Experimental methods

A primer for economists

DANIEL FRIEDMAN
University of California, Santa Cruz

and

SHYAM SUNDER
Carnegie Mellon University

CAMBRIDGE
UNIVERSITY PRESS

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
40 West 20th Street, New York, NY 10011–4211, USA
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

First published 1994

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data

Friedman, Daniel, 1947–
Experimental methods : a primer for economists / Daniel Friedman, Shyam Sunder.
p. cm.
Includes bibliographical references and index.
ISBN 0-521-45068-3. – ISBN 0-521-45682-7 (pbk.)
1. Economics – Methodology. 2. Economics – Simulation methods. 3. Economics – Research. I. Sunder, Shyam. II. Title.
HB131.F75 1994
330'.01'1 – dc20 93–8005
CIP

A catalog record for this book is available from the British Library.

ISBN 0-521-45068-3 hardback
ISBN 0-521-45682-7 paperback

CONTENTS

List of figures and tables — *page* ix
Preface — xi
Acknowledgments — xiii

1 Introduction — 1
1.1 Economics as an experimental discipline — 1
1.2 The engine of scientific progress — 2
1.3 Data sources — 3
1.3.1 Some evidence — 6
1.4 Purposes of experiments — 7

2 Principles of economics experiments — 10
2.1 Realism and models — 10
2.2 Controlled economic environments — 12
2.3 Induced-value theory — 12
2.4 Parallelism — 15
2.5 Practical implications — 17
2.6 Application: The Hayek hypothesis — 18

3 Experimental design — 21
3.1 Direct experimental control: Constants and treatments — 22
3.2 Indirect control: Randomization — 22
3.3 The within-subjects design as an example of blocking and randomization — 25
3.4 Other efficient designs — 26

3.5 Practical advice 29
3.5.1 Chronic nuisances 29
3.5.2 Disposition of variables 30
3.5.3 Phases of experimentation 31
3.6 Application: New market institutions 32
3.6.1 Performance testing 33
3.6.2 Development testing 34

4 Human Subjects 38
4.1 Who should your subjects be? 39
4.1.1 Students 39
4.1.2 Professionals 40
4.1.3 Classroom experiments 43
4.1.4 Gender 44
4.2 Subjects' attitudes toward risk 44
4.3 How many subjects? 47
4.4 Trading commissions and rewards 48
4.4.1 Commissions 48
4.4.2 Rewards 49
4.4.3 The bankruptcy problem 51
4.5 Instructions 52
4.5.1 Purpose 52
4.5.2 Examples 53
4.5.3 Privacy 53
4.5.4 Realistic story 53
4.5.5 Duration of an experimental session 54
4.6 Recruitment and maintaining subject history 54
4.7 Human subject committees and ethics 55
4.8 Application: Bargaining experiments 56

5 Laboratory facilities 61
5.1 Choosing between manual and computer modes 61
5.1.1 Partial computerization 63
5.2 Manual laboratory facilities 63
5.3 Computerized laboratory facilities 64
5.3.1 Space 64
5.3.2 Layout 64
5.3.3 Furniture 65
5.3.4 Hardware and communications 65
5.3.5 Software 66
5.4 Random number generation 67

5.5 Application: Experiments with monetary overlapping generations economies 68
5.5.1 Equilibrium selection in a simple OLG economy 69
5.5.2 Hyperinflationary monetary economies 71
5.5.3 "Sunspot" economies 71

6 Conducting an experiment 74
6.1 Lab log 74
6.2 Pilot experiments 74
6.3 Lab setup 75
6.4 Registration 75
6.5 Conductors 76
6.6 Monitors 76
6.7 Instruction 77
6.8 Handling queries from subjects 77
6.9 Dry-run periods 78
6.10 Manual conduct of markets 78
6.11 Recording the data 79
6.12 Termination 79
6.13 Laboratory termination of infinite-period economies 80
6.14 Debriefing 80
6.15 Payment 81
6.16 Bankruptcy 81
6.17 Bailout plan 82
6.18 Application: Committee decisions under majority rule 82

