New Perspectives on Statistical Decisions under Ambiguity

Jörg Stoye*

Cornell University

September 4, 2011

Abstract

This survey summarizes and connects recent work in the foundations and applications of statistical decision theory. Minimax models of decisions making under ambiguity are identified as a thread running through several literatures. In axiomatic decision theory, they motivated a large literature on how to model ambiguity aversion. Some findings of this literature are reported in a way that should be directly accessible to statisticians and econometricians. In statistical decision theory, they inform a rich theory of estimation and treatment choice, which was recently extended to take into account partial identification and, thereby, ambiguity that does not vanish with sampe size. This literature is illustrated by discussing global, finite sample admissible and minimax decision rules for a number of stylized decision problems with point and partial identification.

Keywords: Statistical decision theory, ambiguity, minimax, minimax regret, treatment choice.

JEL code: C44, D81.

*I thank Chuck Manski for helpful comments. Address: Jörg Stoye, Department of Economics, Cornell University, Uris Hall, Ithaca, NY 14853, stoye@cornell.edu.

1 Introduction

Statistical decision theory touches upon questions that lie at the core of economics in several ways. In its application, it informs estimation and treatment choice and hence, how we learn from data and what advice we give. In its theoretical foundation, it is itself an application of economics, namely of (microeconomic) decision theory. This survey connects both aspects and aims to link three recently active literatures. First, there is renewed interest in statistical decision theory, whether frequentist or Bayesian, as the natural framework to think about estimation or treatment choice.[1] Second, the literature on partial identification is now rather broad.[2] Third, the foundations of statistical decision theory have received some attention in economic decision theory.

These literatures share the theme of decision making under ambiguity. Recall that ambiguity (or "Knightian uncertainty," after Knight (1921)) obtains when the outcomes corresponding to different actions are uncertain in a manner than cannot be described by a probability distribution. This contrasts with risk, where a probability distribution is available and where statisticians (more so than decision theorists) agree that actions should be assessed by expected utility. A typical problem in statistical decision theory combines aspects of both. Given a true data generating process or parameter value, the performance of a statistical decision rule is completely described by a probability distribution, and all prominent approaches agree that it can be assessed by expected utility or loss. However, one must choose a decision rule before learning the true parameter value, and no probability distribution over such values is easily available. Different approaches to statistical decision theory differ in how they deal with the resultant ambiguity.

This contrast becomes especially prominent in the presence of partial identification. Recall that in settings of conventional identification, called "point identification" henceforth, knowledge of the population distribution of observables implies knowledge of parameters of interest. In well-behaved cases, the latter are, therefore, revealed with arbitrary precision as samples expand. All reasonable estimators or decision rules then agree "in the limit," and different approaches to ambiguity are effectively about how to deal with finite sample problems. While this is already a challenging problem and the subject of a large literature, its importance is exacerbated in settings of partial identification. If true parameter values are not revealed in the limit, reasonable estimators or decision rules need not converge to each other as samples grow large, and ambiguity attitude can greatly affect decisions even for very large samples. Indeed, the issues raised by ambiguity "in the limit" are the subject of an entire

[1] See the literature starting with Chamberlain (2000) and Manski (2000) and continuing with references cited later. For surveys that integrate decision theory with econometrics, see Manski (2007b) for a frequentist and Geweke, Koop, and van Dijk (eds., 2011) for a Bayesian perspective.

[2] I will presume knowledge of basic concepts in partial identification. Manski (2003) and Tamer (2011) are good entry points to this literature.

survey in these pages (Manski (2011b)) that abstracts from finite sample issues. One contribution of the present paper is to link this work to the extensive literature that uses the opposite simplification, dealing with finite sample problems but assuming point identification.

The problem of ambiguity "in the limit" also lends additional salience to normative justification of decision criteria that differ in their handling of ambiguity. One potential source of such justification is axiomatic decision theory. Non-Bayesian approaches to ambiguity, in particular models of *ambiguity aversion*, recently received much attention in axiomatic decision theory, but findings are rarely communicated to statisticians or econometricians. To remedy this, this survey contains a summary of relevant work that is geared toward, or immediately applies to, the specific questions asked by statistical decision theorists. This renders the survey somewhat idiosyncratic from an economic theorist's point of view: I use statistical (as opposed to decision theoretical) notation and focus on models that embed expected utility in a specific manner that is common to all models in statistical (but not general) decision theory.

This paper is structured as follows. The next section introduces notation and defines numerous optimality criteria. Section 3 uses axiomatic characterizations to highlight abstract trade-offs involved in choosing among them. Section 4 illustrates more practical trade-offs by training the criteria on some canonical decision problems, as well as extensions that are characterized by partial identification. In both sections, attention is exclusively on exact finite sample results; indeed, this survey completely avoids asymptotic approximations. Section 5 concludes and hints at connections to other literatures.

2 Setting the Stage

2.1 Environment

The setting is Wald's (1950), with notation following Ferguson (1967).[3] There is a space (Θ, Σ_Θ) of possible values of a parameter θ. The true value of θ (typically labeled θ_0 in econometrics) is unknown, but the decision maker observes a realization ω of a random variable distributed according to probability distribution P_θ on sample space (Ω, Σ_Ω). The subscript on P indicates that the distribution of ω varies with θ, thus ω is informative about which θ occurred. I will also call ω a "signal." In most applications, ω is generated by a sampling process.

The decision maker must choose an action a from an action space $(\mathcal{A}, \Sigma_\mathcal{A})$. This action can be conditioned on the signal, thus the decision maker really picks a (Σ_Ω-measurable) function $\delta : \Omega \to \mathcal{A}$ that is henceforth called a *statistical decision rule*. We will assume that decision makers can randomize. This is either captured by an understanding that δ can be picked randomly or by explicitly

[3] See also Berger (1985) and in economics Hirano (2009) and Manski (2007b, 2011b).

incorporating randomization in δ. I will occasionally do the latter and then write δ as mapping from Ω into $\Delta\mathcal{A}$, the set of distributions on $(\mathcal{A}, \Sigma_{\mathcal{A}})$. Alternatively, one could extend the domain of δ to be $\Omega \times [0,1]$, where the second argument of δ corresponds to a realization of a uniformly distributed random variable (the randomization device).

Example 1: Estimating a binomial proportion

Denote by $Y_1 \in \{0,1\}$ the possible outcomes of a coin toss, where an outcome of 1 will also be called a "success." Consider the problem of estimating $\theta = \mu_1 = E(Y_1)$ from n i.i.d. tosses. Then $\Theta = [0,1]$, $\Omega = \{0,1\}^n$ with obvious sigma-algebra, a typical realization ω is a binary sequence $(0,1,...,1)$ of length n, and P_θ is characterized by $\Pr(\omega = (0,1,...,1)) = \mu_1^{n\overline{y}} \cdot (1-\mu_1)^{n-n\overline{y}}$, where $\overline{y}$ is the sample average and $n\overline{y}$ therefore the count of successes observed.

If the aim is to estimate θ, then the natural set of actions is $\mathcal{A} = [0,1]$, and any function $\delta : \{0,1\}^n \to \Delta[0,1]$ is a possible decision rule. A salient loss function is square loss, i.e. $L(a, \mu_1) = (a - \mu_1)^2$, leading to risk function

$$r_1(\delta, \theta) = \int (\delta(\omega) - \mu_1)^2 dB(\mu_1, n),$$

where $B(\mu_1, n)$ is the binomial distribution with indicated parameters.

Example 2: Adoption of an innovation.

A social planner must decide whether to adopt a treatment when the performance of a status quo treatment is well understood. This will be modelled by presuming that there are a binary treatment $t \in \{0,1\}$ and two expectations of interest, $(\mu_0, \mu_1) = (E(Y_0), E(Y_1))$. Here, Y_t are *potential outcomes* that would obtain under treatment t, $\mu_0 \in [0,1]$ is known, and $\mu_1 \in [0,1]$ is unknown but can be (implicitly) estimated from the same experiment as in example 1. Thus, Ω is as before, and a treatment rule is a function $\delta : \Omega \to [0,1]$ that associates every possible sample realization with an induced probability of assigning treatment 1.[4]

Assume that evaluation of treatment rules conditional on $\theta = (\mu_0, \mu_1)$ is by expected outcome $\int \mu_1 \delta(\omega) + \mu_0(1 - \delta(\omega)) dB(\mu_1, n)$. Risk is the negative of welfare, so the risk function is

$$r_2(\delta, \theta) = -\int [\mu_1 \delta(\omega) + \mu_0(1 - \delta(\omega))] \, dB(\mu_1, n).$$

In interesting decision problems, no decision rule will have lowest possible risk for all possible parameter values. While some risk functions may be dominated by others, there will typically be a

[4] Only expectations taken over the treatment population matter for any of the decision criteria discussed later. Hence, it is immaterial whether said probability applies to all subjects simultaneously (all get assigned to the same treatment, this being treatment 1 with probability $\delta(\omega)$) or individually (by randomly assigning a fraction of $\delta(\omega)$ subjects into treatment 1, a method that Manski (2009) called *fractional treatment assignment*). Treatment rules that are not randomized (nor fractional) only take values in $\{0,1\}$.

large menu of nondominated ones. Choosing one of them then is to (at least implicitly) commit to a specific attitude toward ambiguity. Some systematic ways of doing so will now be explored.

2.2 Decision Criteria

The following optimality criteria will be considered.

Admissibility

A decision rule δ is inadmissible (dominated) if there exists δ' s.t. $r(\delta, \theta) \geq r(\delta', \theta)$ for all θ, with strict inequality for some θ. A decision rule is admissible otherwise.

Admissibility seems incontrovertible – for sure, nobody would insist (in the sense of strict preference) on choosing an inadmissible decision rule if a dominating one is known. Unfortunately, so many decision rules will be admissible in any interesting decision problem that in practice, admissibility should be thought of more as carving out a consensus than as guiding one's ultimate decision making. Optimality criteria that lead to sharper recommendations are as follows.

Bayes

Let π be a probability on (Θ, Σ_Θ). A decision rule δ is Bayes against prior π if $\delta \in \arg\min_{d \in \mathcal{D}} \int r(d, \theta) d\pi$.

Minimax Loss

A decision rule δ attains minimax loss if $\delta \in \arg\min_{d \in \mathcal{D}} \max_{\theta \in \Theta} r(d, \theta)$.

