Problem Solving and Hypothesis Testing Using Economic Experiments

Darren Hudson

The roots and uses of economic experiments in problem solving and hypothesis testing are explored in the present article. The literature suggests that the primary advantage of economics experiments is the ability to use controlled stimuli to test economic hypotheses. Other literature also suggests that experiments are useful in problem solving settings. The advantages and disadvantages of experiments are discussed.

Key Words: experiments, problem solving, hypothesis testing

JEL Classifications: C81, C91, C92, C93

Economics is a science, and, like any other field of science, is primarily interested in the creation of knowledge. There are a variety of types of knowledge and a variety of means of creating knowledge (Ethridge). However, science is most concerned with knowledge creation through the systematic development and testing of hypotheses. It is in this vein that economics draws the most criticism from the “hard” sciences.

Economics has witnessed a tremendous amount of analytical development in the past 100 years, moving it beyond a field of conjecture to more rigorous, systematic testing. Even Nobel Prize–winning physicist Richard Feynman stated that, of the social sciences, economics is by far the most advanced. Yet there remains a common criticism that economics lacks the rigor to classify it as a hard science. Yes, economists use mathematically rigorous techniques to develop theory. Yes, there is empirical testing of that theory. However, empirical tests are derived from data that embody many complex relationships that are often confounded, forcing the researcher to adopt a ceteris paribus assumption that may not be justified.

More pragmatically, the world we live in is changing rapidly. Concentration and consolidation in agriculture are occurring at a feverish pace (Drabenstott). New technologies are being released into the market quickly, and the desire of food firms to differentiate products is increasing (Lusk and Hudson 2002a). In addition, policymakers desire information about new product labeling and food safety rules. All of these factors point to a glaring problem in economics: the data we need for analysis and policy prescriptions often either do not exist or are not timely enough to be useful. This data problem significantly reduces the usefulness of any results obtained and diminishes the importance of economists in problem solving and policy formulation.

The growing field of experimental econom-
ics offers a means to counter both the criticisms about the rigor of economics and the data problems we face. By using known stimuli, controlled environments, and proper experiment design, researchers can isolate effects and test specific hypotheses. Experimental methods can be used to generate data in situations where no data exist (e.g., a demand function for a new product that is not currently on the market) and can also be used to test responses to potential public policy programs. Given these opportunities, the present article seeks to explore the advantages and disadvantages of experimental methods in economics. However, this article should only be viewed as an introduction to these concepts, because there is a vast literature on economic experiments.

Why Do Experiments?

There are a number of prominent reasons why economists choose to perform economic experiments (Davis and Holt). Smith (1994) provides a list of seven primary reasons. Without listing the justifications for each, they are:

1. Test a theory or discriminate between theories,
2. Explore the reasons for a theory’s failure,
3. Establish empirical regularities as a basis for new theory,
4. Compare environments,
5. Compare institutions,
6. Evaluate policy proposals, and
7. Use the laboratory as a testing ground for institutional design.

Reasons 1–3 are often used in disciplinary research in an attempt to affect theoretical developments in economics, although problemsolving research may be interested in theory tests as well. Reasons 4–5 are often combinations of disciplinary and problem-solving applications (see Lusk et al., and Ruström for recent examples using auctions). Reasons 6–7 are most closely associated with problem-solving research.

Roth provides three classifications based on the motivation for conducting an experiment: (1) speaking to theorists, (2) searching for facts, and (3) whispering in the ears of princes. Speaking to theorists involves reasons 1–2 on Smith’s (1994) list and focuses primarily on theory testing. Searching for facts is most closely associated with reason 3 on the list, although comparing environments and institutions is also a fact searching process. Finally, whispering in the ears of princes involves policy evaluation and institutional design. This could also be extended to providing recommendations to businesses about the impact of product introductions, for example, on demand and/or profitability.

As applied economists, we are most often concerned with the latter of these motivations. Problem-solving and policy prescriptions demand a considerable amount of our attention and research time. However, one should not dismiss the other motivations as well. Experiments provide a convenient means to test theory and search for empirical regularities that can serve as the basis for theory development. These tasks, while seemingly irrelevant for problem-solving and policy research, can enrich the quality of the research and increase the publishability of results. More generally, experiments provide a means to control for confounded factors that exist in market data (if it exists) and more rigorously test premises, whether they be theoretical or problem-solving oriented.

What Is Experimental Economics?