7 Data analysis 85
7.1 Graphs and summary statistics 86
7.2 Statistical inference: Preliminaries 93
7.2.1 Basic concepts 94
7.2.2 Good samples and bad samples 95
7.3 Reference distributions and hypothesis tests 99
7.3.1 Internal reference distributions 99
7.3.2 External reference distributions 103
7.3.3 More statistical tests 104
7.4 Practical advice 106
7.5 Application: First-price auctions 107

8 Reporting your results 110
8.1 Coverage 110
8.2 Organization 112

8.3 Prose, tables, and figures 113
8.4 Documentation and replicability 114
8.5 Project management 115
8.6 Application: Asset-market experiments 116

9 The emergence of experimental economics 121
9.1 Economics as an experimental science 122
9.2 Games and decisions up to 1952 124
9.3 Two pioneers 125
9.4 Experimental economics in Germany 127
9.5 Early classroom markets 128
9.6 Building theoretical foundations, 1960–76 129
9.7 Joining the economics mainstream 131
9.8 Divergence from experimental psychology 132
9.9 Application: Laboratory games 134

Appendixes: Supplemental materials
I Readings in experimental economics 143
II Instructions and procedures 174
III Forms 201
IV *Econometrica* guidelines 205
V List of experimental economics laboratories 207

Glossary 211
References 215
Index 227

FIGURES AND TABLES

Figures

1.1	Theory and empirics.	*page* 3
1.2	Examples of data sources.	4
3.1	Independent variation of treatment variables.	23
3.2	Mean outcomes in a hypothetical factorial experiment.	27
7.1	Time graph for Day 18 of Das2.	89
7.2	Certainty equivalents and expected values.	90
7.3	Efficiency under the clearinghouse (CH) and quantity-only clearinghouse (CHQ) institutions.	93
9.1	Evolutionary game Exp5, Run2.	138
9.2	Evolutionary game Exp19, Run13.	139

Tables

7.1	Data from trading period Day18 of double-auction asset-market experiment Das2.	*page* 87

1

Introduction

1.1 Economics as an experimental discipline

> One possible way of figuring out economic laws . . . is by *controlled experiments*. . . . Economists [unfortunately] . . . cannot perform the controlled experiments of chemists or biologists because they cannot easily control other important factors. Like astronomers or meteorologists, they generally must be content largely to observe. (Samuelson and Nordhaus, 1985, p. 8)

Samuelson and Nordhaus echo a widely shared view that some disciplines are inherently experimental, but others (including economics) are not. History has not been kind to this view. In Aristotle's day some 2,000 years ago, even physics was considered nonexperimental. About 400 years ago, innovators such as Bacon and Galileo established a tradition of controlled experiments, mostly in physics. Experiments in related disciplines such as chemistry followed. For a long time biology was considered inherently nonexperimental because its subject was living organisms, but Mendel, Pasteur, and others introduced new experimental techniques in the nineteenth century. Modern biology certainly is an experimental science. Even psychology, whose mental subject matter might seem least accessible to laboratory study, has evolved a distinctive experimental tradition over the last century.

History suggests that a discipline becomes experimental *when innovators develop techniques for conducting relevant experiments.* The process can be contagious, with advances in experimental technique in one discipline inspiring advances elsewhere. Still, each discipline must innovate for itself. Even closely related disciplines differ in their intellectual focus, so wholesale transfer of experimental technique across disciplinary boundaries is seldom possible.

It took a long time but economics has finally become an experimental science. Most economists have heard about the experimental work of Vernon Smith, Charles Plott, Reinhard Selten, and others in the last three decades. (Indeed, in later editions of their text Nordhaus and Samuelson edited out the remarks we quoted.) Experiments are now commonplace in industrial organization, game theory, finance, public choice, and most other microeconomic fields. Some aspects of macroeconomic theory recently have been examined experimentally, although full-scale macroeconomic experiments do not seem feasible for budgetary and political reasons. (We refer to true, controlled experiments; uncontrolled macroeconomic "experiments" are all too common in recent years!) Perhaps macroeconomics too, like meteorology and astronomy, will become an indirectly experimental discipline, one that relies on experimentally verified results in constructing its central theories, although the central theories themselves are not amenable to direct experimental examination.