Minimax Regret

A decision rule δ attains minimax regret if $\delta \in \arg\min_{d \in \mathcal{D}} \max_{\theta \in \Theta} \{r(d, \theta) - \min_{d^* \in \mathcal{D}} r(d^*, \theta)\}$.

Minimax loss is the best known alternative to the Bayesian approach.[5] It evaluates decision rules by imputing a worst-case scenario as opposed to aggregating loss with respect to a prior. Minimax regret has a similar intuition, but evaluates the worst-case scenario with respect to ex-post efficiency loss. While psychological interpretations of criteria are not a main concern here, this ex-post loss may be loosely intuited as the regret experienced by a decision maker who chose d and then learned that parameter value θ obtained.

Minimax loss became prominent in statistics through the work of Wald (1945, 1950). Minimax regret was how Savage (1951) initially (but mistakenly, see Savage (1954)) interpreted Wald. It was independently suggested by Niehans (1946) and recently emphasized by Manski (2004); see Stoye

[5] It is often called just minimax; I call it minimax loss to clarify the distinction from minimax regret. Also, assumptions made up to here do not ensure existence of minima and maxima. Existence obtains in all of this paper's examples though, so I avoid sup and inf notation for readability.

(2009b) for recent references. The two criteria agree if $\min_{d^* \in \mathcal{D}} r(d^*, \theta) = 0$ for all θ. This is the case in many applications, notably estimation under conventional loss functions, including example 1. It is not the case in example 2, where risk is r_2 as defined above but regret is

$$\max\{\mu_0, \mu_1\} - \int \mu_1 \delta(\omega) + \mu_0 (1 - \delta(\omega)) dB(\mu_1, n),$$

leading to very different solutions.

These minimax criteria may appear extremely pessimistic, especially if some – but perhaps vague – prior information about parameter values is available. One might, therefore, want to find a middle ground between them and the Bayesian approach. This motivates robust Bayesian models, i.e. imposition of (convex) sets of priors Γ. Such a set generates a set of admissible decision rules that will be smaller than the previous such set but not a singleton. The question of which decision rule to pick will arise again, and it can be resolved in analogy to the above minimax criteria. Thus, one gets:

Γ-Admissibility

A decision rule δ is Γ-inadmissible (with respect to a set of priors Γ) if there exists δ' s.t. $\int r(\delta, \theta) d\pi \geq \int r(\delta', \theta) d\pi$ for all $\pi \in \Gamma$, with strict inequality for some π. A decision rule is Γ-admissible otherwise.

Γ-Minimax Loss

A decision rule δ attains Γ-minimax loss if $\delta \in \arg\min_{d \in \mathcal{D}} \max_{\pi \in \Gamma} \int r(d, \theta) d\pi$.

Γ-Minimax Regret

A decision rule δ attains Γ-minimax regret if $\delta \in \arg\min_{d \in \mathcal{D}} \max_{\pi \in \Gamma} \int \left[r(d, \theta) - \min_{d^* \in \mathcal{D}} r(d^*, \theta) \right] d\pi$.

By identifying Γ with the set of all distributions over Θ, these criteria can be seen to nest the preceding ones as special cases. At the other extreme, if Γ is a singleton, they recover Bayesianism. In between, they generate smooth transitions between Bayesianism and "extreme minimax" approaches.

The name "Γ-admissibility" is new to the literature but would seem obvious for terminological consistency. The underlying idea corresponds to Walley's (1990) maximality, is close in spirit to Levi's (1980) E-admissibility, and was recently axiomatized (see below). Γ-minimax loss will appear familiar to readers acquainted with the decision theoretic literature on ambiguity aversion: Reading loss as the negative of utility, it is the maxmin expected utility model axiomatized by Gilboa and Schmeidler (1989). Indeed, axiomatic characterizations of all of the above criteria are available, and it is to them that we now turn.

3 Statistics and Axiomatic Decision Theory

3.1 Introduction

One way to understand the trade-offs involved in choosing an optimality criterion is axiomatic analysis. To do so, think of decision criteria as arising from maximization of a preference $\succsim$ over the risk functions induced by different decision rules. For example, a minimax loss decision maker prefers decision rule δ_1 to δ_2 iff $\max_{\theta \in \Theta} r(\delta_1, \theta) \leq \max_{\theta \in \Theta} r(\delta_2, \theta)$. Different optimality criteria correspond to different preferences. Axiomatic analysis proceeds by listing a number of consistency properties of preferences within and across choice problems. In principle, these can have normative (as desirable features of choice behavior) or descriptive (as conjectured empirical regularities) interpretation. A characterization result then shows that certain of these properties are fulfilled iff preferences are described by a specific model. In the normative interpretation, this is read as an argument justifying the model; in the descriptive interpretation, it is a parameter-free elucidation of its empirical content. We will be concerned with normative interpretations. A famous example in statistics is Savage's (1954) normative (in original intent, not always in subsequent use) characterization of Bayesianism.

I will actually depart from Savage and frame this paper's discussion by taking for granted the device of risk functions. In particular, it is presumed that risk functions are sufficient descriptions of decision rules' performance: Two decision rules that have the same risk function are considered equivalent, and dominance with respect to risk functions is to be respected. Under this premise, an Anscombe-Aumann (1963) framework is an obvious choice. Furthermore, by agreeing that a risk function completely characterizes a decision rule's performance, virtually all statisticians agree on a number of axioms – for example, that expected utility is appropriate in the absence of ambiguity – that have received much discussion on decision theory. To focus attention on those trade-offs that are important in comparing the above optimality criteria, I will leave all those controversies implicit and take for granted that we compare transitive and monotonic (in the sense of respecting dominance) orderings of risk functions.[6] Indeed, since risk functions fully describe decision rules, I will temporarily

[6] This footnote is for decision theorists who want to better understand how the settings connect. Standard notation for an Anscombe & Aumann (1963) setting would begin with a space of ultimate outcomes $\mathcal{X}$ and a state space $(\mathcal{S}, \Sigma_{\mathcal{S}})$. Acts f are $\Sigma_{\mathcal{S}}$-measurable mappings from $\mathcal{S}$ to the set of finite lotteries over $\mathcal{X}$, thus they map states onto objective probability distributions over ultimate outcomes. An act is constant if $f(s)$ does not depend on s.

All statistical criteria considered here embed expected utility in the sense that preference over constant acts $p \in \mathcal{X}$ is expected utility, thus it is represented by an integral $\int U(x)dp$, where $U : \mathcal{X} \to \mathbb{R}$ is a utility function. Second, acts are completely summarized by the corresponding *utility acts* $u \circ f$ which map states of the world s onto expected utilities $u \circ f(s) = \int U(x)df(s)$.

Utility acts are not primitive objects in decision theory; they play a central role only given specific and perhaps not innocuous axioms. However, these axioms are fulfilled by all criteria considered here, and utility acts are crucial in transferring results from economic theory to statistics because they correspond to risk functions. Thus, statisticians drop δ from notation and argue in terms of a potentially menu-dependent family of orderings $\succsim_M$ of risk functions that are being compared in a specific menu $M = \{r_1, r_2, ...\}$.[7] The axioms that differentiate between this paper's optimality criteria are:

von Neumann-Morgenstern Independence. For any risk functions r_1, r_2, r_3 and menu M containing $\{r_1, r_2\}$: $r_1 \succsim_M r_2$ iff $\lambda r_1 + (1-\lambda) r_3 \succsim_{\{\lambda r + (1-\lambda) r_3 : r \in M\}} \lambda r_2 + (1-\lambda) r_3$ for all $\lambda \in (0,1)$.

C-independence. Like independence, but the axiom is only imposed if r_3 is constant.

C-betweenness in special menus. Consider any menu M s.t. $\arg\min_{r \in M} r(\theta) = 0$ for every $\theta \in \Theta$ and any act $r_1 \in M$ and constant act $r_2 \in M$. Then $r_1 \sim_M r_2$ iff $\lambda r_1 + (1-\lambda) r_2 \sim_M r_2$ for all $\lambda \in (0,1)$.

Independence of irrelevant alternatives (IIA). For any risk functions r_1, r_2 and menus M, N containing $\{r_1, r_2\}$: $r_1 \succsim_M r_2$ iff $r_1 \succsim_N r_2$.

Independence of never strictly optimal alternatives (INA). For any risk functions r_1, r_2, r_3 and menu M containing $\{r_1, r_2\}$: If it is not the case that $\arg\min_{r \in M} r(\theta) = \{r_3\}$ for some $\theta \in \Theta$, then $r_1 \succsim_M r_2$ iff $r_1 \succsim_{M \cup \{r_3\}} r_2$.

Ambiguity aversion. For any risk functions r_1, r_2 and menu M containing $\{r_1, r_2\}$: $r_1 \sim_M r_2$ implies $\lambda r_1 + (1-\lambda) r_2 \succsim_M r_2$ for all $\lambda \in (0,1)$.

Symmetry. Consider any risk functions r_1, r_2, menu M containing $\{r_1, r_2\}$, and disjoint nonempty events $E, F \in \Sigma_\Theta$ s.t. any risk function in M is constant on both E and F. For any risk function r in M, let r' be the risk function generated from r by interchanging the consequences of events E and F. Then $r_1 \succsim_M r_2$ iff $r'_1 \succsim_{\{r' : r \in M\}} r'_2$.

Completeness. For any risk functions r_1, r_2 and menu M containing $\{r_1, r_2\}$: $r_1 \succsim_M r_2$ or $r_2 \succsim_M r_1$.

Mixture continuity. For any risk functions r_1, r_2, r_3 and menu M containing $\{r_1, r_2, r_3\}$: The sets $\{\lambda \in [0,1] : \lambda r_1 + (1-\lambda) r_3 \succsim_M r_2\}$ and $\{\lambda \in [0,1] : \lambda r_1 + (1-\lambda) r_3 \precsim_M r_2\}$ are closed.[8]

evaluate constant acts by von Neumann-Morgenstern expected utility (and by specifying a loss function, even provide U). More subtly, it ensures that if the decision maker randomizes over two decision rules, the objective probabilities generated by her randomization are compounded with the state-dependent lotteries induced by these acts. This is a nontrivial commitment – see Seo (2009) for a recent analysis that drops it – but, like other commitments that are reflected in the device of risk functions, will be left implicit.

[7] I take for granted that statisticians can randomize, thus $\succsim_M$ actually orders all mixtures over elements of M.

