The Journal of Behavioral Finance
2003, Vol. 4, No. 1, 41–48

Economic Data, Conventions, and Replication in Laboratory Asset Markets

Matthew J. Anderson and Eugene R. Blue

Data extracted from naturally occurring markets and other economic environments often suffer from problems like data confounds and intercorrelations. We report results from a series of experimental markets that suggest some of the data problems can be overcome by using experimental techniques. We use predetermined (videotaped) draws to replicate results from prior research. This is contrary to the conventional wisdom in the literature, which holds that draws should be "live." Our results suggest that the price and allocation behavior of markets can be replicated using predetermined draws to initiate trading. Furthermore, the primary strength of the experimental method—control—is maintained.

Empirical economic analyses often include implicit or explicit caveats about data problems. Data may be unavailable, or it may be altered by various forces, which renders tests and theories incomplete. Data problems arise for many reasons, but problems often start with the data extraction process. Data from naturally occurring economic environments (such as markets) are often extracted at single points in time from essentially dynamic systems. It may be unclear where in a cycle or process a measurement was taken. Measurements may also be taken from multiple systems or locations, which may yield data with a laundry list of problems.

For example, in many settings multiple variables may be present simultaneously, making individual variables difficult to disentangle and identify. This may include variables not central to the question of interest. Regardless of their importance, however, variables are typically correlated and difficult to measure. Such problems generally introduce noise and/or bias into the analysis. The correlated omitted variable is the bane of many economic analyses. Measurements may also ignore factors important for economic measurement, such as inflation effects.

Statistical analysis may be significantly affected by factors like these. For example, data correlations, between variables or over time with the same variable, may result in a decreased ability to draw inferences about outcomes or the nature of relationships among variables. Analysis problems may be exacerbated because the relative number of observations, and the number of variables under test, are likely to be reduced where significant correlations exist.

Laboratory or experimental markets provide a mechanism for mitigating some of these problems. In particular, laboratory techniques provide analysts with some control over the variables under test and over the data extracted. We use experimental techniques to replicate a set of results from the literature, while also addressing some of the concerns raised by the use of laboratory techniques. We address issues from research related to the invisible hand argument of Adam Smith.

In his original treatise, *The Wealth of Nations*, Adam Smith makes perhaps the most well-known claim in economic theory:

> But it is only for the sake of profit that any man employs his capital...in the support of that industry of which the produce is likely...of...greatest value.... He is in this led...by an invisible hand to promote an end which was no part of his intention (Hirshleifer [1976]).

Adam Smith implies that market participants are motivated only by potential profit, and have no knowledge of their real effect on the market. In seeking to better themselves, market participants are somehow led to make everyone better off.

Since Adam Smith's time, there have been numerous efforts to understand the role of individual actions

Matthew J. Anderson received his Ph.D. and MBA in Accounting from Michigan State University, where he is currently on the faculty. His research is concerned with the use of accounting information in firm valuation at the individual level, and related, the role or interplay between or among individual expertise, individual decisions and actions, and market level outcomes. He has published in a variety of journals.

Eugene R. Blue received his Ph.D. in Accounting from Michigan State University. He is currently on the faculty at Howard University. His research is concerned with the use of accounting information in firm valuation. He has also investigated the role or interplay between individual decisions and actions, and market level outcomes.

Requests for reprints should be sent to: Matthew J. Anderson, Ph.D., Accounting Department, N229 Business College Complex, Michigan State University, East Lansing, MI 48824. Email: ander130@msu

(characteristics) on market prices and allocations, using factors such as participant market experience, knowledge, and endowment (wealth) level, among others. Such assessments are difficult in naturally occurring markets, because observing individual traders is problematic for many reasons, and costly. Furthermore, economic theory generally suggests that such analyses are of limited importance regardless; the assumption is that in most economic environments individuals' characteristics are similar to random noise. The effects of such factors cancel each other out when participants interact and actions are examined as outcomes at the aggregate market level.

Consistent with Smith's invisible hand observation, economic theorists also argue that the structure of economic environments like markets actually causes behavior that is in line with the theory. This outcome obtains regardless of the intent, motivation, or actions of individual traders. The argument again suggests that individual characteristics can be safely ignored, and that markets work in spite of individual idiosyncrasies.

