CHAPTER 6 EXPERIMENTAL FINANCE

ROBERT BLOOMFIELD
Nicholas H. Noyes Professor of Management and Professor of Accounting, Cornell University,

ALYSSA ANDERSON
Ph.D. Student in Finance, Cornell University

October 9, 2009. Forthcoming in Behavioral Finance, Kent Baker and John Nofsinger, Eds. Wiley (NY),

ABSTRACT

This chapter provides a guide for those interested in experimental research in finance. The chapter emphasizes the role experiments play in a field governed largely by modeling and archival data analysis; discusses the basic methods and challenges of experimental finance; explores the close connection between experiments and behavioral finance; and comments on how to think about experimental design. First, the chapter begins by discussing the relationship between experiments and archival data analysis. Experiments are useful because they allow researchers to circumvent common econometric issues such as omitted variables, unobserved variables and self-selection. Next, the chapter examines the contributions that experiments can make beyond theoretical models, either by relaxing certain assumptions or by addressing settings that are too complex to be modeled analytically. Lastly, the chapter discusses the difference between experiments and demonstrations, and emphasizes the critical role of controlled manipulation.

INTRODUCTION

Experiments are useful in finance because they allow researchers to isolate and manipulate one variable at a time, thereby illustrating its causal effects without resorting to complex and imperfect econometric techniques to filter out effects of other variables. Experiments also allow researchers to observe independent and dependent variables that might be unobservable outside the laboratory setting, and to avoid the complications of self-selection by assigning subjects randomly to different treatments.

Second, a key challenge in experimental finance is to construct experiments that can test economic models in settings that are true to the models' assumptions, but in which alternative hypotheses are sufficiently plausible such that the experimental results are not foregone conclusions. One way to do so is to relax the structural, behavioral or equilibrium assumptions underlying the model being tested; another is to examine settings that are too complex to be definitively modeled.

Third, experimentalists in finance and economics must distinguish more carefully between experiments and demonstrations. A true experiment entails the controlled manipulation of a specific variable, while holding all other variables constant. A demonstration simply examines behavior within a single setting. The lack of controlled manipulation leaves a demonstration susceptible to criticisms that any feature of the setting (such as the wording of instructions, labeling of strategies, or even the color of the laboratory) is driving observed behavior. Experiments are more robust to such criticisms because the feature in question does not vary across treatments, and thus is unlikely to drive the *difference* in behavior across settings. Researchers should conduct demonstrations only when experiments are impractical, which happens rarely.

The chapter proceeds as follows. Section 2 provides a discussion of how experiments complement theoretical and archival (econometric) research in finance. Section 3 describes the basic methods of experimental economics and discusses ways in which experiments can provide contributions above and beyond the models they are testing. Section 4 focuses on one of the most important streams in experimental finance that relates directly to behavioral finance: the ability of markets to aggregate information and eliminate individual biases. Section 5 compares methods of experimental economics and experimental psychology. Section 6 provides a summary and conclusions.

THEORY, ECONOMETRICS AND EXPERIMENTS

Financial economics is grounded in analytical modeling, which uses mathematical methods to derive the implications of some fundamental assumptions about individual or aggregate behavior. Many of these models provide testable predictions about the behavior of markets, firms, and investors.

Archival data analysis tests financial theories using data that are generated and archived for another purpose. For example, asset pricing tests typically use CRSP data that is generated as a natural outcome of trade in large stock exchanges, perhaps combined with accounting data from Compustat that is generated from SEC filings. A key challenge in archival data analysis is that the data are drawn from settings created for a purpose other than answering the research question at hand. As a result, almost any interpretation of the results can be challenged as ignoring other features that have changed. Key problems include omitted variables biases, self-selection biases, unobservable independent variables, and unobservable dependent variables.

Some examples will clarify how well-designed experiments can avoid these problems:

- ***Experimentalists avoid omitted-variables biases*** by creating settings that differ from one another in exactly one independent variable, controlling all other variables of the setting to eliminate alternative explanations for observed differences in the dependent variable. For example, Bloomfield and O'Hara (1999) address the role of transparency regulations in an experimental setting by having traders trade with market makers in three different market settings. In the "transparent" setting, all quotes and trades are publically disclosed. In the "semiopaque" setting, quotes are publically disclosed but individual trades are not disclosed to any participants. In the "opaque" setting, quotes are disclosed only to traders while trades are again not disclosed. Cohorts of traders are assigned to trade in each of the different market settings in a random order. These market settings are identical in all aspects except the degree of market transparency.

Therefore, any differences across settings will be due strictly to transparency differences. Bloomfield and O'Hara (2000) and Flood, Huisman, Koedijk, and Mahieu (1999) use similar techniques.

- ***Experimentalists avoid self-selection problems*** by randomly assigning subjects to treatments. For example Tosi, Katz, and Gomez-Mejia (1997) perform an experiment on the effects of monitoring and incentive alignment on corporate decision making. Subjects are randomly assigned to one of six treatments: high-incentive alignment (CEO pay was linked to a profit-maximizing strategy), low-incentive alignment (CEO pay was linked to a sales-growth strategy), high-monitoring, low-monitoring, long-term-CEO (the subject had been the CEO, and therefore was responsible, for previous investment decisions) and short-term-CEO (the subject was recently appointed CEO and therefore was not responsible for previous investment decisions).. The subjects then acted as the CEO of a firm and had to make an investment allocation decision, given information about the firm's prior poor investment decision. The authors find that incentive alignment is more effective than monitoring in making sure management acts in the interests of the shareholders. By randomly assigning subjects to these different treatment groups, the authors avoid the issues caused by self-selection and are able to directly observe the role of different governance practices on earnings.
- ***Experimentalists avoid problems of unobservable independent variables*** by creating settings themselves, so that they can observe all variables. For instance, the degree of risk aversion that investors have is unobservable in archival data sets. It is also unobservable in an experimental setting because experimenters cannot directly elicit this information from their subjects. However, Bossaerts and Plott (2004) demonstrate how good experimental design can help avoid this issue in their study addressing the equilibration of large-scale financial markets. By using the Capital Asset Pricing Model (CAPM) framework, Bossaerts and Plott can measure how far the market

is from equilibrium at any point without directly knowing the level of risk aversion among the participants. They simply need to know the true expected return, which the experimental setting makes observable. They predict that risk premia will be proportional to the covariances between the risky assets and the market portfolio, as predicted by CAPM, and test this prediction using Sharpe ratios. Because expected returns and variances are directly measurable in the experimental setting, Sharpe ratios can be calculated and the problem of unobservable independent variables is avoided.

