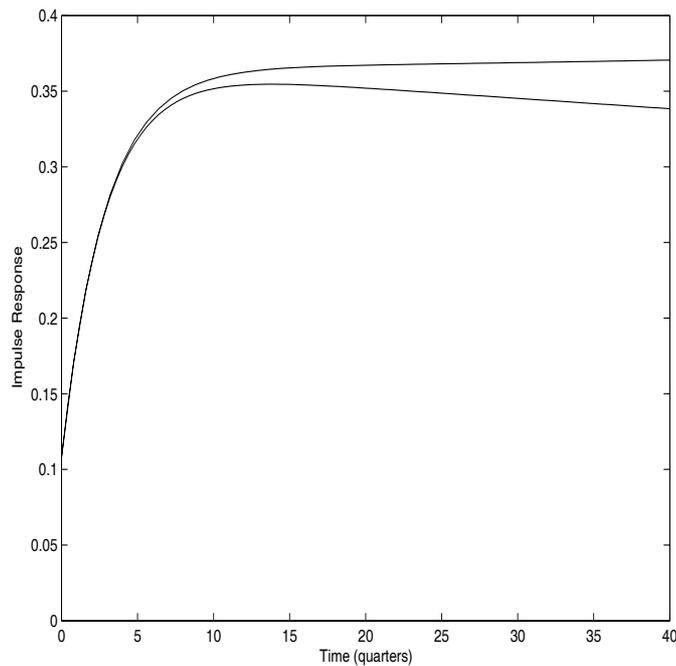
Figure 9.5: This figure gives impulse response functions for the persistent income process under the approximating model and the model for which mpu = .16. The more enduring response is from the distorted model. The implied risk-free rate is constant and identical across the two models.

---

[33] See HST and HSW for a proof in a discrete time setting of an observational equivalence proposition that identifies a locus of (δ, θ) pairs that are observationally equivalent for equilibrium quantities.

mpu $(= \lvert \hat{g} \rvert)$	$\frac{1}{\theta}$	probability bound	simulated probabilities
.02	1.76	.495	.446
.04	3.51	.479	.388
.06	5.27	.453	.334
.08	7.02	.416	.282
.10	8.78	.372	.231
.12	10.53	.323	.185
.14	12.29	.271	.143
.16	14.04	.220	.099
.18	15.80	.171	.072
.20	17.56	.128	.054
.30	26.33	.015	.004
.40	35.11	.0004	.000

Table 9.2: Prices of Model Uncertainty and Detection-Error Probabilities for the Permanent Income Model, $N = 200$. The probability bounds were computed numerically and optimized over the choice of α. The simulated probabilities were computed by simulating discrete-approximations with a time interval of length .01 of a quarter and with 1000 replications.

Table 9.2 gives the detection-error probabilities corresponding to those reported in Table 9.1. In this case, the Chernoff entropy is state dependent, leading us to compute the right side of (8.11) numerically in Table 9.2.[34] The values of θ are set so that the market prices of uncertainty match those in Table 9.1.

Despite the different structures of the two models, for mpus up to about .12, the results in Table 9.2 are close to those of Table 9.1, though the detection-error probabilities are a little bit lower in Table 9.2. However, the probabilities do decay faster in the tails for the permanent income model. As noted, the model in Table 9.1 has no dependence of $\hat{g}$ on the state x. As shown by HST, the fluctuations in $\lvert \hat{g} \rvert$ are quite small relative to its level, which contributes to the similarity of the results for low mpus. It remains the case that statistical detection is difficult for market prices of uncertainty up to about half that of the empirically estimated magnitude.

---

[34] See Hansen and Sargent (2004) for computational details. In computing the probability bounds, we chose the following initial conditions: we set the initial marginal utility of consumption to 15.75 and the mean zero components of the income processes to their unconditional means of zero.

We conclude that a preference for robustness that is calibrated to plausible values of the detection error probabilities can account for a substantial fraction, but not all, of estimated equity premia, and that this fraction is impervious to details of the model specification. Introducing other features into a model can help to account more fully for the steep slope of the frontier, but they would have to work through $\bar{g}$ and not the robustness component $\hat{g}$. For example, market frictions or alterations to preferences affect the $\bar{g}$ component.[35]

10. Concluding remarks

In this paper we have applied tools from the mathematical theory of continuous-time Markov processes to explore the connections between model misspecification, risk, robustness, and model detection. We used these tools in conjunction with a decision theory that captures the idea that the decision maker views his model as an approximation. An outcome of the decision making process is a constrained worst-case model that the decision maker uses to create a robust decision rule. In an equilibrium of a representative agent economy with such robust decision makers, the worst case model of a fictitious robust planner coincides with the worst case model of a representative private agent. These worst case models are endogenous outcomes of decision problems and give rise to model uncertainty premia in security markets. By adjusting constraints or penalties, the worst case model can be designed to be close to the approximating model in the sense that it is difficult statistically to discriminate it from the original approximating model. These same penalties or constraints limit model uncertainty premia. Since mean returns are hard to estimate, nontrivial departures from an approximating model are difficult to detect. As a consequence, within limits imposed by statistical detection, there can still be sizable model uncertainty premia in security prices.