Experimental economics is a field in economics that attempts to test hypotheses about economic behavior in a controlled environment. It is not a new field, having begun in earnest in the 1940s and 1950s. Chamberlin became interested in recreating a market environment in the classroom (Davis and Holt). Since that time, experimental economics has grown rap-

---

1 Roth suggests that perhaps a first example of experiments can be attributed to Bernoulli in 1738. In that paper, Bernoulli elicited opinions of experts about the St. Petersburg Paradox. This informal method is somewhat like hypothetical elicitation techniques used today.
idly. Now, experimental papers are widely published in all major journals, and this field even has its own journal (*Experimental Economics*), which is published by the Economic Sciences Association.

At its heart, experimental economics is an amalgamation of economics and experimental psychology that closely resembles experiments in the hard sciences. One starts with (or derives) a theoretical prediction or hypothesis. Then, the researcher develops an experiment to test that hypothesis. Next, an appropriate subject group is identified and the experiment conducted. Finally, data are analyzed and conclusions are drawn. More specifically, Smith (2001) says:

"Quite generally, we can think of experimental outcomes (the observed replicable order in the final allocations) as the consequence of individual choice behavior, driven by the economic environment and mediated by the language and rules governing interactions supplied by the institution."

In this, there are three key elements: environment, institution, and behavior (Smith 1994). The environment defines the means of exchange, initial endowments, preference rewards, etc. Every economic situation, whether in the laboratory or the real world, has an environment. How the environment is defined will affect the experimental outcome. Second, institutional factors, or the rules of the game, are important as well. The institution determines how messages are sent between actors (bids, offers, acceptances, etc.) as well as how information can be exchanged and how binding contracts are established. Finally, there is behavior. Behavior is a function of both the environment and institutions, and behavior is the experimental outcome alluded to by Smith (2001).

The term "experimental economics" is defined a number of ways. I use a broad definition here but will focus on different segments later in the article. A general description can be seen in Figure 1. Generally, experiments can be divided into hypothetical and nonhypothetical experiments, although there is considerable overlap (for a discussion of different methods in the context of willingness to pay, see Lusk and Hudson 2002a).

"Hypothetical" in this context means that respondents are not expected to pay, nor is remuneration provided on the basis of the experiment. That is, participation fees may be paid or respondents may get some intrinsic value from participation, but payments are not based on the outcome of the experiment. Some authors argue that the hypothetical nature of these experiments yields hypothetical bias (Cummings, Harrison, and Rutström; Cummings and Taylor; Fox et al.; List and Gallet; Fox et al.; List and Gallet;)

---

2 It is important to note that the methods listed under hypothetical in Figure 1 can also be used in a non-hypothetical setting (Lusk; Lusk and Hudson 2002a). This distinction is made here because these methods are most commonly used in a hypothetical setting.
List and Shogren 1998, 1999; List; Loomis et al.; Neill et al.; Rutström). That is, because the respondent is not actually responsible for their choice, they may respond in manner different than if they were actually held accountable for choices.

By contrast, participants are expected to face the consequences of their choices in non-hypothetical experiments. For example, in an auction, participants are asked to place a bid on a product. If theirs is the winning bid according to the auction mechanism, they are expected to pay the bid price (or winning price). If the elicitation mechanism is incentive-compatible, the individual's dominant strategy is to truthfully reveal their preference for the good in question. The nonhypothetical nature of these experiments is intended to induce incentive compatibility, thus revealing true demand and mitigating hypothetical bias.

We, as agricultural economists, are probably most familiar with hypothetical experiments and, more specifically, contingent valuation methods. Contingent valuation methods are popular methods of eliciting willingness to pay, especially in environmental applications (Hanemann; Hite, Hudson, and Intarapapong; Hudson and Hite). There are a variety of contingent techniques, most of which have been well documented and tested in the literature. Perhaps less well-known are choice-based methods, often called conjoint models (Adamowicz et al., 1997, 1998; Hudson and Lusk 2002a; Louviere, Hensher and Swait; Lusk, Roson, and Fox). This method attempts to elicit marginal utilities for attributes of products in a controlled experimental design. Choice-based methods have several advantages over traditional contingent valuation techniques, including the fact that it is based on random utility theory, which is consistent with Lancaster’s theory of utility maximization (Louviere, Hensher, and Swait), is more consistent with realistic choices made by respondents on a daily basis (Lusk and Hudson 2002a), and has been shown to be consistent with revealed preferences that respondents actually exhibit (Adamowicz et al., 1997; Carlsson and Martinsson).

The second arm in Figure 1 is probably less familiar to most agricultural economists: non-hypothetical experiments. If one were conversing in a crowd of general economists, this would more likely be the visualization conjured up by the phrase “experimental economics.” There are two general types of experiments: induced value and “homegrown” value, which differ in the objective of the experiment.