The methods as well as the substance of experimental economics are new in some respects. In the last few years the substantial findings of experimental economics have been expertly surveyed; see the annotated bibliography in Appendix I, pp. 143–74. However, no readily accessible, self-contained summary of experimental method and technique has yet been written for students and researchers in economics. The purpose of this primer is to bridge that gap.

Chapters 2 through 8 examine specific methods and techniques for economic experiments. The final chapter takes a look at the emergence of experimental economics in the last thirty years. The present chapter touches on some preliminary but fundamental issues: the interaction between theory and empirics, the differences between experimental and nonexperimental data for empirical work, and the diverse purposes of experiments. Since this book is a primer and not a theoretical treatise, we barely skim the surface of the deeper philosophical issues.

1.2 The engine of scientific progress

Theory organizes our knowledge and helps us predict behavior in new situations. In particular, theory tells us what data are worth gathering and suggests ways to analyze new data. As theory progresses, it guides us in refining our use of data and in selecting questions we should ask.

Conversely, data collection and analysis often turn up regularities that are not explained by existing theory. Such empirical regularities spur refinement of theory, usually as minor adjustments and sometimes as revolutionary changes. Kuhn (1970) and Lakatos (1978) discuss how

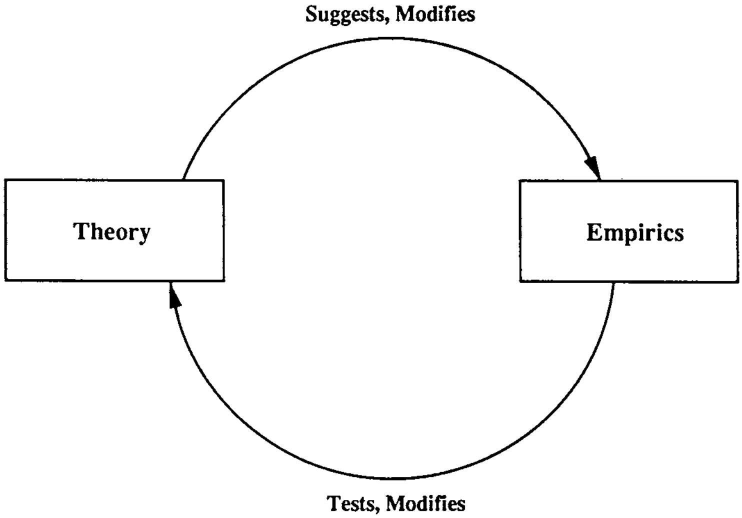

Fig. 1.1 Theory and empirics.

data and theory interact over time. The alternation of theory and empirical work, each refining the other, is the engine of progress in every scientific discipline. (See Figure 1.1.) Economics is no exception. Traditionally, observations from naturally occurring economic phenomena were the only source of data to stimulate revision of theory. If data relevant to an economic proposition could not be captured from naturally occurring conditions, then the proposition went without benefit of empirical refinement. In recent years, experimental methods have given economists access to new sources of data and have enlarged the set of economic propositions on which data can be brought to bear.

1.3 Data sources

Data for empirical work can be drawn from several types of sources, each with distinctive characteristics, advantages, and disadvantages. A key distinction is between *experimental data,* which are deliberately created for scientific (or other) purposes under *controlled* conditions, and *happenstance data,* which are a by-product of ongoing *uncontrolled* processes. A less important but still useful distinction can be drawn between *laboratory data,* which are gathered in an artificial environment designed for scientific (or other) purposes, and *field data,* which are gathered in a naturally occurring environment.