There is another maybe less radical interpretation of our calculations. A rational expectations econometrician is compelled to specify forcing processes, often without much guidance except from statistics, for example, taking the form of a least squares estimate of an historical vector autoregression. The exploration of worst case models can be thought of as suggesting an alternative way of specifying forcing process dynamics that is inherently more pessimistic, but that leads to statistically similar processes. Not surprisingly, changing forcing processes can change equilibrium outcomes. More interesting is the fact that

---

[35] We find it fruitful to explore concern about model uncertainty because these other model modifications are themselves only partially successful. To account fully for the market price of risk, Campbell and Cochrane (1999) adopt specifications with substantial risk aversion during recessions. Constantinides and Duffie (1996) accommodate fully the high market prices of risk by attributing empirically implausible consumption volatility to individual consumers (see Cochrane, 1997). Finally, Heaton and Lucas (1996) show that reasonable amounts of proportional transaction costs can explain only about half of the equity premium puzzle.

sometimes subtle changes in perceived forcing processes can have quantitatively important consequences on equilibrium prices.

Our altered decision problem introduces a robustness parameter that we restrict by using detection probabilities. In assessing reasonable amounts of risk aversion in stochastic, general equilibrium models, it is common to explore preferences for hypothetical gambles in the manner of Pratt (1964). In assessing reasonable preferences for robustness, we propose using large sample detection probabilities for a hypothetical model selection problem. We envision a decision-maker as choosing to be robust to departures that are difficult to detect statistically. Of course, using detection probabilities in this way is only one way to discipline robustness. Exploration of other consequences of being robust, such as utility losses, would also be interesting.

We see three important extensions to our current investigation. Like builders of rational expectations models, we have side-stepped the issue of how decision-makers select an approximating model. Following the literature on robust control, we envision this approximating model to be analytically tractable, yet to be regarded by the decision maker as not providing a correct model of the evolution of the state vector. The misspecifications we have in mind are small in a statistical sense but can otherwise be quite diverse. Just as we have not formally modelled how agents learned the approximating model, neither have we formally justified why they do not bother to learn about potentially complicated misspecifications of that model. Incorporating forms of learning would be an important extension of our work.

The equilibrium calculations in our model currently exploit the representative agent paradigm in an essential way. Reinterpreting our calculations as applying to a multiple agent setting is straightforward in some circumstances (see Tallarini, 2000), but in general, even with complete security market structures, multiple agent versions of our model look fundamentally different from their single-agent counterparts (see Anderson, 1998). Thus, the introduction of heterogeneous decision makers could lead to new insights about the impact of concerns about robustness for market outcomes.

Finally, while the examples of section 9 all were based on diffusions, the effects of concern about robustness are likely to be particularly important in environments with large shocks that occur infrequently, so that we anticipate that our modelling of robustness in the presence of jump components will be useful.

References

Anderson, E. W. (1998), "Uncertainty and the Dynamics of Pareto Optimal Allocations," University of Chicago Dissertation.

Anderson, E. W., L. P. Hansen and T. J. Sargent (1998), "Risk and Robustness in Equilibrium," manuscript.

Bansal, R. and A. Yaron (2000), "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," NBER Working Paper.

Blackwell, D. and M. Girshick (1954), *Theory of Games and Statistical Decisions*, Wiley: New York.

Bray, M. (1982), "Learning, Estimation, and the Stability of Rational Expectations," *Journal of Economic Theory*, 26:2, 318-39.

Breeden, D. T. (1979), "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities," *Journal of Financial Economics* 7, 265-296.

Brock, W. A. (1979), "An Integration of Stochastic Growth Theory and the Theory of Finance, Part I: The Growth Model," *General Equilibrium, Growth and Trade*.

Brock, W. A, (1982), "Asset Pricing in a Production Economy," In *The Economics of Information and Uncertainty*, J. J. McCall, editor, Chicago: University of Chicago Press, for the National Bureau of Economic Research.

