

# Robustness, Estimation, and Detection

Lars Peter Hansen

University of Chicago and NBER

Thomas J. Sargent

New York University and Hoover Institution

October 3, 2009

## Abstract

We propose adjustments for model ambiguity that survive in continuous-time limits. Our formulations emerge from continuous-time versions of two discrete-time recursive models that Hansen and Sargent (2007) used for a decision maker who knows neither models nor distributions of unobserved states. We use expectations of likelihood ratios (*relative entropies*) to measure concerns about model misspecification and exploit the role of relative entropies in statistical methods for using historical data to discriminate between probability models. Statistical detection motivates our adjustment for model ambiguity and also helps us calibrate parameters that measure model ambiguity.

## 1 Introduction

This paper proposes two continuous-time recursive specifications of robust control problems with hidden state variables such as unknown parameters. Each is a continuous-time limit of a discrete-time problem proposed by Hansen and Sargent (2007), the first being a limiting version of recursions (17)-(18) in that paper, the second being a version of recursion (19). The first is a recursive counterpart to formulations in the robust control literature that in effect condition a continue value function on hidden states, while the second reinterprets the recursive utility model of Kreps and Porteus (1978) and Klibanoff et al. (2005, 2009) in terms of concerns about robustness of a

continuation value function that conditions only on observed information. Both formulations suggest ways to quantify a decision maker’s model ambiguity by constructing a worst-case model and assessing its statistical discrepancy from a benchmark model. A benchmark model is the decision maker’s working or approximating model. Because he does not trust it, the decision maker subjects his benchmark model to a robustness analysis by considering other statistically nearby models too. Promising empirical results in Ju and Miao (2007), Collard et al. (2008), and Hansen and Sargent (2008a) make us want to understand more fully what degrees of robustness or ambiguity aversion are *a priori* reasonable.

An important part of our analysis is how we modify and apply the continuous-time methods developed by Newman and Stuck (1979) and Anderson et al. (2003) for characterizing statistical discrepancies via an approach invented by Chernoff (1952). Important by-products of these statistical model detection calculations are ways to formulate entropy penalties that can allow us to represent preferences that capture model ambiguity and robustness concerns in continuous-time models. Our approach opens the way to using statistical measures of model discrimination to calibrate concerns about ambiguity within economic models.

As in Hansen and Sargent (2007), there is a relation between the smooth ambiguity models of Klibanoff et al. (2005, 2009) and our continuous time version of the discrete time recursion (19) of Hansen and Sargent (2007). By design, model ambiguity survives in our limiting formulations. As a consequence, our continuous-time formulations provide a response to a challenge that Skiadas (2009) posed for recursive formulations of smooth ambiguity. Skiadas (2009) showed how a smooth concern for ambiguity vanishes in a continuous-time limit that is distinct from ours. While our statistical detection motivation leads us to explore a different continuous-time limit than the one criticized by Skiadas, the sense in which our limiting preferences remains smooth is delicate. It *is* smooth, in the sense of taking a weighted average of a *derivative* of a continuation value (but not its *level*) using probabilities as weights.<sup>1</sup>

Sections 2 and 3 describe the stochastic model, information structure, and a linear filtering problem. Section 4 describes alternative representations of entropy and establishes the important result that entropy over alternative posterior distributions

---

<sup>1</sup>The smoothness in Klibanoff et al. (2005, 2009) refers to averaging using probabilities as weights. Expressing ambiguity in this way avoids the kinks in indifference curves that are present when ambiguity is expressed as in Gilboa and Schmeidler (1989).

is the concept pertinent for decision making. A peculiar consequence is that corrections representing smooth ambiguity over hidden states vanish in a continuous time limit. We trace this outcome to how different contributions to entropy are scaled with respect to the passage of time. Section 5 describes two robust control problems that sustain concerns about model ambiguity even in their continuous-time limits. A key here is how we rescale the contributions to entropy coming from unknown distributions of hidden states with increments in time. Section 6 extends the analysis to a non-linear filtering setting. Section 7 appeals to outcomes from the statistical detection literature to justify our way of rescaling contributions to entropy. Section 8 then describes how our proposal for rescaling relates to characterizations of smooth ambiguity. Section 9 offers some concluding remarks while appendix A poses and solves a discrete-time entropy problem.

## 2 Model specification

We consider a continuous-time specification with model uncertainty. The state dynamics are

$$dY_t = \mu(Y_t, \iota)dt + \sigma(Y_t)dW_t. \quad (1)$$

The parameter  $\iota$ , which indexes a model, is hidden from the decision-maker, who does observe  $Y$ . The matrix  $\sigma(Y)$  is nonsingular, implying that the Brownian increment  $dW_t$  would be revealed if  $\iota$  were known. Initially, we assume that the unknown model  $\iota$  is in a finite set  $\mathcal{I}$ , a restriction we impose mainly for notational simplicity, but we will also consider examples with a continuum of models, in which case we shall think of  $\iota$  as an unknown parameter.

For pedagogical convenience, we suppose that  $Y$  is beyond the control of the decision maker. We use this setting to explore the preferences of the decision maker under uncertainty.

## 3 Filtering

A filtering problem in which the model is unknown to the decision-maker is central to our analysis. A key ingredient is the log-likelihood conditioned on the unknown model  $\iota$ . For the moment, we ignore concerns about model misspecification.

### 3.1 Log-likelihoods

Log-likelihoods are depicted relative to some measure over realized values of the data. Consider a counterpart to (1) in which the drift is zero, but the same  $\sigma$  matrix governs conditional variability. This process has a local evolution

$$dY_t = \sigma(Y_t)dW_t,$$

the solution to which implies an induced probability measure over vector valued continuous functions on  $[0, t]$  and take  $Y_0$  as given. Call this measure  $\tau$ . The likelihoods we consider are built from densities with respect to  $\tau$ . In particular, the local evolution of the log-likelihood is

$$d \log L_t(\iota) = \mu(Y_t, \iota)' [\sigma(Y_t)\sigma(Y_t)']^{-1} dY_t - \frac{1}{2} \mu(Y_t, \iota)' [\sigma(Y_t)\sigma(Y_t)']^{-1} \mu(Y_t, \iota) dt.$$

We initialize the likelihood by specifying  $L_0(\iota)$ . We stipulate that  $L_0(\iota)$  includes contributions both from the density of  $Y_0$  conditioned on  $\iota$  and from a prior probability  $\pi_0(\iota)$  assigned to model  $\iota$ .

### 3.2 Posterior

As a consequence, the date  $t$  posterior for  $\iota$  is

$$\pi_t(\iota) = \frac{L_t(\iota)}{\sum_{\iota} L_t(\iota)}.$$

The date  $t$  posterior for the local mean of  $dY_t$  is

$$\bar{\mu}_t = \sum_{\iota} \pi_t(\iota) \mu(Y_t, \iota).$$

Then

$$dY_t = \bar{\mu}(Y_t)dt + \sigma(Y_t)d\bar{W}_t, \tag{2}$$

and

$$d\bar{W}_t = \sigma(Y_t)^{-1} [\mu(Y_t, \iota) - \bar{\mu}_t]dt + dW_t \tag{3}$$

is a multivariate standard Brownian motion relative to a filtration  $\mathcal{Y}$  generated by  $Y$  that does not include knowledge of the model  $\iota$ . Since he does not know the model  $\iota$ ,

the increment  $d\bar{W}_t$  is the date  $t$  innovation pertinent to the decision-maker. Equation (2) gives the stochastic evolution of  $Y$  in terms of this innovation.

The resulting likelihood evolution for the mixture (of means) model:

$$\begin{aligned} d \log \bar{L}_t &= \bar{\mu}(Y_t)' [\sigma(Y_t) \sigma(Y_t)']^{-1} dY_t - \frac{1}{2} \bar{\mu}(Y_t)' [\sigma(Y_t) \sigma(Y_t)']^{-1} \bar{\mu}(Y_t) dt \\ &= \bar{\mu}(Y_t)' [\sigma(Y_t) \sigma(Y_t)']^{-1} d\bar{W}_t + \frac{1}{2} \bar{\mu}(Y_t)' [\sigma(Y_t) \sigma(Y_t)']^{-1} \bar{\mu}(Y_t) dt. \end{aligned}$$

### 3.3 Example

Consider an additively separable specification of the drift  $\mu(Y, \iota) = \nu(Y) + \iota$  so that

$$dY_t = \nu(Y_t) dt + \iota dt + \sigma dW_t.$$

Suppose that  $\mathcal{I}$  is an entire Euclidean space and impose a normal prior  $\pi_0$  on  $\iota$ , so that

$$\log \pi_0(\iota) = -\frac{1}{2}(\iota - \bar{\iota}_0)' \Lambda_0 (\iota - \bar{\iota}_0) + \text{constant}$$

where  $\Lambda_0$  is the prior precision matrix and  $\bar{\iota}_0$  is the prior mean.

For this example,

$$d \log L_t(\iota) = [\nu(Y_t) + \iota]' (\sigma \sigma')^{-1} dY_t - \frac{1}{2} [\nu(Y_t) + \iota]' (\sigma \sigma')^{-1} [\nu(Y_t) + \iota] dt.$$

We form a likelihood conditioned on  $Y_0$  and use the prior on  $\iota$  to initialize  $\log L_0(\iota)$ . Since the log-likelihood increment and the logarithm of the prior are both quadratic in  $\iota$ , it follows that the posterior density for  $\iota$  is normal.