- ***Experimentalists avoid problems of unobservable dependent variables*** by creating tasks that elicit them. For instance, Bloomfield and Hales (2006) conduct an experiment to study the role of mutual observation in analysts' forecasts. They find that, when analysts' that are able to see each other's forecasts, the consensus forecast is more extreme but more accurate. By conducting this study experimentally, Bloomfield and Hales are able to observe the analysts' prior beliefs. Additionally, they impose a structure that eliminates performance-based incentives and provides analysts with flexibility in adjusting their estimates. Therefore, their study can draw more precise results about the specific question at hand—whether mutual observation leads analysts to engage in free-riding or excessive extremity. In traditional archival data studies, extracting that role of mutual observation and differentiate between potential reasons that analysts may change their forecasts is difficult.

The most common form of experimentation is to construct a highly controlled setting in the laboratory. Laboratory experiments allow for extremely simple settings that facilitate clear inferences. For example, securities can have easily known values with simple distributions. The ability to control the variables in the experiment provides for a greater degree of assessing causality. Laboratory settings allow very clear inferences about causal relationships within the experiment (internal validity), but allow doubts about how well behavior in the laboratory will generalize to the outside world (external validity).

Field experiments have recently become popular in economics. In the field experiment, the researcher goes into a natural setting, with all of its messiness, but will manipulate variables one at a time (usually with the cooperation of someone with appropriate authority). As one example, Thaler and Benartzi (2004) conducted a field experiment on their Save More Tomorrow plan, which offered workers at actual firms the chance to commit to devoting more of their future income increases to their retirement savings. The goal was to avoid loss aversion by increasing savings only when the employee gets a raise, and then to take advantage of inertia and the status quo bias to keep people in the program and at the progressively high savings rates. Thaler and Benartzi implemented this program at several different firms and, despite some uncontrollable differences across implementations, obtained similar, positive results in all cases.

While field experiments are a promising direction, they are rare enough in finance that the remainder of the paper focuses exclusively on laboratory experiments.

THE FUNDAMENTAL METHOD AND CHALLENGE IN EXPERIMENTAL FINANCE

The fundamental method of experiment economics is to create a setting that implements some institutional features of interest, and then provide participants with incentives to maximize utility within that setting. Smith (1982), who won the Nobel Prize in Economics for his work in experimental economics, places great emphasis on providing participants with incentives similar to those that economists would model, without unwanted distortion. For example, Smith requires that participants have a reward of real value they can pursue. He also argues that the incentive to pursue that reward should never be satisfied (no ceiling on incentives), that rewards are entirely private (to avoid the possibility of social pressures that would lie outside an economic model), and that the monetary payoffs are so large that they dominate any non-monetary rewards. This last requirement is called the principle of ‘dominance.’

In Smith's view, the aim of experiments is to test economic theory by implementing the assumptions of the theory as faithfully as possible. However, Smith's vision of the economic experiment does present one very serious challenge to researchers—it is not easy to ensure that the experimental data actually provide a contribution beyond the economic model being tested. To clarify the challenge, imagine an archetypal pricing experiment, in which a single trader is presented with two assets, A and B, each paying a single liquidating dividend. The dividends are distributed normally with identical means, but the variance of A's dividend is lower than the variance of B's dividend. The experimenter induces a negative exponential utility function using the Berg, Daley, Dickhaut, and O'Brien (1986) mechanism, by providing the payout in the form of lottery tickets, with each additional lottery ticket increasing the probability of a payout by slightly less than the previous one. Economic theory makes very clear predictions about the optimal choice: every participant should prefer A to B, if each costs the same. In a large market of traders, assuming the risk-free rate is zero, the price of B should be lower according to the following formulas:

$$S = \frac{E(R_A)}{\sqrt{Var(R_A)}} = \frac{E(R_B)}{\sqrt{Var(R_B)}} \tag{1}$$

$$\frac{\frac{D}{P_A}}{\sqrt{Var(R_A)}} = \frac{\frac{D}{P_B}}{\sqrt{Var(R_B)}} \tag{2}$$

$$P_B = \frac{\sqrt{Var(D_A)}}{\sqrt{Var(D_B)}} \cdot P_A \tag{3}$$

Given this description of the model and experiment, the question can be asked: What is learned from conducting the experiment? As Kachelmeier (1996, p. 83) puts it:

> If observed behavior is consistent with a model that is predicated on induced values, the skeptic may ask what behavioral insights we learn other than the demonstrated strategic preference for more money over less. However, if results contradict the model, the skeptic will be just as quick to raise the usual objections available whenever hypothesized findings are not observed.

What are the usual objections? Typically, these entail a failure to ensure that the assumptions underlying the theory (such as expected utility maximization, negative exponential utility functions, and competitive markets) actually hold in the environment. But if experimentalists take this approach, they are simply viewing economic models as tautologies, which experiments could not possibly refute.

