Chapter One

Introduction

Econometrics done as a productive enterprise deals with the interaction between economic theory and statistical analysis. Theory provides an organizing framework for looking at the world and in particular for assembling and interpreting economic data. Statistical methods provide the means of extracting interesting economic information from data. Without economic theory or statistics all that is left is an overwhelming flow of disorganized information. Thus, both theory and statistics provide methods of data reduction. The goal is to reduce the mass of disorganized information to a manageable size while retaining as much of the information relevant to the question being considered as possible. In economic theory, much of the reduction is done by reliance on models of optimizing agents. Another level of reduction can be achieved by considering equilibrium. Thus, many explanations of behavior can be ruled out and need not be analyzed explicitly if it can be shown that economic agents in the same setting can do better for themselves. In statistical analysis, the reduction is through sufficiency—if the mass of data can be decomposed so that only a portion of it is relevant, the inference problem can be reduced to analysis of only the relevant data.

Stochastic models are important in settings in which agents make choices under uncertainty or, from a completely different point of view, in models that do not try to capture all the features of agents' behavior. Stochastic models are also important for models involving measurement error or approximations. These models provide a strong link between theoretical modeling and estimation. Essentially, a stochastic model delivers a probability distribution for observables. This distribution can serve as a natural basis for inference. In static models, the assumption of optimization, in particular of expected utility maximization, has essentially become universal. Methods for studying expected utility maximization in dynamic models are more difficult, but conceptually identical, and modeling and inference methods for these models are developing rapidly.

1.1 EXPECTED UTILITY THEORY

Expected utility maximization is a widely accepted hypothesis and a widely applied tool in economic theory. A behavioral model in which the activities of an economic agent are analyzed in a setting of uncertainty would naturally be analyzed in an expected utility framework. There are two reasons for this widespread acceptance. First, the axiom systems that can be used to produce utility representations of preferences and subjective probability distributions over events are compelling. Second, pragmatically, without a universal principle to follow in constructing models, there are too many candidate ad hoc models to make for a productive research program. The optimization principle is the principle that has been widely accepted.

The key reference on expected utility theory is L.J. Savage (1954), and it is this set of axioms that we shall sketch. Pratt, Raiffa, and Schlaifer (1964) give a simple but rigorous development in the finite case. DeGroot (1970) provides an insightful textbook discussion. We stress again that our presentation is from Savage. Kiefer and Nyarko (1995) argue for the wide use of expected utility theory in dynamic economics, specifically in the study of the economics of information and learning.

A state s is a complete description of all items over which there is uncertainty. The world is the set of all states S. An event is a collection of states. The true state is the state that pertains. There is a set of consequences Q. An action f is a mapping from the set of states into the set of consequences, $f\colon S \to Q$. In particular, an action specifies the consequence in each state. Let A denote the set of all actions.

Agents have preferences $\preceq$ over actions. The strict preference $\prec$ is defined in the usual manner: $f \prec g$ if $f \preceq g$ but not $g \preceq f$. The axioms placed on the preferences are

P1. The preference order $\preceq$ is complete (for all f and $g \in A$, either $f \preceq g$ or $g \preceq f$) and transitive (for all $f, g, h \in A$, $f \preceq g$ and $g \preceq h$ implies $f \preceq h$). A complete transitive ordering is said to be a weak, or simple, order.

P2. $\preceq$ obeys the sure-thing principle; i.e., let f, f', g, and g' be acts and let B be a subset of S such that

(i) on B^c (the complement of B in S), $f = g$ and $f' = g'$;

(ii) on B, $f = f'$ and $g = g'$; and

(iii) $f \preceq g$.

Then $f' \preceq g'$.

To interpret P2, suppose acts g and f agree outside of the set B and g is (weakly) preferred to f. Modify f and g outside of B, but ensure that they are still the same outside of B, and maintain their values on B. Then the modified g is (weakly) preferred to the modified f.

We say that action g is weakly preferred to action f given the event B if, when g and f are modified outside of B so that they are the same outside of B, then the modified g is weakly preferred to the modified f regardless of how the modification outside of B is done. Under the sure-thing principle the manner of modification outside of B is irrelevant. An event B is said to be a null event if for all acts f and g, $f \preceq g$ given B. As an example, suppose you are ranking two candidates for elective office. They agree on tax policy. Then your ranking does not depend on the tax policy on which they agree.

P3. Let f and g be constant actions (i.e., independent of s) and let B be an event that is not null. Then $f \preceq g$ given B if and only if $f \preceq g$.