[8] Here is a technical subtlety. In the original statement of most results presented below, continuity is replaced with an Archimedean property: If $r_1 \succ_M r_2 \succ_M r_3$, then there exists $\lambda, \gamma \in (0,1)$ s.t. $\lambda r_1 + (1-\lambda) r_3 \succ_M r_2 \succ_M \gamma r_1 + (1-\gamma) r_3$. Given completeness, this condition is marginally weaker than mixture continuity, and the original results therefore marginally stronger than the versions stated here. However, this nesting depends on completeness, and the incomplete orderings characterized below fail the Archimedean property. I will, therefore, use mixture continuity throughout for unity of presentation.

Most of these axioms have received much discussion; see Gilboa (2009) and Binmore (2011) for recent examples. I will add my own take on a few. Independence is usually defended with the following thought experiment: Imagine you like r_1 more than r_2, but then you are told that your decision only matters if a previous coin toss came up head; else, you will get r_3 anyway. Then you should still like r_1 over r_2. In judging whether this intuition is appealing, the reader should be aware that r_3 need not be constant – in the extreme, it might be that r_3 perfectly hedges against r_2, so that $\frac{1}{2}r_2 + \frac{1}{2}r_3$ is constant. If the intuition is considered compelling but only for constant r_3, then it is an intuition for c-independence only, a subtlety that will make a big difference. In any case, note that in this thought experiment, the previous coin toss occurs whatever one's choice, so it should be thought of as r_3 being mixed into the entire menu. This is reflected in the axiom's statement.

Independence of irrelevant alternatives is self-explanatory. Independence of never strictly optimal alternatives weakens it in a specific way: Changing a menu may change preferences within it, but only if it affects the lowest possible risk that can be achieved for at least one θ. Thus, the channel through which menus may affect preferences are restricted in a way that is intuitively connected to regret and will be instrumental in characterizing the according criteria.

Symmetry (Arrow & Hurwicz (1972)) reflects an unwillingness to implicitly assign different likelihoods to different events. This is undesirable if one has access to prior information about states; this is as it should be, because the axiom will prevent the imposition of priors. However, if no such information exists, symmetry makes sense since a decision criterion might otherwise be sensitive to arbitrary manipulations of the state space, either by relabeling states or by duplicating some via conditioning on trivial events.

Ambiguity aversion (Schmeidler (1989); see also Milnor (1954)) plays a central role in much recent decision theoretic work. It mandates a weak preference for randomization, the intuition being that such randomization may constitute a hedge against ambiguous states. Note that the mandated preference is only weak, however. Independence-type axioms can be seen as limiting the scope for this preference to be strict. Most notably, independence and IIA jointly imply ambiguity aversion but with an indifference in the conclusion, an attitude that should probably be called ambiguity *neutrality*. C-independence (Gilboa & Schmeidler (1989); see also Milnor (1954)) denies a hedge only if r_3 is constant. Intuitively, c-betweenness is even weaker (though non-nested as it is imposed within menus, not across them): It says that if the best possible ex-post risk is the same in every state, then mixing an act with a constant act to which it is indifferent anyway – i.e., its certainty equivalent – cannot constitute a hedge. The axiom is instrumental in characterizing Γ-minimax regret.

For an intuition as to why ambiguity aversion might clash with the conjunction of independence and IIA, consider an example inspired by Ellsberg (1961). An urn contains red and blue balls in unknown proportion, and you are invited to buy either a red or blue ticket. A ticket wins $2 if its color matches the next ball being drawn and nothing otherwise. Let's assume risk neutrality, thus you are willing to pay \$1 for a lottery that is known to yield \$2 with probability 1/2. Nonetheless, given the ambiguity about the urn, your willingness to pay for either the red or the blue ticket might be less than \$1, say 50c for either. This seems intuitive enough. Now imagine that you may toss your own coin, which you know to be fair, and purchase your ticket only after seeing the coin toss (but before the ball is drawn). In particular, you are free to buy the red ticket iff the coin came up heads. Independence and IIA jointly imply that you are willing to pay precisely 50c for this randomized lottery ticket as well. However, despite feeling ambiguous about the proportion of red balls in the urn, you might be confident that this proportion does not change in reaction to extraneous coin tosses.[9] In this case, the randomized lottery ticket has a 50% chance of winning irrespective of the true proportion of red balls, and you should be willing to pay \$1 for it. While the perfect hedge implied here is an extreme example, the general behavior is typical of both minimax loss and minimax regret decision makers.

3.2 Main Result

The following summarizes some classic and some recent findings.[10]

Result

Assume that a family of orderings $\succsim_M$ over risk functions fulfils ambiguity aversion and mixture continuity. Then:

(i) If it fulfils independence of irrelevant alternatives, von Neumann-Morgenstern independence, and symmetry, then it is the (incomplete) ordering corresponding to admissibility.

(ii) If it fulfils independence of irrelevant alternatives and von Neumann-Morgenstern independence, then it is the (possibly incomplete) ordering corresponding to Γ-admissibility.

(iii) If it fulfils independence of irrelevant alternatives, symmetry, and completeness, then it corresponds to minimax loss.

(iv) If it fulfils independence of irrelevant alternatives, c-independence, and completeness, then it corresponds to Γ-minimax loss.

(v) If it fulfils independence of never strictly optimal alternatives, von Neumann-Morgenstern independence, symmetry, and completeness, then it corresponds to minimax regret.

[9] Indeed, the use of risk functions mandates this confidence, perhaps explaining why not all decision theorists find it completely innocuous.

[10] All of these results hold with an infinite state space Θ, though strictly speaking some of them have been established only for risk functions that are finite step functions from Θ to $\mathbb{R}$ and menus that contain (randomizations over) finitely many acts. An extension to well-behaved infinite functions along the lines of Gilboa & Schmeidler (1989) would seem feasible throughout.

Another subtlety is that priors characterized in the below results can be finitely (as opposed to countably) additive. See Ghirardato et al. (2004, remark 1) for how to remedy this with an additional axiom.

	adm	**Γ-adm**	**ML**	**Γ-ML**	**MR**	**Γ-MR**	**Bayes**
IIA	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$-$	$-$	$\oplus$
INA	$+$	$+$	$+$	$+$	$\oplus$	$\oplus$	$+$
independence	$\oplus$	$\oplus$	$-$	$-$	$\oplus$	$\oplus$	$\oplus$
c-independence	$+$	$+$	$+$	$\oplus$	$\oplus$	$\oplus$	$+$
c-betweenness	$+$	$+$	$+$	$+$	$+$	$\oplus$	$+$
symmetry	$\oplus$	$-$	$\oplus$	$-$	$\oplus$	$-$	$-$
completeness	$-$	$-$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$

Table 1: A visual representation of axiomatic characterizations.

(vi) If it fulfils independence of never strictly optimal alternatives, von Neumann-Morgenstern independence, c-betweenness for special menus, and completeness, then it corresponds to Γ-minimax regret.

(vii) If it fulfils independence of irrelevant alternatives, von Neumann-Morgenstern independence, and completeness, then it corresponds to Bayesianism.

(viii) It cannot simultaneously fulfil independence of irrelevant alternatives, von Neumann-Morgenstern independence, symmetry, and completeness.

Part (i) is from Stoye (2011e, motivated by the writing of this survey); (ii) is due to Gilboa et al. (2010, building on Bewley (2002)); (iii), (v), and (viii) are from Stoye's (2011a) update of Milnor (1954); (iv) is due to Gilboa and Schmeidler (1989), (vi) is a corollary of results about choice correspondences in Stoye (2011b; Hayashi (2008) earlier provided a similar result); and (vii) is due to Anscombe & Aumann (1963; see Fishburn (1970) for a version more obviously alike the present one).

The result is summarized in table 1, which is inspired by Milnor (1954). In this table, $\oplus$ denotes that an axiom is part of a criterion's characterization. A $+$ indicates that this is not so, yet the axiom is fulfilled anyway, frequently because a stronger version is imposed. A $-$ means that the axiom is not in general fulfilled (though it may be in special cases; notably, the rows corresponding to Γ-minimax criteria are best read as referring to sets of priors Γ that are neither singletons nor the maximal set).

Here is an attempt to translate the result into English. I would suggest that most axioms, with the notable exception of completeness, are substantive rationality axioms, i.e. consistency conditions – typically inspired by thought experiments – that we would ideally like to see fulfilled. Their conjunction characterizes admissibility, which therefore lies at the intersection of the other decision criteria not only intuitively but in a specific axiomatic sense.[11] In particular, the fact that the conjunction of all axioms characterizes what can be thought of as baseline consensus or "greatest common denominator" across

[11] To be precise, the weak part of the admissibility ordering is the intersection of the weak parts of all other orderings.

decision makers might give us confidence that there is some sense to the axioms. But it also means that imposing all of them at once does not allow us to go beyond the obvious in comparing risk functions.

To arrive at sharp decision recommendations, we will generally need a complete ordering. Imposition of completeness is motivated by this pragmatic concern and not by any claim of its centrality to substantive rationality. Unfortunately, it induces a contradiction. Thus, we face a genuine trade-off: No optimality criterion that fulfils all desirable consistency conditions is just waiting to be discovered, and in order to arrive at a complete ordering, we must privilege some axioms over others. The quick summary is as follows: Relaxing independence leads to minimax loss. Relaxing independence of irrelevant alternative leads to minimax regret. Relaxing symmetry leads to Bayesianism. In addition, symmetry plays a special role because it excludes any prior information whatsoever. Both of the minimax criteria are turned into their robust Bayesian counterparts by relaxing it.

3.3 Further Remarks

3.3.1 More on Interpretation

What to make of these results? If one has a strong preference for a specific subset of axioms, then one could just pick the according decision rule. For example, this is how Binmore (2011) defends his preferred decision rules. But things might not be as easy for all readers. All axioms have some appeal to them, and yet they cannot be simultaneously fulfilled. More subtly, their appeal sometimes stems from the fact that they are related to appealing features of certain optimality criteria, so intuitions in favor of certain criteria are not well separated from intuitions in favor of certain axioms. In this author's view, the point of the axiomatizations is not, therefore, to replace the consideration of specific examples, but to extract in the sharpest possible form the issues raised by such examples.

