Approximate Models and Robust Decisions: A Discussion*

Lars Peter Hansen
Departments of Economics and Statistics, University of Chicago

Massimo Marinacci
Department of Decision Sciences and IGIER, Università Bocconi

April 26, 2016

1 Introduction

In their paper, Watson and Holmes (2015) follow the statistical decision approach pioneered by Wald (1950). Under Wald’s perspective, the aim of the analysis shifts from the discovery of a statistical “truth” (for instance, as revealed by the correct statistical model), to making decisions that are defensible according to posited objective functions that trade off alternative aims. We also draw on decision theory in our discussion because it provides a formal framework for confronting uncertainty.

Decades ago, Arrow (1951) distinguished two sources of uncertainty: (i) risk within a model, where the uncertainty is about the outcomes that emerge in accordance to a model that specifies fully the outcome probabilities; and (ii) ambiguity among models, where the uncertainty is about which alternative model, or convex combination of such models, should be used to assign the probabilities. If the true model is not assumed to be among the original set of models under consideration, a third source of uncertainty emerges, (iii) model misspecification, here uncertainty is induced by the approximate nature of the models under consideration to use in assigning probabilities. These different sources of uncertainty

*We thank Pierpaolo Battigalli, Thomas Sargent, and Stephen Stigler for helpful comments.
are inherent to any analysis that includes decision makers who have (probabilistic) theories about the outcomes and form beliefs over their relevance.

How to accommodate potential model misspecification is a challenging topic. On the one hand, if we have very precise information about the nature of the misspecification, then presumably we would fix or repair the model. On the other hand, if we allow for too large of a set of possible ways for a model to be misspecified, we may find that little can be said of value in confronting the decision problem. The interplay between tractability and conceptual appeal is a central consideration when producing tools that aid in statistical decision making. By using formal decision theory to frame and analyze misspecification, the advances described by Watson and Holmes (2015) help to nurture further connections between statistics and economics, and they allow for the incorporation of further insights from statistics into economic analyses.

In their essay, Watson and Holmes (2015) draw connections to some research in economics. Our comment will describe other important advances in decision theory within the economics discipline that are designed to confront uncertainty conceived broadly to include an aversion to ambiguity and a concern about model misspecification. We will also delineate some special challenges for applications in the social sciences.

Statisticians and econometricians estimate parameters and assess the specification of alternative models. But economics and other social sciences also use decision theory for a second purpose. People inside the models that we build must speculate about the future when making investment and other forward-looking decisions. The people inside the models could be individuals from the private sector of the market economy or they could be policy makers who employ econometricians to help them in making better decisions with either self-serving or social objectives. It has long been understood within economics that beliefs about the future are important inputs into model construction, and this opens the door to letting people inside the economic models to use data and statistical methods to help shape their beliefs. Economists have long debated what degree of statistical sophistication we should ascribe to the people inside our models. Thus, there is a role for depicting decision making under uncertainty to help capture behavior inside the models we build. Doing so can have substantial consequences for equilibrium outcomes of the economic analysis. Statistical challenges, thus, show up in two places. They are present inside the models we build to capture the behavior of astute decision makers and outside these models when econometricians adapt and apply statistical methods to models that interest them. See, for instance, Hansen (2014) for more discussion of this point.
Statistical challenges inside the models we build influence how we shape and apply
decision theoretic concerns about model misspecification. Moreover, dynamics are central
to many economic analyses. This interest in dynamics has had a substantial impact in the
development of decision theory within economics at least since Koopmans (1960) and also
within control theory. The interplay between concerns about model misspecification in a
dynamic, stochastic environment along with the desire for tractable recursive formulations
has led to some recent advances in decision theory within economics that we will briefly
describe.¹

2 Decision theory

Decision theory aims to describe how a person should behave in an uncertain environment.
For a statistician, decision theory typically guides a choice of a model or an estimator of
an underlying parameter vector captured by the possibly infinite dimensional parameter
vector $\theta$. Econometrics often adopts this same perspective. This is captured by an unknown
parameter vector $\theta$ that resides in a set $\Theta$. Given $\theta$, a random vector $Y \in \mathcal{Y}$ is described
by a probability density $\psi(y|\theta)$ relative to a measure $\tau$ over $\mathcal{Y}$, where $y$ denotes a generic
element of $\mathcal{Y}$. A decision maker observes a realization $y$ and takes an action $a \in \mathcal{A}$ that
can depend on $y$. Formally, an action (or decision) rule is a suitably measurable function
$A : \mathcal{Y} \rightarrow \mathcal{A}$.

Represent the decision maker’s preferences in terms of a utility function $U(a, y, \theta)$.
Integrate over $y$ to construct expected utility conditioned on $\theta$:

$$U(A | \theta) = \int_{\mathcal{Y}} U(A(y), y, \theta) \psi(y|\theta) \tau(dy),$$

which will be an important ingredient of our decision theories. The expected utility, as we
have computed it, conditions on the parameter $\theta$ that is typically unknown to the decision
maker. Statistical decision theory often regards $-U$ as a loss function and calls $-U$ a risk
function. Consistent with this label, we view the integration over $y$ conditioned on $\theta$ in
equation (1) as adjusting for risk.

