IDENTIFICATION AND STATISTICAL DECISION THEORY

Charles F. Manski
Department of Economics and Institute for Policy Research
Northwestern University

Draft: February 3, 2022

Abstract

Econometricians have usefully separated study of estimation into identification and statistical components. Identification analysis aims to place a tractable and informative upper bound on what may be learned about a true state of nature with sample data. Statistical decision theory has studied decision making with sample data without reference to identification. Nevertheless, identification analysis can be useful to statistical decision theory, placing a tractable and informative upper bound on the performance of decision making with sample data. The argument is straightforward when the true state of nature is point identified. This paper calls attention to a subtlety that arise when the true state is partially identified, and a decision must be made under ambiguity. Then the performance of some criteria, particularly minimax regret, is enhanced by permitting randomized choice of an action, which requires availability of sample data. An upper bound on the performance of decision making holds when one combines the knowledge assumed in identification analysis with sample data enabling randomized choice. Using sample data to randomize choice is conceptually distinct from its ordinary use to infer the true state of nature.
Econometricians have long found it instructive to separate the study of estimation into identification and statistical components. Koopmans (1949, p. 132) put it this way in the article that introduced the term identification into the literature:

“In our discussion we have used the phrase “a parameter that can be determined from a sufficient number of observations.” We shall now define this concept more sharply, and give it the name identifiability of a parameter. Instead of reasoning, as before, from “a sufficiently large number of observations” we shall base our discussion on a hypothetical knowledge of the probability distribution of the observations, as defined more fully below. It is clear that exact knowledge of this probability distribution cannot be derived from any finite number of observations. Such knowledge is the limit approachable but not attainable by extended observation. By hypothesizing nevertheless the full availability of such knowledge, we obtain a clear separation between problems of statistical inference arising from the variability of finite samples, and problems of identification in which we explore the limits to which inference even from an infinite number of observations is suspect.”

Koopmans recognized that statistical and identification problems limit in distinct ways the conclusions that may be drawn in empirical research. Statistical problems may be severe in small samples but diminish as the sampling process generates more observations. Identification problems cannot be solved by gathering more of the same kind of data. They can be alleviated only by invoking stronger assumptions or by initiating new sampling processes that yield different kinds of data.

Whereas study of identification has been central to econometrics, the concept does not appear in statistical decision theory as developed by Wald (1950). Indeed, estimation does not appear. Statistical decision theory shares with econometrics the concept of a true parameter value (state of nature) assumed to lie in a specified parameter space (state space). As in econometrics, statistical decision theory supposes that
one observes sample data drawn from a probability (sampling) distribution. However, statistical decision theory does not presume that a decision maker uses the sample data to estimate the true state of nature. Wald’s core concept of a statistical decision function (SDF) embraces all mappings from sample data to an action, including ones that do not involve estimation of the true state of nature. Prominent decision criteria— including maximin, minimax-regret, and maximization of subjective average welfare—are do not perform estimation.

Given that estimation is not required to compute and evaluate statistical decision functions, it is natural to ask whether the study of estimation is relevant to statistical decision theory. One may further ask whether separation of study of estimation into statistical and identification components is useful.

The first question has a straightforward positive answer. Although SDFs need not perform estimation, they often do so. Many SDFs have the form [data → estimation → action], first performing estimation and then using the estimate to choose an action. Koopmans and Hood (1953) posed this idea early on. Considering “The Purpose of Estimation,” they focused on the use of estimates to make predictions and wrote that (p. 127): “estimates . . . can be regarded as raw materials, to be processed further into solutions of a wide variety of prediction problems.” Manski (2021) recently studied the prominent class of SDFs that perform as-if optimization, computing a point estimate of the true state of nature and then choosing an action that optimizes welfare as if the estimate is the true state. I also considered SDFs that use set-valued estimates, such as estimates of the identification regions of partially identified states of nature.

This paper addresses the second question, whose answer is more subtle. Aiming to make study of identification useful to empirical research, econometricians have connected identification to statistical inference through development of asymptotic statistical theory, particularly through theorems on consistent estimation and rates of convergence. These theorems formalize the idea that, as sample size increases, the hypothetical knowledge of the probability distribution of observations assumed in identification analysis is increasingly well-approximated.

On the other hand, Wald directly studied decision making with sample data, evaluating SDFs by their
risk. He did not study limit experiments of asymptotic estimation theory, where a finite sample size hypothetically increases toward infinity. A fortiori, he did not study the even more hypothetical setting of identification, where one is assumed to know the probability distribution generating observations.