To illustrate, consider an example from Berger's (1985, page 372 and figure 5.8, reproduced here as figure 1) classic textbook. The minimax loss risk function is the dashed one in the left-hand panel and the solid one in the other panel. Berger finds it objectionable that adding one and the same function – in the example, the indicator of a particular event – to both risk functions can cause a reversal of preference. He goes on to somewhat favorably discuss minimax regret, to which his particular example does not apply. But this raises some questions: Is minimax regret really immune to the fundamental criticism expressed in the example, or does it just happen to perform well in one specific example? In other words, could one tailor a similar example that is unfavorable to minimax regret versus minimax loss? If no, are there fine-tunings of minimax loss that avoid the problem but do not go all the way to minimax regret and, for example, retain independence of irrelevant alternatives?

The above result resolves these questions because Berger's criticism essentially asserts von Neumann-Morgenstern independence. Formally, it can be shown that a criterion respects independence iff it is

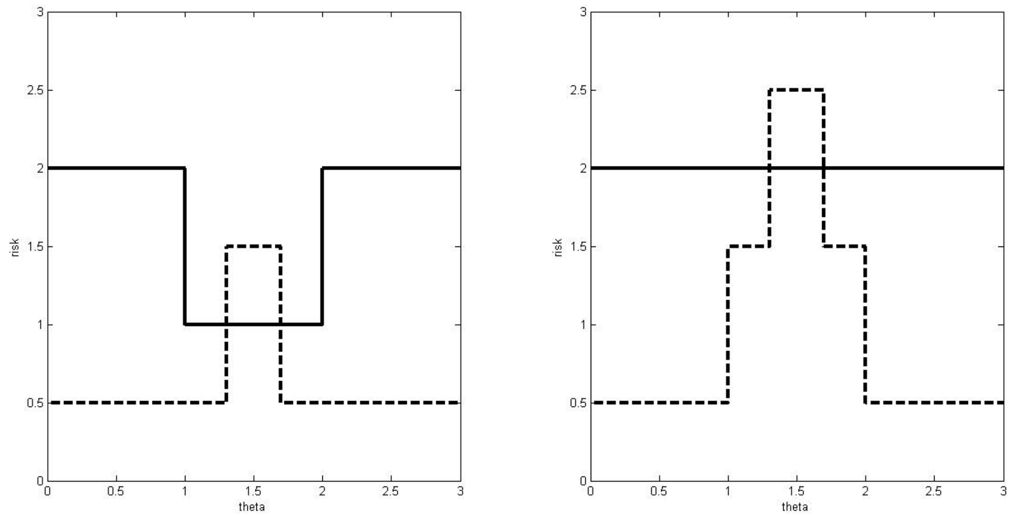

Figure 1: An example due to J.O. Berger (1985). The risk functions in the right-hand panel differ from their counterparts in the left-hand panel by the addition of $1\{\theta \in [1, 2]\}$. The solid risk function attains minimax loss in the right-hand panel but not the left-hand one. It attains minimax regret in both.

invariant under the addition of risk functions. Minimax regret possesses this robustness, and there is no simple modification of minimax loss that does so while maintaining some other of its core features, e.g., menu independence, avoidance of priors, and completeness. Indeed, one cannot combine all of the features just mentioned.

3.3.2 Revealed Preference and (Limits of) Axiomatizations

Axiomatic analysis depends on considering not only the specific decision problem at hand, but a multitude of potential decision problems. The premise is that some form of consistency of behavior across these problems is desirable. This premise is subject to both pragmatic and methodological criticism.

On the pragmatic side, axiomatizations are historically connected to revealed preference theory. Recall the origins of this literature: Influenced by the philosophy of science of their time, economists set out to purge from the foundations of their theorizing objects that are not directly observable. Hence, they did not assume preferences as primitive, but derived their existence (as "as-if" representations) from assumptions about choice behavior.[12] The criticism that preferences might not be observable

[12] To illustrate, Little (1949, p. 97) argued that "the new formulation is scientifically more respectable" because "if

is especially salient for regret-based preferences because these are menu-dependent. Indeed, if r_3 is chosen from $\{r_1, r_2, r_3\}$, then the statement "$r_1 \succ_{\{r_1,r_2,r_3\}} r_2$" has no choice theoretic interpretation. Some theorists consider this a decisive objection to axiomatization of regret preferences as opposed to the corresponding choice functions.

I nonetheless presented results in terms of preferences, for a number of reasons. With menu-independent preferences, an easy dualism between preference and choice from binary menus (i.e., $r_1 \succsim r_2$ iff r_1 is chosen from $\{r_1, r_2\}$) seemingly resolves the issue right away (Arrow (1959)). With regard to minimax regret, a less obvious such dualism has been identified (Stoye (2011b)). In the latter case, it turns out that the required modifications of axioms are subtle but do not affect the grand message contained in the above result, so the use of preferences is at least justifiable as a heuristic device. I would like to push the argument further, however. If one were truly consequential about pursuing a revealed preference approach, there would be many more problems to consider.[13] To begin, results hinge on domain assumptions, e.g. about the richness of the set of possible risk functions, that are unlikely to hold. For example, essentially every measurable mapping from Θ to $[-1, 1]$ must be a conceivable risk function.[14] This is not expected to hold if one restricts attention to risk functions that arise in actual statistical decision problems. In addition, Arrow's (1959) dualism might not apply so easily because statisticians can and do randomize, meaning that choice from binary menus is not necessarily observable. In particular, the statement "$r_1 \succ_{\{r_1,r_2\}} r_2$" becomes vulnerable to the criticism levelled against "$r_1 \succ_{\{r_1,r_2,r_3\}} r_2$" if r_3 is a proper mixture of r_1 and r_2 and, therefore, available to the statistician whenever r_1 and r_2 are.[15]

In sum, not only are relatively few choices ever observable in practice, but many do not even correspond to observable statistical decision problems in principle. I therefore suggest that to apply axiomatic arguments to statistical decisions, one should avoid a rigid notion of revealed preference. From this point of view, a statement like $r_1 \succ_{\{r_1,r_2,r_3\}} r_2$ is useful as long as one finds comprehensible the notion that "in choice problem $\{r_1, r_2, r_3\}$, r_1 would appear more attractive than r_2."

A more fundamental critique was recently voiced by Manski (2011b). He proposes a notion of "actualist rationality," according to which a decision maker should not consider relevant choice sce-

an individual's behaviour is consistent, then it must be possible to explain that behaviour without reference to anything other than behaviour." Savage (1954, p. 17) held similar views: "I think it of great importance that preference, and indifference, between [acts] be determined, at least in principle, by decisions between acts and not by response to introspective questions."

[13] In this author's view, most of these affect revealed preference theory more generally, but this survey is not the place to argue the point.

[14] For technical discussions of this point, see Fishburn (1970, for Bayesian models) and Puppe & Schlag (2008, for minimax models). The interval can be changed to be $[0, 1]$, but then it is necessary to slightly strengthen independence of never strictly optimal alternatives to the precise axiom used in Stoye (2011a).

[15] See Stoye (2011b, 2011d) for more on the ramifications of this observation for revealed preference theory.

narios other than the one she is actually facing. While some of the above axioms have implications for choice from any given menu (e.g., ambiguity aversion implies that the choice set must be closed under probabilistic mixture), it is hard to see how this criticism would leave scope for axiomatic analyses to inform statistical decision making.

3.3.3 Dynamic Consistency

All decision problems considered in this survey are static in the sense that planners commit to statistical decision rules before seeing any signals. In contrast, much of economics is primarily concerned with dynamic problems, where relevant information becomes available between (and possibly caused by) agents' decisions. These settings raise the question of dynamic consistency. A decision criterion is *dynamically consistent* if a decision maker with full commitment and one who can re-optimize as she moves down a decision tree exhibit the same behavior. Bayesians are dynamically consistent – δ is Bayes against π iff it coincides with the behavior of a Bayesian agent with prior π who takes an optimal decision only after seeing the data (and accordingly updating π). In general, other agents are not dynamically consistent and therefore are subject to "multiple selves" issues.

This complication is peripheral to this survey for two reasons: The problems considered are static anyway, and evaluation of decision rules according to risk corresponds to "full commitment" scenarios, hence it would be clear how to resolve "multiple selves" issues. But some readers might find the potential for such issues an unsettling aspect of minimax criteria. Also, the question arises naturally in extensions of this paper's examples, e.g. Manski's (2009) multi-period version of example 2.[16]

With the non-Bayesian decision criteria presented here, dynamic settings raise the question of how to update sets of priors. The decision theoretic literature has explored numerous updating rules, partly in a quest to recover dynamic consistency.[17] Among these, pointwise updating of all priors is probably the most immediately appealing, and it is also taken for granted in robust Bayesian inference. Γ-minimax with pointwise updating is dynamically consistent under very specific conditions on Γ that were elucidated by Epstein & Schneider (2001). Interestingly, these conditions will obtain in the next section's examples, hence Γ-minimax loss does not actually encounter the problem in this paper. However, this fact is specific to this paper's analysis and does not translate to Γ-minimax loss more generally, nor to (prior-free) minimax loss. It is also not expected to generally hold for (Γ-)minimax regret, which (with pointwise updating) is dynamically consistent only under very restrictive conditions identified by Hayashi (2011).

[16] Manski (2009) proposes a "dynamic adaptive minimax regret" decision rule that can be intuitited as period-by-period myopic minimax regret. He does not claim that it solves the (very complicated) multi-period minimax regret problem.

[17] See Gilboa & Schmeidler (1993) and Hanany & Klibanoff (2009) for two of many possibilities not further investigated here.

4 Examples in Estimation and Treatment Choice

There is a large literature on minimax estimation, including scores of textbook exercises (e.g., Ferguson (1967), Berger (1985)). Nominally, most of this literature is on minimax loss as opposed to minimax regret. However, this might be due largely to the fact that most of it also is on estimation problems, and minimax loss and minimax regret agree in many such problems. Stark differences between the criteria will become apparent for treatment choice problems, which are somewhat specific to econometrics and where minimax regret recently became prominent.

