Experiments and Econometrics

Daniel E. Houser

"Experimetrics" refers to formal procedures used in designed investigations of economic hypotheses. Fundamental experimetric contributions by Ronald A. Fisher provided the foundation for a rich literature informing the design and analysis of economics experiments. Key components of this foundation include the concepts of randomization, independence and blocking. Experimetric analysis plays a central role in advancing economic models, and will gain further importance as scholars adopt increasingly sophisticated designed research programs to illuminate positive economic theory.

1. Introduction

"Experimetrics" refers to formal procedures used in designed investigations of economic hypotheses. A series of pathbreaking experimetric contributions by Ronald A. Fisher, written largely during the 1920s and early 1930s, elucidated fundamental concepts in the design and analysis of experiments (see, e.g., Box, 1980 for a survey). He was first to obtain rigorous experimetric results on the importance of randomization, independence and blocking, and he created many powerful analysis tools that remain widely used, including Fisher's non-parametric Exact Test (Fisher, 1926; see also Fisher, 1935).

Controlled experiments allow compelling scientific inferences with respect to hypotheses of interest. Many economic experiments inform hypotheses regarding primitives assumed to be constant within an experiment (e.g., preferences or decision strategies), or the effects on economic outcomes of changes in institutions (e.g.,

comparing different auction rules or unemployment regulations, see Vernon Smith entry on Experimental Economics). One conducts controlled experiments to inform economic hypotheses because relevant naturally occurring data typically include noise of unknown form and magnitude outside the investigator's control. Econometric procedures can go some distance towards solving this problem, but even sophisticated approaches often allow only limited conclusions.

For example, suppose one wanted to investigate the (causal) effect of caffeine on heart rhythms. One approach is to obtain a random sample of "heavy" coffee drinkers and compare them to a random sample of people who do not use caffeine. Because it is not possible with naturally occurring data to control the reason a person falls into a category, discovering that people with greater caffeine consumption have more cardiac episodes need not imply a causal caffeine effect. The reason is that a preference for coffee may stem from a biological characteristic that is itself causally tied to irregular cardiac events.

An advantage of designed investigations is that they allow cogent inference regarding causal effects through the appropriate use of randomization, independence and blocking.

1.1. Randomization. Experiments with randomized designs allow compelling causal inference. The reason is that randomly assigning participants to treatments, and randomly assigning treatments to dates and times, minimizes the possibility of systematic error. In the caffeine example, intentionally assigning heavy caffeine drinkers exclusively to a caffeine treatment generates a systematic error and invalidates causal inference.

However, an experiment where subjects are randomly assigned to caffeine and no-caffeine treatments independent of their typical caffeine use allows one to draw appropriate inferences regarding causal relationships.

1.2. Independence. Randomization also helps to ensure independence both within and between treatments' observations. Loosely speaking, observations are independent if information about one does not provide information about another. Independence is critical for many experimetric analyses, and its failure can lead to misleading conclusions. An objective randomization procedure for treatment assignments insures against the possibility that participants in one treatment might unintentionally systematically vary from other treatments' participants.

1.3. Blocking. Causal relationships can be assessed with greater precision through "blocking." Blocking is a design procedure with which an experimenter can separate treatment effects from nuisance sources of data variation. In the above, heart rhythms might be affected by both caffeine and anxiety over the process of measuring heart rhythms. Especially because it is expected to differ between participants, anxiety is a source of nuisance variation that clouds inferences regarding caffeine effects. To address this one could "block" by participant. This involves measuring each subject both with and without caffeine (in separate, randomly ordered trials.) Caffeine effects are measured as the difference between trials, thus mitigating noise due to individual anxiety effects.

2. Experimetrics Toolbox

Although many specialized experimetric tools have been developed, the experimetrics toolbox also includes a large number of general purpose procedures that have become standard in the experimental economics literature. A regular concern is that independence is not satisfied. The failure of independence can occur because of "session" effects, meaning that there is less behavioral variation within than between sessions. Violations of independence can also occur if repeated measurements are taken on the same individual due to individual effects. Standard procedures can address this. Sessions can be treated as fixed effects, and random effects can be used to control for individual differences. The resulting "mixed effect" model can be analyzed using standard parametric, panel-data procedures (see, e.g., Frechette, 2005).