When it comes to applying decision theories to economics (and other fields), the un-
known parameter $\theta$ may be an intermediate target. For instance, consider a decision maker
facing uncertainty captured by a future payoff relevant state. Represent this state as a ran-

¹See e.g. Gilboa and Marinacci (2013) and Marinacci (2015) for recent overviews.
dom vector $S$ with realized values $s$ in a set $S$. Let $\psi^*(a, y, \theta)$ denote the density relative to a measure $\tau^*$ over alternative $s$’s in $S$ conditioned on the current period action $a$ and observed data $y$. Consider a next period utility function $U^*$ that depends on $(s, a)$ and integrate over $s$ to construct:

$$U(a, y, \theta) = \int_{\mathcal{S}} U^*(s, a) \psi^*(a, y, \theta) \tau^*(ds).$$

Even when not directly payoff relevant, the $\theta$ dependence of $U$ is induced by the dependence of $\psi^*$ on $\theta$.

This formulation is amenable to dynamic programming methods where value functions are incorporated into the specification of $U$ and $U^*$, and $S$ is a future Markov state or shock that determines this Markov state given current information.\(^2\)

As posed so far, this representation of decision theory is incomplete. Following de Finetti (1937) and Savage (1954), we include a subjective prior probability $\pi$, and integrate over the posited $\theta$. With this, we complete the specification:

$$\int \mathcal{U}(\theta|A) \pi(d\theta)$$

and can use this integral to rank alternative action rules $A$. Since the action rule depends on $y$, given $\pi$ we may rank alternative actions $a$ conditioned on $y$ by

$$\int U(a, y, \theta) \pi^*(d\theta)$$

where $\pi^*$ is the familiar Bayesian posterior

$$\pi^*(d\theta|y) \propto \psi(y|\theta) \pi(d\theta).$$

A Bayesian statistician’s job is to construct the posterior $\pi^*$.

As stated, however, in this specification there is no obvious scope for the expression of an aversion to model misspecification or to model ambiguity. As noted by Watson and Holmes (2015), both de Finetti and Savage acknowledge the challenge in using subjective probability to address such aversions.

\(^2\)The parameter $\theta$ itself might be extended to evolve in a Markovian fashion as in a hidden state Markov process.
3 Misspecification

We share the Watson and Holmes (2015) interest of exploring the impact of model misspecification. Issues that we discuss in this section are already relevant for uncertainty induced by model ambiguity, but concerns about model misspecification magnify their importance potentially by expanding substantial the set of models under consideration. There has been a rich set of extensions of decision theory that has emerged to confront uncertainty broadly conceived. These advances include some mentioned by Watson and Holmes (2015), but there are also others. The alternative approaches alter the inputs into a Bayesian decision problem in a variety of ways. Some follow Wald (1950)’s approach by relying on the game theoretic analysis of Von Neumann and Morgenstern (1944) to shape an approach to uncertainty.³

Consider now a convex function $C$ to model ambiguity about the prior $\pi$ and also the decision maker’s response to that ambiguity. The decision maker solves:

**Problem 3.1.**

$$\max_{A \in A} \min_{\pi} \int_{\Theta} U(A|\theta)\pi(d\theta) + C(\pi).$$

The cost function imposes a penalty on the choice of prior. Penalization methods are well known in both statistics and control theory. The preferences implicit in this decision problem are what Maccheroni et al. (2006b) call variational preferences. Such preferences nest the multiple priors specification of Gilboa and Schmeidler (1989), where the cost function takes on the extreme form of being equal to infinity if the priors are outside a convex set of priors $\Pi$ and zero inside. It extends the usual maximin approach to accommodate penalization, thus including formulations with a reference prior and a relative entropy penalty as proposed by Hansen and Sargent (2001). It accommodates robust Bayesian analysis in its systematic exploration of prior sensitivity with the use of a utility or loss function. The choice to minimize represents the aversion to ambiguity over the choice of prior or a concern about prior misspecification.

We may think of Problem 3.1 as a zero-sum game. When the order of extremization can be reversed without altering the objective, then we may obtain a so-called worst-case prior under which the decision maker optimizes by taking this prior as given. Bayesians such as Good (1952) argue for assessing the plausibility of this (restrained) worst-case prior. While

³Decision theories that extend (1) also violate the sure-thing principle, which has been viewed as a basic normative principle. However, the appeal of this principle becomes questionable under model uncertainty as Ellsberg (1961), pp. 653-654, famously illustrated with the three-color paradox.
this forges a link to Bayesian decision theory, notice that this “choice” of prior depends on
the utility or loss function rather than on a subjective introspection.