Brock, W. A. and M. J. P. Magill (1979), "Dynamics Under Uncertainty," *Econometrica* 47, 843-868.

Brock, W. A. and L.J. Mirman (1972), "Optimal Growth Under Uncertainty: The Case of Discounting," *Journal of Economic Theory*, 4:3, 479-513.

Cameron, H. R. and W.T. Martin (1947), "The Behavior of Measure and Measurability under Change of Scale in Wiener Space," *Bulletin of the American Mathematics Society*, 53, 130-137.

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

Chen, Z. and L. G. Epstein (2001), "Ambiguity, Risk and Asset Returns in Continuous Time," manuscript.

Chernoff, H. (1952), "A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on the Sum of Observations," *Annals of Mathematical Statistics*, 23, 493-507.

Cho I.-K., Williams. N. and T.J. Sargent (2002), "Escaping Nash Inflation," *Review of Economic Studies*, 69:1, 1-40.

Chow, C. K. (1957), "An Optimum Character Recognition System Using Decision Functions," *Institute of Radio Engineers (IRE) Transactions on Electronic Computers*, 6, 247-254.

Cochrane, J. H. (1997), "Where is the Market Going? Uncertain Facts and Novel Theories," *Federal Reserve Bank of Chicago Economic Perspectives*, 21:6, 3-37.

Cochrane, J. H. and L. P. Hansen (1992), "Asset Pricing Lessons for Macroeconomics," 1992 NBER Macroeconomics Annual, O. Blanchard and S. Fischer (eds.), 1115-1165.

Constantinides, G. M. and D. Duffie (1996), "Asset Pricing with Heterogeneous Consumers," *Journal of Political Economy*, 104, 219-240.

Cox, J. C., J. E. Ingersoll, Jr. and S. A. Ross (1985), "An Intertemporal General Equilibrium Model of Asset Prices," *Econometrica*, 53, 363-384.

Csiszar, I. (1991), "Why Least Squares and Maximum Entropy? An Axiomatic approach to Inference for Linear Inverse Problems," *The Annals of Statistics* 19:4, 2032-2066.

Dow, J. and S. R. C. Werlang (1994), "Learning under Knightian Uncertainty: The Law of Large Numbers for Non-additive Probabilities, manuscript

Duffie, D. and L.G. Epstein (1992), "Stochastic Differential Utility," *Econometrica*, 60:2, 353-394.

Duffie, D. and P.L. Lions (1992), "PDE Solutions of Stochastic Differential Utility," *Journal of Mathematical Economics*, 21, 577-606.

Duffie, D. and C. Skiadas (1994), "Continuous-Time Security Pricing: A Utility Gradient Approach," *Journal of Mathematical Economics*, 23, 107-131.

Dupuis, P. and R. S. Ellis (1997), *A Weak Convergence Approach to the Theory of Large Deviations*, New York: John Wiley and Sons.

Dynkin, E. B. (1956), "Markov Processes and Semigroups of Operators," *Theory of Probability and its Applications* 1, 22-33.

Elliot, R. J., L. Aggoun, and J. B. Moore (1995), *Hidden Markov Models: Estimation and Control*, Springer: New York.

Ellsberg, D. (1961), "Risk, Ambiguity, and the Savage Axions," *Quarterly Journal of Economics*, 75:4, 643-669.

Epstein, L.G. and T. Wang (1994), "Intertemporal Asset Pricing Under Knightian Uncertainty," *Econometrica*, 62:3, 283-322.

Epstein, L.G. and S.E. Zin (1989), "Substitution, Risk Aversion and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework," *Econometrica*, 57:4, 937-969.

Ethier, S.N and T.G. Kurtz (1986), *Markov Processes: Characterization and Convergence*, New York: John Wiley and Sons.

Evans, G. W. and S. Honkapohja (2001), *Learning and expectations in macroeconomics*, Princeton: Princeton University Press.

Fleming, W. H. and H. M. Soner (1991), *Controlled Markov Processes and Viscosity Solutions*, New York: Springer-Verlag

Fudenberg, D. and D. K. Levine (1998), *The Theory of Learning in Games*, Cambridge: MIT Press.

Gilboa, I. and D. Schmeidler (1989), "Maxmin Expected Utility with Non-unique Prior," *Journal of Mathematical Economics* 18, 141-153.

Hansen, L.P. and R. Jagannathan (1991), "Implications of Security Market Data for Models of Dynamic Economies," *Journal of Political Economy*, 99:2, 225-262.