P4. Let f, f', g, and g' be constant actions such that $f' \prec f$ and $g' \prec g$ (so that we may think of f and g as "good" constant actions and f' and g' as "bad"constant actions). Let B and C be any two events. Define the act f_B to be equal to f on B and f' outside of B and define g_B to be g on B and g' outside of B. Define f_C and g_C analogously, with C replacing B. Then $f_B \preceq f_C$ implies $g_B \preceq g_C$.

Assumption P4 breaks out probabilities from utilities by asserting essentially that the ranking of events according to their probabilities does not depend on the payoff for getting the ranking right. That is, the choice of which side one will take in a bet does not depend on the payoff. With our constant actions f and f' with $f' \prec f$, let $f_A = f$ on A and f' on A^c and let $f_B = f$ on B and f' on B^c. Then Savage defines A as not more probable than B iff $f_A \preceq f_B$.

P5. There exists a pair of consequences or constant actions f and f' such that $f' \prec f$.

Assumption P5 seems genuinely innocuous in any nontrivial application: It says that there are at least two actions that are strictly ranked.

P6. (Continuity Axiom) Let g and h be two actions with $g \prec h$ and let f be any consequence (or constant action). Then there exists a partition of S such that if g or h is modified to become g' or h' with the modification taking place on only one element of the partition and being equal to f there with the other values being unchanged, then $g' \prec h'$.

Assumption P6 essentially requires a continuous state space. Alternative assumptions deliver Theorem 1 for the discrete state space case (Pratt, Raiffa, and Schlaifer, 1964, and Gul, 1992).

P7. Let f and g be any two actions and let B be an event. Let $g(s)$ denote the constant action equal to the consequence $g(s)$ regardless of the state. Then $f \preceq g(s)$ given B for all $s \in B$ implies $f \preceq g$ given B.

Assumption P7 allows the result to apply to actions having infinitely many consequences (e.g., fairly general functions of a continuous state). The result of Savage (1954) is

Theorem 1 (*Savage*) *If $\preceq$ obeys axioms P1–P7, then there exists a* (*utility*) *function $u : Q \to \mathbb{R}$ and a* (*prior*) *probability measure P over S such that for all acts f and g,*

$$f \preceq g \textit{ if and only if } \int u\left(f\left(s\right)\right) P\left(ds\right) \leq \int u\left(g\left(s\right)\right) P\left(ds\right).$$

The lesson of theorem 1 is that expected utility maximization need not be a primitive assumption. The result that preference systems can be represented by expected utilities follows from axioms on preferences. These axioms can be considered separately; each appears fairly weak, but taken together they have very strong implications. Variations on the axioms suffice (Fishburn, 1970). Note that there is nothing in the axiom system restricting its applicability to the dynamic case (this assertion does require some argument; Kreps and Porteus (1979) give details and proofs). The logic compelling expected utility maximization modeling in static economic models applies equally to dynamic models.

1.2 UNCERTAINTY AVERSION, ELLSBERG AND ALLAIS

The Savage axioms are individually compelling, and the implications are strong and useful for economic modeling. Of course, the system is itself a model of rational decision making, and therefore its axioms and hence implications cannot be expected to hold for every individual or for every situation. There are, however, serious objections that became apparent early, which are also compelling and which indicate that the area, even now, remains unsettled. The principal objection is the possibility of uncertainty aversion, as illustrated in examples by Ellsberg (1961) and Allais (1953). We consider two versions of Ellsberg's example. In the first, you see two urns, urn *I* containing 100 red and black balls in an unknown proportion and urn *II* containing 50 red and 50 black balls. You will win \$100 if your chosen event occurs. The events being compared are red from urn *I*, *RI*, black from urn *I*, *BI*, etc. Typical preferences are $RI = BI$, $RII = BII$ (perhaps these make sense), $RI \prec RII$, and $BI \prec BII$. It is impossible to find probabilities consistent with these rankings, since if *RI* is regarded as less likely than *RII*, then *BI* must be more likely than *BII*, as the red and black probabilities must sum to 1 for each urn. This pattern of choices violates either the complete ordering assumption (P1) or the sure-thing principle (P2).