To obtain the evolution of posterior probabilities, note that

$$dL_t(\iota) = L_t(\iota) [\nu(Y_t) + \iota]' (\sigma \sigma')^{-1} dY_t.$$

Integrating with respect to  $\iota$ , we obtain

$$d\bar{L}_t = \left[ \int L_t(\iota) [\nu(Y_t) + \iota] d\iota \right]' (\sigma \sigma')^{-1} dY_t.$$

Then the posterior probability density

$$\pi_t(\iota) = \frac{L_t(\iota)}{\bar{L}_t}$$

evolves as

$$\begin{aligned} d\pi_t(\iota) &= \frac{L_t(\iota)}{\bar{L}_t} \left( [\nu(Y_t) + \iota]' - \int \frac{L_t(\iota)}{\bar{L}_t} [\nu(Y_t) + \iota]' d\iota \right) (\sigma\sigma')^{-1} dY_t \\ &\quad - \frac{L_t(\iota)}{\bar{L}_t} [\nu(Y_t) + \iota]' (\sigma\sigma')^{-1} \left( \int \frac{L_t(\iota)}{\bar{L}_t} [\nu(Y_t) + \iota]' d\iota \right)' dt \\ &\quad + \frac{L_t(\iota)}{\bar{L}_t} \left[ \int \frac{L_t(\iota)}{\bar{L}_t} [\nu(Y_t) + \iota]' d\iota \right] (\sigma\sigma')^{-1} \left( \int \frac{L_t(\iota)}{\bar{L}_t} [\nu(Y_t) + \iota]' d\iota \right) dt \\ &= \pi_t(\iota) (\iota - \bar{\iota}_t)' (\sigma\sigma')^{-1} [dY_t - \nu(Y_t)dt - \bar{\iota}_t dt] \end{aligned} \quad (4)$$

where  $\bar{\iota}_t$  is the posterior mean of  $\iota$ . Integrating (4) with respect to  $\iota$ , the posterior mean evolves as

$$d\bar{\iota}_t = \Sigma_t (\sigma\sigma')^{-1} [dY_t - \nu(Y_t)dt - \bar{\iota}_t dt]$$

where  $\Sigma_t$  is the posterior covariance matrix for  $\iota$ .

To get a formula for  $\Sigma_t$ , first note that the evolution for the logarithm of the posterior density implied by (4) is

$$\begin{aligned} d \log \pi_t(\iota) &= (\iota - \bar{\iota}_t)' (\sigma\sigma')^{-1} [dY_t - \nu(Y_t)dt - \bar{\iota}_t dt] \\ &\quad - \frac{1}{2} [dY_t - \nu(Y_t)dt - \bar{\iota}_t dt]' (\sigma\sigma')^{-1} [dY_t - \nu(Y_t)dt - \bar{\iota}_t dt]. \end{aligned}$$

Integrating between zero and  $t$ , the log-density is quadratic in  $\iota$ , and hence  $\iota$  is normally distributed. The date  $t$  precision matrix  $\Lambda_t = (\Sigma_t)^{-1}$  is

$$\Lambda_t = \Lambda_0 + t(\sigma\sigma')^{-1},$$

where  $\Lambda_0$  is the prior precision matrix.

## 4 Relative entropy

Statistical discrimination and large-deviation theories<sup>2</sup> underlie relative entropy's prominent role in dynamic stochastic robust control theory. In this section, we construct relative entropies and discuss some of their implications in the continuous-time stochastic setting of section 2. This will set the stage for the discussion of alternative ways of scaling different contributions to entropy that will preoccupy us in subsequent sections.

### 4.1 Factorization

Recall that  $L_t(\iota)$  is the likelihood function conditioned on  $\iota$  and scaled to incorporate the prior probability over  $\iota$ . This likelihood depends implicitly on the record  $Y_u, 0 \leq u \leq t$  of observed states between zero and  $t$ . Let  $f_t(y, \iota)$  be the density constructed so that  $f_t(Y_u, 0 \leq u \leq t, \iota) = L_t(\iota)$ , where  $y$  is a hypothetical realization of the  $Y$  process between dates zero and  $t$ . Similarly, let

$$g_t(y) = \sum_{\iota} f_t(y, \iota),$$

implying that  $g_t(Y_u, 0 \leq u \leq t) = \bar{L}_t$ . Let  $\tilde{f}_t(\iota)$  be an alternative joint density for  $Y$  observed between dates zero and  $t$  and  $\iota$ . Then relative entropy is

$$\text{ent} = \sum_{\iota} \int \left[ \log \tilde{f}_t(y, \iota) - \log f_t(y, \iota) \right] \tilde{f}_t(y, \iota) d\tau(y).$$

We can also factor the joint density  $f_t$  as

$$f_t(y, \iota) = g_t(y) \left[ \frac{f_t(y, \iota)}{g_t(y)} \right] = g_t(y) \psi_t(\iota|y),$$

and similarly for the alternative  $\tilde{f}$  densities. Notice that  $\psi_t(\iota|y) = \pi_t(\iota)$ . These density

---

<sup>2</sup>See Dupuis and Ellis (1997) for a development of these tools.

factorizations give rise to an alternative measure of entropy:

$$\begin{aligned} \text{ent}(\tilde{g}_t, \tilde{\psi}_t) &= \int \left( \sum_{\iota} \tilde{\psi}_t(\iota|y) \left[ \log \tilde{\psi}_t(\iota|y) - \log \psi_t(\iota|y) \right] \right) \tilde{g}_t(y) d\tau(y) \\ &+ \int [\log \tilde{g}_t(y) - \log g_t(y)] \tilde{g}_t(y) d\tau(y). \end{aligned}$$

For a fixed  $t$ , consider the following *ex ante* decision problem

**Problem 4.1.**

$$\max_d \min_{\tilde{g}_t, \tilde{\psi}_t} \int \sum_{\iota} U[d(y), y, \iota] \tilde{\psi}_t(\iota|y) \tilde{g}_t(y) dy + \theta \text{ent}(\tilde{g}_t, \tilde{\psi}_t)$$

where  $U$  is a concave function of  $d$ .

This is a static *robust control* problem. We refer to it as an *ex ante* problem because the objective averages across *all* data that could *possibly* be realized. But the decision  $d$  depends only on the data will *actually* be realized. Just as in Bayesian analysis, for determining  $d$  and  $\tilde{\psi}_t$ , we can solve the *conditional problem*

**Problem 4.2.**

$$\max_d \min_{\tilde{\psi}_t} \sum_{\iota} U[d(y), y, \iota] \tilde{\psi}_t(\iota|y) + \theta \sum_{\iota} \tilde{\psi}_t(\iota|y) \left[ \log \tilde{\psi}_t(\iota|y) - \log \psi_t(\iota|y) \right]$$

separately for each value of  $y$  without simultaneously computing  $\tilde{g}_t$ . Solving this conditional problem for each  $y$  gives the robust-control solution for  $d$ . In the conditional problem, we perturb only the posterior  $\psi_t(\iota, y)$ .<sup>3</sup> This simplification led Hansen and Sargent (2007) to perturb the *outcome* of filtering, namely, the posterior distribution, rather than the likelihood and prior.

Thus, we have established that for purposes of decision making, it suffices to consider relative entropy over the posterior:

$$\text{ent}^*(\tilde{\pi}_t) = \sum_{\iota} \tilde{\pi}_t(\iota) [\log \tilde{\pi}_t(\iota) - \log \pi_t(\iota)] \quad (5)$$

where  $\pi_t(\iota) = \psi_t(\iota|Y_u, 0 \leq u \leq 1)$ .

---

<sup>3</sup>In practice it suffices to compute  $d(y)$  for  $y$  given by the realization of  $Y$  between dates zero and  $t$ .

## 4.2 A forwarding-looking concern for misspecification

We follow Hansen and Sargent (2007) in considering forward-looking model misspecifications (of *future* shocks to states conditional on the entire state) as well as backward-looking misspecifications (of the distribution of hidden states arising from filtering). We say forward looking because of how the worst-case distortions of this type depend on forward-looking value functions. We say backward-looking because the orientation of filtering is toward processing historical data.

As for forward-looking misspecifications, suppose that the original Brownian motion increment  $dW_t$  is altered by appending a drift. This follows the full information specification of distortions in continuous time discussed by James (1992), Chen and Epstein (2002), Anderson et al. (2003), and Hansen et al. (2006). That is,  $dW_t = h_t(\iota)dt + d\tilde{W}_t$ , where  $d\tilde{W}_t$  is a standard Brownian increment. Conditional relative entropy is  $\frac{1}{2}|h_t(\iota)|^2dt$ . This measure of entropy scales linearly with the time increment  $dt$ .

Looking backwards, we argued that it suffices to focus on the relative entropy of the date  $t$  “posterior” (conditioned on information available at time  $t$ ) defined in (5). A key observation for us is that because it reflects the history of data up to  $t$ , this measure of entropy does not scale linearly with the time increment  $dt$ .

## 4.3 The scaling problem

Associated with the two distortions  $h_t(\iota)$  and  $\tilde{\pi}_t(\iota) - \pi_t(\iota)$  is a distorted conditional mean:

$$\tilde{\mu}_t = \sum_{\iota} \tilde{\pi}_t(\iota) [\sigma(Y_t)h_t(\iota) + \mu(Y_t, \iota)].$$

Notice that  $\tilde{\mu}_t$  is influenced by both  $h_t(\iota)$  and  $\tilde{\pi}_t$ .