The knowledge of the probability distribution of observations assumed in identification analysis is qualitatively more informative than any data sample. Knowledge of the probability distribution enables one to shrink the state space, eliminating states of nature that are inconsistent with the known distribution. Sample data may be useful to decision making, but they do not enable one to shrink the state space. Le Cam (1986) presents an asymptotic form of statistical decision theory in which the state space is presumed to shrink toward the true state of nature at a specified rate as sample size increases. Le Cam’s analysis differs fundamentally from Wald’s, where the state space is specified ex ante and does not shrink with observation of sample data.

Exploring the connection between identification and statistical decision theory, I find an intriguing distinction between settings with point and partial identification. When the true state of nature is point identified, study of decision making with the knowledge assumed in identification analysis yields an upper bound on the quality of decisions that may be achievable with sample data. The reasoning is straightforward. Knowledge of the true state of nature implies that the decision maker can choose an action that maximizes welfare. It is logically impossible to achieve more with sample data.

I show that the situation differs if the true state is partially identified and the planner faces a problem of decision making under ambiguity; that is, no feasible action dominates all alternatives. Then a decision maker cannot maximize welfare. He can at most use some reasonable criterion to choose an action. It is known that the performance of some criteria for choice under ambiguity, particularly minimax regret, is enhanced by permitting randomized choice of an action. Randomized choice is performed by specifying a probability distribution on a sample space and a function mapping the sample space into an action. One draws a realization at random and applies the function to choose an action. Thus, randomized choice uses sample data to make a decision.
I observe that using sample data to randomize choice is conceptually distinct from its ordinary use to infer the true state of nature. When randomized choice is beneficial, decision making combining knowledge of the identification region with sample data can outperform decision making using knowledge of the identification region alone. This is so even though the data yield no new knowledge of the state of nature.

Section 2 reviews basic statistical decision theory. Section 3 formalizes decision making with the knowledge of the probability distribution of observations assumed in identification analysis, but without accompanying sample data. This is a simple extension of Section 2. Section 4 considers decision making when the true state of nature is partially identified, and one combines knowledge of the identification region with sample data. Section 5 presents analytical findings for the important special case of choice between two actions. Section 6 concludes.

2. Basic Statistical Decision Theory

To begin I review basic elements of statistical decision theory, which makes no reference to identification. The presentation draws on Manski (2021).

2.1. Decisions without Sample Data

First consider decisions without sample data. Hence, randomized choice is not possible. A decision maker, called a planner for short, faces a choice set $C$ and believes that the true state of nature lies in a specified state space $S$. An objective function $w(\cdot, \cdot): C \times S \rightarrow \mathbb{R}^1$ maps actions and states into welfare. The planner wants to maximize true welfare, but he does not know the true state. Decision theorists have proposed various ways of using $w(\cdot, \cdot)$ to form functions of actions alone, which can be optimized. I use
max and min notation below, without concern for subtleties that may make it necessary to use sup and inf operations.

One approach places a subjective probability distribution $\pi$ on $S$, computes average state-dependent welfare with respect to $\pi$, and maximizes subjective average welfare over $C$. The criterion solves

\[
\max_{c \in C} \int w(c, s) d\pi.
\]

Another approach seeks an action that, in some sense, works uniformly well over all of $S$. This yields the maximin and minimax-regret (MMR) criteria. These respectively solve the problems

\[
\begin{align*}
(2) & \quad \max_{c \in C} \min_{s \in S} w(c, s). \\
(3) & \quad \min_{c \in C} \max_{s \in S} \left[ \max_{d \in C} w(d, s) - w(c, s) \right].
\end{align*}
\]

In (3), $\max_{d \in C} w(d, s) - w(c, s)$ is the *regret* of action $c$ in state $s$; that is, the degree of suboptimality.

Given a specified welfare function and state space, enlargement of the choice set weakly improves the optimized value of each decision criterion (1) – (3). Thus, a decision maker using each criterion always prefers to expand the choice set.

Given a specified welfare function and choice set, obtaining new information that shrinks the state space weakly increases the maximum value of minimum welfare (2) and weakly decreases the minimum value of maximum regret (3). Thus, a decision maker using these criteria always prefers to know more about the true state. Shrinking the state space may increase or decrease the maximum value of subjective expected welfare, depending on the welfare function, the prior distribution $\pi$ on $S$, and the posterior distribution on the shrunken subset of $S$. 
2.2. Statistical Decision Problems

Statistical decision problems suppose that the planner observes data generated by a sampling distribution. Knowledge of the sampling distribution is generally incomplete. To express this, one extends state space $S$ to list the feasible sampling distributions, denoted $(Q_s, s \in S)$. Let $\Psi_s$ denote the sample space in state $s$; $\Psi_s$ is the set of samples that may be drawn under distribution $Q_s$. The literature typically assumes that the sample space does not vary with $s$ and is known. I do likewise and denote the sample space as $\Psi$.