Attention will be focused on a sequence of stylized decision problems. I solve examples 1 and 2 and also consider extensions, notably by introducing partial identification through missing data. The problems will be similar in that identified quantities can be expressed in terms of probabilities, and sample frequencies will be available as signals of these probabilities. This reflects numerous simplifications. First, all outcomes – or at least their utilities – are bounded. The further restriction to $[0, 1]$ is a normalization. Second, outcomes will be taken to be binary, that is, $Y \in \{0, 1\}$. Given boundedness, this restriction loses less generality than one might at first think because for many purposes, it can be justified by appealing to a "binary randomization" technique due to Schlag (2007).[18] To apply this technique, generate treatment rules for non-binary outcomes by first replacing every observed outcome with one realization of a binary, mean-preserving spread of itself and then operating rules defined for binary outcomes. This deliberate discarding of information might be counterintuitive and raises questions about admissibility, with inadmissibility being demonstrable in one example. However, it leaves intact minimax values with respect to both loss and regret. At the very least, it therefore reveals nonparametric minimax efficiency bounds, as well as decision rules that attain them.

A more subtle, but important simplification of the treatment choice problem is that outcomes experienced in the sample do not directly enter the planner's welfare. Without this, the problem would be a *bandit problem* (e.g., Bergemann (2008)) with the attendant tension between exploration and exploitation motives, i.e. the conflicting aims of treating current treatment recipients as well as possible and generating informational externalities down the road. Minimax analysis of such problems is extremely hard; see Schlag (2003) for some results using minimax regret.

Γ-minimax criteria require specification of a set of priors. To provide a tractable example, I will take inspiration from recent work by Kitagawa (2011) and use single but noninformative priors for identified parameters while imposing no prior information at all for unidentified parameters. Regarding the choice of noninformative prior for identified quantities, Bayesian inference for binomial proportions has been much analyzed (e.g., Geisser (1984)). While different principles of generating noninformative priors

[18] More precisely, the technique was first applied to statistical treatment choice by Schlag (2007); it had been independently discovered for related problems by Cucconi (1968), Gupta and Hande (1992), and Schlag (2003).

lead to different priors and hence posteriors, the way in which they do so is well understood, and I hope that the issue can be relegated to the background. Having said that, I will use the (improper) beta prior with parameters $(0,0)$, hence posterior expectations will coincide with sample averages. Of course, this implementation of Γ-minimax leads to interesting departures from conventional Bayesian analysis only in cases of partial identification below. An intriguing feature of it is that in those cases, it recovers dynamic consistency.

4.1 How to find minimax rules?

Finding minimax rules is notoriously difficult, and algorithms that allow for automated discovery exist only for very specific subclasses of problems. An important technique to discover minimax rules or to establish minimaxity is game theoretic analysis (Wald (1945)). Consider a fictitious zero-sum game in which the decision maker picks a decision rule $\delta \in \mathcal{D}$ (possibly at random; note, however, that $\mathcal{D}$ is closed under probabilistic mixture) and Nature picks θ (possibly at random, meaning that her strategies can be identified with priors π). Nature's payoff is $r(\delta, \theta)$ in the "minimax loss" game and $r(\delta, \theta) - \min_{d \in \mathcal{D}} r(d, \theta)$ in the "minimax regret" game. Any Nash equilibrium (δ^*, π^*) of the game characterizes a minimax treatment rule δ^*, and Nature's equilibrium strategy π^* can be interpreted as the *least favorable prior* that the rule implicitly selects.[19] The other important technique stems from a corollary of this approach: Risk equalizing rules that are Bayes with respect to some prior are minimax. This claim can be augmented: Admissible risk equalizers are minimax as well.[20]

4.2 Estimation

We begin by solving example 1.

Example 1: Solution.

An estimator δ is admissible iff it can be written as

$$\delta = \begin{cases} 0, & n\overline{y} \leq r, \\ \frac{\int_0^1 \theta^{n\overline{y}-r}(1-\theta)^{s-n\overline{y}-1} d\pi(\theta)}{\int_0^1 \theta^{n\overline{y}-r-1}(1-\theta)^{s-n\overline{y}-1} d\pi(\theta)}, & r < n\overline{y} < s, \\ 1, & n\overline{y} \geq s, \end{cases}$$

[19] The following simplification is particularly helpful in finding minimax regret rules:

$$\int r(\delta, \theta) - \min_{d \in \mathcal{D}} r(d, \theta) d\pi = \int r(\delta, \theta) d\pi - \int \min_{d \in \mathcal{D}} r(d, \theta) d\pi,$$

where the last integral is constant in δ. Thus, given a prior, regret minimization equals risk minimization (and utility maximization), meaning that verification of the decision maker's best-response condition is the same as in the "minimax loss" game.

[20] See Berger (1985) or Ferguson (1967) for proofs and further extensions of these claims.

where $\overline{y}$ is a sample average, $-1 \leq r < s \leq n+1$, and π is a probability measure on $[0,1]$ s.t. $\pi(\{0\} \cup \{1\}) < 1$ (Johnson (1971)).

The minimax loss (as well as regret) estimator is

$$\delta_1^{MU} = \delta_1^{MR} = \frac{n\overline{y} + \sqrt{n}/2}{n + \sqrt{n}}.$$

To prove this (Hodges & Lehmann (1950)), one can verify that δ_1^{MU} is a risk-equalizing Bayes rule supported by the beta-prior with parameters $(\sqrt{n}/2, \sqrt{n}/2)$. As μ is identified, the class of priors used for Γ-criteria is a singleton, and $\delta_1^{\Gamma-MU} = \delta_1^{\Gamma-MR} = \overline{y}$. (While not completely obvious from inspection of Johnson's result, these estimators are also admissible.)

The minimax estimator implies strong shrinkage of the sample average toward 1/2 and is frequently presented as illustrating deficiencies of minimax (e.g., Strawderman (2000)). To see why $\overline{y}$ is not a minimax estimator, note that it is unbiased for μ_1, hence its risk is its variance and therefore maximized by setting $\mu_1 = 1/2$. By shrinking $\overline{y}$ toward 1/2, the planner adjusts this risk in a manner that is the more favorable, the closer the true μ_1 is to 1/2. The shrinkage hurts in most of parameter space, but helps around $\mu_1 = 1/2$ and therefore improves the estimator's worst-case performance.

A planner using δ_1^{MU} behaves as if she had prior knowledge to the effect that μ_1 is very likely to be close to 1/2. Once it is made explicit, such a prior may appear unappealing, and depending on the sample observed, it may also be implausible with hindsight. On the other hand, one should keep in mind that minimax loss and regret are not motivated by a quest to find priors, and their defenders might argue that judging the epistemic appeal of the prior they implicitly optimize against makes as much sense as criticizing a particular prior for failing to minimize worst-case regret. I will avoid discussions of this kind henceforth.[21]

We will also solve example 2 but first consider a variation of example 1 that introduces the theme of partial identification.

Example 3: Binomial proportion with missing data.

Suppose that the random sample is taken from a sampling population that is a subset (with known relative mass p) of the population of interest. This induces partial identification through missing data. Letting $Z \in \{0,1\}$ indicate membership of the sampling population, one can write

$$\mu_1 = E(Y_1) = pE(Y_1|Z=1) + (1-p)E(Y_1|Z=0),$$

[21] If Y could have any distribution on $[0,1]$, then binary randomization in conjunction with δ_1^{MU} would achieve both minimax loss and minimax regret. However, direct application of δ_1^{MU} to the sample mean results in an estimator that has the same expectation and a variance that is weakly lower for all distributions of Y and strictly lower for some, meaning it dominates in terms of square loss. Thus, binary randomization is demonstrably inadmissible in this example. I am not aware of a similar demonstration for later examples.

which yields best possible bounds on μ_1 as follows:

$$\mu_1 \in [pE(Y_1|Z=1), pE(Y_1|Z=1) + (1-p)].$$

This same setting will be used to model missing data throughout. Assuming that p is known is optimistic, but risk is easily seen to decrease in p, so it is unclear how to interestingly embed estimation of p in a minimax approach.[22] The parameter space now is $\Theta = [0,1]^2$ with typical element $\theta = (\mu_{10}, \mu_{11}) = (E(Y_1|Z=0), E(Y_1|Z=1))$. It is easy to see that μ_{11} is identified, whereas the data are completely uninformative about μ_{10}. For a Bayesian, there is no fundamental difficulty here: She will place a prior on both μ_{11} and μ_{10}, and the latter will simply not be updated from the data. To save space, I will not work out an example.[23] However, I will demonstrate the use of Γ-ML and Γ-MR, necessitating a definition of Γ. As mentioned before, I will use the $\beta(0,0)$ prior for μ_{11} but place no restrictions whatsoever on the collection of conditional distributions of $(\mu_{10}|\mu_{11})$ on $[0,1]$.[24]

Example 3: Solution.

The maximin utility and minimax regret estimators still agree, but are hard to evaluate. On the other hand,

$$\delta_2^{\Gamma-MU} = \delta_2^{\Gamma-MR} = p\overline{y} + (1-p)/2,$$

that is, both $\Gamma - MU$ and $\Gamma - MR$ effectively impute $\mu_{10} = 1/2$. To see this, one can use unbiasedness of sample averages to verify that

$$r(\delta_2^{\Gamma-MU}, \theta) = p^2 \int (\overline{y} - \mu_{11})^2 dP(Y|Z=1) + (1-p)^2 \int (1/2 - \mu_{10})^2 dP(Y|Z=0).$$

As $\int (\overline{y} - \mu_{11})^2 dP(Y|Z=1)$ is determined through the prior on μ_{11}, only $\int (1/2 - \mu_{10})^2 dP(Y|Z=0)$ varies over Γ. To maximize it, one may restrict attention to $\mu_{10} \in \{0,1\}$, corresponding to degenerate distributions of $(Y|Z=0)$. Within the accordingly restricted parameter space, $\delta^2_{\Gamma-MU}$ is a risk equalizer. The argument is concluded by observing that $\delta_2^{\Gamma-MU}$ is Bayes against a prior which renders μ_{10} equally likely to be 0 or 1 for any μ_{11}.

The example illustrates that discovery of global minimax rules can become difficult even for problems that are simple to state. I suspect that example 3 can be solved for these rules, but the contribution would go beyond this paper's scope.

---

[22] In a local asymptotics context, Song (2010) conditions on different values of p and finds statements that are true for all of them. This approach might be interesting to import.

[23] See Moon and Schorfheide (2010), Poirier (1998), and Yuan & Jian (2010) for more on Bayesian inference without identification.

[24] I avoid mentioning the joint distribution of (μ_{10}, μ_{11}) to avoid subtleties due to the improper prior on μ_{11}. For the same reason, theoretical results in Kitagawa (2011) do not apply here, but this is not a problem for implementation of the method.