Because of the different ways that the two entropies are scaled, to achieve a given distorted mean  $\tilde{\mu}_t$  at a minimum cost in terms of weighted entropy

$$\sum_{\iota} \tilde{\pi}_t(\iota) \left( \lambda \frac{1}{2} |h_t(\iota)|^2 dt + (1 - \lambda) [\log \tilde{\pi}_t(\iota) - \log \pi_t(\iota)] \right)$$

for any weight  $\lambda \in (0, 1)$ , we would set  $h_t(\iota) = \sigma(Y_t)^{-1}(\tilde{\mu}_t - \bar{\mu}_t)$  and leave the model probabilities unchanged, thus setting  $\tilde{\pi}_t(\iota) = \pi_t(\iota)$ . Doing this makes the resulting minimized entropy be scaled linearly by  $dt$ .

This observation has important consequences for recursive specifications of decision problems designed to express a decision maker’s concerns about misspecification. When a single entropy restriction is imposed over the joint distribution of the data and model and when the objective is to distort the drift  $\bar{\mu}_t$ , it is too costly in terms of entropy to induce *any* distortion to the current-period posterior as an intermediate step aimed ultimately at distorting the drift  $\bar{\mu}$ .

We will explore two responses to this problem, one that reduces the cost of perturbing the posterior, another that enhances the benefits from distortions that accrue to the fictitious evil agent imagined to be choosing the minimizing probabilities. The first weights the entropy contribution from  $\tilde{\pi}_t$  by  $dt$ ; the second makes changes in  $\tilde{\pi}_t$  more consequential to the decision-maker by altering the forward-looking objective function. These two responses lead to continuous-time versions of discrete-time formulations in Hansen and Sargent (2007).

## 5 Robust alternatives to expected utility

Our continuous-time versions of the two recursive specifications in Hansen and Sargent (2007) differ in terms of the information sets used to compute baseline value functions.

### 5.1 Continuation-values based on reduced information

Consider a continuation-value process  $V$  adapted to  $\mathcal{Y}$ . Represent this as

$$dV_t = \bar{\eta}_t dt + \varsigma_t \cdot d\bar{W}_t.$$

For discounted utility, the drift satisfies

$$\bar{\eta}_t = \delta V_t - \delta U(C_t)$$

where  $\delta$  is the subjective rate of discount and  $U$  is the instantaneous utility function.

Recall that

$$d\bar{W}_t = \sigma(Y_t)^{-1}[\mu(Y_t, \iota) - \bar{\mu}_t]dt + dW_t.$$

It is convenient to view this as a compound lottery where  $dW_t$  is the risk conditioned on a model and the unknown  $\iota$  indexes uncertainty about the model. Consider in turn two distortions to  $d\bar{W}_t$ .

First introduce a drift distortion  $h_t(\iota)dt$  so that  $dW_t = h_t(\iota)dt + d\hat{W}_t$ , where  $\hat{W}_t$  is a standard Brownian increment. Recall that local entropy is  $\frac{|h_t(\iota)|^2}{2}dt$ . Find a worst-case model by solving

**Problem 5.1.**

$$\min_{h_t(\iota)} \frac{\theta_1}{2} |h_t(\iota)|^2 + \varsigma_t \cdot h_t(\iota),$$

where  $\theta_1$  is an entropy penalty parameter.

The solution

$$\tilde{h}_t = -\frac{1}{\theta_1} \varsigma_t,$$

is independent of  $\iota$ . Our first adjustment for robustness of the continuation value drift is thus

$$\bar{\eta}_t = \delta V_t - \delta U(C_t) + \frac{|\varsigma_t|^2}{2\theta_1}.$$

As discussed in Hansen et al. (2006), this distorted drift outcome is a special case of the variance multiplier specification of Duffie and Epstein (1992), where  $\frac{1}{\theta_1}$  is a multiplier on the conditional variance of the continuation value.<sup>4</sup>

We also distort  $d\bar{W}$  in (5.1) by altering the posterior  $\pi_t(\iota)$  again penalized in terms of relative entropy. Since changing  $\pi_t$  alters the drift, we adopt a common scaling for both. That is, in constructing the penalty for the minimizing agent, we scale  $\text{ent}^*$  in equation (5) by  $dt$ . This sharply distinguishes the treatment of the two entropy penalties that Hansen and Sargent (2007) allow in their discrete-time specification, one that measures unknown dynamics, the other that measures unknown states. We compute our second drift distortion by solving:

**Problem 5.2.**

$$\min_{\tilde{\pi}_t} \varsigma_t \cdot \left( \sigma(Y_t)^{-1} \left[ \sum_{\iota} \tilde{\pi}_t(\iota) \mu(Y_t, \iota) \right] \right) + \theta_2 \text{ent}^*(\tilde{\pi}_t)$$

where  $\theta_2$  is a penalty parameter.

The minimized objective is

$$(\varsigma_t)' \sigma(Y_t)^{-1} \bar{\mu}(Y_t) + \theta_2 \log \left( \sum_{\iota} \pi_t(\iota) \exp \left[ -\frac{1}{\theta_2} (\varsigma_t)' \sigma(Y_t)^{-1} \mu(Y_t, \iota) \right] \right).$$

---

<sup>4</sup>Duffie and Epstein (1992) develop continuous-time limiting versions of the recursive utility specification of Kreps and Porteus (1978).

**Result 5.3.** *The drift  $\bar{\eta}_t$  and local Brownian exposure vector  $\varsigma_t$  for the continuation value process  $\{V_t\}$  satisfy*

$$\begin{aligned} \bar{\eta}_t = & \delta V_t - \delta U(C_t) + \frac{|\varsigma_t|^2}{2\theta_1} + \varsigma_t \cdot [\sigma(Y_t)^{-1} \bar{\mu}(Y_t)] \\ & + \theta_2 \log \left[ \sum_{\iota} \pi_t(\iota) \exp \left( -\frac{1}{\theta_2} \varsigma_t \cdot [\sigma(Y_t)^{-1} \mu(Y_t, \iota)] \right) \right]. \end{aligned} \quad (6)$$

The contribution from the two distortions, namely,

$$\frac{|\varsigma_t|^2}{2\theta_1} + \varsigma_t \cdot \left( \sigma(Y_t)^{-1} \bar{\mu}(Y_t) + \theta_2 \log \left[ \sum_{\iota} \pi_t(\iota) \exp \left( -\frac{1}{\theta_2} \varsigma_t \cdot [\sigma(Y_t)^{-1} \mu(Y_t, \iota)] \right) \right] \right)$$

is necessarily nonnegative.

Consider the filtering problem in example 3.3. Given that the baseline posterior is normal, formulas for exponentials of normals imply

$$\begin{aligned} & \theta_2 \log \left[ \int \pi_t(\iota) \exp \left( -\frac{1}{\theta_2} \varsigma_t \cdot [\sigma(Y_t)^{-1} \mu(Y_t, \iota)] \right) d\iota \right] \\ = & -\varsigma_t \cdot [\sigma(Y_t)^{-1} \bar{\mu}(Y_t)] + \frac{1}{\theta_2} \varsigma_t' [\sigma(Y_t)^{-1}] \Sigma_t [\sigma(Y_t)^{-1}]' \varsigma_t. \end{aligned} \quad (7)$$

For this example, the composite penalty can be decomposed as

$$\varsigma_t' \left( \begin{array}{cc} \frac{1}{2\theta_1} I & + \frac{1}{2\theta_2} [\sigma(Y_t)^{-1}] \Sigma_t [\sigma(Y_t)^{-1}]' \\ \uparrow & \uparrow \\ \text{misspecified} & \text{misspecified} \\ \text{dynamics} & \text{state estimation} \end{array} \right) \varsigma_t.$$

## 5.2 Continuation-values based on full information

Next we consider a discounted, recursive counterpart to formulations more common in the robust control literature.<sup>5</sup> Suppose that the continuation value is constructed knowing  $\iota$ . Thus, we write

$$dV_t(\iota) = \eta_t(\iota) dt + \varsigma_t(\iota) \cdot dW_t.$$

---

<sup>5</sup>See Hansen and Sargent (2005) for a distinct commitment-based solution that is explicitly linked to the robust control literature.

We again append a drift to the Brownian increment subject to an entropy penalty. The solution to problem 5.1 applies again except now the optimized  $h_t(\iota)$  can depend on  $\iota$ :

$$\eta_t(\iota) = \delta V_t(\iota) - \delta U(C_t) + \frac{|\varsigma_t(\iota)|^2}{2\theta_1}.$$

We compute the continuation value subject to these restrictions on the local mean and variance, imposing an appropriate terminal condition.

To adjust for robustness in estimation, we solve:

**Problem 5.4.**

$$\hat{V}_t = \min_{\tilde{\pi}_t} \sum_{\iota} \tilde{\pi}_t [V_t(\iota) + \theta_2 (\log \tilde{\pi}_t(\iota) - \log \pi_t(\iota))]$$

**Result 5.5.** *The date  $t$  continuation value  $\hat{V}_t$  that solves problem 5.4 is*

$$\hat{V}_t = -\theta_2 \log \left( \sum_{\iota} \exp \left[ -\frac{1}{\theta_2} V_t(\iota) \right] \pi_t(\iota) \right).$$

where the drift  $\eta_\tau(\iota)$  and shock exposure  $\varsigma_\tau(\iota)$  for the complete information continuation value process  $\{V_\tau(\iota) : \tau \geq t\}$  are restricted by

$$\eta_\tau(\iota) = \delta V_\tau(\iota) - \delta U(C_\tau) + \frac{|\varsigma_\tau(\iota)|^2}{2\theta_1}.$$

We use the continuation value  $\hat{V}_t$  to rank alternative consumption processes from the perspective of time  $t$ . Since we apply the robustness adjustment for estimation to the continuation value and not its stochastic increment, we do not scale relative entropy linearly by  $dt$  in the minimization problem on the right side of the equation defining problem 5.4.