Also in the toolbox is the McKelvey and Palfrey (1995) "Quantal Response Equilibrium" (QRE) framework (see entry on Quantal Response Equilibrium). QRE is a parametric procedure for analyzing data from finite games. The key idea is to incorporate errors into players' best response functions, thus creating "quantal response" functions. This results in an extremely flexible model that can rationalize a wide variety of behaviors. Haile, et. al., (2006) point out that this flexibility comes with a cost: in general, QRE can rationalize any distribution of behavior in any normal form game, and imposes no falsifiable restrictions without additional assumptions on the stochastic components of the model. Thus, those who wish to implement QRE analyses face the experimetric challenge of creating designs within which such assumptions are defensible.

For reasons including sample size and robustness, the experimetrics toolbox includes many nonparametric procedures (see Siegel and Castellan, 1988, for a user-

friendly textbook treatment of popular nonparametric approaches). For example, Mann-Whitney tests, and their k-sample generalization due to Jonckheere (1954), are frequently used to compare medians among treatments' data. Also common is Fisher's Exact Test, which uses all the information in the data and is the most powerful non-parametric approach to inference with respect to differences among treatments. Its use is limited by the fact that it can be computationally cumbersome to implement when the numbers of treatments or observations are large.

3. External Validity

An experiment's conclusions are "externally valid" if they can be extrapolated to other environments. To rigorously address external validity requires that the source of treatment effects can be identified, which in turn implies a fundamental rule of experiment design: Within any good experiment, any treatment can be matched with another that differs from it in exactly one way.

External validity is both important and subtle. For example, consider the well-known "dictator game" where one participant is assigned the role of "dictator", and the other "receiver" (see entry on Dictator games). The dictator is given $20, and the receiver nothing. The dictator is told to split the $20 between herself and her receiver in any way she likes, after which the experiment ends. A widely replicated result is that a large fraction of dictators send half ($10) to an anonymous stranger, and one might question whether this finding is externally valid. In particular, there is no evidence that this behavior is prevalent among winners of naturally occurring lotteries.

There are clear similarities between the situations of lottery winners and dictators. Still, the fact that actions of dictators in laboratory games do not match actions of lottery winners does not necessarily mean that dictator games lack external validity. The reason is that identical decision strategies can imply different decisions in different environments. For example, recent research provides compelling evidence that dictators' decisions are tightly connected to their beliefs regarding the decisions of others who have faced this same situation: dictators give because they believe other dictators give (Bicchieri and Xiao, 2007). This mechanism plausibly guides decisions in naturally occurring environments. In particular, lottery winners do not give because they believe other lottery winners do not give large fractions of their winnings to anonymous strangers. Thus, external validity does not require that one be able to match actions in an experiment to actions in another environment. Rather, an experiment is externally valid if one can extrapolate to novel contexts its conclusions with respect to individual or strategic decision processes.

4. Applied Experimetrics Research

An important application of experimetrics is to discriminate between many competing theories of learning that have emerged. (See entry on Individual Learning in Games). Doing this includes significant experimetric challenges, as it requires one to account for heterogeneity in the way subjects learn. The reason is that not doing so will tend to bias fit statistics in favor of reinforcement (and hybrid) models. Wilcox (2006) shows the reason is that reinforcement models condition behavior on informative functions of past choices, and in the presence of learning heterogeneity these choices will carry

idiosyncratic parameter information not otherwise incorporated into the specification. Having said this, it is also the case that many data sets from typical learning experiments can be roughly equally well described by many different learning models (Salmon, 2001). Consequently, the "best" model can be highly sensitive to the particular criterion one uses for model selection, as well as the particular experiment under consideration (Feltovich, 2000). As a result, in-sample fit is often good, but this does not necessarily imply that much has been learned about the way in which people actually learn and make choices (Salmon, 2001).