One way around this difficulty is to think more clearly about the nature of assumptions in economic models. Following the discussion in Bloomfield, Tayler, and Zhou (2009), experimentalists usually classify models as having three types of assumptions: *Structural* assumptions describe the institutions in which agents interact, including the distribution of information, possible actions, and incentives; *Behavioral* assumptions characterize agents' preferences and decision-making abilities (such as expected utility maximization and the form of the utility function); and *Equilibrium* assumptions that describe the solution concepts used to predict behavior (such as Bayesian Nash Equilibrium, rational expectations, or arbitrate free pricing).

The pricing experiment described above fails to contribute beyond the model on which it is based because it imposes behavioral assumptions and (certainly in the one-person case) provides no plausible alternative to the equilibrium assumption. However, imagining slight relaxations that would make the experiment more interesting is not diffiicult. The remainder of this section is devoted to discussing various examples of studies that relax structural, behavioral, and equilibrium assumptions in different ways to allow experiments to make novel contributions.

Testing Behavioral Assumptions

Benartzi and Thaler (1999) conduct an experiment that relaxed, and thus tested, the behavioral assumption that people perfectly process information about risk. The experiment provides participants with information about the historical performance of debt and equity investments, and manipulates whether participants are informed of the yearly return for each of thirty years, or just summary information about returns over a 30-year period. Results indicate that the yearly feedback makes the more volatile equity investment seem far more risky, while the summary information reduces participants concerns about volatility and highlights the higher expected return. Therefore, participants who are given the summary information are much more likely to invest in equities. By explicitly testing the behavioral assumption of perfect information processing in the model, Benartzi and Thaler are able to show that people in fact suffer from myopic loss aversion, and thus make a contribution beyond a simple pricing experiment.

Forsythe, Lundholm, and Reitz (1999) test a behavioral assumption in a different setting. Their paper incorporates voluntary disclosure—the potential buyers learn about the possible dividends from the sellers, who report a range of values that must include the true value. In equilibrium, the buyers should assume that the value is the lowest element of the reported range, to protect themselves from sellers who attempt to inflate prices by including higher elements. The paper is often known by the title 'Half a sucker is born every minute' because sellers do engage in such attempts at price inflation, and it works—even though the same people alternate between roles as buyers and sellers. Thus, people seem to make shrewd reporting decisions, but are gullible in interpreting others' reports, in a clear violation of the behavioral assumptions behind most cheap-talk models.

Equilibrium Assumptions: Multiple Equilibria

Other papers relax equilibrium assumptions. The most natural direction is to examine contexts with multiple equilibria. Most modelers are fairly cavalier about their equilibrium assumptions. In rational expectations models with heterogeneous information, a standard conjecture is that demand is a linear function of expectation, but in fact there could be other equilibria. Thus, there is benefit in testing models of information aggregation by conducting laboratory markets—the experimental data demonstrate that the equilibrium assumption is in fact accurate, which need not be the case.

Equilibrium assumptions are particularly important in signaling models. Cadsby, Frank, and Maksimovic (1990) use a series of experiments to test the theoretical predictions of Myers and Majluf (1984) regarding the signaling of firms seeking investors. Participants are divided into two groups, firms and investors. Firms are told they are either of type H or type L, then make a decision to undertake a new project or not. Investors are then informed of the firms' decisions, but not their types, and participate in an auction to fund the projects. When theory predicts a unique equilibrium, subjects attain this equilibrium in all cases. However, if there are multiple equilibria predicted, whether participants should pool or separate is unclear. The authors argue that experiments provide an important tool to address these ambiguous cases. In all versions of the experiment in which theory predicted multiple equilibria, subjects pooled, possibly because the pooling equilibrium is always Pareto superior to the separating and semi-separating equilibria. Therefore, experimental means can offer predictions in cases when theoretical approaches are insufficient.

Equilibrium Assumptions: Convergence

Even in settings with a unique equilibrium, there is no guarantee that the equilibrium can be achieved. An equilibrium is a fixed point—an outcome that, if obtained, will leave no participant wishing to deviate from it. Still, experimentalists are quickly forced to think about the process that would drive participants toward equilibrium if they are not already there (convergence), and also about the process that might drive participants away from an equilibrium if, having attained it, they drift infinitesimally far away (instability).

Laboratory studies of information aggregation are natural settings in which to think about the dynamics leading to equilibrium. Studies by Plott and Sunder (1982, 1988) provide a classic example. In these studies, security values are determined by which state of nature occurs, and each trader is given information about a state that has not occurred. Collectively, traders know the state that must have occurred. For example, in some markets the possible states are X, Y and Z, and some traders know the state is not X, while others know the state is not Z; therefore, collectively traders know the state is Y.

An equilibrium analysis would predict that prices fully reflect the information held collectively. Plott and Sunder (1982) show that this is in fact the case in a simple one period market in which some traders are informed of the true state. In this experiment, a security pays a state-dependent dividend that differs across individual traders. Some traders know the realized state but it is unknown which traders are informed and which are uninformed. Double oral auctions between the traders should result in complete information aggregation in this setting. However, through another series of experiments, Plott and Sunder (1988) show that this is not always the case in more complex settings. Markets aggregate information much more effectively when the securities are Arrow-Debreu securities: that is, the market includes one security that pays off only in state X, another that pays off only in state Y, and a third that pays off only in state Z. Apparently, this setting allows participants to extract information from the trades they observe more easily. Additionally, if all traders have identical preferences, information can be fully aggregated even if there is only one security. In a market with only one security and diverse preferences, however, information aggregation is incomplete.

Bloomfield (1996) provides some additional insight into the process of information aggregation. In his study, the value of each security is the sum of four random numbers. In one setting, every random number is seen by two traders and every trader sees one number. In another setting, every random number is seen by four traders and every trader sees two numbers.