4.3 Treatment Choice

We will now move from estimation to treatment choice. Example 2 resembles the problem of deciding whether to adopt a new medical treatment. In the real world, established practice is to conduct hypothesis tests. For example, the FDA circulates precise guidelines regarding the number and size of rejections of $H_0 : \mu_0 \geq \mu_1$ required for approval of a new treatment. Decision rules based on hypothesis tests can be reconstructed as attaining minimax loss as well as minimax regret. However, the loss functions needed for this reconstruction very heavily weight type II errors over type I errors (see in particular Tetenov (2011)). It is, therefore, expected that this paper's optimality criteria will produce rather different rules. I will also report the (trivial) results corresponding to Bayesian inference with a $\beta(0,0)$ prior. Readers interested in deeper insight about Bayesian treatment choice should consult Chamberlain's (2011) theoretical analysis and Dehejia's (2005) execution of a Bayesian treatment choice analysis on a famous real-world data set.

Example 2: Solution.

A decision rule is admissible iff it can be written as

$$\delta = \begin{cases} 0, & n\overline{y} < n^* \\ \lambda^*, & n\overline{y} = n^* \\ 1, & n\overline{y} > n^* \end{cases}, \tag{1}$$

where $n^* \in \{0, 1, ..., n\}$ and $\lambda^* \in [0, 1]$ (Karlin & Rubin (1954)). The minimax loss decision rule is $\delta_2^{ML} = 0$, that is, the status quo treatment is assigned for any sample outcome. This rule is uniquely minimax unless $\mu_0 = 0$, in which case any rule is minimax. Γ-ML and Γ-MR again coincide and best respond to the $\beta(0,0)$ prior on μ_1, thus they amount to the *empirical success rule* (Manski (2004))

$$\delta_2^{\Gamma-ML} = \delta_2^{\Gamma-MR} = 1\{\overline{y} > \mu_0\}.$$

This rule is unique except if $\overline{y} = \mu_0$, in which case any treatment assignment is optimal. Finally, δ_2^{MR} is as in (1), where n^* and λ^* are characterized by

$$\begin{aligned} & \max_{a\in[0,\mu_0]} (\mu_0 - a) \left[\sum_{m>n^*} \binom{n}{m} a^m (1-a)^{n-m} + \lambda^* \binom{n}{n^*} a^{n^*} (1-a)^{n-n^*} \right] \qquad (2) \\ = \; & \max_{b\in[\mu_0,1]} (b - \mu_0) \left[\sum_{m<n^*} \binom{n}{m} b^m (1-b)^{n-m} + (1-\lambda^*) \binom{n}{n^*} b^{n^*} (1-b)^{n-n^*} \right]. \end{aligned}$$

The last result is from Stoye (2009; see Manski and Tetenov (2007) and Tetenov (2011) for generalizations). A game theoretic intuition goes as follows: A least favorable prior must have the feature that both $\mu_1 > \mu_0$ and $\mu_0 > \mu_1$ are possible a priori; else, the planner could achieve zero regret

	$\mathbf{n=1}$	$\mathbf{n=2}$	$\mathbf{n=3}$	$\mathbf{n=4}$	$\mathbf{n=5}$	$\mathbf{n=10}$	$\mathbf{n=20}$	$\mathbf{n=50}$	$\mathbf{n=100}$	$\mathbf{n=500}$
$\mu_0 =$ **.05**	0.33	0.48	0.59	0.44	0.74	1.22	1.18	2.68	5.18	25.18
$\mu_0 =$ **.25**	0.64	0.82	1.07	1.33	1.58	2.82	5.32	12.82	25.32	125.32
$\mu_0 =$ **.50**	1.0	1.5	2.0	2.5	3.0	5.5	10.5	25.5	50.5	250.5
$\mu_0 =$ **.75**	1.36	2.18	2.91	3.67	4.42	8.18	15.68	38.18	75.68	375.68
$\mu_0 =$ **.95**	1.67	2.52	3.41	4.56	5.26	9.78	19.82	48.32	95.82	475.82

Table 2: Testing an innovation: The minimax regret decision rule.

	$\mathbf{n=1}$	$\mathbf{n=2}$	$\mathbf{n=3}$	$\mathbf{n=4}$	$\mathbf{n=5}$	$\mathbf{n=10}$	$\mathbf{n=20}$	$\mathbf{n=50}$	$\mathbf{n=100}$	$\mathbf{n=500}$
$\mu_0 =$ **.05**	1.0	1.5	1.68	1.79	1.87	2.48	3.43	5.81	9.38	33.78
$\mu_0 =$ **.25**	1.8	2.2	2.76	3.02	3.61	5.48	8.85	18.19	32.78	141.58
$\mu_0 =$ **.50**	1.9	2.8	3.6	4.2	4.88	8.11	14.21	31.35	58.74	268.89
$\mu_0 =$ **.75**	1.93	2.91	3.88	4.84	5.79	10.11	18.62	42.90	82.51	391.29
$\mu_0 =$ **.95**	1.95	2.94	3.94	4.94	5.94	10.92	20.86	50.35	98.84	483.27

Table 3: Testing an innovation: The classical decision rule (5 percent significance, one-tailed test).

through a no-data rule. Hence, this prior must be supported on some parameter value a as well as some parameter value b, where $b > \mu_0 > a$. This, in turn, requires that both a and b are equally regret maximizing given δ_2^{MR}. These conditions are simultaneously captured in (2).

Table 2 illustrates δ_2^{MR} for a selection of sample sizes n and values of μ_0. Cell entries display $\alpha \equiv n^* + 1 - \lambda^*$, a smooth index of the treatment rule's conservatism: The number to the left of the decimal point is the critical number of observed successes that leads to randomized treatment assignment, and the number to its right gives the probability with which this randomization will pick treatment 0.

It turns out that δ_2^{MR} closely resembles the empirical success rule for rather small samples, rendering it similar to Bayesian decision rules using noninformative priors, but very different from hypothesis tests. To illustrate, table 3 is analogous to table 2, but describes a decision maker who chooses treatment 1 if the data reject $H_0 : \mu_1 \leq \mu_0$ at 5% significance, with randomization on the threshold to maximize the test's power. This decision rule is much more conservative because it emphasizes avoidance of type I errors over avoidance of type II errors. In contrast, δ_2^{MR} treats type I and type II errors roughly symmetrically. To be sure, while the classical decision rule will incur high worst-case regret, the minimax regret rule will fail on classical terms, with the size of the implicit hypothesis test rapidly approaching 50% as n grows. Which consideration matters depends on one's objective function; the point here is to demonstrate that the choice matters.

In stark contrast to all of these solutions, minimax loss is achieved by a "no-data rule" which entirely disregards the sample. This is because the least favorable prior is entirely concentrated on $\mu_1 = 0$. Minimax loss incurs such problems throughout treatment choice examples, including all such examples in this paper. This is a major reason why econometricians recently considered minimax regret.

Example 4a: Treatment Choice when Neither Treatment is Well Understood.

Assume now that the effect of neither treatment is well understood, thus both μ_0 and μ_1 could be anywhere in $[0,1]$. This is captured by extending the parameter space to $\Theta = [0,1]^2$ with typical element $\theta = (\mu_0, \mu_1)$. It raises the question of sample design: To which treatment should sample subjects be assigned? The most obvious sampling scheme is probably by matched pairs, that is, $n/2$ sample subjects are assigned to each treatment (assuming that n is even). This assignment scheme will be analyzed first. The sample space then is $\Omega = [\{0,1\} \times \{0,1\}]^n$ with typical element $\omega = (t_i, y_i)_{i=1}^n$. Here t_i is treatment assigned to sample subject i and, for concreteness' sake, will be taken to equal $1\{i \text{ is even}\}$; y_i is the outcome experienced by that same subject and is distributed independently Bernoulli with parameter μ_{t_i}. For the prior on (μ_0, μ_1), the two quantities are independent $\beta(0,0)$.

Example 4a: Solution

I am not aware of a characterization of admissible treatment rules. Minimax regret is achieved by

$$\delta_{4a}^{MR} = \begin{cases} 0, & \overline{y}_1 < \overline{y}_0 \\ 1/2, & \overline{y}_1 = \overline{y}_0 \\ 1, & \overline{y}_1 > \overline{y}_0 \end{cases},$$

where $\overline{y}_t$ is the sample average of Y_t (Canner (1970); see also Schlag (2007), Stoye (2009)). The game theoretic argument goes as follows. If the planner uses δ_{4a}^{MR}, then parameter values which can be generated from each other by interchanging treatment labels achieve the same regret. Hence, if $(\mu_0, \mu_1) = (a, b)$ is a best response to δ_1^{MR}, then the prior π_{4a}^{MR} that randomizes evenly over $\{(a,b),(b,a)\}$ is one as well. But it is easy to see that δ_{4a}^{MR} is Bayes against π_{4a}^{MR}.

The same decision rule is also optimal under both the Γ-ML and Γ-MR criterion, though for different reasons: Both continue to respond to a singleton (improper) prior, which does not coincide with π_{4a}^{MR} but induces the same Bayes act. Minimax loss again encounters a pathology because *any* treatment rule achieves minimax loss. The least favorable prior supporting this is concentrated on $(\mu_0, \mu_1) = (0, 0)$.

Example 4b: Same Example with a Different Sample Design

Assume now that within-sample treatment assignment is by independent tosses of a fair coin. This is also reasonably close to some real-world procedures. At the same time, it generates different results for minimax regret and allows for certain extensions. A typical sample is $\omega = (t_i, y_i)_{i=1}^n$ as before, but t_i is now distributed as independent toss of a fair coin.

Example 4b: Solution

The minimax regret treatment rule is

$$\delta_{4b}^{MR} = \begin{cases} 0, & I_n < 0 \\ 1/2, & I_n = 0 \\ 1, & I_n > 0 \end{cases},$$

where

$$\begin{aligned} I_n &\equiv \sum_{i=1}^n (2t_i - 1)(y_i - 1/2) = n_1(\overline{y}_1 - 1/2) - n_0(\overline{y}_0 - 1/2) \\ &\propto [\#\,(\text{observed successes of treatment 1}) + \#\,(\text{observed failures of treatment 0})] \\ &\quad - [\#\,(\text{observed successes of treatment 0}) + \#\,(\text{observed failures of treatment 1})] \end{aligned}$$

with n_t the number of sample subjects assigned to treatment t and the convention that $n_t(\overline{y}_t - 1/2) = 0$ if $n_t = 0$ (Stoye (2009)). All other decision rules are as in example 4a.

