Ambiguity Aversion and Model Misspecification: An Economic Perspective

Lars Peter Hansen and Massimo Marinacci

1. INTRODUCTION

In their paper, Watson and Holmes (2016) 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 (probability) 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, where uncertainty is induced by the approximate nature of the models under consideration to use in assigning probabilities. These different sources of uncertainty 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 (2016) help to nurture further connections between statistics and economics, and they allow for the incorporation of additional insights from statistics into economic analyses.

In their essay, Watson and Holmes (2016) 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 economics.

Statisticians and econometricians use decision theory to guide how they 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. These 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 debate 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 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 formula
tions has led to some recent advances in decision
theory within economics that we will briefly describe
[see, e.g., Gilboa and Marinacci (2013) and Marinacci
(2015) for recent overviews].

2. DECISION THEORY

Decision theory aims to describe how a person
should behave in an uncertain environment. We follow Watson and Holmes (2016) in our use of decision
theory, except that we extend the notation in ways that
will support our discussion of potential model misspec-
fication in dynamic environments.

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 dimen-
sional 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 \( X \) with realized values \( x \) in a set \( \mathcal{X} \) is described by a probability model \( f(x|\theta) \), a density
relative to a measure \( \tau \) over \( \mathcal{X} \) that provides a proba-
bilistic specification of \( X \). A decision maker observes
a realization \( x \) and takes an action \( a \in A \) that can de-
pend on \( x \). Formally, an action (or decision) rule is a suitably measurable function \( A : \mathcal{X} \rightarrow A \).

Represent the decision maker’s preferences in terms of
a utility function \( U(a, x, \theta) \). Integrate over \( x \) to con-
struct expected utility conditioned on \( \theta \):

\[
U(A|\theta) = \int_{\mathcal{X}} U(A(x), x, \theta)f(x|\theta)\tau(dx),
\]

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 the-
tory often regards \(-U\) as a loss function and calls \(-U\)
a risk function. Consistent with this label, we view the
integration over \( x \) conditioned on \( \theta \) in equation (1) as
adjusting for risk.

When it comes to applying decision theories to eco-
nomics and other fields, the unknown parameter \( \theta \) may
be an intermediate target. For instance, consider a de-
cision maker facing uncertainty captured by a future
payoff relevant state. Represent this state as a random
vector \( S \) with realized values \( s \) in a set \( S \). Let
\( f^*(s|a, x, \theta) \) denote the density relative to a mea-
ure \( \pi^* \) over alternative \( s \)'s in \( S \) conditioned on the
current period action \( a \) and observed data \( x \). Con-
sider a next period utility function \( U^* \) that depends on
\( (s, a) \) and integrate over \( s \) to construct
\( U(a, x, \theta) = \int_S U^*(s, a)f^*(s|a, x, \theta)\pi^*(ds) \). Even when not di-
rectly payoff relevant, the \( \theta \) dependence of \( U \) is in-
duced by the dependence of \( f^* \) 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 a shock
that determines this Markov state given current informa-
tion.

As posed so far, this representation of decision the-
ory 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:

\[
\int_\Theta U(A|\theta)\pi(d\theta)
\]

and can use this integral to rank alternative action
rules \( A \). Since the action rule depends on \( x \), given
\( \pi \) we may rank alternative actions \( a \) conditioned
on \( x \) by
\[
\int_\Theta U(a, x, \theta)\pi^*(d\theta|x),
\]
where \( \pi^* \) is the fam-
iliar Bayesian posterior \( \pi^*(d\theta|x) \propto f(x|\theta)\pi(d\theta) \).
A Bayesian statistician’s job is to construct the poste-
rior \( \pi^* \).

As stated, however, in this specification there is no obvi-
ous scope for the expression of an aversion to model misspecification or to model ambiguity. As noted by Watson and Holmes (2016), both de Finetti
and Savage acknowledge the challenge in using sub-
jective probability to address such aversions.

3. MISSPECIFICATION

We share the Watson and Holmes (2016) interest of exploring the impact of model misspecification. Is-
ues that we discuss in this section are already relevant
for uncertainty induced by model ambiguity, but con-
cerns about model misspecification magnify their im-
portance potentially by expanding substantially the set
of models under consideration. A rich set of extensions of decision theory has emerged to confront uncertainty broadly conceived. These advances include some mentioned by Watson and Holmes (2016), but there are also others. The alternative approaches alter the inputs into a Bayesian decision problem in a variety of ways. Some follow Wald’s (1950) approach by relying on the game theoretic analysis of von Neumann and Morgenstern (1944) to shape an approach to uncertainty.