In this formulation, the value function for the fictitious evil agent can be constructed recursively, but the induced preferences for the decision-maker change over time. For decision problems, as in Peleg and Yaari (1973), we manage this preference change by having our decision-maker play a dynamic game with future versions of himself.

### 5.3 Examples

We illustrate the two recursions when the state dynamics in example 3.3 are specialized so that  $U(C_t) = \log C_t = H \cdot Y_t$  where

$$dY_t = \Delta Y_t dt + \iota dt + \sigma dW_t.$$

Consider first the subsection 5.1 formulation in which the continuation value depends on the reduced information. Guess that

$$V_t = \lambda \cdot Y_t + \kappa \cdot \bar{v}_t + \phi_t.$$

Then

$$\begin{aligned} \eta_t &= (\Delta Y_t + \bar{v}_t) \cdot \lambda + \frac{d\phi_t}{dt} \\ \varsigma_t &= \sigma' \lambda + \sigma^{-1} \Sigma_t \kappa. \end{aligned}$$

Moreover, since the posterior for  $\iota$  is normal, it follows from (6) and (7) that

$$\eta_t = \delta V_t - \delta H Y_t + \varsigma_t' \left[ \frac{1}{2\theta_1} I + \frac{1}{2\theta_2} (\sigma^{-1}) \Sigma_t (\sigma^{-1})' \right] \varsigma_t.$$

Combining these findings,

$$\begin{aligned} \Delta' \lambda &= \delta \lambda - \delta H \\ \lambda &= \delta \kappa \\ \frac{d\phi_t}{dt} &= \delta \phi_t + \varsigma_t' \left[ \frac{1}{2\theta_1} I + \frac{1}{2\theta_2} (\sigma^{-1}) \Sigma_t (\sigma^{-1})' \right] \varsigma_t. \end{aligned}$$

The first equation can be solved for  $\lambda$  and then the second for  $\kappa$ . Then  $\phi_t$  can be computed by solving a first-order differential equation forward:

$$\phi_t = - \int_0^\infty \exp(-\delta u) \left( \varsigma_{t+u}' \left[ \frac{1}{2\theta_1} I + \frac{1}{2\theta_2} (\sigma^{-1}) \Sigma_{t+u} (\sigma^{-1})' \right] \varsigma_{t+u} \right) du.$$

Notice that  $\phi_t$  is negative. This reflects an aversion to uncertainty (a.k.a. a concern about misspecified estimation and models). The welfare implications of this aversion can be quantified using the methods of Barillas et al. (2008) and Cerreia et al. (2008).

Now consider the subsection 5.2 formulation in which the continuation value depends on the model  $\iota$ . Guess

$$V_t(\iota) = \lambda \cdot Y_t + \kappa \cdot \iota + \phi$$

where

$$\begin{aligned} \Delta' \lambda &= \delta \lambda - \delta H \\ \lambda &= \delta \kappa \\ \delta \phi &= -\frac{1}{2\theta_1} \lambda' \sigma \sigma' \lambda. \end{aligned}$$

Then

$$\hat{V}_t = \lambda \cdot Y_t + \kappa \cdot \bar{\iota}_t \quad -\frac{1}{2\delta\theta_1} \lambda' \sigma \sigma' \lambda \quad -\frac{1}{2\theta_2} \kappa' \Sigma_t \kappa.$$

$\uparrow$   
misspecified  
dynamics

$\uparrow$   
misspecified  
state estimation

The robust-adjusted continuation value includes two negative terms, one that adjusts for model misspecification and another that adjusts for estimation based on a possibly misspecified model.

## 6 Hidden-state Markov models

This section extends the previous Markov setup by letting  $\iota$  itself be governed by Markov transitions. This motivates the decision maker to learn about a moving target  $\iota_t$ . Bayesian learning carries an asymptotic rate of learning connected to the tail behavior of detection error probabilities in interesting ways. For expositional convenience, we use special Markov settings that imply quasi-analytical formulas for the solution to filtering problems.

## 6.1 Kalman filtering

Consider a linear model in which the time-varying Markov states are hidden, but designed so that we have quasi-analytical formulas for the filtering problem

$$\begin{aligned}dX_t &= AX_t dt + BdW_t \\dY_t &= DX_t dt + Fdt + GdW_t\end{aligned}$$

where  $dY_t$  is observed. The random vector  $DX_t + F$  plays the role of  $\mu(Y_t) + \iota_t$  and is partially hidden from the decision-maker. The Kalman filter provides a recursive solution to the filtering problem. We abstract from one aspect of time variation by letting the prior covariance matrix for the state vector  $\Sigma_0$  equal its limiting value. The recursive filtering solution gives the innovation representation for  $Y_t$  and the conditional mean  $\bar{X}_t$  of  $X_t$

$$\begin{aligned}d\bar{X}_t &= A\bar{X}_t dt + \bar{B}d\bar{W}_t \\dY_t &= D\bar{X}_t dt + Fdt + \sigma d\bar{W}_t\end{aligned}\tag{8}$$

where  $\sigma$  is nonsingular,  $\mu_t = D\bar{X}_t + F$  is the drift for the signal increment  $dY_t$ ,

$$\begin{aligned}\sigma\sigma' &= GG' \\ \bar{B} &= (\Sigma D' + BG')(GG')^{-1}\sigma = (\Sigma D' + BG')(\sigma')^{-1} \\ d\bar{W}_t &= \bar{G}^{-1} (dY_t - D\bar{X}_t - Fdt) = \sigma^{-1} [GdW_t + D(X_t - \bar{X}_t)dt].\end{aligned}$$

The matrix  $\Sigma$  is the limiting covariance matrix, which we assume exist and is non-singular. Log consumption is given by  $\log C_t = H \cdot Y_t$  and the benchmark preferences are discounted expected logarithmic utility as in section 5.3.

### 6.1.1 Continuation values that do not depend on hidden states

Guess a solution

$$V_t = \lambda \cdot \bar{X}_t + \phi.$$

The innovation for the continuation value is

$$\begin{aligned} (\lambda'\overline{B} + H'\overline{G})d\overline{W}_t &= (\lambda'\overline{B} + H'\overline{G})\overline{G}^{-1} [GdW_t + D(X_t - \overline{X}_t)dt], \\ &= [\lambda'(\Sigma D' + BG')(GG')^{-1} + H'] [GdW_t + D(X_t - \overline{X}_t)dt] \end{aligned}$$

and the drift for the continuation value under our guess is

$$\eta_t = \lambda \cdot (A\overline{X}_t) + H \cdot (D\overline{X}_t) + H \cdot F. \quad (9)$$

From the specification of preferences, the drift  $\eta$  satisfies

$$\begin{aligned} \eta_t = & \delta V_t - \delta \log C_t + \frac{1}{2\theta_1} [\lambda'(\Sigma D' + BG')(GG')^{-1} + H'] GG' [(GG')^{-1}(D\Sigma + GB')\lambda + H] \\ & + \frac{1}{2\theta_2} [\lambda'(\Sigma D' + BG')(GG')^{-1} + H'] D\Sigma D' [(GG')^{-1}(D\Sigma + GB')\lambda + H] \end{aligned} \quad (10)$$

Equating coefficients on  $\overline{X}_t$  as given from (9) and (10) gives,

$$A'\lambda + D'H = \delta\lambda$$

Hence,  $\lambda = (\delta I - A')^{-1}D'H$ .

As inputs into constructing detection-error probabilities, we require the worst-case distortions. The worst-case drift for  $dW_t$  is

$$\tilde{h}_t = -\frac{1}{\theta_1}G' [(GG')^{-1}(D\Sigma + GB')\lambda + H]$$

and the distorted mean  $\tilde{X}_t$  for  $X_t$  is

$$\tilde{X}_t = \overline{X}_t - \frac{1}{\theta_2}\Sigma D' [(GG')^{-1}(D\Sigma + GB')\lambda + H].$$

Combining these distortions gives

$$d\overline{W}_t = -(\overline{G})^{-1} \left( \frac{1}{\theta_1}GG' + \frac{1}{\theta_2}D\Sigma D' \right) [(GG')^{-1}(D\Sigma + GB')\lambda + H] + d\tilde{W}_t \quad (11)$$

where  $d\tilde{W}_t$  is multivariate standard Brownian motion under the distorted probability

law. The drift distortions are constant. Substituting (11) into (8) gives the implied distorted law of motion for the reduced information structure generated by the signal history. In particular, the distorted drift  $\tilde{\mu}_t$  for  $dY_t$  is:

$$\tilde{\mu}_t = D\bar{X}_t - G \frac{1}{\theta_1} [(D\Sigma + GB')\lambda + (GG')H] - \frac{1}{\theta_2} D\Sigma D' [(GG')^{-1}(D\Sigma + GB')\lambda + H].$$

### 6.1.2 Continuation values that depend on the hidden states

Guess a solution

$$V_t = \lambda \cdot X_t + H \cdot Y_t + \phi.$$

The innovation to the continuation value is

$$(B'\lambda + G'H) \cdot dW_t,$$

and the drift is

$$\eta_t = \lambda \cdot (AX_t) + H \cdot (DX_t + F).$$

This drift satisfies

$$\eta_t = \delta V_t - \delta H \cdot X_t + \frac{1}{2\theta_1} |B'\lambda + G'H|^2.$$

The vector  $\lambda$  is the same as in the limited information case, and the worst-case model prescribes the following drift to the Brownian increment  $dW_t$ :

$$\tilde{h}_t = -\frac{1}{\theta_1} (B'\lambda + G'H).$$

The robust state estimate  $\tilde{X}_t$  is

$$\tilde{X}_t = \bar{X}_t - \frac{1}{\theta_2} \Sigma \lambda.$$

Thus, the combined distorted drift for  $dY_t$  is

$$\tilde{\mu}_t = D\bar{X}_t - \frac{1}{\theta_1} G(B'\lambda + G'H) - \frac{1}{\theta_2} D\Sigma \lambda.$$

The drift distortions in both robustness specifications are constant. While this parameterization of a hidden-state Markov model is convenient, the constant distor-

tions make it empirically limiting.<sup>6</sup> We next consider environments that imply state dependent distortions in the probabilities.