Although traditional theory predicts a fully-revealing equilibrium, Bloomfield predicts and finds that the markets will impound information incompletely, and impound less completely when the information is less widely distributed. The reasoning is that traders make decisions to buy or sell on the basis of the information they personally hold, the information they extract from the market, and their preferences. At the same time, traders invert that function to infer information from others' trades. This inversion and inference process is far more difficult when traders are endowed with less information individually, even though collectively the total amount of information is the same in both cases. Thus, the market may not be able to fully incorporate all of the available information; rather, the degree of convergence is dependent on how that information is distributed among individual traders.

As a final example of non-convergence to equilibrium, Bhojraj, Bloomfield, and Tayler (2009) construct a model with a single security, which all human participants are told will pay a liquidating dividend of 500 laboratory dollars. The market uses a robot specialist who sets a price of 500 + k(D-S), where D-S is the net cumulative demand for the security. The market also includes a robot buyer who buys shares steadily in each period of trade until a known ending time, and thus would drive prices upward. Bhojraj et al. (2009) assume that the market price constitutes a Nash equilibrium, and show that a simple backward induction argument should keep prices at 500 in each period of trade, as long as the traders have access to enough capital.

Bhojraj et al. (2009) develop an alternative hypothesis by noting that the structure of the setting includes a social dilemma. Participants collectively make far more money by a strategy of front-running: buying shares at the beginning of trade to force the robot to buy at even higher prices, and selling those shares only after the robot has driven the price up. Extensive research on prisoners' dilemmas shows that people will initially choose disequilibrium strategies that would be socially optimal if everyone did so, and learn to play equilibrium strategies only as they gain experience. In the setting of Bhojraj et al. (2009), this shows that traders will initially engage in front-running to take advantage of the robot trader's positive sentiment. This is particularly true when investors have a small initial share endowment. Additionally, the authors find that looser margin restrictions, and therefore more short-selling, will result in delayed convergence to equilibrium because traders face the risk of a margin call if they go against the crowd and attempt to arbitrage the deviation from equilibrium too early.

Testing Models that Cannot Be Solved

A final way to avoid the problem of having no plausible alternative hypothesis is to construct a setting in which theory is simply unable to provide a unique prediction. This alternative is particularly relevant in market microstructure, which deals with extremely complex settings and strategic problems. The most common market used in laboratory settings is the double auction, which is simple in execution, but extremely difficult to model (see Friedman, 1984). Modelers in microstructure will often look at simpler settings, such as Kyle-type or Glosten-Milgrom-type experiments. For example, Bloomfield (1996) uses a Glosten-Milgrom-type setup in which investors and market makers simultaneously submit the best bid and ask at which they are willing to trade. Crossing trades are then executed, all other orders are canceled and trade moves on to the next period. Orders that are not immediately marketable do not have any impact on trade as they would in a limit order market. These types of models of quote- or order-driven markets are much less relevant these days, when most trade takes place in electronic limit order markets.

Bloomfield, O’Hara, and Saar (2005, 2009) examine the behavior of informed and noise traders in an electronic limit order market. The first study addresses the liquidity provision strategies of different types of traders. In this experiment, informed traders and uninformed liquidity traders with specific trading targets trade in a market with many features of actual electronic markets. This enables the authors to determine how both market characteristics (such as the depth of the limit order book) and security characteristics (such as volatility) affect traders’ strategies depending on trader type in a way that is impossible in a strictly theoretical framework. Therefore, this setting is much more robust and less restrictive than typical theory work in this field must necessarily be. Results show that order submission strategies are dependent on trader type and evolve over the trading period. While informed traders begin trading with market orders to capitalize on their informational advantages, they switch to limit orders as the period progresses. In this way, informed traders act as a dealer and provide liquidity to the market because they know the security's true value. They benefit from this strategy by profiting from the bid-ask spread.

The latter paper by Bloomfield, O'Hara, and Saar (2009) examines the behavior of noise traders, who have no exogenous reason to trade, in an electronic limit order market. The markets in these experiments contain informed traders with individually imperfect information but perfect information in the aggregate, liquidity traders with fixed trading targets, and, in some cases, uninformed noise traders. By comparing markets with and without noise traders, their role can be better understood. The authors find that these traders benefit the market by increasing volume and liquidity through their contrarian strategies, but also hinder the market's ability to incorporate new information.

INDIVIDUAL BIAS AND AGGREGATE MARKET BEHAVIOR

A large literature from psychology shows individual errors in judgment. Some of the biases most relevant in the financial setting, are outlined by Tversky and Kahneman (1974). These biases include representativeness, the tendency to assume commonality between similar objects; availability, which causes probabilities to be assigned based on how easily similar examples can be brought to mind; and anchoring, the reliance on a single piece of information or starting point when making an estimate.

A key tenet of traditional finance is that markets eliminate these errors. It is suggested that people learn to avoid these mistakes through experience and incentives,. As people learn, either directly from investment professionals or indirectly through their own experience, irrational investors will be flushed out of the market. Additionally, even if individual biases are present in the market, they will likely cancel each other out so that, on aggregate, the market will be unbiased. Yet, experimental work has shown that the markets' ability to do so is somewhat limited.

Camerer (1987) conducted the initial experimental work on markets' ability to eliminate individual biases. He reports 15 experiments asking people to predict from which urn a series of three balls are drawn. First, an urn, X or Y, is chosen by picking a random number between 1 and 10 from a third urn. There is a 60 percent chance of choosing urn X, and a 40 percent chance of urn Y. Subjects know these probabilities but do not know which urn is chosen. Then three balls are chosen with replacement from either X or Y. X contains 1 red and 2 black balls and Y contains 2 red and 1 black balls. Participants are assigned to be either type I or II and then trade an asset that pays a state-dependent dividend in a double-oral auction. Camerer finds that the prices of the assets do tend toward the values predicted by Bayesian theory, but there is also evidence of statistically significant representative bias. This bias decreases as participants gain experience and as incentives are increased. This study may not generalize to real world financial markets, but is still instrumental in demonstrating some of the potential shortcomings of markets' ability to eliminate individual biases.