Watson and Holmes (2016) point to a rich literature on robust Bayesian methods that explores prior sensitivity in systematic ways. Motivated by applications in economic dynamics and control theory, decision theorists have found valuable to adopt the following starting point. Consider a convex function $C$ to assess a decision maker’s response to ambiguity about the prior $\pi$. The decision maker solves the following problem:

$$\max_{A \in A} \min_\pi \int \theta \cdot U(A|\theta) \pi(d\theta) + C(\pi). \tag{3}$$

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, Marinacci and Rustichini (2006b) call variational preferences. Such preferences nest the multiple priors specification of Gilboa and Schmeidler (1989), where the cost function $C$ takes on the extreme form of being equal to infinity if $\pi$ is 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 selection of the prior or a concern about prior misspecification.

We may think of Problem (3) 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 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.

The theory of constrained optimization uses Lagrange or Kuhn–Tucker multipliers as a convenient way to enforce constraints, say on a convex family of priors. This suggests a direct connection between the more special Gilboa and Schmeidler (1989) approach and that of Maccheroni, Marinacci and Rustichini (2006b) in many applications, including those explored by Hansen and Sargent (2001) and Hansen et al. (2006) and mentioned in Watson and Holmes (2016). From the perspective of the axiomatic analysis of Maccheroni, Marinacci and Rustichini (2006b) and of dynamic implementation, there are important conceptual distinctions, however.

There is a seemingly different approach to ambiguity about a prior that also features ambiguity aversion. 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 $\Phi$ as in the smooth ambiguity model of Klibanoff, Marinacci and Mukerji (2005). The decision maker then solves the following problem:

$$\max_{A \in A} \Phi^{-1} \left( \int \Phi(U(A|\theta)) \pi(d\theta) \right). \tag{4}$$

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 $\Phi$. 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 $g$ denote a probability density with respect to $\pi_o$. If $\mathcal{G}$ denotes the family of such densities, then

$$\min_{g \in \mathcal{G}} \int \theta \cdot U(A|\theta) g(\theta) \pi_o(d\theta)$$
$$+ \lambda \int \theta \cdot g(\theta) \log g(\theta) \pi_o(d\theta)$$
$$= -\lambda \log \int \exp \left[ -\frac{1}{\lambda} U(A|\theta) \right] \pi_o(d\theta).$$

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 and essentially follows from Theorem 4.1 in Watson and Holmes (2016). Thus, a particular form of a smooth ambiguity model emerges from a search over alternative prior densities subject to a penalization.

We close this section by observing that decision theories that extend (2) 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), pages 653–654, famously illustrated with the three-color paradox.

4. DYNAMICS

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

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 $f^*$ for the future state, as well as the density $f$ used to represent the observed data, depend on an underlying parameter $\theta$. This gives us the opportunity to learn using Bayesian updating, even though potential model misspecification can impede this process. But if past data are only revealing for some of the components of $\theta$, there may remain uncertainty about how the economic environment will evolve in the future that is not fully captured by random shocks. This type of uncertainty motivated the work of Hansen and Sargent (2001), Chen and Epstein (2002) and Anderson, Hansen and Sargent (2003). It is also a central motivation for the dynamic extension of the class of variational preferences as discussed in Maccheroni, Marinacci and Rustichini (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. This aim can shape the formulation of decision problems. Indeed the penalization approach suggested by Hansen and Sargent (2001) and the generalization developed by Maccheroni, Marinacci and Rustichini (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 the ones used by Watson and Holmes (2016), to constrain ways in which models could be misspecified. The resulting rectangular embeddings are so large as to lead to degenerate outcomes [see Hansen and Sargent (2016) for a formal discussion]. This phenomenon has led Hansen et al. (2006) 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, James and Dupuis (2000). The added flexibility of the dynamic version of variational preferences characterized Maccheroni, Marinacci and Rustichini (2006a) gives an alternative attractive way to support the use of statistical discrepancy measures as noted originally by Hansen and Sargent (2001) and developed more fully in Hansen et al. (2006).

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, Marinacci and Mukerji (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 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 (2016) and our discussion will encourage further synergistic research linking statistical challenges to economic model building and analysis.
ACKNOWLEDGEMENTS

We thank Pierpaolo Battigalli, Thomas Sargent, Stephen Stigler and an anonymous referee for helpful comments.

REFERENCES


