Risk, Ambiguity, and Misspecification

Decision Theory, Robust Control, and Statistics

Lars Peter Hansen (University of Chicago) and
Thomas J. Sargent (NYU)

Statisticians' wisdom

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

Now it would be very remarkable if any system existing in the real world could be exactly represented by any simple model. However, cunningly chosen parsimonious models often do provide remarkably useful approximations. George Box (1979)

Why are we interested in this topic?

John F. Muth, "Rational Expectations and the Theory of Price Movements," 1961.

Rational Expectations Hypothesis:

> *... expectations of firms (or, more generally, the subjective probability distribution of outcomes) tend to be distributed, for the same information set, about the prediction of the theory (or the "objective" probability distributions of outcomes).*

What we want to do

- Build models that include people (investors and entrepreneurs) who themselves use statistical models and who, like econometricians, have doubts about their specifications.
- Explore how to acknowledge model uncertainty when statistically evaluating economic policies in the tradition of Haavelmo, Koopmans, Marschak, Hurwicz, Lucas, . . .
- Include econometric and statistical challenges formally.
- Formulate "confidence" in ways new to "behavioral economics".

Decision theories

- Statistical decision theory - Wald, Savage, Ferguson
  - axiomatic - Savage
  - partial ordering - admissibility
  - complete class theorem
- Extensions of axiomatic decision theory (within economics) that formalize ambiguity aversion as distinct from risk aversion
  - max-min expected utility - Gilboa and Schmeidler
  - smooth ambiguity aversion - Klobanoff, Marinacci, and Mukerji
  - variational preferences - Maccheroni, Marinacci and Rustichini
  - objective and subjective rationality - Gilboa, Maccheroni, Marinacci and Schmeidler

Our starting point

Likelihoods and priors are central objects in the statistics literature but are obscure in the economics literature that is motivated by revealed preferences.

Findings

- modifications of Savage-style axiomatic formulations in the economics literature open doors to extending notions of uncertainty beyond risk in ways that make contact with applied econometric challenges
- we distinguish concerns about potential misspecifications of likelihoods from concerns about robustness of alternative priors

Anscombe-Aumann (AA)

- preferences defined over acts
- act: maps state $\rightarrow$ probabilities (lotteries) over outcomes

In a static setting, we assume:

- a state is a parameter vector that indexes a statistical model
- each statistical model induces a probability distribution over outcomes
- a probability distribution over “states” is a prior distribution

Remark: For us, a statistical model conditions on an unknown parameter vector.

Static decision theory

Consider a parameterized model of a random vector with realization w:

$$\ell(w \mid \theta) d\tau_o(w)$$

where

$$\int_W \ell(w \mid \theta) d\tau_o(w) = 1$$

and $\theta \in \Theta$ and Θ is a parameter space. Put a baseline prior distribution π_o over Θ and consider a "decision rule" $\gamma(w)$.

Θ can be infinite dimensional.

Decision rules are constrained to belong to a restricted set of functions, e.g., $\gamma(w) = \Gamma(d, w)$ for $d \in D$.

$\gamma(w)$ is an uncertain outcome of a decision.

Subjective expected utility

Order preferences over γ by

$$\int_{\Theta} \left[\int_{W} u[\gamma(w)] \ell(w \mid \theta) d\tau_o(w) \right] d\pi_o(\theta).$$

Ambiguity?

> *... if I knew of any good way to make a mathematical model of these phenomenon [vagueness and indecision], I would adopt it, but I despair of finding one. One of consequences of vagueness is that we are able to elicit precise probabilities by self-interrogation in some situations but not others.*

Personal communication from L. J. Savage to Karl Popper in 1957

Sets of likelihoods and priors

- ▷ Likelihoods: Let $m(w \mid \theta) \geq 0$ in $\mathcal{M}$ satisfy

$$\int_W m(w \mid \theta)\ell(w \mid \theta) d\tau_o(w) = 1.$$

- ▷ Priors: Let $n(\theta) \geq 0$ in $\mathcal{N}$ satisfy

$$\int_\Theta n(\theta) d\pi_0(\theta) = 1.$$

Use these to explore two forms of possible misspecifications.