Ganguly, Kagel, and Moser (1994) show that information aggregation depends on market structure. They conduct two versions of an experiment to test the persistence of individual biases in the form of the base-rate fallacy; one version is a market in which unbiased traders have the highest expected payoffs and the other is a market in which biased traders have the highest expected payoffs. The *base-rate fallacy* is the tendency for people to overweight current information rather than the initial base rate. As hypothesized, prices are biased when biased traders have the highest expected payoffs. On the other hand, when unbiased trades have the highest expected payoffs, prices move toward the unbiased level but remain biased. This is because so few traders are actually unbiased; the majority fall victim to the base-rate fallacy. When there are short-sale constraints, as there are in this market, people with unbiased views cannot push prices back to their true values because they cannot sell short.

If there is a high probability of traders being biased, even a competitive market cannot remove individual biases if its structure limits certain types of trading.

Gode and Sunder (1993) show that smart institutions can compensate for dumb traders, even if none of the traders in the market are smart. Gode and Sunder create ‘zero-intelligence’ (ZI) programs that submit random orders, and report the results of their simulations. These machine-based traders submit random, independent, uniformly distributed orders; they do not try to maximize profits and they do not remember or learn from past orders. There is obviously a large discrepancy between the behavior of human traders and unconstrained ZI traders. Gode and Sunder attempt to determine how much of this difference is due to learning and profit incentive, and how much can simply be attributed to market discipline. To address this, they run three versions of each market: one with human traders, one with unconstrained ZI traders, and one with ZI traders with budget constraints. The authors find that, while there is no learning with constrained ZI traders, the price series with these traders has much less volatility than that of unconstrained ZI traders and converges to equilibrium within each trading period. In this case, the market is able to eliminate individual irrationality and converge to equilibrium despite the randomness in the traders’ strategies. These methods do not work very well in slightly more complex settings.

More complex information structures also interfere with information aggregation in markets with human traders. As discussed above, Plott and Sunder (1988) and Bloomfield (1996) show how the distribution of information can block full revelation. Forsythe and Lundholm (1990) show that information aggregation requires both experienced traders and common knowledge of the dividend structure. O’Brien and Srivastava (1991) also conduct a series of experiments in which information is not fully aggregated. They argue that the more complex the market is, in terms of the number of securities and the number of trading periods, the more difficult it is for information to be aggregated. In these complex settings, fully arbitraging away these inefficiencies is impossible. Lundholm (1991) shows that information aggregation is less complete when traders would face uncertainty even if they knew all information available in the market.

Aggregation is even more difficult when traders' information comes from their own knowledge, rather than from a fact passed on by the experimenter. Bloomfield, Libby, and Nelson (1996) conduct markets in which the value of a security is based on the answer to an "almanac-style" business-related question. The experiment has two treatments; one in which participants can see the number of traders who traded above, below, and at the posted share price; and one in which they can see the number of shares traded above, below, and at the posted share price. They show that individuals who have higher accuracy in assessing the value of the security trade a larger number of shares. This implicitly allows traders to determine each others' confidence levels. Therefore, when traders see the number of shares rather than just the number of traders, the market price is more accurate. This process is likely to be imperfect in real financial markets due to well known limits of calibration. This paper does not demonstrate that people are often unaware of their own biases so they will trade aggressively even when they are biased. Therefore, biased judgments will remain prevalent in the market price as well.

Finally, some research examines the possibility that markets can create biases that would not exist at the individual level. Seybert and Bloomfield (2009) examine this issue in the context of wishful thinking. People often trade on their optimistic biases, which may in turn lead others to overestimate these probabilities too because people often infer others' beliefs from their actions. In this study, participants traded multiple assets simultaneously and were all endowed with a long position in half of the assets or a short position in the other half. The assets' prices are a function of cumulative demand, so if traders buy shares, the price increases, and vice versa. Each trader also had imperfect information about the value of the asset. Seybert and Bloomfield find that, while traders do not initially engage in wishful thinking (their beliefs are unbiased), they do engage in wishful betting. They are more likely to buy the assets in which they have an initial long endowment. This results in a contagion of wishful thinking because other traders cannot differentiate between wishful betting and actual information about the value of the security. Thus, while other studies show that markets can eliminate some individual biases, Seybert and Bloomfield show that markets can create and magnify biases as well.

INSIGHTS FROM COMPARING EXPERIMENTAL PSYCHOLOGY AND EXPERIMENTAL ECONOMICS

Experimentation has a very long tradition in psychology, but is relatively new to economics. In fact, Chamberlain performed the first experiment in 1948 and Smith laid out the tenets of experimental economics in 1976. Camerer (1997) describes many of the differences between the styles of those who conduct experiments based in economics (the E's) and psychology (the P's). Camerer notes a number of key differences in the experiments run by E's and P's. Some of the most important are:

- E's insist that participants receive incentive compensation, while P's rarely do.
- E's typically have experiments with groups of people who interact with one another, while P's often look at individual participants' beliefs and decisions.
- E's typically focus on participant actions that affect aggregate outcomes (e.g., market price) and payoffs, while P's often focus on stated beliefs.
- E's typically remove context from their settings, while P's rarely do. For example, a test of probability assessment by an E would be described in terms of balls drawn randomly from urns with replacement, while a P would more likely present a question like "how likely is it that the population of Miami is greater than the population of Paris?"
- E's often include extremely complex tables of raw data and econometric estimates of parameters, while P's will usually provide only summary statistics (such as means, variances, F- and t-statistics, and p-values).