Our training likely leads us to think in terms of homegrown values, which are the values that respondents take with them into the experiment. We, as experimenters, have no control over homegrown values; they are implicit in the respondent’s experience, preferences, upbringing, etc. Rather than control for these values, we are often interested in eliciting these values and attempting to explain why these values exist on the basis of sociodemographic variables or other latent variable proxies for a person’s implicit preference ordering.

By contrast, induced-value experiments are not interested in homegrown values. Rather, induced-value experiments attempt to completely control for valuation by carefully and completely (as much as possible) defining both the environment and institutions of an experiment. By providing payoff possibilities, exchange mechanisms, and experimental contexts (environments), the experimenter can isolate whether respondents behave consistently with theory. In fact, induced-value experimenters would argue that, by allowing homegrown values to enter the analysis, one cannot effectively control for all effects and, thus, loses control of the experiment. Without taking a position on this argument, it is interesting to note that there is a tremendous amount of experiments in the literature that are based on homegrown values.

Under each of these in Figure 1, there are

---

3 To some extent, induced value experiments are what many people define as experimental economics. Traditional experimenters operate almost exclusively in the world of induced value.

4 Smith (2001) notes that one cannot completely control for homegrown values in an experiment. However, if the experiment is carefully designed in an induced-value setting, homegrown values may be a means of explaining theory failure.
the same three basic types of experiments. The first are market-based experiments, the most famous of which are Chamberlin's discussion of imperfect markets and Smith's (1962, 1964) double-auction institution. Since that time, market-based experiments have been used numerous times to test the robustness of competitive price theory under alternative institutions and exchange mechanisms (Menkhau et al.) and remains a popular method of illustrating market mechanisms to students (Holt). Second, there are experiments with games (or bargaining experiments). This branch of literature began in the 1950s with interest in the "prisoner's dilemma" problem (Tucker) by psychologists, game theorists, and business economists (Davis and Holt). Based in game and bargaining theory, these experiments attempt to construct a situation that contains the payoffs and sequences of a bargaining relationship in which the strategic behavior between participants is of importance. Faced with this situation, the participants attempt to strategically optimize their payoffs. Experimenters observe participant behavior and compare it with derived theoretically optimal behavior (e.g., Lusk and Hudson 2002b). Game experiments have been applied widely in the industrial organization literature (e.g., Benson and Faminow; Brown-Kruse; Courrèse, Isaac, and Smith; Davis and Williams).

Finally, there are individual choice experiments, which are used in situations in which respondents only need to optimize their own behavior and strategic interaction is not important. These experiments are often used in the context of tests of expected utility maximization but have also been used in the context of testing the rationality of market forecasts (Williams). An example is the use of the iterated Becker, DeGroot, and Marschak procedure to elicit a subject's von Neumann-Morgenstern utility function.

To summarize, experimental economics is concerned with examining behavior in a systematic, controllable fashion. Viewed broadly, experimental economics embodies both hypothetical and nonhypothetical approaches. More narrowly, experimental economics is generally considered to be those approaches in the nonhypothetical realm that deal with both induced and homegrown values. Experiments can be viewed as being market based, game or bargaining based, or individual choice based.

Advantages and Disadvantages

As with any other laboratory science, experimental economics has the advantages of replicability and control (see Davis and Holt for a thorough treatment of replicability and control). Replicability in this context does not refer to the ability to reproduce the same results with the same data set, as we are accustomed to thinking about in econometrics. Rather, replicability refers to the ability of one researcher to reproduce the same experiment as another researcher, to verify results. In traditional econometric analysis of natural markets, data are recorded at particular intervals, and unobserved spatial and temporal factors are constantly changing and, therefore, nonreplicable.

Control refers to the ability to manipulate either the environment or institutions within an experiment to cleanly evaluate theoretical predictions. As was noted above, data from natural markets are confounded with many complex relationships that cannot be adequately controlled. Alternatively, data may simply not exist in a natural setting to adequately test a new theory or to gain insight from empirical regularities. Experiments allow researchers control over stimuli and allow for the isolation of particular effects of interest.