A statistical decision function, $c(\cdot): \Psi \to C$, maps the sample data into a chosen action.

An SDF is a deterministic function after realization of the sample data, but it is a random function ex ante. Hence, welfare is a random variable ex ante. Wald’s theory evaluates the performance of SDF $c(\cdot)$ in state $s$ by $Q_s \{w[c(\psi), s]\}$, the ex-ante distribution of welfare that it yields across realizations $\psi$ of the sampling process. In particular, Wald measured the performance of $c(\cdot)$ in state $s$ by its expected welfare across samples; that is, $E_s \{w[c(\psi), s]\} = \int w[c(\psi), s] dQ_s$. Not knowing the true state, a planner evaluates $c(\cdot)$ by the expected welfare vector $(E_s \{w[c(\psi), s]\}, s \in S)$.

Statistical decision theory has mainly studied the same decision criteria as has decision theory without sample data. Let $\Gamma$ denote the set of feasible SDFs, which map $\Psi \to C$. The statistical versions of criteria (1), (2), and (3) are

\[
\begin{align*}
(4) \quad \max_{c(\cdot) \in \Gamma} \int E_s \{w[c(\psi), s]\} \, d\pi, \\
(5) \quad \max_{c(\cdot) \in \Gamma} \min_{s \in S} E_s \{w[c(\psi), s]\}, \\
(6) \quad \min_{c(\cdot) \in \Gamma} \max_{s \in S} \left( \max_{d \in C} w(d, s) - E_s \{w[c(\psi), s]\}\right).
\end{align*}
\]

Observing sample data is not informative in the sense of shrinking the state space. Indeed, incomplete
knowledge of the sampling distribution may enlarge the state space. Nevertheless, sample data may enhance decision making through two mechanisms.

First, observing data enlarges the choice set, which now is composed of SDFs $c(\cdot): \Psi \rightarrow C$ rather than simply actions $c \in C$. Relatedly, sample data enable randomized choice of an action. Second, data may be informative about an incompletely known sampling distribution, as studied in research on statistical inference. These mechanisms jointly suggest that observing sample data may be useful to decision making. The task of statistical decision theory is to operationalize this suggestion.

### 3. Decisions with Knowledge of the Sampling Distribution

From here on, I develop connections and distinctions between identification analysis and statistical decision theory. This section considers decision making with the knowledge that econometricians have assumed in identification analysis. Thus, supposing that some $s \in S$ is the true state of nature, I assume that the planner can learn the probability distribution $Q_s$ that generates sample data. However, the planner does not observe sample data.

Knowledge of $Q_s$ is useful to decision making to the extent that it shrinks the state space. Let $S$ be the original state space of Section 2.1, without knowledge of $Q_s$. Let $S(Q_s) \subset S$ be the shrunken state space obtained with knowledge of $Q_s$. Presuming no error in determination of $S(Q_s)$, it is necessarily the case that $s \in S(Q_s)$. The true state is point-identified if $S(Q_s) = s$. It is partially identified if $S(Q_s)$ is a proper non-singleton subset of $S$. Note that $\bigcup_{s \in S} S(Q_s) = S$ in all identification settings.

Given that the true state may be any element of $S$, it is useful to also define uniform point and partial identification. The true state is uniformly point-identified if $S(Q_s) = s$, $\forall s \in S$. It is uniformly partially-identified if $S(Q_s)$ is a proper non-singleton subset of $S$, $\forall s \in S$. Manski (1988, Section 1.1) distinguished
identification and uniform identification, with identification implicitly meaning point identification. See also the discussion in Molinari (2020, Section 3.1).

I consider decision making from two timing perspectives: ex post, after \( Q_s \) becomes known, and ex ante, before it is known. Section 3.1 formalizes the ex-post problem, which is analogous to decision making without sample data. Section 2.1 described this problem with state space \( S \). Now the planner chooses an action with knowledge of the shrunken state space \( S(Q_s) \).

Section 3.2 formalizes the ex-ante decision problem, which is analogous to Wald’s study of decision making before sample data have been observed. Wald considered a planner who chooses an SDF \( c(\cdot): \Psi \rightarrow C \), specifying the action that the planner would choose should any sample data \( \psi \) be observed. Now the planner chooses a decision function \( c(\cdot): [S(Q_s), s \in S] \rightarrow C \), specifying the action that the planner would choose should any shrunken state space \( S(Q_s) \) become known.

The two timing perspectives generate ex ante and ex post versions of the decision criteria considered in Section 2. I show that the ex-ante and ex post criteria are dynamically consistent.