Demonstrations vs. Experiments

Understanding the causes of these differences provides an excellent vehicle for understanding how to conduct better experiments in finance. Camerer (1997argues that some of the differences are driven by the variation in psychological and economic theory. In particular, compensation plays a large role in E studies because the theory assumes there is a payoff to be maximized. Much P research needs only to ensure that participants pay attention to the task and take it seriously, therefore requiring minimal or no payment. Similarly, economic theory is usually devoid of context, so therefore most E's see little benefit to enhancing the numerical settings with irrelevant content and, in fact, many see costs to doing so (because the context might matter for some unknown reason). In contrast, the context in which decisions are made often largely drives psychological models. In general, E's favor precise mathematical formulations of a concise abstraction of reality, while P's prefer verbal descriptions that illustrate the big picture. While the two groups have significant stylistic differences, Camerer argues that, substantively, they are quite similar.

While differences between economic and psychological theory drive some of the differences, a far more important force is likely to be that many experiments in economics are not actually experiments—they are demonstrations. A defining characteristic of an experiment is that the researcher manipulates a single variable, while holding all other aspects of the setting constant. Rarely does an experiment appear in a top psychology journal that does not manipulate a variable. By contrast, many papers published by E's, even in top journals, contain no manipulated variables.

Perhaps the most famous of these is Smith, Suchanek, and Williams (1988), which conducted a series of markets for a security that paid a constant dividend D for each of N periods. Although the value declines from ND to 0 over the periods, Smith et al. conjectur that market prices might form a bubble, with participants buying for more than the fundamental value, in hopes of reselling at a higher price in the future. Smith et al. (p. 1129) report a ‘success’ as follows: “We observed our first full scale bubble – a boom followed by a market crash. Replication of this experiment (19x) with experienced subjects failed to extinguish a boom-bust pattern of trading.”

This paragraph suggests that the study is more a demonstration of a behavior than an experiment. The authors have shown that bubbles arise in the setting they created. Demonstrations have a long tradition in the physical sciences. For example, chemistry researchers will often publish papers in which they describe how they created a molecule that had not been previously synthesized. The authors will do what is necessary to create the molecule, through months or years of trial and error, and then report the method ended up working. Trial and error often takes the form of controlled experiments, but researchers rarely report the results of those experiments, While the success of the demonstration may be consistent with existing theory, and prompt conjecturing refinements to existing theory, the demonstration does not emphasize controlled manipulations used to test a specific theory in a focused way.

In contrast, a true experiment identifies a variable that theory predicts will alter behavior, and then manipulates that variable to test the theory. For example, Smith et al. (1988) conjecture that, even if traders are rational and have common initial beliefs about the value of the asset, they may be uncertain about how other traders will react to the same information. Therefore, speculation can arise if traders think they can profit by trading with someone else who interprets the market information differently, and thus bubbles and crashes can occur. A P would likely suggest an experiment in which half of the participants were randomly assigned to a market in which speculation is allowed, while the remainder is assigned to a market in which they are solely a buyer or a seller, thus making speculation impossible. Lei, Noussair, and Plott (2001) conducted such an experiment, which shows that bubbles can arise even when speculation is not allowed. Traders do in fact make judgment errors and the theory that traders are rational and bubbles are caused strictly by speculation cannot be the whole story. By conducting experiments with controlled manipulation rather than simply demonstrating a market characteristic, true sources of causality for that characteristic can be better identified. The introduction of a controlled manipulation dramatically changes the costs and benefits of various choices in design and analysis. Consider, for example, the role of context. For psychologists, the benefits of context are to provide a natural setting for a decision that is similar to one that participants might make in real life.

Experimental economics often worry about the extra-theoretical implications of context. Yet, recognizing that concerns about such 'baggage' are far more serious for a study that has no manipulations is important. Without a manipulation, *any* aspect of the setting could be important in driving bubble formation. Thus, the presence of context makes attributing the presence or absence of bubbles to economic factors very difficult. While concerns about 'baggage' are a problem, context is only one relatively obvious noneconomic factor that might drive results. The color and temperature of the room, the background and intelligence of the participants, and details of the trading interface and the noise generated by trading could all affect pricing. Moreover, economic factors not being considered by the research could also matter, such as the length of trading periods, nominal price levels, or the nature of the pricing mechanism (e.g., double auctions vs. clearinghouse markets). Quite simply, Smith et al. (1988) can only provide conjectures of why they observe bubbles and any deviation from their setting may change their results.

In contrast, imagine that Smith et al. (1988) had in fact manipulated a variable such as the amount of cash in the market. In fact, Caginalp, Porter, and Smith (2001) conducted a study like this by indicating that large cash endowments do indeed make bubbles more likely. Thus, assume that this alternative to Smith et al. would generate a similar result. In this case, finding such a difference across treatments to be driven by the features of the two settings that are held constant would be very unlikely. As a result, there is little reason to be concerned about the fact that context and meaningful labels would detract from the inferences one can take from the study. After all, the context and labels do not drive bubbles when cash endowments are low. Thus, context and labels by themselves are unlikely to drive bubbles. Also highly unlikely is that the presence of context and labels drives the difference between the two treatments. This would require an interaction between context and cash endowments that, at least at first glance, does not seem plausible.

The power of controlled manipulations also provides experimentalists with defenses against many common criticisms, such as the types of participants in the study and the levels of compensation. While more training and greater incentives might reduce the formation of bubbles, the level of training and incentives in this particular experiment is very unlikely to explain differences across settings in which training and incentives are identical. Thus, those reviewing experiments should be extremely cautious in criticizing experiments in which they believe participants had too little experience or were not paid enough, unless they have a specific reason to believe that experience or incentives will interact with the manipulated variable.