To gain some intuition for δ_{4b}^{MR}, note that $E(I_n/n) = \mu_1 - \mu_0$, i.e. I_n/n is an unbiased estimator of the average treatment effect $(\mu_1 - \mu_0)$, and its sign accordingly has a reasonable interpretation. Also, the contrast to δ_{4a}^{MR} is not as strong as might initially appear: If coin tosses come out such that half of sample subjects are assigned into each treatment, then $\delta_{4b}^{MR} = \delta_{4a}^{MR}$. However, minimax regret now disagrees with Γ-ML and Γ-MR because $\delta_{4b}^{\Gamma-MR} = \delta_{4b}^{\Gamma-ML} = \delta_{4a}^{MR}$ for all ω.

Finally, treatment choice can be taken to missing data problems, but minimax regret analysis becomes much more involved. I here present a special case of a finding in Stoye (2011c).

Example 5: Example 4b but with Missing Data

Extend example 4b to feature missing data in the same manner used in example 3. Thus, the sample population is a (mass p) subset of the treatment population. The experiment continues to be otherwise ideal. A real-world analogy are laboratory experiments performed on volunteers: These are internally valid, but self-selection into the volunteer population may affect their external validity.

In this example, there are four expectations of interest, namely $(\mu_{00}, \mu_{01}, \mu_{10}, \mu_{11})$, where $\mu_{tz} = E(Y_t|Z = z)$ and $Z \in \{0, 1\}$ is as before. The parameter space is $\Theta = [0, 1]^4$ with typical element $\theta = (\mu_{00}, \mu_{01}, \mu_{10}, \mu_{11})$. The sampling scheme, and hence the sample space, are as in example 4b, though one should keep in mind that sample outcomes are only informative about (μ_{01}, μ_{11}). No direct learning about (μ_{00}, μ_{10}) occurs from the data.

Example 5: Solution

First, $\delta_5^{\Gamma-ML} = \delta_{4a}^{\Gamma-ML} = 1\{\overline{y}_1 > \overline{y}_0\}$ because for any (μ_{01}, μ_{11}), loss is maximized by setting $\mu_{00} = \mu_{10} = 0$. The ML decision rule continues to be trivial. The Γ-MR decision rule is hard to find, and the question is left open here. Finally, let β_m denote δ_{4b}^{MR} but applied on the first m data points only. Let $n' = \max\{m \leq n : m \text{ odd}\}$ and $\gamma_n = 2^{-n'} \sum_{t>n'/2} \binom{n'}{t}(2t - n')$, noting that $\gamma_n \geq 1/2$, with equality if $n = 1$. Then

$$\delta_5^{MR} = \begin{cases} \beta_n & \text{if } p \geq \frac{1}{1+2\gamma_n}, \\ \alpha_{n^*}^* \beta + (1-\alpha^*)\beta_{n^*-2} & \text{if } \frac{1}{1+2\gamma_n} > p > \frac{1}{2}, \\ \frac{p}{1-p} \cdot \beta_1 + \left(1 - \frac{p}{1-p}\right) \cdot \frac{1}{2} & \text{if } p \leq \frac{1}{2}, \end{cases} \tag{3}$$

where

$$\begin{aligned} n^* &= \min\left\{m : \gamma_m > \frac{p}{2-2p}\right\}, \\ \alpha^* &= \frac{p/(2-2p) - \gamma_{n^*-2}}{\gamma_{n^*} - \gamma_{n^*-2}}. \end{aligned}$$

The minimax regret treatment rule is somewhat involved but has some interesting features. Perhaps the most important insight is that for p close enough to 1, $\delta_5^{MR} = \delta_{4b}^{MR}$, thus small but positive amounts of missing data are completely ignored.[25] Here, "small but positive" is quantified by $f(n) = (1 + 2\gamma_n)^{-1}$. This expression can be read in two ways: As a function of n, it gives the largest proportion of missing data that, at sample size n, will be completely ignored by minimax regret treatment rules. Note that $f(n) = O(n^{-1/2})$, i.e. this proportion vanishes with sample size. The inverse function f^{-1} is a function of p that, for any given proportion of missing data, returns a maximum sample size up to which minimax regret will ignore these missing data. This threshold sample size is greater than 1 iff $p > 1/2$. This interpretation of f^{-1} also yields an intuition for what happens if there are too many missing data: δ_5^{MR} then discards sample points until the effective sample size is just $f^{-1}(p)$, with randomization to accommodate integer problems.

This deliberate discarding of data may appear puzzling at first but has a clear intuition. Depending on parameter values, the fictitious game has exactly one of two kinds of equilibria. If the amount of missing data is small relative to sample size, Nature's best response to δ_{4b}^{MR} much resembles the least favorable prior from example 4*b*. Expected outcomes in the missing data are positively correlated with expected outcomes in the sample population, increasing regret relative to example 4b but not affecting the decision rule. But as missing data become prevalent, Nature might eventually want to choose parameter values s.t. $(\mu_1 - \mu_0)$ and $(\mu_{11} - \mu_{01})$ have opposing signs, giving the signal ω a misleading tendency. The problem is obvious in an extreme case: If p is close to 0 yet n is very large,

[25] This feature qualitatively extends to the analogous extension of example 4a, but the cutoff will be different.

n	1	3	5	10	20	50	100	200	500	1000
$\mathbf{1-f(n)}$	0.5	0.4	0.348	0.289	0.221	0.151	0.112	0.082	0.053	0.038

Table 4: Numerical illustration: Largest amount of missing data that will be ignored my minimax regret decision rules in example 5.

then $\max_{\theta\in\Theta} R(\delta_{4b}^{MR},\theta) \approx 1$, this value being attained if $\mu_{11}-\mu_{01}=1$ yet $\mu_1-\mu_0\approx -1$. This is far from an equilibrium outcome – the planner can always enforce an expected regret of $1/2$ by just tossing a coin. The minimax regret decision rule avoids the trap because by discarding data until the effective sample size leads to the previous equilibrium.

The threshold $p=f(n)$ also is a switching point with regard to the decision problem's minimax regret value. In the equilibrium that involves discarding data, the least favorable prior implies $\mu_{01}=\mu_{11}$, thus sample data are noise. This firstly clarifies why the discarding of data can happen in equilibrium. Also, it means that adding data points beyond $n=f^{-1}(p)$ does not affect the problem's minimax regret value. This value therefore displays unusual asymptotics: It does not converge to zero and precisely attains its limiting value (of $(1-p)/2$) beyond some finite sample size n.[26] Another consequence is that δ_5^{MR} is unique (up to subtleties about tie-breaking) only as long as it coincides with δ_{4b}^{MR}. Once the least favorable prior renders the data uninformative, there are many minimax regret treatment rules. For example, one could use any appropriately sized subset of sample points rather than the first ones, or also reduce the variance of δ by computing $\delta(\omega)$ for every possible, appropriately sized subset of the sample and then taking the average.[27]

Table 4 displays f for selected sample sizes. For example, if $n=10$, δ_5^{MR} ignores up to 29% missing data; the corresponding number for $n=100$ still exceeds 10%. Thus, under the minimax regret criterion, it is finite sample optimal to ignore rather substantial amounts of missing data.

5 Conclusion

Statistical decision theory has recently seen a resurgence in econometrics. This survey attempted to draw attention to relevant issues from the perspective of a microeconometrician who attempts to

[26] This implication was emphasized by Tetenov (2009), who discovered a similar decision rule in a related scenario.

[27] Schlag (2007) proposes and further explains such variance reduction tools.

A case of special interest occurs whenever $p>1/2$, thus $\delta_5^{MR}=\frac{a}{b}\cdot\delta_1+\left(1-\frac{a}{b}\right)\cdot\frac{1}{2}$. Here $E(\delta_5^{MR}(\omega))$ is linear in $(\mu_1-\mu_0)$, thus any unbiased estimator of $(\mu_1-\mu_0)$ can substitute for δ_1. Using $\frac{1+I_n/n}{2}$, the decision rule becomes an unbiased estimator of a rule discovered by Manski (2007a) for the case of known (μ_{01},μ_{11}). Furthermore, Manski (2007a) observed that under conditions very similar to $p>1/2$ (in particular, the identity of the better treatment is known to be unlearnable), this specific rule is finite sample minimax regret optimal. A close analog to Manski's finding is, therefore, embedded in results reported here.

(re-)integrate the languages of decision theory and econometrics. Decision making under ambiguity was identified as common theme connecting statistical and axiomatic decision theory. In particular, minimax decision rules correspond to decision models that were thoroughly analyzed in the decision theoretic literature on ambiguity aversion. I tried to make these connections explicit, for example by translating decision theoretic results into statistical notation and suppressing many subtleties that have attracted legitimate attention in economic decision theory. I also investigated minimax decision rules in numerous estimation and treatment choice problems, including some classic ones but perhaps most importantly, examples that feature partial identification and thereby ambiguity that does not vanish as samples grow.

The themes discussed here connect to other, important literatures in many ways, and this survey does not claim to be encyclopedic. Within econometrics proper, a notable limitation was the focus on globally minimax approaches. There is also a rich, recent literature on local asymptotic theory for estimation and treatment choice problems with point and partial identification. To give some examples that are directly related to this survey, Hirano & Porter (2009) establish that in the treatment choice examples without missing data, the empirical success rules $\delta = 1\{\overline{y}_1 > \mu_0\}$ (for example 3) and $\delta = 1\{\overline{y}_1 > \overline{y}_0\}$ (for examples 4a and 4b) are locally asymptotically minimax regret optimal. For the specific setting of example 2, i.e. estimating a mean (binomial or not) with missing data, Song (2010) shows that the Γ-ML rule $\delta_2^{\Gamma-ML} = p\overline{y}_1 + (1-p)/2$ is locally asymptotically minimax regret optimal as well.

Beyond econometrics, many of the themes of partial identification connect to issues explored in the interval probabilities community. Levi's (1980) and Walley's (1990) approaches to decision making are highly complementary to what was discussed here, and I could easily extend this paper's list of references by separately citing a dozen contributions in Augustin et al. (eds., 2009) and earlier proceedings of the same conference. In a similar vein, a complete discussion of axiomatics and statistical decision making should mention a long line of work by (in alphabetical order) Kadane, Schervish, and Seidenfeld (e.g., Seidenfeld et al. (1994)). While it would take another paper to fully explore these connections, I hope that this survey will encourage conversation not only across fields within the discipline of economics, but also across disciplines.