All combinations are possible. For example, an experimenter may

	Happenstance	Experimental
Field	Rate of Inflation in U.S.	Income Maintenance Experiments
Laboratory	Discovery of Penicillin	Laboratory Asset Markets

Fig. 1.2 Examples of data sources.

intervene in a naturally occurring process and record the outcomes; such data are field-experimental (FE). An economic example is the income-maintenance experiments in Denver, Seattle, and elsewhere (see Killingsworth, 1983; Pencavel, 1986). Traditionally, almost all empirical work in economics has used field-happenstance (FH) data such as national income accounts, commodity prices, or corporate financial statements. The story goes that penicillin was discovered in a laboratory when controls failed in a nutrient experiment, so this is an example of rare laboratory-happenstance (LH) data. Of course, this primer focuses on the last type of data, laboratory-experimental (LE). In this and later chapters, we often loosely refer to LE data as laboratory data or as experimental data and often ignore LH and FE data, but we make the finer distinctions when necessary.

Experimental data (LE or FE) are especially valuable for scientific purposes because they are relatively easy to interpret. If outcome Y (say, highly efficient allocations) is always associated with institution X (say, a certain kind of auction market) as institutional and other environmental variables are manipulated in a well-designed experiment, then we can confidently conclude that X causes Y. Happenstance data can't support such confident causal conclusions. Given the absence of control, an observed correlation between X and Y may be due to Y indirectly causing X, or may be due to some unobserved variable Z causing both X and Y. Leamer (1983, p. 31) makes the point while satirizing Monetarists and Keynesians in his delightful "Luminist versus Aviophile" parable. Aviophiles explain the higher crop yields found under trees in terms of bird droppings, while Luminists explain the same finding in terms of light intensity. Their quarrel is unresolvable with the "field" data because the two explanatory variables are completely confounded – that is, shade and bird droppings go together. The process-control example in Box, Hunter, and Hunter (1978, p. 487ff) provides a more elaborate discussion of the same point. We defer discussion of the underlying statistical issues until Chapter 7.

The other main issues in comparing experimental and happenstance data are cost and validity. Flexible, controllable laboratory environments usually are expensive to build, maintain, and operate, and each experiment requires further costs such as payments to human subjects. Thus both fixed (or sunk) costs and marginal costs may be significant for laboratory experiments, and typically are even higher for field experiments. Of course, it is also costly to obtain new field-happenstance data. The costs of gathering FH data on individual choice behavior, for example, are about the same as for LE data. Obviously it is least expensive to use data previously collected by someone else, such as a government agency.

Validity (or relevance) is a crucial issue for all data sources. When the field environment is of direct interest, FH and FE data are automatically relevant. On the other hand, FH data are normally,

> collected by government or private agencies for non-scientific purposes. . . . [By contrast,] astronomers are directly responsible for the scientific credibility of their data in a way that economists have not been. In economics, when things appear not to turn out as expected the quality of the [FH] data is more likely to be questioned. . . . (Smith, 1987, p. 242)

Specifically, the validity of FH data often is impaired by the omission of the really interesting variables (necessitating use of crude proxies), by measurement error of unknown magnitude, or by skewed coverage.

Laboratory data pose different validity questions. First, there is the question of internal validity: Do the data permit correct causal inferences? As we will see in later chapters, internal validity is a matter of proper experimental controls, experimental design, and data analysis. Second, there is the question of external validity: Can we generalize our inferences from laboratory to field? The issue of external validity or relevance often troubles economists who are unfamiliar with experimental work, and it remains a concern for experimentalists. Chapter 2 begins with a discussion of the gentle art of designing relevant experiments. Parallelism, the last substantive topic in Chapter 2, deals directly with the general question of external validity. For now, suffice it to say that, in economics as in other experimental disciplines, external validity has been firmly established in a diverse set of laboratory studies.

