

IZA DP No. 3273

## Field Experiments in Economics: Palgrave Entry

John A. List  
David Reiley

January 2008

# Field Experiments in Economics: Palgrave Entry

**John A. List**

*University of Chicago,  
NBER, RFF and IZA*

**David Reiley**

*University of Arizona*

Discussion Paper No. 3273  
January 2008

IZA

P.O. Box 7240  
53072 Bonn  
Germany

Phone: +49-228-3894-0

Fax: +49-228-3894-180

E-mail: [iza@iza.org](mailto:iza@iza.org)

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post World Net. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

## **ABSTRACT**

### **Field Experiments in Economics: Palgrave Entry**

Field experiments occupy a middle ground between laboratory experiments and naturally occurring field data. The idea is to perform a controlled experiment that captures important characteristics of the real world. Relative to traditional empirical economics, field experiments provide an advantage by creating exogenous variation in the variables of interest, allowing us to establish causality rather than mere correlation. Relative to a laboratory experiment, a field experiment gives up some of the control that a laboratory experimenter may have over her environment in exchange for increased realism.

JEL Classification: C72

Keywords: field experiments

Corresponding author:

John A. List  
Department of Economics  
University of Chicago  
1126 E. 59th Street  
Chicago, IL 60637  
USA  
E-mail: [jlist@uchicago.edu](mailto:jlist@uchicago.edu)

## Field Experiments in Economics: Palgrave Entry

Field experiments occupy a middle ground between laboratory experiments and naturally occurring field data. The idea is to perform a controlled experiment that captures important characteristics of the real world. Relative to traditional empirical economics, field experiments provide an advantage by creating exogenous variation in the variables of interest, allowing us to establish causality rather than mere correlation. Relative to a laboratory experiment, a field experiment gives up some of the control that a laboratory experimenter may have over her environment in exchange for increased realism.

The distinction between the laboratory and the field is much more important in the social sciences and the life sciences than it is in the physical sciences. In physics, for example, it appears that every hydrogen atom behaves exactly alike. Thus, when astronomers find hydrogen's signature wavelengths of light coming from the Andromeda Galaxy, they use this information to infer the quantity of hydrogen present there. By contrast, living creatures are much more complex than atoms and molecules, and they correspondingly behave much more heterogeneously. Despite the use of "representative consumer" models, we know that not all consumers purchase the same bundle of goods when they face the same prices. With complex, heterogeneous behavior, it is important to sample populations drawn from many different domains—both lab and field. This permits stronger inference and one can also provide an important test of generalizability: whether results from the laboratory continue to hold in the chosen field environment (see also List, 2006, on this argument).

We find an apt analogy in the study of pharmaceuticals, where randomized experiments scientifically evaluate new drugs to treat human diseases. Laboratory experiments evaluate whether drugs have desirable biochemical effects on tissues and proteins *in vitro*. If a drug looks promising, it would next be tested *in vivo* on several species of animals, to see whether it gets absorbed by the relevant tissues, whether it produces the desired effects on the body, and whether it produces undesirable side effects. If it still looks promising after those tests, it will then be tested in human clinical trials to test for efficacy and side effects.

Even after being tested thoroughly in human clinical trials and approved by regulators, a drug may sometimes reveal new information in large-scale use. For example, *effectiveness* may be different than the *efficacy* measured in clinical trials: if a drug must be taken frequently, for example, patients may not remember to take it as often as they're supposed to or as often as they did in closely-supervised clinical trials. Furthermore, rare side effects may show up when the drug is finally exposed to a large population.

In economics, there are a number of reasons why a field experiment could produce different results from a laboratory experiment (see Bohm, 1972, Harrison and List, 2004, and Levitt and List, 2007). First, different types of subjects might behave differently; university students in the laboratory might not exhibit the same behavior as financial traders or shopkeepers. In particular, the people who undertake a given economic

activity have selected into that activity; you might expect regular bidders to have more skill and interest in auctions than a randomly selected laboratory subject.

A second reason why a field experiment might differ from a laboratory experiment is that the laboratory environment might not be fully representative of the field environment. For example, a typical donor asked to give money to charity might behave quite differently than if the same person were asked to participate by choosing how much money to contribute to the public fund in a public-goods game. The charitable-giving context could provide familiar cognitive cues that make the task easier than an unfamiliar laboratory task. Even the mere fact of knowing that one's behavior is being monitored, recorded, and subsequently scrutinized might alter choices (Orne, 1962).

Perhaps most important is the fact that any theory is an approximation of reality. In the laboratory, experimenters usually impose all the structural modeling assumptions of a theory (induced preferences, trading institutions, order of moves in a game) and examine whether subjects behave as predicted by the model. In a field experiment, one accepts the actual preferences and institutions used in the real world, jointly testing both the structural assumptions (such as the nature of values for a good) and the behavioral assumptions (such as Nash equilibrium).