Statistical Divergences

Use two convex functions ϕ_m and ϕ_p for constructing divergence between probability measures. Each ϕ is a convex function with $\phi(1) = 0$ and $\phi''(1) = 1$ (normalization).

- ▷ For each θ, form statistical divergence

$$\int \phi_m[m(w \mid \theta)]\ell(w \mid \theta) d\tau_o(w) \geq 0.$$

- ▷ For priors over Θ, form

$$\int \phi_p[n(\theta)] d\pi_o(\theta) \geq 0.$$

We often use relative entropy, e.g. $\phi_m(m) = m \log m$.

Max-min expected utility: Gilboa Schmeidler

- Recall that π_o is a baseline prior
- Form $\mathcal{N}_o \subset \mathcal{N}$, a convex set; each $\mathcal{N}_o$ induces a prior $n(\theta) d\pi_o(\theta)$ over Θ
- Construct

$$Q = \left\{ q : q(w) = \int_\Theta \ell(w \mid \theta) n(\theta) d\pi_o(\theta), n \in \mathcal{N}_o \right\}.$$

  This is a convex set of predictive distributions.
- Represent preferences over γ with

$$\min_Q \left[\int_W u[\gamma(w)] q(w) d\tau_o(w) \right]$$

Axiomatic generalization: Maccheroni, Marinacci, and Rustichini

- Weak Certainty Independence: If $f, g \in \mathcal{A}$, $h, k \in \mathcal{A}_o$, and $\alpha \in (0,1)$, then

$$\alpha f + (1-\alpha)h \succsim \alpha g + (1-\alpha)h \Rightarrow \alpha f + (1-\alpha)k \succsim \alpha g + (1-\alpha)k$$

Observations:

- Weakens certainty independence by holding fixed α when making comparisons.
- Introduces a smooth tradeoff between the expected utility gain and the divergence loss.

Prior divergence

- Let π_o be a baseline prior.
- Consider alternative priors of the form $d\pi(\theta) = n(\theta) d\pi_o(\theta)$ for $n \in \mathcal{N}$.

Represent preferences over γ with:

$$\min_{n \in \mathcal{N}} \int_{\Theta} \left(\int_{W} u[\gamma(w)] \ell(w \mid \theta) d\tau_o(w) \right) n(\theta) d\pi_o(\theta) + \xi_p \int_{\Theta} \phi_p[n(\theta)] d\pi_o(\theta)$$

for $\xi_p > 0$.

Comments about prior divergences

- Reinterpret previous contributions to decision theory literature as representing a prior ambiguity instead of potential model misspecifications
- With relative entropy divergence, the implied preference ordering agrees with smooth ambiguity preferences but is rationalized in a fundamentally different way

Variational preferences for model misspecification concerns

- ▷ Condition on a specific θ.
- ▷ Replace $\ell(w \mid \theta) d\tau_o(w)$ with $m(w \mid \theta)\ell(w \mid \theta) d\tau_o(w)$ and explore consequences.
- ▷ Rank alternative γ's conditioned on θ by solving:

$$\min_{m \in \mathcal{M}} \int_W \left(u[\gamma(w)] m(w \mid \theta) + \xi_m \phi_m[m(w \mid \theta)] \right) \ell(w \mid \theta) d\tau_o(w)$$

for $\xi_m > 0$.

Observations:

- ▷ A prior distribution is not imposed over a space of alternative models
- ▷ Links to parts of robust control theory.

Robust Bayes with model misspecification, I

Represent preferences over γ using:

$$\min_{n\in\mathcal{N}_o}\min_{m\in\mathcal{M}}\int_\Theta\left(\int_W u[\gamma(w)]m(w\mid\theta)\ell(w\mid\theta)d\tau_o(w)\right)n(\theta)d\pi_o(\theta)$$
$$+\,\xi_m\int_\Theta\left(\int_W \phi_m[m(w\mid\theta)]\ell(w\mid\theta)d\tau_o(w)\right)d\pi(\theta)$$

The contribution of the divergence is zero whenever $m=1$ for some $n\in\mathcal{N}_o$.