## 6.2 Wonham filtering

Suppose that

$$dY_t = \Delta Y_t dt + \iota_t dt + \sigma dW_t$$

where  $\iota_t = \Gamma Z_t$  and  $Z_t$  follows a discrete-state Markov chain with intensity matrix  $A$ . The realized values of  $Z_t$  are coordinate vectors. This is a version of a Wonham filtering problem where the signal increment is now  $dY_t - \Delta Y_t dt$ . Running the local regression of  $Z_t$  onto  $dX_t - \Delta X_t dt - \Gamma \bar{Z}_t dt$  yields the conditional regression coefficient

$$K_t = \left[ \text{diag}(\bar{Z}_t) - \bar{Z}_t \bar{Z}_t' \right] G' (\sigma \sigma')^{-1}.$$

Then the recursive solution to the filtering problem is

$$\begin{aligned} dY_t &= \Delta Y dt + \Gamma \bar{Z}_t dt + \sigma d\bar{W}_t \\ d\bar{Z}_t &= A' \bar{Z}_t dt + K_t \sigma d\bar{W}_t, \end{aligned}$$

where the innovation

$$d\bar{W}_t = \sigma^{-1} \Gamma (Z_t - \bar{Z}_t) dt + dW_t$$

is an increment to a multivariate standard Brownian motion.

### 6.2.1 Continuation-values

When continuation values do not depend on hidden states, the continuation value function must be computed numerically. The full information value function, however, is of the form

$$V_t = \lambda \cdot Y_t + \kappa \cdot Z_t + \xi.$$

Appendix A derives formulas for  $\kappa$  and  $\lambda$  when the logarithm of consumption is given by  $H \cdot X_t$  and the instantaneous utility function is logarithmic (a unitary elasticity of substitution.) Given solutions for  $\lambda$  and  $\kappa$ , appendix A also provides formulas

---

<sup>6</sup>This constancy motivated Hansen and Sargent (2008a) and Hansen (2007) to explore other specifications in their empirical investigations.

for the worst-case drift distortion  $\tilde{h}$  and the worst-case intensity matrix  $\tilde{A}$  for the continuous-time Markov chain.

### 6.2.2 Worst-case state estimate

It remains to compute the worst-case state estimate. For this, we solve

$$\begin{aligned} & \min_{\{\tilde{Z}_{i,t}\}_i} \sum_i \tilde{Z}_{i,t} \left[ \kappa_i + \theta_2 \left( \log \tilde{Z}_{i,t} - \log \bar{Z}_{i,t} \right) \right] \\ & \text{subject to } \sum_i \tilde{Z}_{i,t} = 1. \end{aligned}$$

The minimizer

$$\tilde{Z}_{i,t} \propto Z_{i,t} \exp \left( -\frac{\kappa_i}{\theta_2} \right)$$

“tilts”  $\tilde{Z}_t$  towards states with smaller continuation values.

### 6.2.3 Combined distortion

Given the initial robust state estimate and the worst-case dynamics, we again apply the Wonham filter to obtain:

$$\begin{aligned} dY_t &= \Delta Y_t dt + \Gamma \tilde{Z}_t dt + \sigma \tilde{h} dt + \sigma d\tilde{W}_t \\ d\tilde{Z}_t &= \tilde{A}' \tilde{Z}_t dt + \tilde{K}_t (dY_t - \Delta Y_t dt - \Gamma \tilde{Z}_t dt - \sigma \tilde{h} dt) \end{aligned}$$

where

$$\tilde{K}_t = \left[ \text{diag} \left( \tilde{Z}_t \right) - \tilde{Z}_t \tilde{Z}_t' \right] \Gamma' (\sigma \sigma')^{-1}.$$

## 7 Statistical discrimination

This section uses insights from the statistical detection literature to defend the rescaling of entropy contributions recommended in our sections 5, 8, and 6. Chernoff (1952) used likelihood ratios to discriminate among competing statistical models. Newman and Stuck (1979) extended Chernoff’s analysis to apply to observable continuous-time Markov processes. We follow Anderson et al. (2003), Hansen (2007) and Hansen and Sargent (2008b, ch. 9) in using such methods to quantify how difficult it is to distinguish worst-case models from the decision maker’s benchmark models. We modify

the analysis of Anderson et al. (2003) to allow for hidden Markov states.

For the hidden-state Markov models of section 6, the log-likelihood ratio between the worst-case and benchmark models evolves as

$$\begin{aligned} d \log \tilde{L}_t - d \log \bar{L}_t &= (\tilde{\mu}_t - \mu_t)' (\sigma \sigma')^{-1} dY_t \\ &\quad - \frac{1}{2} (\tilde{\mu}_t)' (\sigma \sigma')^{-1} \tilde{\mu}_t dt \\ &\quad + \frac{1}{2} (\mu_t)' (\sigma \sigma')^{-1} \mu_t dt, \end{aligned}$$

where the specification of  $\mu_t$  and  $\tilde{\mu}_t$  depends on the specific hidden state. Equivalently, this likelihood ratio evolution can be written as

$$\begin{aligned} d \log \tilde{L}_t - d \log \bar{L}_t &= (\tilde{\mu}_t - \mu_t)' (\sigma')^{-1} d\bar{W}_t \\ &\quad - \frac{1}{2} (\tilde{\mu}_t - \mu_t) (\sigma \sigma')^{-1} (\tilde{\mu}_t - \mu_t) dt, \end{aligned}$$

which makes the likelihood ratio a martingale under the reduced information filtration under the benchmark model. The alternative specifications in section 6 imply different formulas for the conditional means of the hidden states and worst-case adjustments, but for all of them the continuous-time likelihood ratio has this common structure.

Chernoff (1952) used the expected value of the likelihood ratio to a power  $0 < \alpha < 1$  to bound the limiting behavior of the probability that the likelihood ratio exceeds alternative thresholds. His was one of the initial applications of Large Deviation Theory. In our setting, this approach leads us to study the behavior of the conditional expectation of  $M_t(\alpha) = \left( \tilde{L}_t / L_t \right)^\alpha$ . The logarithm of  $M_t(\alpha)$  evolves as:

$$\begin{aligned} d \log M_t(\alpha) &= \alpha (\tilde{\mu}_t - \mu_t)' (\sigma')^{-1} d\bar{W}_t \\ &\quad - \frac{\alpha}{2} (\tilde{\mu}_t - \mu_t) (\sigma \sigma')^{-1} (\tilde{\mu}_t - \mu_t) \\ &= \alpha (\tilde{\mu}_t - \mu_t)' (\sigma')^{-1} d\bar{W}_t - \frac{\alpha^2}{2} (\tilde{\mu}_t - \mu_t) (\sigma \sigma')^{-1} (\tilde{\mu}_t - \mu_t) \\ &\quad + \left( \frac{\alpha^2 - \alpha}{2} \right) (\tilde{\mu}_t - \mu_t) (\sigma \sigma')^{-1} (\tilde{\mu}_t - \mu_t). \end{aligned}$$

This shows that  $M_t(\alpha)$  can be factored into two components identified by the last

two lines. The first component  $M_t^1(\alpha)$  evolves as

$$d \log M_t^1(\alpha) = \alpha(\tilde{\mu}_t - \mu_t)'(\sigma')^{-1}d\overline{W}_t - \frac{\alpha^2}{2}(\tilde{\mu}_t - \mu_t)(\sigma\sigma')^{-1}(\tilde{\mu}_t - \mu_t)$$

and is a local martingale in levels. The second component  $M_t^2(\alpha)$  evolves as

$$d \log M_t^2(\alpha) = \left( \frac{\alpha^2 - \alpha}{2} \right) (\tilde{\mu}_t - \mu_t)(\sigma\sigma')^{-1}(\tilde{\mu}_t - \mu_t)dt.$$

Since it has no instantaneous exposure to  $d\overline{W}_t$ , this second component is locally predictable.