Sometimes data from computer simulations or surveys are improperly labeled as experimental economic data. Computer simulations of a theoretical model (no human decision makers involved except in writing the computer code) are best regarded as a type of theoretical results rather than as empirical data. Traditionally the investigator uses deductive logic and mathematical derivations to discover the implications of a theoretical model. You may resort to simulation because you have an intractable theoretical model so you can't derive the relevant theorems. As computing power becomes cheaper and more convenient, computer simulations become increasingly attractive relative to formal derivations as a discovery method. Survey data (human responses to hypothetical questions) are empirical but, unless responses are economically motivated, their reliability as economic data is questionable. This last point is developed in Section 2.3.

1.3.1 Some evidence

Econometricians have devised many ingenious techniques to deal with the weaknesses of happenstance data. Direct opportunities to test the effectiveness of these techniques are rare, LaLonde (1986) being the prime example. (See Cox and Oaxaca, 1991, for a different kind of effectiveness test.) LaLonde obtained field-experimental earnings data on former participants and nonparticipants in a job-training program. Experimental control had been achieved by random assignment of individuals as participants or nonparticipants; this important technique is discussed in Section 3.2. Straightforward statistical procedures showed that participants' mean annual earnings were about $900 higher, a statistically significant difference.

LaLonde then treated the data as if it were happenstance and the "control group" of nonparticipants did not exist. He used standard data sources and several multiequation specifications (some involving self-selection) and several econometric procedures to estimate the earnings effect. Estimates of the job-training effect on earnings varied considerably and some even had the wrong sign. He concludes

> This study shows that many of the econometric procedures and comparison groups used to evaluate employment and training programs would not have yielded accurate or precise estimates of the impact of the National Supported Work Program. The econometric estimates often differ significantly from the experimental results. Moreover, even when the econometric estimates pass conventional specification tests, they still fail to replicate the experimentally determined results. (LaLonde, 1986, p. 617)

The point is that, when obtainable at comparable cost, experimental data allow more reliable inferences than happenstance data. There are many cases where happenstance data are adequate and cheap; then experiments are not worthwhile. In many other cases happenstance data are inadequate and experimental data can be obtained at reasonable cost. Such cases present the best opportunities for experimental work.

Different types of data can be complementary. You can combine evidence from computer simulations, field, and laboratory to get sharper conclusions than those obtainable from a single data source.

1.4 Purposes of experiments

Experiments have many possible purposes. The proper way to design and to conduct your experiment depends on your purpose. Before proceeding further, a review of the purposes of experiments is in order (see Plott, 1982, 1987).

Some experiments have been conducted to generate data that might influence a specific decision. For example, Grether and Plott (1984) report an experiment designed to provide evidence in an antitrust case. Hong and Plott's (1982) research arose from a case considered by Interstate Commerce Commission. Alger (1988), Alger, O'Neill, and Toman (1987a,b), Plott (1988), and Rassenti, Reynolds, and Smith (1988) discuss the experiments conducted to assist Federal Energy Regulatory Commission. Roth (1987a) refers to experimentation designed to influence policymakers as "whispering in the ears of princes."

Influencing authorities is not the only persuasive purpose for experiments. Innumerable laboratory and field experiments have been conducted in order to provide data on how best to influence the decisions of consumers, voters, and managers. Cohen (1992) reports that white American consumers are more responsive to advertisements for stereo equipment featuring Asian models. This responsiveness of demand for stereos, where Asian manufacturers have dominated the U.S. market, is not discernible in advertisements for pickup trucks. Recently several popular business magazines have discussed new field technology that allows accurate measurement of market response to product innovations or advertising campaigns. In U.S. presidential campaigns at least since 1988, laboratory studies of voter response to proposed television messages and campaign slogans have played an important part in the strategies of most major candidates. For example, Torry and Stencel (1992) report in the *Washington Post* that the Bush–Quayle campaign confirmed through focus groups that bashing trial lawyers was an effective vote-getting theme; see Payne (1992) for another typical example. The large (and apparently increasing) sums of money devoted to such marketing applications suggests that they do provide commercially valuable data.