Robust Bayes with model misspecification, II

Represent preferences over γ with:

$$\min_{n\in\mathcal{N}}\min_{m\in\mathcal{M}}\int_\Theta\left(\int_W u[\gamma(w)]m(w\mid\theta)\ell(w\mid\theta)d\tau_o(w)\right)n(\theta)d\pi_o(\theta)$$
$$+\xi_m\int_\Theta\left(\int_W \phi_m[m(w\mid\theta)]\ell(w\mid\theta)d\tau_o(w)\right)n(\theta)d\pi_o(\theta)$$
$$+\xi_p\int_\Theta \phi_p[n(\theta)]d\pi_o(\theta)$$

Joint divergence over (m,n).

Dynamics

- Hansen, and Sargent, *American Economic Review* , 2001
- Epstein and Schneider *Journal of Economic Theory*, 2003
- Maccheroni, Marinacci, and Rustichini *Journal of Economic Theory*, 2006
- Hansen and Miao, *Proceedings of the National Academy of Sciences*, 2018 and *Economic Theory*, 2022
- Hansen and Sargent, *Journal of Economic Theory*, 2022

Dynamics

Use conditional counterparts to the previous analysis

- explore consequences of misspecifying Markov transition dynamics by representing potential changes in probabilities as nonnegative martingales
- explore consequences of misspecifying priors/posteriors over alternative parameters
- address dynamic consistency
  - recursive construction of possible conditional probabilities over parameterized models
  - recursive construction of statistical divergences and their set counterpart

Hansen and Sargent (JET, 2022) confront a tension between dynamic consistency and admissibility

A dynamic discrete-time formulation

Three preference aggregator recursions

$$\widehat{V}_t = \frac{1}{1-\rho} \log \left[(1-\beta) \exp \left[(1-\rho) \widehat{C}_t \right] + \beta \exp \left[(1-\rho) \overline{R}_t \right] \right]$$

$$\widehat{R}_t = -\xi_m \log \mathbb{E} \left[\exp \left(-\frac{1}{\xi_m} \widehat{V}_{t+1} \right) \mid \mathfrak{A}_t, \theta \right]$$

$$\overline{R}_t = -\xi_p \log \mathbb{E} \left[\exp \left(-\frac{1}{\xi_p} \widehat{R}_t \right) \mid \mathfrak{A}_t \right]$$

where

- first one adjusts for discounting and intertemporal substitution
- second one adjusts for model misspecification (or risk aversion)
- third one adjusts for prior misspecification (or smooth ambiguity)

Other divergence measures

$$\widehat{R}_t = \min_{M_{t+1} \geq 0, \mathbb{E}(M_{t+1} | \mathfrak{A}_t, \theta) = 1} \mathbb{E}\left[M_{t+1} V_{t+1} + \xi_m \phi_m \left(M_{t+1}\right) \mid \mathfrak{A}_t, \theta\right]$$

$$\overline{R}_t = \min_{N_t \geq 0, \mathbb{E}(N_t | \mathfrak{A}_t) = 1} \mathbb{E}\left[N_t \widehat{R}_t + \xi_n \phi_n \left(N_t\right) \mid \mathfrak{A}_t\right]$$

Apply the Envelope Theorem and conclude

- $M\widehat{V}_{t+1} = M^*_{t+1}$
- $M\widehat{R}_t = N^*_t$

Stochastic discount factor

Recall that the marginal utilities:

$$MC_t = (1-\beta)\exp\left[(\rho-1)\widehat{V}_t\right](C_t)^{-\rho}$$
$$M\overline{R}_t = \beta\exp\left[(\rho-1)\widehat{V}_t\right]\exp\left[(1-\rho)\overline{R}_t\right]$$

By using the various marginal utility formulas, the stochastic discount factor ratio is

$$\begin{aligned}\frac{S_{t+1}}{S_t} &= \frac{MC_{t+1}M\overline{R}_t M^*_{t+1}N^*_t}{MC_t}\\ &= \beta\left(\frac{C_{t+1}}{C_t}\right)^{-\rho}\exp\left[(\rho-1)\left(\widehat{V}_{t+1}-\overline{R}_{t+1}\right)\right]M^*_{t+1}N^*_t\end{aligned}$$

where

- ▷ M^*_{t+1} adjusts for potential model misspecification
- ▷ N^*_t adjusts for potential prior misspecification