For example, Vickrey (1961) assumes that in an auction, there is a fixed, known number of bidders who have valuations for the good drawn independently from the same (known) probability distribution. He uses these assumptions, along with the assumption of a risk neutral Nash equilibrium to derive the "revenue equivalence" result: that Dutch, English, first-price, and second-price auctions all yield the same expected revenue. However, in the real world the number of bidders might actually vary with the good or the auction rules, and the bidders might not know the probability distribution of values. These exceptions do not mean that the model should be abandoned as "wrong"; it might well still have predictive power if it is a reasonable approximation to the truth. In a field experiment (such as Lucking-Reiley (1999), for this example), we approach the real world; we don't take the structural assumptions of a theory for granted.

Such an example raises the natural question related to the actual difference between lab and field experiments. Harrison and List (2004) propose six factors that can be used to determine the field context of an experiment: the nature of the subject pool, the nature of the information that the subjects bring to the task, the nature of the commodity, the nature of the task or trading rules applied, the nature of the stakes, and the environment in which the subjects operate. Using these factors, they discuss a broad classification scheme that helps to organize one's thoughts about the factors that might be important when moving from the lab to the field.

A first useful departure from laboratory experiments using student subjects is simply to use "non-standard" subjects, or experimental participants from the market of interest. Harrison and List (2004) adopt the term "artefactual" field experiment to denote such studies. While one might argue that such studies are not "field" in any way, for consistency of discussion, we denote such experiments as artefactual field experiments

for the remainder of the study, since they do depart in a potentially important manner from typical lab studies. This type of controlled experiment represents a useful type of exploration beyond traditional laboratory studies.

Moving closer to how naturally-occurring data are generated, Harrison and List (2004) denote a “framed field experiment” as the same as an artefactual field experiment but with field context in the commodity, task, stakes, or information set of the subjects. This type of experiment is important in the sense that a myriad of factors might influence behavior, and by progressing slowly toward the environment of ultimate interest one can learn about whether, and to what extent, such factors influence behavior in a case by case basis.

Finally, a “natural field experiment” is the same as a framed field experiment but where the environment is one where the subjects naturally undertake these tasks and where the subjects do not know that they are participants in an experiment. Such an exercise represents an approach that combines the most attractive elements of the lab and naturally-occurring data: randomization and realism. In this sense, comparing behavior across natural and framed field experiments permits crisp insights into whether the laboratory environment in and of itself unduly influences behavior.

All of these field experiment types represent means of collecting data. Why do we collect data, in any science?

First, we use data to collect enough facts to help construct a theory. Several prominent broader examples illustrate this point. After observing the anatomical and behavioral similarities of reptiles, one may theorize that reptiles are more closely related to each other than they are to mammals on the evolutionary tree. Watson and Crick used data from Rosalind Franklin’s X-ray diffraction experiment to construct a theory of the chemical structure of DNA. Careful observations of the motions of the planets in the sky led Kepler to theorize that planets (including Earth) all travel in elliptical orbits around the Sun, and Newton to theorize the inverse-square law of gravitation. After observing with a powerful telescope that the fuzzy patches called “spiral nebulae” are really made up of many stars, one may theorize that our solar system is itself part of its own galaxy, and the spiral nebulae are external to our Milky Way Galaxy. Robert Boyle experimented with different pressures using his vacuum pump in order to infer the inverse relationship between the pressure and the volume of a gas. Rutherford’s experiments of shooting charged particles at a piece of gold foil led him to theorize that atoms have massive, positively charged nuclei.

Second, we use data to test theories’ predictions. Galileo experimented with balls rolling down inclined planes in order to test his theory that all objects have the same rate of acceleration due to gravity. Pasteur rejected the theory of spontaneous generation with an experiment that showed that microorganisms grow in boiled nutrient broth when exposed to the air, but not when exposed to carefully filtered air. Arthur Eddington measured the bending of starlight by the Sun during an eclipse in order to test Einstein’s theory of general relativity.

Third, we use data to make measurements of key parameters, assuming a theory to be correct. Assuming that the electron is the smallest unit of electric charge, Robert Millikan experimented with tiny, falling droplets of oil to measure the charge of the electron. Assuming that radioactive carbon-14 decays at a constant rate, archaeologists have been able to provide dates for various ancient artifacts. Similarly, scientists have assumed theory to be true and designed careful measurements of many other parameters, such as the speed of light, the gravitational constant, and various atomic masses.

Field experiments can be a useful tool for each of these data purposes. For example, Anderson and Simester (2003) collect facts useful for constructing a theory about consumer reactions to \$9 endings on prices. They explore the effects of different price endings by conducting a controlled experiment with a retail catalog merchant. Randomly selected customers receive one of three catalog versions that show different prices for the same product. Systematically changing a product's price varies the presence or absence of a \$9 price ending. For example, a cotton dress may be offered to all consumers, but at prices of \$34, \$39, and \$44 in each catalog version. They find a positive effect of a \$9 price on quantity demanded, large enough that a price of \$39 actually produced higher quantities than a price of \$34. Their results reject the theory that consumers turn a price of \$34 into \$30 by either truncation or rounding. This finding gives hard evidence on an interesting topic and demonstrates the need for a better theory of how consumers process price endings (for an earlier field experimental example in the area of environmental economics see, e.g., List and Shogren, 1998).