## 7.1 Local discrimination

Anderson et al. (2003) used the  $d \log M_t^2(\alpha)$  component for a fully observed Markov process to define the local rate of statistical discrimination. This local rate gives a statistical measure of how easy it is to discriminate among competing models using historical data, in the sense that it bounds the rate at which the probability of making mistake in choosing between two models decreases as the sample size grows. The counterpart of this local rate for a hidden state model is

$$\frac{1}{8}(\tilde{\mu}_t - \mu_t)(\sigma\sigma')^{-1}(\tilde{\mu}_t - \mu_t) \tag{12}$$

This rate attains

$$\max_{\alpha} \left( \frac{\alpha - \alpha^2}{2} \right) (\tilde{\mu}_t - \mu_t)(\sigma\sigma')^{-1}(\tilde{\mu}_t - \mu_t)$$

where the objective is maximized by setting  $\alpha = 1/2$ . Since Anderson et al. (2003) consider complete information models, they focus only on the drift distortion to an underlying Brownian motion. But for us, states are hidden from the decision-maker, so robust estimation necessarily plays a role gives rise to an additional contribution to the local rate of statistical discrimination between models.<sup>7</sup>

For the stochastic specification with an invariant  $\iota$  and continuation values that

---

<sup>7</sup>The term

$$\sigma^{-1}(\tilde{\mu}_t - \mu_t)$$

entering (12) is also the uncertainty component to the price of local exposure to the vector  $d\overline{W}_t$ . See Anderson et al. (2003) and Hansen and Sargent (2008a).

do not depend on the unknown model,

$$\tilde{\mu}_t - \mu_t = \underbrace{-\frac{1}{\theta_1} \sigma(Y_t) \varsigma_t}_{\substack{\uparrow \\ \text{misspecified} \\ \text{dynamics}}} + \underbrace{[\sum_{\iota} \tilde{\pi}_t(\iota) \mu(Y_t, \iota)] - \bar{\mu}(Y_t)}_{\substack{\uparrow \\ \text{misspecified} \\ \text{model estimation}}}$$

where  $\tilde{\pi}_t$  solves problem 5.2. This formula illustrates the two contributions to statistical discrimination. In particular, the second term on the right shows how a concern about misspecified model estimation alters the local rate of statistical discrimination. Recall that in our recursive formulation, we scaled the contribution of entropy from the posterior over the unknown model by  $dt$ . Apparently this scaling balances the contributions to the detection error rate so that both components are of comparable magnitudes. The impact of the misspecified model estimation will vanish over time as the decision maker learns  $\iota$ , however.

Consider next the Kalman filtering model with continuation values that do not depend on the hidden state. In this case we showed that

$$\tilde{\mu}_t - \mu_t = \underbrace{-\frac{1}{\theta_1} G(B'\lambda + G'H)}_{\substack{\uparrow \\ \text{misspecified} \\ \text{dynamics}}} \quad \underbrace{-\frac{1}{\theta_2} D\Sigma\lambda}_{\substack{\uparrow \\ \text{misspecified} \\ \text{state estimation}}}$$

Now the second term on the right shows how a concern about misspecified state estimation alters the local rate of statistical discrimination. Since the hidden state evolves over time the impact of the second term will not vanish. Both contributions, however are time invariant.

Finally, consider the Wonham filtering model with continuation values that do not depend on the hidden state, which are discrete. Now both distortions must be computed numerically. They depend on the vector of vector  $Y_t$  of observables and on probabilities over the hidden states, persist through time, and have comparable magnitudes.

An analogous set of results can be obtained when continuation values depend on

the unknown model or hidden states.

## 7.2 Long-run discrimination

For the Kalman filtering model, the local discrimination rate is constant and necessarily coincides with its long-term counterpart. For the Wonham filtering model with hidden states that are discrete, the local discrimination rate is state dependent. However, it has a limiting discrimination rate that is state independent. Newman and Stuck (1979) construct this rate for a fully observed Markov state. The solution to a filtering problem associated with a hidden state Markov process gives an alternative Markov process in which the hidden state is replaced by the posterior density of the hidden state given the history of signals. The recursive representation of the solution gives the central component to the Markov evolution by showing how the posterior distribution for the hidden states evolves in response to new information that arrives to the decision-maker. We apply the approach suggested by Newman and Stuck (1979) to the Markov process constructed from the recursive solution to the filtering problem.

We know the likelihood ratio and the evolution under the benchmark model. In particular, notice that we can construct  $\tilde{Z}_t$  as a function of  $\bar{Z}_t$  given the vector  $\kappa$ . The long-run rate is given by maximizing the limit

$$\rho(\alpha) = \lim_{t \rightarrow \infty} \frac{1}{t} \log E [M_t(\alpha) | Y_0 = y, \bar{Z}_0 = z],$$

by choice of  $\alpha$ . The rate  $\rho(\alpha)$  can be characterized through an alternative approach by finding the dominant eigenvalue for the generator of a semigroup of operators. Operator  $t$  in this semigroup maps a function of  $(y, z)$  into another function of  $(y, z)$  defined by computing

$$E \left[ \frac{M_t(\alpha)}{M_0(\alpha)} \varphi(Y_t, \bar{Z}_t) | Y_0 = y, \bar{Z}_0 = z \right].$$

The eigenvalue problem is

$$E \left[ \frac{M_t(\alpha)}{M_0(\alpha)} \varphi(Y_t, \bar{Z}_t) | Y_0 = y, \bar{Z}_0 = z \right] = \exp[-\rho(\alpha)t] \varphi(y, z)$$

where  $\varphi$  is restricted to be a positive function.<sup>8</sup> Since this equation must hold for all  $t$ , there is a local counterpart that requires solving a second-order partial differential equation. Again, concerns about both misspecified dynamics and misspecified state estimation contribute to the resulting asymptotic rate. Following Newman and Stuck (1979) the long-term counterpart to Chernoff entropy is:

$$\max_{0 \leq \alpha \leq 1} \rho(\alpha).$$

## 8 Smooth adjustments for model uncertainty

Thus far, we have described links between entropy-based robust control problems and recursive utility in continuous time. In this section, we explore an analogous link of robust control to the smooth ambiguity decision model of Klibanoff et al. (2005, 2009). Hansen and Sargent (2007) provided this link for discrete-time models. Here we explore the link in the context of continuous-time specifications. To isolate the link to decision-theoretic models with smooth ambiguity, we set  $\theta_1$  to infinity, and for simplicity feature the case in which the learning is about an invariant parameter or model indicator,  $\iota$ .

Consider a sequence of stochastic environments indexed by a parameter  $\epsilon$  that indexes the time gap between observations. It is perhaps simplest to think of  $\epsilon = 2^{-j}$  for nonnegative integers  $j$  and for a fixed  $j$  construct a stochastic process of observations at dates  $0, \epsilon, 2\epsilon, \dots$ . An increment from  $j$  to  $j + 1$  divides the sampling interval in half. To simplify notation and approximation we preserve the two continuous time information structures set out in subsection 5.1 with continuation values based on reduced information, but we consider consumption choices made on the  $\epsilon$ -spaced grid just described.<sup>9</sup>

Given knowledge of  $\iota$ , the decision maker uses expected utility preferences:

$$\begin{aligned} V_t(\iota) &= [1 - \exp(-\epsilon\delta)] E \left[ \sum_{k=0}^{\infty} \exp(-k\epsilon\delta) U[C_{t+k\epsilon}] | \mathcal{Y}_t, \iota \right] \\ &= [1 - \exp(-\epsilon\delta)] U(C_t) + \exp(-\epsilon\delta) E [V_{t+\epsilon}(\iota) | \mathcal{Y}_t, \iota]. \end{aligned}$$

---

<sup>8</sup>There may be multiple solutions to this eigenvalue problem but there are well known ways to select the appropriate solution. See Hansen and Scheinkman (2009).

<sup>9</sup>An alternative approach would be to sample the  $\{Y_t\}$  for each choice of  $\epsilon$  and solve the corresponding filtering problem in discrete-time for each  $\epsilon$ .

Thus

$$\begin{aligned} E[V_{t+\epsilon}(\iota)|\mathcal{Y}_t, \iota] - V_t(\iota) &= [\exp(\epsilon\delta) - 1]V_t(\iota) - [\exp(\epsilon\delta) - 1]U(C_t) \\ &\approx \epsilon\delta V_t - \epsilon\delta U(C_t). \end{aligned} \tag{13}$$

To extend these preferences to accommodate smooth ambiguity as in Klibanoff et al. (2009), we first consider preferences over one step-ahead continuation plans. The one step-ahead construction begins with a consumption process at date  $t$ :  $(C_t, C_{t+\epsilon}, C_{t+2\epsilon}, \dots)$  and forms a new stochastic sequence:  $(C_t, C_{t+\epsilon}, C_{t+\epsilon}, \dots)$ . This new sequence is constant over time from time period  $t + \epsilon$  forward. Therefore, the date  $t + \epsilon$  continuation value  $V_{t+\epsilon}$  is

$$V_{t+\epsilon} = U(C_{t+\epsilon}).$$

No expectations are needed to compute this continuation value so knowledge of  $\iota$  plays no role in the valuation from date  $t + \epsilon$  forward. For a given  $\iota$ , Klibanoff et al. (2009) define the one-step ahead certainty equivalent  $\hat{C}_t(\iota)$  to be the solution to

$$[1 - \exp(-\epsilon\delta)]U(C_t) + \exp(-\epsilon\delta)U[\hat{C}_t(\iota)] = [1 - \exp(-\epsilon\delta)]U(C_t) + \exp(-\epsilon\delta)E[V_{t+\epsilon}|\mathcal{Y}_t, \iota].$$

Therefore,

$$\hat{C}_t(\iota) = U^{-1}(E[V_{t+\epsilon}|\mathcal{Y}_t, \iota]).$$

For future reference, define

$$\hat{V}_t(\iota) = U[\hat{C}_t(\iota)] = E[V_{t+\epsilon}|\mathcal{Y}_t, \iota]. \tag{14}$$

Klibanoff et al. (2009) refer to  $\hat{C}_t(\iota)$  as the second order act associated with  $(C_t, C_{t+\epsilon}, C_{t+2\epsilon}, \dots)$ , and they impose an assumption of subjective expected utility over second-order acts where the probability measure is the date  $t$  posterior for  $\iota$ . (See their Assumption 7.)