3.1. Ex Post Decisions

The new analogs of decision criteria (1) – (3) choose actions with knowledge of the shrunken state space \( S(Q_s) \) rather than the original space \( S \). The criteria are

\[
(7) \quad \max_{c \in C} \int w(c, s) d\pi[s|S(Q_s)].
\]

\[
(8) \quad \max_{c \in C} \min_{s \in S(Q_s)} w(c, s).
\]

\[
(9) \quad \min_{c \in C} \max_{s \in S(Q_s)} [\max_{d \in C} w(d, s) - w(c, s)].
\]
In (7), the subjective expectation is taken with respect to the posterior distribution $\pi[s|S(Q_s)]$ rather than the prior distribution $\pi$.

If the true state is point-identified, criteria (7) and (8) both reduce to the deterministic optimization problem $\max_{c \in C} w(c, s)$. Criterion (9) reduces to $\min_{c \in C} [\max_{d \in C} w(d, s) - w(c, s)]$, which is equivalent to the problem $\max_{c \in C} w(c, s)$.

3.2. Ex-Ante Decisions

The new analogs of decision criteria (4) – (6) choose decision functions $c(\cdot): [S(Q_s), s \in S] \rightarrow C$ that select actions for all potential sampling distributions. The feasible decision functions $\Gamma$ map $[S(Q_s), s \in S] \rightarrow C$. The criteria are

\[
(10) \quad \max_{c(\cdot) \in \Gamma} \int w\{c[S(Q_s)], s\} \, d\pi,
\]

\[
(11) \quad \max_{c(\cdot) \in \Gamma} \min_{s \in S} \{w[c[S(Q_s)], s]\},
\]

\[
(12) \quad \min_{c(\cdot) \in \Gamma} \max_{s \in S} \left( \max_{d \in C} w(d, s) - w[c[S(Q_s)], s]\right).
\]

If the true state is uniformly point identified, $\Gamma$ maps $(s \in S) \rightarrow C$ and the criteria reduce to

\[
(10') \quad \max_{c(\cdot) \in \Gamma} \int w[c(s), s] \, d\pi,
\]

\[
(11') \quad \max_{c(\cdot) \in \Gamma} \min_{s \in S} w[c(s), s],
\]

\[
(12') \quad \min_{c(\cdot) \in \Gamma} \max_{s \in S} \left( \max_{d \in C} w(d, s) - w[c(s), s]\right).
\]
Selecting the decision function to be \( c^*(s) = \arg\max_{d \in C} w(d, s), \forall s \in S \) solves each of problems (10'). \( c^*(\cdot) \) is the unique solution to the MMR criterion (12'), yielding maximum regret equal to zero. Deviating from \( c^*(s) \) for any \( s \in S \) makes regret positive for that state. Hence, maximum regret is positive.

\( c^*(\cdot) \) is not necessarily the unique solution to problem (10'). Let \( S_0 \) denote any subset of \( S \) such that \( \pi(S_0) = 0 \). The value of the integral \( \int [w(c(s), s)] d\pi \) does not vary with the chosen actions \( c(s), s \in S_0 \). Hence, the chosen decision function can deviate arbitrarily from \( c^*(\cdot) \) on set \( S_0 \).

\( c^*(\cdot) \) also is not necessarily the unique solution to the maximin problem (11'). The maximin welfare value is \( \min_{s \in S} \max_{d \in C} w(d, s) \). Let \( s' \) be any state of nature and \( c' \) be any action such that \( \max_{d \in C} w(d, s') \geq \min_{s \in S} \max_{d \in C} w(d, s) \). Then the chosen decision function can deviate from \( c^*(\cdot) \) at state \( s' \) by selecting \( c' \) rather than \( c^*(s') \), without changing the maximin welfare value.

3.3. Dynamic Consistency of Ex-Ante and Ex-Post Decisions

The above discussion showed that ex-ante and ex-post decisions are dynamically consistent when the true state is uniformly point identified. Ex ante, the decision function \( c^*(s) = \arg\max_{d \in C} w(d, s), \forall s \in S \) solves each of problems (10') – (12'). Ex post, \( \arg\max_{d \in C} w(d, s) \) solves each of problems (7) – (9).

Ex ante and ex post decisions are also dynamically consistent in the absence of uniform point identification. To see this, recall that \( \bigcup_{s \in S} S(Q_s) = S \). Consider criteria (10) – (12).

In criterion (10), \( \int w[c(S(Q_s))] s] d\pi = \int \int w[c(S(Q_s))] s] d\pi[s|S(Q_s)]d\pi[S(Q_s)] \). The ex post decision criterion (7) maximizes the inner integral \( \int w[c(S(Q_s))] s] d\pi[s|S(Q_s)] \) for each shrunken state space \( S(Q_s) \). Hence, it maximizes the outer integral.