Hansen, L.P. and T.J. Sargent (1993), "Seasonality and Approximation Errors in Rational Expectations Models," *Journal of Political Economy*, 55, 21-55

Hansen, L.P. and T.J. Sargent (1995), "Discounted Linear Exponential Quadratic Gaussian Control," *IEEE Transactions on Automatic Control*, 40:5, 968-971.

Hansen, L.P. and T.J. Sargent (2001), "Time Inconsistency of Robust Control?", manuscript.

Hansen, L.P. and T.J. Sargent (2003), "Decentralizing economies with preferences for robustness," University of Chicago and New York University, manuscript.

Hansen, L.P. and T.J. Sargent (2004), *Robust Control and Economic Model Uncertainty*, Princeton: Princeton University Press.

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

Hansen, L.P., T.J. Sargent, G.A. Turmuhambetova, and N. Williams (2002), "Robustness and Uncertainty Aversion," manuscript.

Hansen, L.P., T.J. Sargent, and N.E. Wang (2002), "Robust Permanent Income and Pricing with Filtering," *Macroeconomic Dynamics*, 6, February, pp. 40–84.

Hansen, L.P. and J.A. Scheinkman (1995), "Back to the Future: Generating Moment Implications for Continuous-Time Markov Processes," *Econometrica*, 63, 767-804.

Hansen, L.P. and J. Scheinkman (2002), "Semigroup Pricing," manuscript.

Heaton, J. C. and D. J. Lucas (1996), "Evaluating the Effects of Incomplete Markets on Risk Sharing and Asset Pricing," *Journal of Political Economy*, 104, 443-487.

Hellman, M. E. and J. Raviv, (1970), "Probability Error, Equivocation, and the Chernoff Bound," *IEEE Transactions on Information Theory*, 16, 368-372.

James, M.R. (1992), "Asymptotic Analysis of Nonlinear Stochastic Risk-Sensitive Control and Differential Games," *Mathematics of Control, Signals, and Systems*, 5, 401-417.

Jorgenson, D.W. (1967), "Discussion," *The American Economic Review*, 57:2, Papers and Proceedings, 557-559.

Knight, F.H. (1921), *Risk, Uncertainty and Profit*, Boston: Houghton Mifflin.

Kreps, D. (1998), "Anticipated Utility and Dynamic Choice," Kalai, E. and Kamien, M. I. (eds.), *Frontiers of research in economic theory: The Nancy L. Schwartz Memorial Lectures*, Econometric Society Monographs, no. 29, Cambridge: Cambridge University Press, 242-74.

Kunita (1969), H. "Absolute Continuity of Markov Processes and Generators," *Nagoya Math. Journal*, 36, pages 1 -26

Kurz, M. (1997), *Endogenous Economic Fluctuations: Studies in the Theory of Rational Beliefs*, New York: Springer.

Lei, C. I. (2000), *Why Don't Investors have Large Positions in Stocks? A Robustness Perspective*, unpublished Ph.D. Dissertation, University of Chicago.

Lucas, R.E., Jr (1976), "Econometric Policy Evaluation: A Critique," *Journal of Monetary Economics*, 1:2, 19-46.

Lucas, R. E., Jr. (1978), "Asset Prices in an Exchange Economy," *Econometrica*, 46:6, 1429-1445.

Lucas, R.E., Jr. and E.C. Prescott (1971), "Investment under Uncertainty," *Econometrica*, 39:5, 659-681.

Maenhout, P. (2001), "Robust Portfolio Rules, Hedging and Asset Pricing", manuscript, INSEAD.

Merton, R. (1973), "An Intertemporal Capital Asset Pricing Model," *Econometrica*, 41, 867-888.

Moscarini, G. and L. Smith (2002), "The Law of Large Demand for Information," manuscript, forthcoming in Econometrica.

Newman. C. M. (1973), "The orthogonality of independent increment processes," in *Topics in Probability Theory*, D.W. Stroock and S.R.S. Varadhan (eds.), 93-111, Courant Institute of Mathematical Sciences, N.Y.U.

Newman, C. M. and B. W. Stuck (1979), "Chernoff Bounds for Discriminating Between Two Markov Processes," *Stochastics*, 2, 139-153.

Petersen, I.R., M. R. James and P. Dupuis (2000), "Minimax Optimal Control of Stochastic Uncertain Systems with Relative Entropy Constraints," *IEEE Transactions on Automatic Control*, 45, 398-412.

Pratt, J. W. (1964), "Risk Aversion in the Small and Large," *Econometrica*, 32:1-2, 122-136.