Knowing how people make choices is critical to advance both economic theory and institution design. Consequently, a significant experimetric literature explores how people make decisions in complex environments, with a focus on characterizing the nature and number of different "decision rules" at use in a population. Most approaches to accomplishing this require pre-specifying the decision rules the researcher believes people could follow, and then using choice data to assign one of those rules to each member of the population (see, e.g., El-Gamal and Grether, 1995). However, in some cases one might be unwilling or unable to pre-specify the decision rules, and it turns out that doing so is not necessary. In particular, Houser et. al. (2004) details a Bayesian experimetric procedure that uses individual choice data to determine endogenously the nature and number of decision rules in a population. The approach requires only that one specify the information relevant to individuals' decisions.

Substantive experimetric advances have been obtained in far too many areas to detail here. Although no general survey is available, Houser et. al., 2004, Ashley et. al.,

2005, and Loomes, 2005, include excellent summaries of experimetric contributions to a variety of widely-studied games and decision problems.

5. Conclusion

Experimetrics continues to evolve as scholars adopt highly sophisticated design and analysis procedures to inform new questions. A ready example is the rapidly expanding research in Neuroeconomics (see entry on Neuroeconomics). The massive spatial-panel data structure that characterizes brain images poses unique inferential problems. Progress on these problems requires significant complementary innovations to both design and analysis strategies. The resulting experimetric advances are sure to have significant impact on economic theory and policy analysis.

References

Ashley, R., Ball, S. and Eckel, C. 2005. Motives for giving. A reanalysis of two classic public goods experiments. Manuscript, Virginia Institute of Technology.

Bicchieri, C., and Xiao, E. 2007. Do the right thing: But only if others do. Manuscript, University of Pennsylvania.

Box, J. F. 1980. R.A. Fisher and the design of experiments, 1922-1926. The American Statistician, 34(1), 1-7.

Dickhaut, J. 2007. Neuroeconomics. New Palgrave Dictionary of Economics and Law.

Feltovich, N. 2000. Reinforcement-based vs. belief-based learning models in experimental asymmetric-information games. Econometrica. 68, 605-641.

Fisher, R.A. 1926. The arrangement of field experiments. Journal of the Ministry of Agriculture of Great Britain, 33, 503-13.

Fisher, R.A. 1935. Design of Experiments. Edinburgh: Oliver and Boyd.

Frechette, G. 2005. Session effects in the laboratory. Manuscript, New York University.

El-Gamal, M.A. and D.M. Grether. 1995. Are People Bayesian? Uncovering Behavioral Strategies. Journal of the American Statistical Association, 90, 1137–1145.

Goeree, J. K., C. A. Holt and T. R. Palfrey. 2007. Quantal Response Equilibrium. New Palgrave Dictionary of Economics and Law.

Haile, P.A., Hortacsu, A., and Kosenok, G. On the empirical content of quantal response equilibrium. Mimeo, Yale University.

Houser, D., Keane, M. and McCabe, K. 2004. Behavior in a dynamic decision problem: An analysis of experimental evidence using a Bayesian type classification algorithm. Econometrica. 72(3), 781-822.

Jonckheere A.R. 1954. A distribution-free k-sample test against ordered alternatives. Biometrika 41: 133-145.

Loomes, G. 2005. Modelling the Stochastic Component of Behavior in Experiments: Some Issues for the Interpretation of Data. Experimental Economics 8, 301-323.

McKelvey, R. D. and Palfrey, T.R. 1995. Quantal Response Equilibria for Normal Form Games. Games and Economic Behavior 10, 6-38.

Salmon, T.C. 2001. An evaluation of econometric models of adaptive learning. Econometrica. 69(6), 1597-1628.

Siegel, S. and Castellan, N. Jr. 1988. Nonparametric Statistics for the Behavioral Sciences. 2nd edition. McGraw Hill, Boston.

Smith, V. 2007. Experimental Economics. New Palgrave Dictionary of Economics and Law.

Wilcox, N. 2006. Theories of learning in games and heterogeneity bias. Econometrica. 74(5), 1271-1292.