In criterion (11), minimization of welfare over \( S \) is the same as minimizing over \( \bigcup_{s \in S} S(Q_s) \). The ex post decision criterion (8) maximizes minimum welfare for each \( S(Q_s) \). Hence, it yields maximin welfare over \( \bigcup_{s \in S} S(Q_s) \).
In criterion (12), maximization of regret over $S$ is the same as maximizing over $\bigcup_{s \in S} S(Q_s)$. The ex post decision criterion (9) minimizes maximum regret for each $S(Q_s)$. Hence, it does so over $\bigcup_{s \in S} S(Q_s)$.

4. Decisions with Knowledge of the Sampling Distribution and with Sample Data

Given knowledge of the sampling distribution $Q_s$, with its implied knowledge of the shrunken state space $S(Q_s)$, one might conjecture that there is no point in observing sample data drawn from $Q_s$. After all, with $S(Q_s)$ known, observation of sample data yields no further information about the true state.

The conjecture is correct if the true state is point-identified. As shown in Section 3, point identification enables the planner to maximize welfare. One cannot improve on that. However, the situation differs if the true state is partially identified and the planner faces a problem of decision making under ambiguity; that is, $S(Q_s)$ is such that no action in $C$ dominates all alternatives. Depending on the welfare function and the decision criterion used to cope with ambiguity, randomized choice of an action may outperform any singleton choice. Sample data enable randomized choice. Thus, decision making combining knowledge of $Q_s$ with sample data may outperform decision making using knowledge of $Q_s$ alone.

The analysis below assumes for simplicity that choice set $C$ contains finitely many actions. Sections 4.1 and 4.2 develop the basic reasoning, which recasts choice of an SDF as selection of choice probabilities for the elements of $C$. Section 5 examines the important special case in which $C$ contains two feasible actions.
4.1. Decision Making as Selection of Choice Probabilities

As described in Section 2.2, Wald’s frequentist statistical decision theory considers a planner who chooses a statistical decision function before observing sample data. The planner evaluates the performance of SDF \( \psi \rightarrow C \) in state \( s \) by expected welfare across samples; \( E_s\{w[\psi(s), s]\} \equiv \int w[\psi(s), s]dQ_s \).

Suppose that the planner chooses an SDF after learning the sampling distribution \( Q_s \), but before observing \( \psi \). Then the relevant versions of the subjective-expected-welfare, maximin, and MMR decision criteria are

\[
\begin{align*}
(13) & \quad \max_{c(\cdot) \in \Gamma} \int E_s\{w[\psi(s), s]\}d\pi[s|S(Q_s)]. \\
(14) & \quad \max_{c(\cdot) \in \Gamma} \min_{s \in S(Q_s)} E_s\{w[\psi(s), s]\}. \\
(15) & \quad \min_{c(\cdot) \in \Gamma} \max_{s \in S(Q_s)} \left[ \max_{d \in C} w(d, s) - E_s\{w[\psi(s), s]\} \right].
\end{align*}
\]

where the feasible SDFs \( \Gamma \) map \( \psi \rightarrow C \).

The decision criteria of basic statistical decision theory, given in (4) – (6), are the polar cases of (13) – (15) in which \( S(Q_s) = S \). When \( S(Q_s) \) is a proper subset of \( S \), the maximin value in (14) is weakly larger than in (5) because expected welfare is minimized over \( S(Q_s) \) rather than \( S \). The MMR value in (15) is weakly smaller than in (6) because regret is maximized over \( S(Q_s) \) rather than \( S \). The maximum of subjective expected welfare in (13) may be larger or smaller than in (4), depending on the welfare function, the prior distribution, and the posterior distribution.

Criteria (13) – (15) can be written in an equivalent form that eases analysis. Knowledge of \( Q_s \) implies that, for any SDF \( c(\cdot) \), the planner can evaluate the choice probabilities \( \{Q_s[c(\psi) = d], d \in C\} \) with which \( c(\cdot) \) selects alternative actions. The expected welfare of \( c(\cdot) \) in state \( s \) is
(16) \( E_s \{ w[c(\psi), s] \} = \sum_{d \in C} Q_s[c(\psi) = d] \cdot w(d, s) \),

which varies with \( c(\cdot) \) only through the choice probabilities. Thus, given knowledge of \( Q_s \), evaluation of the performance of \( c(\cdot) \) is equivalent to evaluation of the choice probabilities it yields.