To form a bridge between the formulation of Klibanoff et al. (2009) and ourselves, we would let the function that they use to represent expected utility over second order acts have the special exponential form<sup>10</sup>

$$U^*(C) = -\exp[-\gamma U(C)]. \tag{15}$$

---

<sup>10</sup>Note that Klibanoff et al. (2005) allow for more general utility functions  $U^*$ .

Then the date  $t$  objective applied to second-order acts is the expected utility

$$-E \left( \exp \left[ -\gamma \hat{V}_t(\iota) \right] \mid \mathcal{Y}_t \right).$$

The continuation value certainty equivalent of  $\hat{V}_t$  is

$$-\frac{1}{\gamma} \log E \left( \exp \left[ -\gamma \hat{V}_t(\iota) \right] \mid \mathcal{Y}_t \right)$$

Piecing things together gives the recursive representation

$$\begin{aligned} V_t &= [1 - \exp(-\epsilon\delta)]U(C_t) - \exp(-\epsilon\delta)\frac{1}{\gamma} \log E \left[ \exp \left[ -\gamma \hat{V}_t(\iota) \right] \mid \mathcal{Y}_t \right] \\ &= [1 - \exp(-\epsilon\delta)]U(C_t) - \exp(-\epsilon\delta)\frac{1}{\gamma} \log E [\exp(-\gamma E[V_{t+\epsilon} \mid \mathcal{Y}_t, \iota]) \mid \mathcal{Y}_t] \end{aligned}$$

where we have substituted for  $\hat{V}_t(\iota)$  from (14). Klibanoff et al. (2009) use dynamic consistency to extend this preference representation beyond these constructed second-order acts.

## 8.1 A smooth ambiguity adjustment that vanishes in continuous time

Consider now a continuous-time approximation. Take a continuation value process  $\{V_t\}$  with drift  $\eta_t(\iota)$  conditioned on  $\iota$ . For example,  $V_\tau = U(C_{t+\epsilon})$   $\tau \geq t + \epsilon$  as in the construction of second-order acts. Using a continuous-time approximation,

$$\hat{V}_t(\iota) = E[V_{t+\epsilon} \mid \mathcal{Y}_t, \iota] \approx V_t + \epsilon\eta_t(\iota)$$

Then

$$-\frac{1}{\gamma} \log E \left( \exp \left[ -\gamma \hat{V}_t(\iota) \right] \mid \mathcal{Y}_t \right) \approx U(C_t) + \epsilon\bar{\eta}_t - \frac{1}{\gamma} \log E \left( \exp \left[ -\gamma\epsilon[\eta_t(\iota) - \bar{\eta}_t] \right] \mid \mathcal{Y}_t \right) \quad (16)$$

where

$$\bar{\eta}_t = E[\eta_t(\iota) \mid \mathcal{Y}_t]$$

is the date  $t$  drift for the process  $\{U(C_\tau) : \tau \geq t\}$  under the filtration  $\{\mathcal{Y}_\tau : \tau \geq t\}$  that omits knowledge of  $\iota$ . Since  $[\eta_t(\iota) - \bar{\eta}_t]$  has conditional mean zero,

$$\frac{1}{\gamma} \log E (\exp [-\gamma \epsilon [\eta_t(\iota) - \bar{\eta}_t] | \mathcal{Y}_t])$$

contributes only an  $\epsilon^2$  term. The limiting counterpart to (13) scaled by  $\frac{1}{\epsilon}$  is

$$\bar{\eta}_t = \delta V_t - \delta U(C_t).$$

The parameter  $\gamma$  drops out of this equation in the limit and there is no adjustment for ambiguity. This calculation reaffirms an insight of Skiadas (2009) that the consequences of a smooth ambiguity adjustment vanish in a continuous-time limit.

## 8.2 A smooth (in a derivative) ambiguity adjustment that survives in continuous time

As an alternative, suppose that we adjust the utility function  $U^*$  over second-order acts simultaneously with  $\epsilon$ . In particular, we replace  $\gamma$  with  $\frac{\gamma}{\epsilon}$  on the right-hand side of (16):

$$U(C_t) + \epsilon \bar{\eta}_t - \frac{\epsilon}{\gamma} \log E [\exp (-\gamma [\eta_t(\iota) - \bar{\eta}_t]) | \mathcal{Y}_t]$$

This leads us to

$$\bar{\eta}_t - \frac{1}{\gamma} \log E [\exp (-\gamma [\eta_t(\iota) - \bar{\eta}_t]) | \mathcal{Y}_t] = \delta V_t - \delta C_t,$$

or

$$\bar{\eta}_t = \delta V_t - \delta C_t + \frac{1}{\gamma} \log E [\exp (-\gamma [\eta_t(\iota) - \bar{\eta}_t]) | \mathcal{Y}_t]. \quad (17)$$

Our alternative adjustment thus makes the concern about ambiguity remain in the continuous-time limit.

To build a link to our analysis of continuous-time versions of preferences for robustness, write the local evolution for the continuation value under the filtration  $\mathcal{Y}$  as:

$$dV_t = \bar{\eta}_t dt + \varsigma_t \cdot d\bar{W}_t$$

and recall that

$$d\bar{W}_t = dW_t + \sigma(Y_t)^{-1} [\mu(Y_t, \iota) - \bar{\mu}(Y_t)] dt.$$

Thus

$$\eta_t(\iota) = \bar{\eta}_t + \varsigma_t \cdot (\sigma(Y_t)^{-1} [\mu(Y_t, \iota) - \bar{\mu}(Y_t)]) dt,$$

and the drift condition for our version of smooth ambiguity can be expressed as:

$$\begin{aligned} \bar{\eta}_t = & \delta V_t - \delta U(C_t) + \varsigma_t \cdot [\sigma(Y_t)^{-1} \bar{\mu}(Y_t)] \\ & + \frac{1}{\gamma} \log \left[ \sum_{\iota} \pi_t(\iota) \exp(-\gamma \varsigma_t \cdot [\sigma(Y_t)^{-1} \mu(Y_t, \iota)]) \right]. \end{aligned}$$

This formulation coincides with (6) when  $\theta_1 = \infty$  and  $\theta_2 = \frac{1}{\gamma}$ .

### 8.2.1 Senses of smoothness

As we showed in (17), the continuous-time limit of our scaling makes a smooth exponential ambiguity adjustment to the *derivative* of the continuation value. The way that this limit  $U^*$  in (16) depends on the sampling interval  $\epsilon$  leads us to raise the important question of the extent to which we can hope to parameterize ambiguity preferences through  $U^*$  in a way that plausibly remains across alternative environments. When we follow Klibanoff et al. (2009) and use  $U^*$  in (16) and expected utility over second-order acts, the ambiguity adjustment vanishes in the continuous time limit computed in subsection 8.1. The reason that the adjustment disappears in the limit is that the impact of uncertainty about  $\iota$  on transition distributions for the Markov vanishes too quickly as we shrink the sampling interval  $\epsilon$  to zero. In order to sustain an ambiguity adjustment in continuous time, we have increased the curvature of  $U^*$  as we have diminished the sampling interval. In terms of the Klibanoff et al. (2009) analysis, we performed this adjustment because of how uncertainty about  $\iota$  is manifested in the constructed second-order acts.

### 8.2.2 Example

In order to illustrate the impact of our proposed adjustment, return to the first example in section 5.3. We showed that the value function has the form

$$V_t = \lambda \cdot Y_t + \kappa \cdot \bar{\iota}_t + \phi_t$$

and reported formulas for  $\lambda$ ,  $\kappa$  and  $\phi_t$ . Under the ambiguity interpretation

$$\phi_t = -\gamma \int_0^\infty \exp(-\delta u) [\zeta_{t+u}'(\sigma^{-1})\Sigma_{t+u}(\sigma^{-1})'\zeta_{t+u}] du$$

scales linearly in the ambiguity parameter  $\gamma$ .

What lessons do we learn from this? The exposure of a continuation value to model uncertainty diminishes proportionally to  $\epsilon$  as  $\epsilon$  shrinks to zero. But the risks conditioned on a model  $\iota$ , namely,  $W_{t+\epsilon} - W_t$ , have standard deviations that scale as  $\sqrt{\epsilon}$ , and these risks come to dominate the uncertainty component. By replacing the ambiguity parameter  $\gamma$  with  $\frac{\gamma}{\epsilon}$ , we can offset this diminishing importance of ambiguity when we move toward approximating a continuous-time specification. The consequences for preferences over consumption processes are apparently not extreme. Even though we drove the ambiguity aversion parameter  $\frac{\gamma}{\epsilon}$  to infinity in the continuous-time limit, the local uncertainty exposure of continuation values and consumption simultaneously diminish at comparable rates. This calculation provokes further thoughts about how separately to calibrate both a decision maker's ambiguity aversion and his risk aversion. The empirical analyses of Ju and Miao (2007) and Collard et al. (2008) study implications of smooth ambiguity models. Our continuous-time limiting investigations suggests that it is problematic to motivate the discrete-time ambiguity aversion parameter  $\gamma$  without reference to the local uncertainty that is prevalent in the environment of the decision-makers.

Our example sets an intertemporal substitution elasticity equal to unity with the consequence that a proportionate change in the consumption process leads to a change in the constant term of the continuation value equal to the logarithm of the proportionality factor. If we modify the substitution elasticity parameter using the Klibanoff et al. (2009) formulation in conjunction with our exponential risk adjustment, this homogeneity property no longer applies. In discrete-time, Ju and Miao (2007) propose an alternative recursion with ambiguity that preserves the homogeneity property just described. Hansen et al. (2009) study the continuous-time limits of this dynamic formulation of smooth ambiguity and relate it to the calculations in this paper.