List and Lucking-Reiley (2000) present an example of a field experiment designed to test a theory. The theory of multi-unit auctions predicts that a uniform-price sealed-bid auction will produce bids that are less than fully demand-revealing, because such bids might lower the price paid by the same bidder on another unit. By contrast, the generalized Vickrey auction predicts that bidders should submit bids equal to their values. In the experiment, List and Lucking-Reiley conduct 2-person, 2-unit auctions for collectible sports cards at a card trading show. The uniform-price auction awards both items to the winning bidder(s) at an amount equal to the third-highest bid (out of four total bids), while the Vickrey auction awards the items to the winning bidder(s) for amounts equal to the bids that they displaced from winning. List and Lucking-Reiley find that, as predicted by the theory of demand reduction, the second-unit bids submitted by each bidder were lower in the uniform-price treatment than in the Vickrey treatment. The first-unit bids were predicted to be equal across treatments, but in the experiment they find that the first-unit bids were anomalously higher in the uniform-price treatment. Subsequent laboratory experiments (see, e.g., Engelmann and Grimm, 2003, and Porter, 2003), have confirmed this finding.

Finally, Karlan and List (2007) is an example of a field experiment designed to measure key parameters of a theory. In their study, they explore the effects of "price" changes on charitable giving by soliciting contributions from more than 50,000 supporters of a liberal organization. They randomize the subjects into several different groups to explore whether solicitees respond to upfront monies used as matching funds. They find that simply announcing that a match is available considerably increases the revenue per

solicitation—by 19%. In addition, the match offer significantly increases the probability that an individual donates—by 22%. Yet, while the match treatments relative to a control group increase the probability of donating, larger match ratios—\$3:\$1 (i.e., \$3 match for every \$1 donated) and \$2:\$1—relative to smaller match ratios (\$1:\$1) have no additional impact.

In closing, we believe that field experiments will continue to grow in popularity as scholars continue to take advantage of the settings where economic phenomena present themselves. This growth will lead to fruitful avenues, both theoretical and empirical, but it is clear that regardless of the increase in popularity, the various empirical approaches should be thought of as strong complements, and combining insights from each of the methodologies will permit economists to develop a deeper understanding of our science.



## References:

- Anderson, Eric T. and Duncan Simester (2003), "Effects of \$9 Price Endings on Retail Sales: Evidence from Field Experiments", *Quantitative Marketing and Economics*, 1 (1), 93-110.
- Bohm, Peter, "Estimating the Demand for Public Goods: An Experiment," *European Economic Review*, June 1972, 3(2): 111-130.
- Engelmann, Dirk, and Veronika Grimm, "Bidding Behavior in Multi-Unit Auctions—An Experimental Investigation and some Theoretical Insights," working paper, CERGE-EI, June 2003.
- Harrison, Glenn W. and John A. List, "Field Experiments," *Journal of Economic Literature*, 2004, 42(4): 1009-1055.
- Karlan, Dean and John A. List. "Does Price Matter in Charitable Giving? Evidence from a Large-Scale Natural Field Experiment," *American Economic Review*, 2007, 97(5): 1774-1793.
- Levitt, Steven D. and John A. List. "What do Laboratory Experiments Measuring Social Preferences tell us about the Real World," *Journal of Economic Perspectives*, (2007), 21 (2): 153-174.
- List, John A. "Field Experiments: A Bridge between Lab and Naturally Occurring Data," *Advances in Economic Analysis and Policy*, (2006), 6(2), Article 2.
- List, John A. and David Lucking-Reiley, "Demand Reduction in a Multi-Unit Auction: Evidence from a Sportscard Field Experiment," *American Economic Review*, 2000, 90(4): 961-972.
- List John A. and Jason F. Shogren, "Calibration of the difference between actual and hypothetical valuations in a field experiment," *Journal of Economic Behavior and Organization*, (1998), 37 (2): 193-205.
- Lucking-Reiley, David, "Using Field Experiments to Test Equivalence Between Auction Formats: Magic on the Internet," *American Economic Review*, December 1999, 89(5): 1063-1080.
- Orne, Martin T. "On the Social Psychological Experiment: With Particular Reference to Demand Characteristics and Their Implications," *American Psychologist*, 1962, 17(10): 776-783.
- Porter, David, and Roumen Vragov, "An Experimental Examination of Demand Reduction in Multi-Unit Versions of the Uniform-Price, Vickrey, and English Auctions," ICES working paper, George Mason University, 2003.
- Vickrey, William (1961), "Counterspeculation, Auctions, and Competitive Sealed Tenders," *Journal of Finance* 16(1):8-37.