Prescott, E.C. and R. Mehra (1980), "Recursive Competitive Equilibrium: The Case of Homogeneous Households," *Econometrica*, 48, 1365-1379.

Revuz D. and M. Yor (1994), *Continuous martingales and Brownian motion*, 2nd ed., New York: Springer Verlag.

Runolfsson, T. (1994), "The Equivalence between Infinite-horizon Optimal Control of Stochastic Systems with Exponential-of-integral Performance Index and Stochastic Differential Games," *IEEE Transactions on Automatic Control*, 39, 1551-1563.

Schroder, M. and C. Skiadas (1999), "Optimal Consumption and Portfolio Selection with Stochastic Differential Utility," *Journal of Economic Theory*, 89, 68-126.

Sims, C. A. (1993), "Rational Expectations Modelling with Seasonally Adjusted Data," *Journal of Econometrics*, 55.

Tallarini, Jr., T.D. (2000), "Risk-Sensitive Real Business Cycles," *Journal of Monetary Economics*, 45:3, 507-32.

A. Entropy Solution

This appendix contains the proof of Theorem 5.1.

Proof. To verify the conjectured solution, first note that the objective is additively separable in g and h. Moreover the objective for the quadratic portion in g is:

$$\theta \frac{g'g}{2} + g'\Lambda' \frac{\partial V}{\partial x}. \tag{A.1}$$

Minimizing this component of the objective by choice of g verifies the conjectured solution. The diffusion contribution to the optimized objective including $(A.1)$ is:

$$\mathcal{G}_d(V) - \frac{1}{2\theta} \left(\frac{\partial V}{\partial x}\right)' \Sigma \left(\frac{\partial V}{\partial x}\right) = -\theta \frac{\mathcal{G}_d\left[\exp\left(-V/\theta\right)\right]}{\exp\left(-V/\theta\right)}$$

where we are using the additive decomposition: $\mathcal{G} = \mathcal{G}_d + \mathcal{G}_n$. Very similar reasoning justifies the diffusion contribution to entropy formula $(5.5c)$.

Consider next the component of the objective that depends on h:

$$\theta \int \left[1 - h(y,x) + h(y,x)\log h(y,x)\right]\eta(dy|x) + \int h(y,x)\left[V(y) - V(x)\right]\eta(dy|x) \tag{A.2}$$

To verify[36] that this objective is minimized by $\hat{h}$, first use the fact that $1 - h + h\log h$ is convex and hence dominates the tangent line at $\hat{h}$. Thus

$$1 - h + h\log h \geq 1 - \hat{h} + \hat{h}\log\hat{h} + \log\hat{h}\left(h - \hat{h}\right).$$

This inequality continues to hold when we multiply by θ and integrate with respect to $\eta(dy|x)$. Thus

$$\theta\epsilon(h)(x) - \theta \int h(y,x)\log\hat{h}(y,x)\eta(dy|x) \geq \theta\epsilon\left(\hat{h}\right)(x) - \theta \int \hat{h}(y,x)\log\hat{h}(y,x)\eta(dy|x).$$

Substituting for $\log[\hat{h}(y,x)]$ shows that $\hat{h}$ minimizes $(A.2)$, and the resulting objective is:

$$\theta \int \left[1 - \hat{h}(y,x)\right]\eta(dy|x) = -\theta \frac{\mathcal{G}_n\exp\left(-V/\theta\right)}{\exp\left(-V/\theta\right)},$$

which establishes the jump contribution to $(5.5b)$. Very similar reasoning justifies the jump contribution to $(5.5c)$

∎

---

[36] In the special case in which the number of states is finite and the probability of jumping to any of these states is strictly positive, a direct proof that $\hat{h}$ is the minimizer is available. Abusing notation somewhat, let $\eta(y_i|x) > 0$ denote the probability that the state jumps to its ith possible value, y_i given that the current state is x. We can write the component of the objective that depends upon h (which is equivalent to equation $(A.2)$ in this special case) as

$$\theta + \sum_{i=1}^{n} \eta(y_i|x)\left[-\theta h(y_i,x) + \theta h(y_i,x)\log h(y_i,x) + h(y_i,x)V(y_i) - h(y_i,x)V(x)\right].$$

Differentiating this expression with respect to $h(y_i,x)$ yields $\eta(y_i|x)\left[\theta\log h(y_i,x) + V(y_i) - V(x)\right]$. Setting this derivative to zero and solving for $h(y_i,x)$ yields $h(y_i,x) = \exp[\frac{V(x)-V(y_i)}{\theta}]$, which is the formula for $\hat{h}$ given in the text.