## 9 Concluding remarks

All versions of max-min expected utility models, including the recursive specifications of robustness developed in Hansen and Sargent (2007), assign a special role to a worst-case model. A theme of this paper is that it is reasonable for a decision maker to think about the statistical plausibility of that model, and that how he thinks about statistically discriminating it from other models should affect how he formulates his robust decision making procedure.

In our formulations, robust decision makers use an approximating model as a benchmark around which they put a cloud of perturbed models that they use to explore the consequences of misspecifications. To characterize how “robust” the decision maker wants to be, Anderson et al. (2003) and Hansen and Sargent (2008b, ch. 19) employed measures of statistical discrepancy between worst-case models and benchmark models. This paper has pushed their use of detection error calculations further by employing them to justify a proposal for a new way of scaling contributions to entropy in continuous time hidden Markov models with robust decision making.

We have explored robustness to two alternative types of misspecifications. The first type is misspecified dynamics as reflected in distributions of current and future states and signals conditioned on current states. The second type is misspecified dynamics of the histories of the signals and hidden state variables entering filtering problems. We study a continuous-time limit as a way so to investigate the impact of both forms of misspecification on decision making and to understand better the impact of relative entropy restraints on a decision-maker’s exploration of model misspecification. We considered parameterizations of preferences for robustness for which both types of misspecification contribute to the problem of designing robust decision rules in the continuous-time limit. We have advocated specific continuous-time formulations for robust decision problems in light of how both types of misspecification contribute to measures of statistical discrepancy.

One of the two discrete-time formulations of robust decision-making with hidden states proposed by Hansen and Sargent (2007) conditions continuation values on information from only signal histories.<sup>11</sup> Unknown hidden states come into play in the evolution of these continuation values over time. Our continuous-time limit has a profound impact on how to parameterize a concern for robustness for the second

---

<sup>11</sup>See formulation (19) in that paper.

type of misspecification. We have shown how to design entropy penalties that sustain concerns about the distribution of hidden states by letting a robustness parameter in Hansen and Sargent (2007) be proportional to the time increment. When applied to dynamic models of smooth ambiguity as in Klibanoff et al. (2009), the analogous scaling allows this concern about ambiguity to be present in a continuous-time limit. In a continuous time environment, the decision-maker makes a smooth ambiguity adjustment to the local mean or drift of the continuation value process conditioned on the hidden state or parameter.

## A Discrete-state entropy problem

In this appendix, we follow Anderson et al. (2003) in computing the worst-case distortion of a discrete-state Markov chain. The continuation value is assumed to be given by

$$V_t = \lambda \cdot Y_t + \kappa \cdot Z_t + \xi.$$

Represent the intensity matrix  $A$  with a matrix  $R$  with nonnegative entries

$$A = R - \text{diag}\{R\mathbf{1}_n\}.$$

Consider an alternative specification of  $R$  given by  $S \otimes R$ , where  $S$  has all positive entries and  $\otimes$  denotes entry-by-entry multiplication.

The combined conditional relative entropy for the drift distortion for the Brownian motion and the distortion to the intensity matrix is

$$\text{ent}(h, S) = \frac{|h|^2}{2} + z \cdot \text{vec} \left[ \sum_j r_{ij} (1 - s_{ij} + s_{ij} \log s_{ij}) \right].$$

The associated distorted drift for the continuation value inclusive of the entropy penalty is

$$\lambda \cdot (\Delta y + \Gamma z + \sigma h) + z \cdot [(R \otimes S)\kappa] - (z \cdot \kappa) (z \cdot [(R \otimes S)\mathbf{1}_n]) + \theta_1 \text{ent}(h, s).$$

To compute the worst-case model for the state dynamics, we minimize this expression by choice of the vector  $h$  and matrix  $S$ .

The worst-case model appends the drift:

$$\tilde{h} = -\frac{1}{\theta_1} \sigma' \lambda$$

to the Brownian increment  $dW_t$  and includes a multiplicative distortion  $\tilde{S}$  to the matrix  $R$ :

$$\tilde{s}_{ij} = \exp \left( -\frac{1}{\theta_1} \kappa_j + \frac{1}{\theta_1} \kappa_i \right).$$

The minimized drift inclusive of the robustness penalty is

$$\lambda \cdot (\Delta y) - \frac{1}{2\theta_1} \lambda' \sigma \sigma' \lambda + \frac{z' \Delta \exp\left(-\frac{1}{\theta_1} \kappa\right)}{z' \exp\left(-\frac{1}{\theta_1} \kappa\right)},$$

where  $\exp\left(-\frac{1}{\theta_1} \kappa\right)$  is a vector with entries given by exponentials of the entries in the vector argument. The drift of the value function must satisfy

$$\lambda \cdot (\Delta y + \Gamma z) - \frac{1}{2\theta_1} \lambda' \sigma \sigma' \lambda + \frac{z' A \exp\left(-\frac{1}{\theta_1} \kappa\right)}{z' \exp\left(-\frac{1}{\theta_1} \kappa\right)} = \delta(\lambda \cdot y + \kappa \cdot z + \xi) - \delta H \cdot y,$$

which gives equations to be solved for  $\lambda$  and  $\kappa$ .

## References

- Anderson, Evan W., Lars Peter Hansen, and Thomas J. Sargent. 2003. A Quartet of Semigroups for Model Specification, Robustness, Prices of Risk, and Model Detection. *Journal of the European Economic Association* 1 (1):68–123.
- Barillas, Francisco, Lars Peter Hansen, and Thomas J. Sargent. 2008. Doubts or Variability? *Journal of Economic Theory* forthcoming.
- Cerreia, Simone, Fabio Maccheroni, Massimo Marinacci, and Luigi Montrucchio. 2008. Uncertainty Averse Preferences. Working paper, Collegio Carlo Alberto.
- Chen, Zengjing and Larry Epstein. 2002. Ambiguity, Risk, and Asset Returns in Continuous Time. *Econometrica* 70:1403 – 1443.
- Chernoff, H. 1952. A measure of the asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Annals of Mathematical Statistics* 23:493–507.
- Collard, Fabrice, Sujoy Mukerji, Kevin Sheppard, and Jean-Marc Tallon. 2008. Ambiguity and the historical equity premium. Universite des Sciences Sociales, University of Oxford and Universite Paris 1-Panthéon-Sorbonne.
- Duffie, Darrell and Larry G. Epstein. 1992. Stochastic Differential Utility. *Econometrica* 60 (2):353–394.
- Dupuis, Paul and Richard S. Ellis. 1997. *A weak convergence approach to the Theory of Large Deviations*. Wiley Series in Probability and Statistics. Wiley-Interscience.
- Gilboa, I. and D. Schmeidler. 1989. Maxmin Expected Utility with Non-Unique Prior. *Journal of Mathematical Economics* 18:141–153.
- Hansen, Lars Peter. 2007. Beliefs, Doubts and Learning: Valuing Macroeconomic Risk. *American Economic Review* 97 (2):1–30.
- Hansen, Lars Peter and Thomas J. Sargent. 2005. Robust estimation and control under commitment. *Journal of Economic Theory* 124 (2):258 – 301. Learning and Bounded Rationality.
- . 2007. Recursive robust estimation and control without commitment. *Journal of Economic Theory* 136 (1):1–27.

- . 2008a. Fragile beliefs and the role of model uncertainty. University of Chicago and New York University.
- . 2008b. *Robustness*. Princeton, New Jersey: Princeton University Press.
- Hansen, Lars Peter and José A. Scheinkman. 2009. Long-term Risk: A Operator Approach. *Econometrica* 77 (1):177–234.
- Hansen, Lars Peter, Thomas J. Sargent, Gauhar Turmuhambetova, and Noah Williams. 2006. Robust control and model misspecification. *Journal of Economic Theory* 128 (1):45–90.
- Hansen, Lars Peter, Jianjun Miao, and Thomas J. Sargent. 2009. Smooth Ambiguity, Robustness and Learning in Continuous Time. In progress, University of Chicago, Boston University and New York University.
- James, M. R. 1992. Asymptotic Analysis of Nonlinear Stochastic Risk Sensitive Control and Differential Games. *Mathematics of Control, Signals and Systems* 5:401–417.
- Ju, Nengjui and Jianjun Miao. 2007. Ambiguity, learning and asset returns. Boston University and Hong Kong University of Science and Technology.
- Klibanoff, Peter, Massimo Marinacci, and Sujoy Mukerji. 2005. A Smooth Model of Decision Making under Ambiguity. *Econometrica* 73 (6):1849–1892.
- . 2009. Recursive Smooth Ambiguity Preferences. *Journal of Monetary Economics* 144:930–976.
- Kreps, D. M. and E. L. Porteus. 1978. Temporal Resolution of Uncertainty and Dynamic Choice. *Econometrica* 46:185–200.
- Newman, C. M. and B. W. Stuck. 1979. Chernoff bounds for discriminating between two markov processes. *Stochastics* 2:139 – 153.
- Peleg, Bezalel and Menahem E. Yaari. 1973. On the Existence of a Consistent Course of Action when Tastes are Changing. *The Review of Economic Studies* 40 (3):391–401.
- Skiadas, Costis. 2009. Smooth Ambiguity Aversion Toward Small Risks and Continuous-Time Recursive Utility. Northwestern University.