This primer emphasizes the scientific purposes of experiments. Persuasion certainly is still in the picture (McCloskey, 1985), but specific immediate decisions are of less concern than the longer run views of the scientific community. One scientific purpose is to discover empirical regularities in areas for which existing theory has little to say. McCabe, Rassenti, and Smith (1993) and Friedman (1993), for example, compare the properties of several market institutions whose theoretical properties are as yet poorly understood. Smith (1982b) calls such experiments *heuristic*. In other areas, by contrast, several competing theories offer differing predictions and experiments can help map the range of applicability for each theory. For example, Fiorina and Plott (1978) study committee decisions in the laboratory and find that only a few of the sixteen models and variants considered are at all consistent with the data. Finally, there are areas for which only one model is applicable. Laboratory work can demonstrate whether there are any conditions under which the theory can account for the data, and if so, can test theory for robustness. "In Search of Predatory Pricing," by Isaac and Smith (1985) is a negative example. Smith (1982b) refers to the last two types of experiments as *boundary* experiments and refers to sets of experiments intended to establish definitive broad laws of behavior as *nomothetic*.

Some experimental economists have hesitated in recent years to describe the purpose of an experiment as a *test of theory*. From a formal point of view, a theory consists of a set of axioms or assumptions and definitions, together with the conclusions that logically follow from them. A theory is formally valid if it is internally consistent – that is, it does not lead to statements that contradict each other – and if the conclusions are indeed provable from the assumptions. What can be learned about theories by conducting experiments? Some experimentalists (including most psychologists) think of experimental data as a means of testing the descriptive validity of the assumptions about human behavior on which the theory is based. Others (including most economists) would readily grant that the behavioral assumptions of most economic theories do not and need not meet the descriptive validity criterion used in psychology. Instead they believe that a theory is of direct practical interest only to the extent that its conclusions provide good approximations (relative to alternative theories) of actual behavior even when its assumptions are not precisely satisfied. See Friedman (1953) and Koopmans (1957) for further discussion.

The proper job of the empirical scientist is to find regularities in observed behavior in a broad range of interesting environments and to see which theories can best account for these regularities. Whether this job is called "testing theories," or more circumspectly referred to as "seeing which theories best organize the data," it is a primary purpose of scientific experiments.

Experimental economists have become increasingly interested in recent years in using laboratory methods (including economic incentives) to measure individual (innate or "home-grown") characteristics in the population, such as willingness to pay for environmental amenities or risk aversion (see Cummings, Harrison, and Rutström, 1992). In a novel application of experimental technique, Forsythe et al. (1992) have introduced a computerized field market for candidate-contingent claims to predict the percentage of total vote received by each candidate in an election. Some experimentalists in previous decades tried to measure behavioral parameters or to simulate natural economic processes in the laboratory. For example, Hoggatt (1959) set out to measure oligopolistic "reaction functions," and Garman (1976) tried to simulate the New York Stock Exchange. Experimental economists now recognize that behavioral parameters usually vary with the institution and the environment, so the external validity of such measurements is questionable. As explained in Section 2.1, experimentalists no longer see simulation (in the sense of replicating a field environment as closely as possible) as a useful goal.

A related but more modest purpose for experiments has recently emerged. Aircraft engineers find it useful to study a small-scale model in a "test bed" before trying to build and fly a new plane. Likewise, economists and policymakers recently have found it useful to study new institutions in the laboratory before introducing them in the field. McCabe et al. (1991) describe "test-bed" experiments of computer-aided markets for composite commodities such as computer resources, and gas and electrical power grids. Given the accelerating pace of transformation in the formerly centrally planned economies and given continuing deregulation in Western economies, the scope for institutional engineering of this sort is large and increasing.

Finally, experiments have an important pedagogical purpose. The first recorded use of economics experiments, by Chamberlin (1948), was primarily pedagogical. Since the 1980s this use of economics experiments has grown steadily. Incorporating experimental demonstration of economic propositions into the high school and college curriculum is a natural accompaniment of the evolution of economics as an experimental science. Walker, Williams, and their colleagues at Indiana University, and Wells and his colleagues at the University of Arizona have developed many pedagogical economics experiments (see Wells, 1991; Williams and Walker, 1993).