Before concluding that experimental economics is a panacea for empirical investigation, one should also note that there are weaknesses of experiments, some of which are more relevant than others. A typical criticism of experiments is that they often are conducted with student subject pools as opposed to more realistic subject pools drawn from relevant decision makers. For example, if one wishes to study the behavior of a particular trading institution in a futures market, would it not be beneficial to use actual market traders as opposed to students? Although this criticism may seem fatal on the surface, several studies have shown that the behavior exhibited by relevant
decision makers is not different than that of students (Dejong, Forsythe, and Uecker; Dyer, Kagel, and Levin; MacCrimmon and Wehrung; Mestelman and Feeny; Smith, Suchanek and Williams). Thus, although there may be concern about the appropriateness of the subject pool, concern about using students to analyze economic behavior appears to be unfounded. The advantage of student subject pools is that they are much less expensive to acquire. However, one should be careful of potential endowment effects (Binswanger; Kahneman, Knetsch, and Thaler) when using these types of pools.

A second common criticism is that we are using simple economic experiments to explain a complex economic world. Although this is true, it would appear, as Davis and Holt pointed out, that this is more a criticism of simple economic theories than of simple economic experiments. Experiments are constructed to be consistent with underlying economic theory. It is true that consistency between experimental results and theoretical predictions will not necessarily translate to the more complex real world, but this suggests that the theory is omitting critical variables that prevent it from effectively predicting reality. At the same time, if the theory does not perform well in an experimental setting, it is not likely to predict well in the real world either.

Despite these relative benign criticisms, there are more important issues to consider in experimental analysis. First, there is always a temptation to “oversell” the implications of experimental findings. As with any deductive/inductive process, there are premises, data collection, analysis, and conclusions inferred from the analysis. Attaching catchy phrases and interpretations to results and attempting to infer broad policy implications may increase attention to the results and generate future grant funding but will do little to further understanding about the underlying economic behavior. Savvy laypeople will see through these broad brush strokes, thus undermining the credibility of the experimental method.

Second, there are real technical impediments to conducting an economic experiment. Economic environments are necessarily complex, and capturing and controlling all of these variables can be quite difficult. For example, can someone effectively understand intertemporal substitution or infinite time horizons? Framing effects, endowment effects, and incentive compatibility are all issues that the experimenter must consider in experimental design. Simple flaws in experimental design can easily negate the credibility of results, limit the usefulness of the results in explaining economic behavior, and render the work unpublishable.

Example—Risk Premiums

As part of a larger study, we were interested in examining different elicitation methods of risk premium measures (see Hudson, Coble, and Lusk). One elicitation method used was a lottery auction (see Hudson and Lusk 2002b for a complete description of the data and methods). A data collection exercise was conducted with 29 agricultural producers in the Delta area of Mississippi. Among other activities, respondents participated in a lottery auction.

The lottery auction was conducted as follows. First, respondents were provided with an envelope and told that there was a 50/50 chance that the envelope contained $10 or nothing. Next, respondents were told that they would have the opportunity to “sell” their envelope back to the experimenter, but that the experimenter wanted to pay the lowest price possible for the envelope. The respondents were provided a bid sheet and asked to provide the lowest amount of money they would be willing to receive in exchange for their envelope. Using a Vickrey second-price auction mechanism, all bids were collected. The lowest bid was deemed the winner, and the win-

---

5 The envelopes were thick manila envelopes, so that the respondents could not see what was inside. Half of the envelopes actually contained a $10 bill. The other half contained a piece of paper that was cut to the same size as a $10 bill so that those feeling the envelope could not distinguish the difference. Respondents were informed of this so that they would not attempt to infer the probability of having $10 by the feel of the envelope.
n aing bidder would receive the second-lowest bid amount.

Respondents were told that there would be five rounds of bidding, with the winning bidder number and second-lowest bid amount posted after each round. Having multiple rounds with posted prices after each round introduced informational effects and allowed for respondent learning during the game (List and Shogren 1999). Respondents were also told that one round would be chosen at random to be "binding" and that the winner of that round would actually be paid the second-lowest bid amount for their envelope. The use of a random binding round mitigated strategic behavior and bidder affiliation (Lusk), because respondents did not know a priori which round would be binding.

The premise behind the lottery auction is simple. If respondents are risk averse, they should be willing to forgo some of the expected income to avoid the gamble. Given that the expected value of the lottery is $5, they should be willing to receive less than $5 to avoid having to face the lottery. Thus, in the lottery auction, they should be willing to "sell" their lottery for less than $5. If, however, they are risk seeking, their bid should be greater than $5.

The resulting average bids and 95% confidence intervals are shown in Figure 2. The results show that, in all rounds, the average bid was greater than $5, which suggests that this group of producers were risk seeking, on average. The average bid declined over the first three rounds. By the third round, results suggest that we cannot conclude a difference from $5 based on the 95% confidence interval. Generally, however, risk seeking appeared to be the average outcome, but there was variability in the bidding behavior. Using a normal cumulative distribution, these results suggest that, on average, 32% of the respondents were risk averse, and 68% were risk seeking.