Now reconsider criteria (13) – (15). Given knowledge of \( Q_s \), evaluation of the performance of all feasible SDFs is equivalent to evaluation of the set \( \Delta_r(Q_s) \equiv \{ Q_s[c(\psi) = d], d \in C \}, c(\cdot) \in \Gamma \) of all feasible vectors of choice probabilities. Let \( \delta(Q_s, d), d \in C \) denote a feasible vector of choice probabilities. Then (13) – (15) are equivalent to the criteria

\[
\begin{align*}
(13') \quad & \max_{\delta(Q_s, \cdot) \in \Delta_r(Q_s)} \int \left[ \sum_{d \in C} \delta(Q_s, d) \cdot w(d, s) \right] d\pi[s|S(Q_s)]. \\
(14') \quad & \max_{\delta(Q_s, \cdot) \in \Delta_r(Q_s)} \min_{s \in S(Q_s)} \sum_{d \in C} \delta(Q_s, d) \cdot w(d, s). \\
(15') \quad & \min_{\delta(Q_s, \cdot) \in \Delta_r(Q_s)} \max_{s \in S(Q_s)} \left[ \max_{d \in C} w(d, s) - \sum_{d \in C} \delta(Q_s, d) \cdot w(d, s) \right].
\end{align*}
\]

Comparison of these decision criteria with those considered earlier shows that combining knowledge of \( Q_s \) with observation of sample data weakly improves the performance of decision making. In Section 2.2, only sample data were available. In Section 3, the only feasible choice probabilities \( \delta(Q_s, \cdot) \) were the vertices of \( \Delta_r(Q_s) \). Now interior choice probabilities may be feasible as well.

If the sample space and sampling distribution are sufficiently rich, \( \Delta_r(Q_s) \) is the entire \(|C|\)-dimensional unit simplex. The richness condition is easily satisfied. It suffices that \( \Psi \) contain an interval on the real line and that \( Q_s \) have positive density on this interval. Then the planner can select an SDF to yield any vector of choice probabilities.
4.2. Fully Ex Ante Decisions

Section 4.1 considered a planner who chooses after learning $Q_s$, as in Section 3.1. A planner might behave in an ex ante manner, as in Section 3.2. He would then select vectors of choice probabilities for all sampling distributions $(Q_s, s \in S)$ that may potentially be observed. The ex-ante versions of (13’ – 15’) are

\[
\begin{align*}
(17) & \quad \max_{\delta(Q_s, \cdot) \in \Delta r(Q_s), s \in S} \int \left[ \sum_{d \in C} \delta(Q_s, d) \cdot w(d, s) \right] d\pi(s), \\
(18) & \quad \max_{\delta(Q_s, \cdot) \in \Delta r(Q_s), s \in S} \min_{s \in S} \sum_{d \in C} \delta(Q_s, d) \cdot w(d, s), \\
(19) & \quad \min_{\delta(Q_s, \cdot) \in \Delta r(Q_s), s \in S} \max_{s \in S} \left[ \max_{d \in C} w(d, s) - \sum_{d \in C} \delta(Q_s, d) \cdot w(d, s) \right].
\end{align*}
\]

Repetition of the analysis in Section 3.3 shows that ex post and ex ante decisions are dynamically consistent.

5. Decisions under Ambiguity with Binary Choice Sets

Section 3 showed that, with $Q_s$ known, observation of sample data is not beneficial if $s$ is point-identified. Point identification enables a planner to maximize welfare. Hence, the randomization of choice made possible by sample data has no value.

Randomization of choice may be beneficial if the true state is partially identified, and the planner chooses under ambiguity. I consider here the important special case where choice set $C$ contains two actions, say \{a, b\} and the welfare values \{w(a, s), w(b, s), s \in S(Q_s)\} have bounded range. The planner faces ambiguity if both actions are undominated; that is, if $w(a, s) > w(b, s)$ for some values of $s$ and $w(a, s) < w(b, s)$ for other values.
Noting that $\delta(Q_s, a) + \delta(Q_s, b) = 1$, I define $\delta(Q_s) \equiv \delta(Q_s, b)$, implying that $\delta(Q_s, a) = 1 - \delta(Q_s)$. I assume for simplicity that $\Delta \tau(Q_s)$ is the entire unit simplex in $\mathbb{R}^2$. Then the planner’s problem is to choose a value $\delta(Q_s) \in [0, 1]$. Criteria (13’) – (15’) reduce to

$$
\begin{align*}
(20) & \quad \max_{\delta(Q_s) \in [0, 1]} \left[ 1 - \delta(Q_s) \right] \cdot \int w(a, s) \, d\pi[s|S(Q_s)] + \delta(Q_s) \cdot \int w(b, s) \, d\pi[s|S(Q_s)], \\
(21) & \quad \max_{\delta(Q_s) \in [0, 1]} \min_{s \in S(Q_s)} \left[ 1 - \delta(Q_s) \right] \cdot w(a, s) + \delta(Q_s) \cdot w(b, s), \\
(22) & \quad \min_{\delta(Q_s) \in [0, 1]} \max_{s \in S(Q_s)} \{ w(a, s), w(b, s) \} - \left[ 1 - \delta(Q_s) \right] \cdot w(a, s) - \delta(Q_s) \cdot w(b, s).
\end{align*}
$$