There is a seemingly different approach to this problem that features ambiguity aversion
but in a different way. Instead of penalizing or constraining a family of priors, it introduces
aversion to prior ambiguity in a way that is conceptually similar to risk aversion by including
a strictly increasing concave function $V$ as in the smooth ambiguity model of Klibanoff et al.
(2005):

**Problem 3.2.**

$$\max_{A \in A} V^{-1} \left( \int_{\Theta} V \left[ U(A|\theta) \right] \pi(d\theta) \right).$$

While this problem is not posed as one with a concern about prior sensitivity, the
informativeness or lack thereof in the prior does play a role in the decision criterion through
curvature in the function $V$. As noted by Hansen and Sargent (2007), for the familiar
and commonly used relative entropy formulation, there is a simple connection between
the penalization approach to prior sensitivity and the smooth ambiguity approach. Let $\pi_o$
denote a reference prior and $f$ denote a probability density with respect to $\pi_o$. If $F$ denotes
the family of such priors, then$^4$

$$\min_{f \in F} \int_{\Theta} U(A|\theta)f(\theta)\pi_o(d\theta) + \theta \int_{\Theta} f(\theta) \log f(\theta) \pi_o(d\theta) = -\theta \log \int_{\Theta} \exp \left[ -\frac{1}{\theta} U(A|\theta) \right] \pi_o(d\theta).$$

Thus, a particular form of a smooth ambiguity model emerges from a search over alternative
prior densities subject to a penalization.

## 4 Dynamics

Incorporating dynamics raises a host of interesting questions and has nurtured the develop-
ment of recursive formulations of decision problems. By design these recursive formulations
are amenable to application of dynamic programming methods which render the computa-
tion and characterization of solutions to economics models tractable. There are a variety
of conceptual changes once we entertain model misspecification.

$^4$This minimization problem is essentially a special case of an optimization problem with a relative
entropy penalty that emerges in a variety of areas of applied mathematics.
4.1 Misspecification in future dynamics

Connecting the future to the past is central to application of time series statistical methods. So far we have presumed that the density $\psi^*$ for the future state, as well as the density $\psi$ used to represent the observed data, depend on the same underlying parameter $\theta$. This gives us the opportunity to learn using Bayesian updating, even though potential model misspecification can impede this process. But even if we know $\psi$, it is of interest to explore uncertainty about $\psi^*$; this motivated the work of Hansen and Sargent (2001), Chen and Epstein (2002), and Anderson et al. (2003). It is also a central motivation for the dynamic extension of the class of variational preferences as discussed in Maccheroni et al. (2006a).

4.2 Dynamic consistency

Dynamic applications of decision making under uncertainty, often, but not always, look for formulations that are dynamically consistent to avoid having decision makers play naive or sophisticated games against future versions of themselves. Such aims can shape formulation of decision problems. Indeed the penalization approach suggested by Hansen and Sargent (2001) and the generalization developed by Maccheroni et al. (2006a) were motivated in part by such concerns.

Dynamic extensions of the multiple priors model of Gilboa and Schmeidler (1989) led Epstein and Schneider (2003) to embed a subjectively specified set of models or priors over models into a potentially larger rectangular set based on considerations of dynamic consistency. Such an approach does not always yield interesting answers, however. Suppose we use dynamic versions of statistical discrepancies, including ones used by Watson and Holmes (2015), to constrain ways in which models could be misspecified. The resulting rectangular embeddings are so large as to lead to degenerate outcomes. This phenomenon has led Hansen et al. (2006) and Hansen and Sargent (2016) to seek recursive implementations of so-called commitment problems solved from an *ex ante* perspective, much like what often occurs in control theory as reflected say in Petersen et al. (2000).

4.3 Robustness and dynamic learning

In dynamic settings, yesterday’s posterior is today’s prior. Bayes’ rule is a wonderful and often convenient recursion. Even under potential misspecification, guesses about the future are typically tied to past evidence. Robust learning and its impact on decisions provides a
fascinating twist to the direct Bayesian approach. There are a variety of approaches that have been suggested, including direct application of Bayes’ rule applied to a reference prior over models with an accompanying adjustment to the recursively generated posteriors (see e.g. Hansen and Sargent (2007) and Klibanoff et al. (2009)). Alternatively, Epstein and Schneider (2003) suggest a prior-by-prior application of Bayes’ rule in a dynamic multiple priors models with a rectangular embedding to enforce consistency. In a model with date zero prior ambiguity, Chamberlain (2000) embraces the prior-by-prior application of Bayes’ rule from an \textit{ex ante} perspective without including a rectangular embedding.

There remains scope for further dialog and discussion of this important topic as there are potentially important tradeoffs between conceptual appeal and computational tractability. We very much hope that the essay by Watson and Holmes (2015) and our discussion will encourage further synergistic research linking statistical challenges to economic model building and analysis.

\textbf{References}


de Finetti, Bruno. 1937. La Prevision: Ses Lois Logiques, Ses Sources Subjectives. \textit{Annales