Analyses

The use of controlled manipulation also partly explains why statistical tests differ so much between E and P studies. Analyses of experiments are far simpler than analyses of archival data for the simple reason that the experimental design eliminates many of the problems that econometricians face using data from uncontrolled settings. The alternative Smith et al. (1988) study proposed above could provide strong evidence on the link between cash endowments and bubbles using simple statistics. For example, one could measure a bubble as the time averaged excess of price over fundamental value in each market, and then conduct a t-test of the difference in means across the two settings. Of course, the usual caveats apply regarding the normality of the dependent variable and the similarity of the variances across cells, so using a nonparametric test might be better. However, testing the theory at hand does not require sophisticated econometrics—good experimental designs are usually followed by simple analyses. This is because the researcher designed the entire experiment to make a small set of particular tests possible in a simple and clean way.

Because demonstrations cannot rely on controlled manipulations, they must follow one of two methods of analysis. The first is to assess how similar the behavior in their setting captures the predictions of theory. This leads to extensive focus on parameter estimation. Theory predicts that price should equal fundamental value in each period, so the mean measure of excess pricing should be zero. Note that this is necessarily a parametric test, so concerns about deviations from normality and outliers are more severe than in a test of the sign of a difference. As a result, such studies traditionally provide tables and figures to permit reviewing the raw data.

The second approach to data analysis in a demonstration is to search for associations within the various dependent variables. For example, Smith et al. (1988) conduct various tests of the adaptive nature of prices, running regressions of the form

$$\bar{P}_t - \bar{P}_{t-1} = \alpha + \beta(B_{t-1} - O_{t-1}), \qquad \beta > 0. \tag{4}$$

Here, P_t is the price at time t and $B_{t\text{-}1}$ — $O_{t\text{-}1}$ is the excess demand at time t-1. Such analyses have some advantages over traditional econometric studies of data from naturally-occurring markets because they can incorporate dependent and independent variables that would be unobservable outside the laboratory (such as investor beliefs and fundamental value). Still, they face the challenge that they rely on measured (rather than manipulated) independent variables. As a result, such analyses are prone to criticism for correlated omitted variables and self-selection biases.

Experimentalists should not avoid analyses relying on measured independent variables. Such analyses are extremely useful in understanding why people behave as they do in settings with many people engaging in repeated interaction. Researchers should rely on these econometric methods as secondary analyses that serve to support treatment effects, which should be the primary focus of the experiment.

SUMMARY AND CONCLUSIONS

Experiments are an underused method in finance and have natural advantages for behavioral finance. Experiments can provide a useful means to circumvent several common econometric issues, such as omitted variables, unobserved variables, and self-selection. Experiments can extend the theoretical models they test by relaxing various assumptions or examining settings that are too complex to be addressed analytically. Whether or not theoretical predictions are clearly known in advance, experiments are most informative when they rely on controlled manipulation, which is the source of their inferential power.

DISCUSSION QUESTIONS

- How can experimental examinations of extremely simple settings shed light on the behavior of far more complex financial settings?
- Assume that an economic model predicts that, under a given set of assumptions, manipulating independent variable X will increase dependent variable Y. What can possibly be learned from an experiment that imposes the set of assumptions and confirms the predictions of the experiment?
- Assume that there is no tractable economic model that can clearly predict that manipulating independent variable X will increase dependent variable Y. How can one interpret the results of an experiment demonstrating such an effect, as it cannot be said to confirm or refute a prediction?
- Why do the authors of this chapter argue that experimentalists in economics and finance worry far too much about the influence of features of experimental tasks that are extraneous to economic models, such as the labels used to describe players and actions?

REFERENCES

Benartzi, Shlomo, and Richard H. Thaler. 1999. "Risk Aversion or Myopia? Choices in Repeated Gambles and Retirement Investments." *Management Science* 45:3, 364-381.

Berg, Joyce E., Lane A. Daley, John W. Dickhaut, and John R. O'Brien. 1986. "Controlling Preferences for Lotteries on Units of Experimental Exchange." *Quarterly Journal of Economics* 101:2, 281-306.

Bhojraj, Sanjeev, Robert J. Bloomfield, and William B. Tayler. 2009. "Margin Trading, Overpricing, and Synchronization Risk." *Review of Financial Studies* 22:5, 2059-2085.

Bloomfield, Robert J. 1996. "Quotes, Prices and Estimates in a Laboratory Market." *Journal of Finance* 51:5, 1791-1808.

Bloomfield, Robert J., and Jeffrey W. Hales. 2006. "Should Analysts Act Alone?" Johnson School Research Paper No. 39-06. Available at http://ssrn.com/abstract=941494.

Bloomfield, Robert J., Robert Libby, and Mark W. Nelson. 1996. "Communication of Confidence as a Determinant of Group Judgment Accuracy." *Organizational Behavior and Human Decision Processes* 68:3, 287-300.

Bloomfield, Robert J., and Maureen O'Hara. 1999. "Market Transparency: Who Wins and Who Loses." *Review of Financial Studies* 12:1, 5-35.

Bloomfield, Robert J., and Maureen O'Hara. 2000. "Can Transparent Markets Survive?" *Journal of Financial Economics* 55:3, 425-459.

Bloomfield, Robert J., Maureen O'Hara, and Gideon Saar. 2005. "The 'Make or Take' Decision in an Electronic Market: Evidence on the Evolution of Liquidity." *Journal of Financial Economics* 75:1, 165-200.

Bloomfield, Robert J., Maureen O'Hara, and Gideon Saar. 2009. "How Noise Trading Affects Markets: An Experimental Analysis." *Review of Financial Studies* 22:6, 2275-2302.

Bloomfield, Robert J., William B. Tayler, and Flora H. Zhou. 2009. "Momentum, Reversal, and Uninformed Traders in Laboratory Markets." *Journal of Finance*, forthcoming.