A variation on this example shows the crucial role of the sure-thing principle. Here, there is one urn containing 30 red and 60 black and yellow balls in an unknown proportion. One ball is chosen at random. Action *I* pays \$100 if a red ball is drawn, \$0 otherwise. Action *II* pays \$100 if a black ball is drawn, \$0 otherwise. Do you prefer *I* to *II* (a frequent preference)? Now consider action *III*, which pays \$100 if a red or yellow ball is drawn, \$0 otherwise, and action *IV*, which pays \$100 if a black or yellow ball is drawn, \$0 otherwise. Frequently, action *IV* is preferred to action *III*. The problem seems to be that the payoff probabilities are known in *I* and *IV*, and these are therefore preferred because of an aversion to uncertainty. However, P2, the sure-thing principle, requires that the ranking between *I* and *II* be the same as the ranking between *III* and *IV* since these comparisons differ only in the payoff in the event that a yellow ball is drawn (B^c in the notation of P2) and these payoffs are the same in *I* and *II* and the same in *III* and *IV*.

The Allais paradox is a little different, in that all the probabilities are specified. Here, suppose you have a choice between action *I*, winning \$1000 with probability .33, \$900 with probability .66, and \$0 with probability .01, and action *II*, which is surely \$900. Many will choose action *II*, in which case the chance of winning nothing is zero. Now consider action *III*, which pays \$1000 with probability .33 and \$0 with probability .67, and action *IV*, which pays \$900 with probability .34 and \$0 with probability .66. Many find it appealing to choose action *III* over action *IV*. These choices are contradictory, in that there is no expected utility rationalization supporting them. Note that there are really three events, occurring with probabilities .33, .66, and .01. Actions *I* and *II* differ from *III* and *IV* only in the payoff for the probability .66 event, and these payoffs are the same for *I* and *II* and the same for *III* and *IV*. Thus, by the sure-thing principle, *I* and *II* should have the same ranking as *III* and *IV*. The Allais paradox illustrates that, for many subjects, the change in probability from .67 to .66 is quite different from the change from .01 to .00. There appears to be a "bonus" associated with achieving certainty. This violates the sure-thing principle.

Efforts to extend the theory to accommodate uncertainty aversion are continuing. At present, the theory of rational decision making via expected utility maximization continues to provide the dominant framework for modeling individual behavior. In applications, it is typically more productive to handle apparent deviations from optimizing behavior by looking for misspecification of preferences or constraints, rather than by searching for alternative predictive theories.

1.3 STRUCTURAL VERSUS REDUCED-FORM METHODS

The distinction between stochastic and deterministic models is a distinction in modeling approaches. A distinction can also be made between the structural and reduced-form approaches. This distinction is concerned more with specification (parametrization) and estimation than with modeling. The distinction is not a sharp dichotomy. Indeed, outside of the linear simultaneous equations framework, "structural" and "reduced form" are typically undefined terms used for praise and criticism, respectively.

A notion that can be sensibly associated with the term *structural* is that of an equation that stands alone, that makes sense by itself, and that has a certain autonomy. In a market system, for example, one can consider the demand equation singly, considering the optimizing behavior of a consumer. This will lead to an equation giving quantity demanded as a function of price and income. Similarly, the supply equation has a life of its own—it can be determined by considering the behavior of a profit-maximizing producer and will give a quantity supplied as a function of price and other factors. These equations are structural in that each can be studied on its own. Adding the equilibrium condition leads to a complete system with two variables determined in the model: price and quantity. The *reduced form* consists of an equation for each of these variables. It is difficult to generate economic insight by considering either of these equations separately. This distinction between structural and reduced-form equations can be made precise in the linear simultaneous equations framework. The general notion that structural equations are autonomous equations and reduced-form equations are not is the position taken by Goldberger (*Econometric Theory* interview, 1989).

It is also common to refer to parameters, not equations, as structural. Structural parameters are the parameters of structural equations if the notion is to make any sense at all. Roughly, structural parameters are parameters that have economic interpretations on their own; reduced-form parameters are each a mishmash of several structural parameters. Unfortunately, reduced-form parameters sometimes do have economic interpretations of sorts (e.g., in the above market system, one reduced-form parameter is the equilibrium effect of income on the market price), and hence one econometrician's structural parameter is another's reduced-form parameter.

We take the view that economic theory and statistics are both guides for organizing data and reducing their dimension. From this point of view the distinction between the structural and reduced-form approaches is not sharp. Structural models rely more on theory for the reduction; reduced-form models rely more on statistics.