Manski (2007, chapter 11; 2009) studied problems (20) – (22) in a different substantive context, where the planner assigns a treatment to each member of a large population of observationally identical persons and $\delta(Q_s)$ is the fraction of persons assigned to treatment $b$. In that context, choosing $0 < \delta(Q_s) < 1$ means deterministic treatment diversification rather than randomized choice of an action. The mathematical problem is the same for both interpretations of $\delta(Q_s)$.

The findings depend on the decision criterion used. Given knowledge of $Q_s$, a planner who maximizes subjective expected welfare (20) does not randomize choice. A planner who uses maximin criterion (21) randomizes when facing some state spaces but not others. One using MMR criterion (22) always randomizes. I summarize the analysis and findings below.

5.1. Maximization of Subjective Expected Welfare

The unique solution to criterion (20) is $\delta(Q_s) = 1$ if $\int w(b, s) \, d\pi[s|S(Q_s)] > \int w(a, s) \, d\pi[s|S(Q_s)]$. It is $\delta(Q_s) = 0$ if $\int w(b, s) \, d\pi[s|S(Q_s)] < \int w(a, s) \, d\pi[s|S(Q_s)]$. All $\delta(Q_s) \in [0, 1]$ are optimal when $\int w(b,$
s)\,d\pi[s|S(Q_s)] = \int w(a, s)\,d\pi[s|S(Q_s)]. These findings hold for all \(w(\cdot, \cdot), \pi,\) and \(S(Q_s).\) Thus, randomized choice is generically sub-optimal.

5.2. Maximin Decisions

To solve maximin problem (21), one first computes the minimum welfare attained by each value of \(\delta(Q_s)\) across all feasible states \(s \in S(Q_s).\) One then chooses \(\delta(Q_s)\) to maximize this minimum welfare.

Let the lower and upper extreme values of the bounded welfares \(w(a, s)\) and \(w(b, s)\) across \(S(Q_s)\) be denoted \(\alpha_L(Q_s) \equiv \min_{s \in S(Q_s)} w(a, s), \beta_L(Q_s) \equiv \min_{s \in S(Q_s)} w(b, s), \alpha_U(Q_s) \equiv \max_{s \in S(Q_s)} w(a, s),\) and \(\beta_U(Q_s) \equiv \max_{s \in S(Q_s)} w(b, s).\) The maximin solution is simple if \([\alpha_L(Q_s), \beta_L(Q_s)]\) is a feasible value of \([w(a, s), w(b, s)]\). Then the unique maximin solution is \(\delta(Q_s) = 0\) if \(\alpha_L(Q_s) > \beta_L(Q_s)\) and \(\delta(Q_s) = 1\) if \(\alpha_L(Q_s) < \beta_L(Q_s).\) All \(\delta(Q_s) \in [0, 1]\) are maximin solutions when \(\alpha_L(Q_s) = \beta_L(Q_s).\) Thus, randomized choice is generically sub-optimal when \([\alpha_L(Q_s), \beta_L(Q_s)]\) is feasible.

Maximin decisions may randomize choice if \([\alpha_L(Q_s), \beta_L(Q_s)]\) is not feasible. Manski (2011, Chapter 11) gives a simple example. Let \(S(Q_s) = \{s_0, s_1\}\) and let \(\{w(a, s_0) = 1, w(b, s_0) = 0, w(a, s_1) = 0, w(b, s_1) = 1\}\). Then 
\[(1 - \delta(Q_s))\cdot w(a, s_0) + \delta(Q_s)\cdot w(b, s_0) = 1 - \delta(Q_s)\text{ and } (1 - \delta(Q_s))\cdot w(a, s_1) + \delta(Q_s)\cdot w(b, s_1) = \delta(Q_s).\]
Setting \(\delta(Q_s) = \frac{1}{2}\) maximizes minimum expected welfare.