Bossaerts, Peter, and Charles R. Plott. 2004. "Basic Principles of Asset Pricing Theory: Evidence from Large-Scale Experimental Financial Markets." *Review of Finance* 8:2, 135-169.

Cadsby, Charles B., Murray Z. Frank, and Vojislav Maksimovic. 1990. "Pooling, Separating, and Semiseparating Equilibria in Financial Markets: Some Experimental Evidence." *Review of Financial Studies* 3:3, 315-342.

Caginalp, Gunduz, David Porter, and Vernon L. Smith. 2001. "Financial Bubbles: Excess Cash, Momentum, and Incomplete Information." *Journal of Behavioral Finance* 2:2, 80-99.

Camerer, Colin F. 1987. "Do Biases in Probability Judgment Matter in Markets? Experimental Evidence." *American Economic Review* 77:5, 981-997.

Camerer, Colin F. 1997. "Rules for Experimenting in Psychology and Economics, and Why They Differ." In Wulf Albers, Werner Güth, Peter Hammerstein, Benny Molduvanu and Eric van Damme (eds.), *Understanding Strategic Interaction: Essays in Honor of Reinhard Selten*, 313-327. Berlin, New York: Springer-Verlag.

Flood, Mark D., Ronald Huisman, Kees G. Koedijk, and Ronald J. Mahieu. 1999. "Quote Disclosure and Price Discovery in Multiple-Dealer Financial Markets." *Review of Financial Studies* 12:1, 37-59.

Forsythe, Robert, and Russell J. Lundholm. 1990. "Information Aggregation in an Experimental Market." *Econometrica* 58:2, 309-347.

Forsythe, Robert, Russell Lundholm, and Thomas Reitz. 1999. "Cheap Talk, Fraud, and Adverse Selection in Financial Markets: Some Experimental Evidence." *Review of Financial Studies* 12:3, 481-518.

Friedman, Daniel. 1984. "On the Efficiency of Experimental Double Auction Markets." *American Economic Review* 74:1, 60-72.

Ganguly, Ananda R., John H. Kagel, and Donald V. Moser. 1994. "The Effects of Biases in Probability Judgments on Market Prices." *Accounting, Organizations and Society* 19:8, 675-700.

Gode, Dhananjay K., and Shyam Sunder. 1993. "Allocative Efficiency of Markets with Zero-Intelligence Traders: Market as a Partial Substitute for Individual Rationality." *Journal of Political Economy* 101:1, 119-137.

Kachelmeier, Steven J. 1996. "Discussion of 'Tax Advice and Reporting Under Uncertainty: Theory and Experimental Evidence'." *Contemporary Accoutning Research* 13:1, 81-89.

Lei, Vivian, Charles N. Noussair, and Charles R. Plott. 2001. "Nonspeculative Bubbles in Experimental Asset Markets: Lack of Common Knowledge of Rationality vs. Actual Irrationality." *Econometrica* 69:4, 831-859.

Lundholm, Russell J. 1991. "What Affects the Efficiency of a Market? Some Answers from the Laboratory." *Accounting Review* 66:3, 486-515.

Myers, Stewart C., and Nicholas S. Majluf. 1984. "Corporate Financing and Investment Decisions When Firms Have Information That Investors Do Not Have." *Journal of Financial Economics* 13:2, 187-221.

O'Brien, John, and Sanjay Srivastava. 1991. "Dynamic Stock Markets with Multiple Assets: An Experimental Analysis." *Journal of Finance* 46:5, 1811-1838.

Plott, Charles R., and Shyam Sunder. 1982. "Efficiency of Experimental Security Markets with Insider Information: An Application of Rational-Expectations Models." *Journal of Political Economy* 90:4, 663-698.

Plott, Charles R., and Shyam Sunder. 1988. "Rational Expectations and the Aggregation of Diverse Information in Laboratory Security Markets." *Econometrica* 56:5, 1085-1118.

Seybert Nicholas, and Robert J. Bloomfield. 2009. "Contagion of Wishful Thinking in Markets." *Management Science* 55:5, 738-751.

Smith, Vernon L. 1976. "Experimental Economics: Induced Value Theory." *American Economic Review* 66:2, 274-279.

Smith, Vernon L. 1982. "Microeconomic Systems as an Experimental Science." *American Economic Review* 72:5, 923-955.

Smith, Vernon L., Gerry L. Suchanek, and Arlington W. Williams. 1988. "Bubbles, Crashes, and Endogenous Expectations in Experimental Spot Asset Markets." *Econometrica* 56:5, 1119-1151.

Thaler, Richard H., and Shlomo Benartzi. 2004. "Save More Tomorrow™: Using Behavioral Economics to Increase Employee Saving." *Journal of Political Economy* 112:S1, S164-S187.

Tosi, Henry L., Jeffrey P. Katz, and Luis R. Gomez-Mejia. 1997. "Disaggregating the Agency Contract: The Effects of Monitoring, Incentive Alignment, and Term in Office on Agent Decision Making." *Academy of Management Journal* 40:3, 584-602.

Tversky, Amos, and Daniel Kahneman. 1974. "Judgment Under Uncertainty: Heuristics and Biases." *Science* 185:4157, 1124-1131.

ABOUT THE AUTHORS

Robert Bloomfield is Nicholas H. Noyes Professor of Management and Accounting at the Johnson Graduate School of Management at Cornell University. He has published experimental research in the *Journal of Finance*, *Journal of Financial Economics*. and *Review of Financial Studies*, as well as in top journals in accounting, economics and psychology. He is the Director of the Financial Accounting Standards Research Initiative, a program of the Financial Accounting Standards Board.

Alyssa Anderson is a Ph.D. student in finance at the Johnson Graduate School of Management at Cornell University. She holds a BA from Dartmouth College. Her research interests include market microstructure, experimental finance and behavioral finance.