References

[1] Arrow, KJ. 1959. Rational choice functions and orderings. *Economica* 26:121-127.

[2] Arrow, KJ, Hurwicz, LJ. 1972. An optimality criterion for decision-making under ignorance. In Carter, CF, Ford, JL (eds.), *Uncertainty and Expectations in Economics: Essays in Honour of G.L.S. Shackle*. Oxford: Basil Blackwell.

[3] Augustin, T, Coolen, FPA, Moral, S, Troffaes, CS (eds.). 2009. *ISIPTA'09: Proceedings of the sixth international symposium on imprecise probability: Theories and applications*. Durham, UK: SIPTA.

[4] Bergemann D. 2008. Bandit problems. In Durlauf, SN, Blume, LE (eds.), *The New Palgrave Dictionary of Economics*.

[5] Berger, JO. 1985. *Statistical decision theory and Bayesian analysis*. New York: Springer Verlag.

[6] Bewley, TF. 2002. Knightian decision theory. Part I. *Dec. Econ. Finance* 25:79-110.

[7] Binmore, K. 2011. *Rational decisions*. Princeton: Princeton University Press.

[8] Canner, PL. 1970. Selecting one of two treatments when the responses are dichotomous. *J Am. Stat. Assoc.* 65:293-306.

[9] Chamberlain, G. 2000. Econometrics and decision theory. *J. Econom.* 95:255-283.

[10] Chamberlain, G. 2011. Bayesian aspects of treatment choice. In Geweke et al. (eds., 2011).

[11] Cucconi, O. 1968. Contributi all'analisi sequenziale nel controllo di accettazione per variabili. *Atti dell' Associazione Italiana per il Controllo della Qualità* 6:171-186.

[12] Dehejia, R. 2005. Program evaluation as a decision problem. *J. Econom.* 125:141-173.

[13] Eozenou, P, Rivas, J, Schlag, KH. 2006. Minimax regret in practice – four examples on treatment choice. Discussion paper. Florence: European University Institute.

[14] Epstein, LG, Schneider, M. 2001. Recursive multiple priors. *J. Econ. Theory* 113:1-31.

[15] Ferguson, TS. 1967. *Mathematical statistics: A decision theoretic approach*. New York: Wiley.

[16] Fishburn, PC. 1970. *Utility theory for decision making*. New York: Wiley.

[17] Geisser, S. 1984. On prior distributions for binary trials. *Am. Statistician* 38:244-247.

[18] Geweke, J, Koop, G, van Dijk, H (eds.). 2011. *The Oxford Handbook of Bayesian Econometrics*. Oxford: Oxford University Press.

[19] Ghirardato, P, Maccheroni, F, Marinacci, M. 2004. Differentiating ambiguity and ambiguity attitude. *J. Econ. Theory* 118:133-173.

[20] Gilboa, I. 2009. *Theory of decision under uncertainty*. Cambridge: Cambridge University Press.

[21] Gilboa, I, Schmeidler, D. 1989. Maxmin expected utility with non-unique prior, *J. Math. Econ.* 18:141-153.

[22] Gilboa, I, Schmeidler, D. 1993. Updating ambiguous beliefs. *J. Econ. Theory* 59:33–49.

[23] Gilboa, I, Maccheroni, F, Marinacchi, M, Schmeidler, D. 2010. Objective and subjective rationality in a multiple prior model. *Econometrica* 78:755–770.

[24] Gupta, SS, Hande, SN. 1992. On some nonparametric selection procedures. In Saleh, AKME (ed.), *Nonparametric Statistics and Related Topics*. Elsevier.

[25] Hanany, E, Klibanoff, P. 2007. Updating preferences with multiple priors, *Theoretical Econ.* 2:261-298.

[26] Hayashi, T. 2008. Regret aversion and opportunity-dependence. *J. Econ. Theory* 139:242-268.

[27] Hayashi, T. 2011. Context dependence and consistency in dynamic choice under uncertainty: the case of anticipated regret. *Theory Dec.* 70:399-430.

[28] Hirano, K. 2009. Econometrics and Decision Theory. In Durlauf, SN, Blume, LE (eds.), *The New Palgrave Dictionary of Economics*.

[29] Hirano, K, Porter, JR. 2009. Asymptotics for statistical treatment rules. *Econometrica* 77:1683–1701.

[30] Hodges, JL, Lehmann, EL. 1950. Some problems in minimax point estimation. *Ann. Math. Stat.* 21:182-197.

[31] Johnson, BM. 1971. On the admissible estimators for certain fixed sample binomial problems. *Ann. Math. Stat.* 42:1579-1587.

[32] Karlin, S, Rubin, H. 1956. The theory of decision procedures for distributions with monotone likelihood ratio. *Ann. Math. Stat.* 27:272–299.

[33] Kitagawa, T. 2011. Estimation and inference for set-identified parameters using posterior lower probability. Discussion paper. London: University College London.

[34] Knight, FH. 1921. *Risk, uncertainty, and profit*. Boston: Houghton Mifflin.

[35] Levi, I. 1980. *The enterprise of knowledge*. Cambridge, MA: MIT Press.

[36] Little, IMD. 1949. A reformulation of the theory of consumer's behaviour. *Oxford Econ. Papers (New Series)* 1:90-99.

[37] Manski, CF. 2000. Identification problems and decisions under ambiguity: Empirical analysis of treatment response and normative analysis of treatment choice. *J. Econom.* 95:415-442.

[38] Manski, CF. 2003. *Partial identification of probability distributions*. New York: Springer Verlag.

[39] Manski, CF. 2004. Statistical treatment rules for heterogeneous populations. *Econometrica* 72:1221-1246.

[40] Manski, CF. 2005. *Social choice with partial knowledge of treatment response*. Princeton: Princeton University Press.

[41] Manski, CF. 2007a. Minimax-regret treatment choice with missing outcome data. *J. Econom.* 139:105-115.

[42] Manski, CF. 2007b. *Identification for prediction and decision*. Cambridge, MA: Harvard University Press.

[43] Manski, CF. 2009. Diversified treatment under ambiguity. *Internat. Econ. Rev.* 50:1013–1041.

[44] Manski, CF. 2011a. Choosing treatment policies under ambiguity. *Annual Rev. Econ.*

[45] Manski, CF. 2011b. Actualist rationality. *Theory Dec.* 71:195–210.

[46] Manski, CF, Tetenov, A. 2007. Admissible treatment rules for a risk-averse planner with experimental data on an innovation. *J. Stat. Plan. Inference* 137:1998-2010.

[47] Milnor, J. 1954. Games against nature. In Thrall, RM, Coombs, CH, Davis, RL (eds.), *Decision Processes*. New York: Wiley.

[48] Moon, HR, Schorfheide, F. 2010. Bayesian and frequentist inference in partially identified models. Discussion paper. Philadelphia: University of Pennsylvania.

[49] Niehans, J. 1948. Zur Preisbildung bei ungewissen Erwartungen. *Schweizerische Zeitschrift für Volkswirtschaft und Statistik*, 84:433-456.

[50] Poirier, DJ. 1998. Revising beliefs in nonidentified models. *Econom. Theory* 14:483-509.

[51] Puppe, C, Schlag, KH. 2009. Choice under complete uncertainty when outcome spaces are state-dependent. *Theory Dec.* 66:1-16.

[52] Savage, LJ. 1951. The theory of statistical decisions. *J. Am. Stat. Assoc.* 46:55-67.

[53] Savage, LJ. 1954. *The foundations of statistics*. New York: Wiley.

[54] Schlag, KH. 2003. How to minimize maximum regret under repeated decision-making. Discussion paper. Florence: European University Institute.

[55] Schlag, KH. 2007. Eleven. Discussion paper. Florence: European University Institute.

[56] Schmeidler, D. 1989. Subjective probability and expected utility without additivity. *Econometrica* 57:571-587.

[57] Seidenfeld, T, Schervish, MJ, Kadane, JB. 1994. A representation of partially ordered preferences. *Ann. Stat.* 23:2168-2217.

[58] Sen, A. 1973. Behaviour and the concept of preference. *Economica* 40:241–259.

[59] Seo, K. 2009. Ambiguity and second-order belief. *Econometrica* 77:1575-1605.

[60] Song, K. 2010. Point decisions for interval-identified parameters. Discussion paper. Philadelphia: University of Pennsylvania.

[61] Stoye, J. 2009a. Minimax regret treatment choice with finite samples. *J. Econom.* 151:70-81.

[62] Stoye, J. 2009b. Minimax regret. In Durlauf, SN, Blume, LE (eds.), *The New Palgrave Dictionary of Economics*.

[63] Stoye J. 2011a. Statistical decisions under ambiguity. *Theory Dec.* 70:129-148.

[64] Stoye J. 2011b. Axioms for minimax regret choice correspondences. *J. Econ. Theory*, in press.

[65] Stoye J. 2011c. Minimax regret treatment choice with covariates and with limited validity of experiments. *J. Econom.*, in press.

[66] Stoye, J. 2011d. Revealed preference when agents can randomize. Discussion paper. Ithaca: Cornell University.

[67] Stoye, J. 2011e. Dominance and admissibility without priors. Discussion paper. Ithaca: Cornell University.

[68] Strawderman, WE. 2000. Minimaxity. *J. Am. Stat. Ass.* 95:1364-1368

[69] Tamer, E. 2011. Partial identification in econometrics. *Annual Rev. Econ.*

[70] Tetenov, A. 2009. Measuring precision of statistical inference on partially identified parameters. Discussion paper. Torino: Collegio Carlo Alberto.

[71] Tetenov, A. 2011. Statistical treatment choice based on asymmetric minimax regret criteria. *J Econom.*, in press.

[72] von Neumann, J, Morgenstern, O. 1947. *Theory of games and economic behavior (2nd edition).* Princeton: Princeton University Press.

[73] Wald, A. 1945. Statistical decision functions which minimize the maximum risk. *Annals Math.* 46:265-280.

[74] Wald, A. 1950. *Statistical decision functions*. New York: Wiley.

[75] Walley, P. 1990. *Statistical reasoning with imprecise probabilities*. New York: Chapman and Hall.

[76] Yuan, L, Jang, W. 2010. Bayesian analysis in moment inequality models. *Ann. Stat.* 38:275-316.