The finding that a large percentage of producers were risk seeking is not inconsistent with prior empirical work (Pennings and Garcia). However, the construction of this experiment offers the opportunity to point out a couple of cautionary notes about conducting experiments. First, endowment effects can be important (Binswanger). That is, the size of the lottery payoff in this case was small relative to the income of the respondents. If the payoff is small enough relative to income, a result of risk-seeking behavior is perhaps not too surprising. Second, experiments are often context specific. That is, we were asking re-
spondents to operate in a world of lotteries where responses may be significantly different than if risk experiments are conducted in other contexts.\textsuperscript{6} Careful consideration must be given to the environment (or context) of the experiment, because environment will affect experimental results (Smith 2001).

Despite these limitations, the results can still be useful. The experiment was conducted in a hotel casino, which offered a unique opportunity to observe the respondent's behavior after the experiment was conducted. Hudson and Lusk (2002b) report the results of an analysis that compared the lottery auction results with postexperiment gambling behavior. We found that individual responses to the lottery auction significantly explained gambling behavior, with a larger average bid being significantly associated with a higher probability of postexperiment gambling. Given that gambling can be inferred as a risk-seeking activity, the results of the lottery auction were an effective predictor of risk-seeking behavior.

Example—A Market Experiment

Menkhaus et al., provided an example published in the Journal of a market experiment. Specifically, these authors were interested in investigating the effects of supply and demand risks on prices and quantities traded in an experimental market. The general hypothesis was that, under conditions of uncertainty, risk increases costs to both buyers and sellers in a market. Thus, in the presence of risk, we should observe higher prices and lower traded quantities than under the condition of certainty.

Menkhaus et al., used a double auction with a computerized trading mechanism. Although the double auction may not exactly resemble traditional trading institutions in agricultural markets, it is informationally rich, prevents tangential issues such as market power from entering the experiment, and can be easily used to isolate the risk factors investigated by the authors.

The double auction used four buyers and four sellers, which is standard in investigations of this type. Three replications were used for each of their risk scenarios plus a control group (resulting in 24 trading sessions). Trading in each session was conducted for 15 3-minute periods, which allowed for learning and a stabilization of behavior over periods. Buyers and sellers were both privately provided with redemption values for their trades (i.e., buyers received the resale values for the products purchased and sellers received costs of production) on a per-unit basis. Both parties were provided an initial endowment of tokens (100 tokens = $1). Earnings accumulated throughout the trading periods, and participants were paid the cash equivalent of their earnings at the end of the trading session.

This trading mechanism is standard and is often used in classroom experiments to demonstrate market mechanisms and equilibrium determination (Holt). However, in each of the treatments, Menkhaus et al., altered the environment by introducing either supply or demand risk and investigated the effects of these risks on price and traded quantities in both spot and forward markets.

The results of that study generally confirmed the hypothesis that supply and demand risk increases the price and decreases the quantities traded in equilibrium. More interesting, however, were their findings that the effect of the trading institution (spot vs. forward markets) had a greater impact on prices and traded quantities. Menkhaus et al., ascribed this finding to there being the added risk of product loss in forward markets. That is, when a contract is signed in advance of production, there is always the risk that the product will be lost before the transaction can be completed. This fact of the trading institution alters the equilibrium outcome relative to both the certain case and the case of supply and demand risk in spot markets. This result is an example of searching for empirical regularities alluded to by Smith (1994) and suggests that the structural changes taking place

\textsuperscript{6} In fact, Hudson, Coble, and Lusk found that there was little consistency in risk-premium measures across different contexts (yield, price, and the lottery). Lottery results were most consistent with open-ended questions in yield and price contexts.
in agriculture will alter equilibrium market outcomes.

Summary and Conclusions

Experimental economics affords a number of opportunities to economists in general as well as applied economists. The experimental method serves as a counter to the criticisms of economics as a science and opens the door to a rich world of controlled, rigorous tests of economic theory. At the same time, experiments can serve as a relatively untapped resource for generating data for situations where no data exist as well as data that are generally free from confounded effects in the real world.

However, experiments are not a problem-solving panacea. There are many important factors to consider when using experiments. Framing effects, endowment effects, and incentive compatibility in experimental design are all critical elements that can inhibit the usefulness of economic experiments. Carefully designed experiments can be a powerful tool in the arsenal of the applied economist, but they should be viewed as only one of the available tools. Experiments supplement, not substitute for, traditional econometric and statistical analysis of natural markets.

References


Etherege, D. *Research Methodology in Applied Economics: Organizing, Planning, and Con-