In one important class of models the sharp distinction can be retained. If the data distribution is in the exponential family class (see the appendix), then a fixed-dimension sufficient statistic exists and the natural parametrization in the linear exponential family form is the reduced-form model. Typically, a structural model would specify a lower-dimensional parametrization imposing restrictions on the reduced form. This curved exponential model is the structural model.

1.4 EXERCISES

1. Verify that Ellsberg's second example and Allais' example violate the sure-thing principle. *Hint:* Construct a table with columns corresponding to states of the world, rows corresponding to actions, and entries giving payoffs.

2. Merton (1973a, 1980) introduces and studies a consumption-based intertemporal capital asset pricing model. Preferences are characterized by an objective function of the form $E(\int_0^H u(c_t)dt)$, where c_t is the rate of consumption at time t, $u(\cdot)$ is a utility function, and H is the planning horizon. The particular case $u(c) = (c^{1-\gamma} - 1)/(1 - \gamma)$ implies that the rate of relative risk aversion (RRA) $-cu_{cc}/u_c$ (subscripts indicate differentiation) is constant at level γ. In the case of constant investment opportunities and constant RRA, the model implies that the excess return on the market portfolio consisting of all risky assets above the riskless interest rate is serially independent and normally distributed with mean proportional to variance, and the factor of proportionality is $\gamma/2$. In particular, if r denotes the continuously compounded riskless rate and $M(t)$ is the market capitalization of all risky assets at time t, then the market excess returns $r_t = \log M(t) - \log M(t-1) - r$, $t = 1, \ldots, T$, are independent and identically distributed (i.i.d.) $N((\gamma/2)\sigma^2, \sigma^2)$, the normal. Assume you are given data $\{r_t\}_{t=1}^T$. Derive the likelihood function for the unknown parameter $\theta = (\gamma, \sigma^2)$, the maximum likelihood estimator (MLE), and its exact and asymptotic distributions. *Hint:* For the exact distribution, note that the MLEs $\hat{\mu}$ and $\hat{\sigma}^2$ in the model $N(\mu, \sigma^2)$ are independent with known distributions involving the normal and central χ^2 distributions. Now use the change-of-variables formula.

3. Hakansson (1971) studies the growth optimal portfolio. Investors choose this portfolio if preferences are given by the Bernoulli logarithmic specification $\gamma = 1$. Derive the likelihood ratio (LR) test of the hypothesis H_0: $\gamma = 1$, and its exact and asymptotic distributions. *Hint:* For the exact distribution, note that $\Sigma_{t=1}^T r_t^2/T = \hat{\sigma}^2 + \hat{\mu}^2$. Show a formula for the density of the LR test based on the change-of-variables formula and marginalization by integration.

4. Derive the MLE of σ^2 under H_0 and its exact and asymptotic distribution. *Hint:* Note that $\Sigma_{t=1}^T r_t^2$ has a noncentral χ^2 distribution.

5. Show that the model under the alternative is a steep (indeed, regular) exponential family (see the appendix) and interpret the model under the null as a curved exponential family. Can the model under the null be rewritten as a regular exponential family with a lower-dimensional canonical statistic? Interpret your results.

1.5 REFERENCES

Utility theory is an enormous field, barely sketched here. Economists are familiar with the path-breaking work of Debreu (1959), giving conditions on preferences that lead to the existence of a utility function. Von Neumann and Morgenstern (1944) give an expected utility representation for decision making given probabilities over states. The fundamental work that establishes a foundation for decision theory is Savage (1954). It provides an axiomatic development that implies the existence of utility functions and subjective probability distributions. Pratt, Raiffa, and Schlaifer (1964) give an accessible but rigorous development for the discrete case. Knight (1921) was an early advocate of the distinction between risk and uncertainty, which many consider artificial (Arrow, 1951). The Ellsberg and Allais paradoxes have given rise to a large literature. Machina (1982) surveys the implications of changing the postulates. Gilboa and Schmeidler (1989) provide a framework for handling uncertainty aversion, essentially by considering ranges of priors. An empirical approach is taken in the work by Hansen and Sargent (2001, 2003). This remains an active research area. Halpern (2003) discusses alternative formal systems for representing uncertainty and updating beliefs. Dreze (1972), in his presidential address to the Econometric Society, reviews the connections between decision theory and econometrics. Classical econometrics textbooks (e.g., Theil (1971), Goldberger (1991)) give informal and very useful conceptual discussions of the distinction between the structural and reduced-form approaches, and Fisher (1966) and Rothenberg (1971) present technical discussions that generalize to nonlinear models.