5.3. MMR Decisions

The MMR solution always randomizes with binary choice sets when the planner faces ambiguity. Let \(S(Q_s)_a\) and \(S(Q_s)_b\) be the subsets of \(S(Q_s)\) on which actions \(a\) and \(b\) are superior. That is, 
\(S(Q_s)_a = \{s \in S(Q_s): w(a, s) > w(b, s)\}\) and \(S(Q_s)_b = \{s \in S(Q_s): w(b, s) > w(a, s)\}\). Both subsets are non-empty when the planner faces ambiguity.
Let $M(Q_s)_a \equiv \max_{s \in S(Q_s)_a} [w(a, s) - w(b, s)]$ and $M(Q_s)_b \equiv \max_{s \in S(Q_s)_b} [w(b, s) - w(a, s)]$ be maximum regret on $S(Q_s)_a$ and $S(Q_s)_b$ respectively. Both $M(Q_s)_a$ and $M(Q_s)_b$ are positive under ambiguity. When a dominant action exists, either $S(Q_s)_a$ and $S(Q_s)_b$ is empty. I then define the corresponding quantity $M(Q_s)_a$ or $M(Q_s)_b$ to equal zero.

Manski (2007, Complement 11A) proves that the MMR value of $\delta(Q_s)$ is always interior to $[0, 1]$ in settings with ambiguity. The result is

\[
\delta(Q_s)_{MR} = \frac{M(Q_s)_b}{M(Q_s)_a + M(Q_s)_b}.
\]

Expressions $M(Q_s)_a$ and $M(Q_s)_b$ simplify when $[\alpha_L(Q_s), \beta_U(Q_s)]$ and $[\alpha_U(Q_s), \beta_L(Q_s)]$ are feasible values of $\{w(a, s), w(b, s)\}$. Then $M(Q_s)_a = \alpha_L(Q_s) - \beta_L(Q_s)$ and $M(Q_s)_b = \beta_U(Q_s) - \alpha_L(Q_s)$ when both actions are undominated. Hence, the MMR solution for choice under ambiguity is

\[
\delta(Q_s)_{MR} = \frac{\beta_U(Q_s) - \alpha_L(Q_s)}{[\alpha_L(Q_s) - \beta_L(Q_s)] + [\beta_U(Q_s) - \alpha_L(Q_s)]}.
\]

When a planner uses choice probability $\delta(Q_s)_{MR}$, maximum regret is $M(Q_s)_aM(Q_s)_b/[M(Q_s)_a + M(Q_s)_b]$. It is interesting to compare this with the maximum regret that would result if the planner were only able to choose between the extreme choice probabilities 0 and 1. The solution then is $\delta(Q_s) = 0$ if $M(Q_s)_a \geq M(Q_s)_b$ and $\delta(Q_s) = 1$ if $M(Q_s)_a \leq M(Q_s)_b$. Maximum regret is $\min[M(Q_s)_a, M(Q_s)_b]$.

The above discussion concerns decision making after the planner learns the sampling distribution but before observation of sample data. Finally, consider the fully ex-ante setting where the planner selects choice probabilities for all possible sampling distributions. From this perspective, maximum regret is the maximum of $M(Q_s)_aM(Q_s)_b/[M(Q_s)_a + M(Q_s)_b]$ over all $Q_s, s \in S$. This quantity provides a computable lower
bound on maximum regret (6) in the setting of basic statistical decision theory, where the planner will observe sample data but will not learn the sampling distribution.

6. Conclusion

Econometricians have usefully separated study of estimation into identification and statistical components. Identification analysis aims to place a tractable and informative upper bound on what may be learned with sample data. Asymptotic theory has been used to connect identification to statistical inference. Consistency theorems show that, in the presence of regularity conditions, the hypothetical knowledge of the probability distribution of observations assumed in identification analysis is increasingly well-approximated as sample size increases.

Statistical decision theory has studied decision making with sample data without reference to identification. Nevertheless, identification analysis can be useful to statistical decision theory, placing a tractable and informative upper bound on the performance of decision making with sample data. The argument is straightforward when the true state of nature is point identified. This paper has called attention to a subtlety that arise when the true state is partially identified, and a decision must be made under ambiguity. Then the performance of some criteria, particularly minimax regret, is enhanced by permitting randomized choice of an action, which requires availability of sample data. Hence, an upper bound on the performance of decision making with sample data holds when one combines the knowledge assumed in identification analysis with sample data enabling randomized choice.

A logical next step beyond the scope of this paper may be to develop asymptotic theory that tightens the connection between the upper bound on decision performance studied here and the performance achievable with sample data. Focusing on settings with point identification of the true state, Manski (2004) noted a concept of regret consistency, wherein the maximum regret of an SDF converges to zero as sample
size increases. Regret consistency in this sense is not achievable in settings where partial identification implies choice under ambiguity. A broader concept of regret consistency that may be achievable holds if the maximum regret of an SDF converges to the MMR value studied in Sections 4 and 5 of this paper, where knowledge of the sampling distribution is combined with sample data enabling randomized choice.
References