Theory and Evidence

Substantial evidence supports the basic contention of these arguments. For example, Jamal and Sunder [1996], using simulations, extend the work of Becker [1962], who suggested that not all traders in a market must be rational to observe theory-consistent prices and allocations. Jamal and Sunder demonstrate that markets populated by biased "statistical" traders invariably result in economic outcomes over time.

Nonetheless, the question of whether individual actions or attributes matter to market outcomes has not been resolved. For example, human subjects (and markets) often have limited time horizons, and may not be "consistently inconsistent" (as is a computer simulation). Experimental research techniques like laboratory asset markets have been used to assess issues like the effect of individual bias on market prices and allocations. Laboratory markets offer two factors of substantive importance: control and replication (Friedman and Sunder [1994]). Variables can be applied and measured with relative precision, and their effect on experimental participant behavior (and on other variables) can be assessed.

A classic replication assumes that all critical features of an experimental setting are held constant across iterations (Keppel [1982, p. 75]). As stated by Friedman and Sunder [1994], replication is central to any experiment. Scientific inquiry depends upon replication to alter the content of the body of knowledge.

One of the issues related to replication is how to determine the critical features of a given experimental setting. Economic experimentalists use devices such as bingo cages and computer algorithms to generate random signals that initiate trading behavior in markets. According to Friedman and Sunder [1994], this ensures that 1) the distribution of possible outcomes is what was originally described to the subjects, and 2) they are able to see and understand how procedures are carried out. Subjects gain a sense of ownership and participation, and any fears regarding deception by the experimenters are removed (Friedman and Sunder [1994]).

Note that the use of bingo cages and computer algorithms (live draws) makes it unlikely that the time order and price structure of the experimental environment will be the same across sessions. However, these features of the market environment are assumed to be irrelevant to theoretical prices, allocations, and control.

Consistent with Friedman and Sunder, economic experimentalists have also been concerned about what constitutes a potentially deceptive practice. For example, Davis and Holt [1993] advise experimenters to show [new] experimental subjects the cash with which they will be paid in order to maintain control of the environment and avoid any presumption of trickery (p. 61). They say that "Many economists believe that deception is highly undesirable in economics experiments...and the results of [such] experiments should not be published" [1993, p. 24]. Davis and Holt also state that failure to follow prescribed procedures "pretty much renders results meaningless."[1]

Friedman and Sunder [1994, p. 56] apply this line of reasoning directly to failure to use the live draw convention: "If a researcher presents a carefully chosen predetermined sequence of numbers to subjects as a randomly drawn series, the subjects have no way of finding out, even ex post, that they have been lied to...." Note that it is not necessary to present such a sequence as random, and it is not clear why subjects might infer any deception.[2]

Nonetheless, consequences of the "live procedures" convention include problems determining what features are critical to results, and what costs are involved in making such judgments. Note that one can interpret the concerns about the nature of experimental features as proof of the importance of individual features to outcomes—even in market settings.

The use of live draws also potentially impacts control of other experimental factors. For example, different experimental environments may result from errors or simple alterations in procedures (Orne [1972]). We next discuss the potential interplay between the use of predetermined draws and other factors as an empirical question, using results from an existing study and a set of additional markets.

Evidence

Anderson and Sunder [1995] examine whether trading experience in a real market can transfer to a similar laboratory environment. In many settings, experience in one area does not readily transfer to another, for example, management success in one industry does not guarantee success in another. Anderson and Sunder compared the market price and allocation behavior for two types of traders: experienced financial analysts and MBA students. The financial analysts had jobs in the investment community, and all had traded securities in real markets such as the New York Stock Exchange for themselves and clients. The MBA students were finance majors at a large midwestern university. All had recent economic and statistical training, and were generally aware of investment and trading theories. Price and allocation behavior was compared across several laboratory markets.

Market participants were assigned one of two trader types, denoted I and II, based on an endowment of certificates that paid state-dependent payoffs (two possible states). Type I traders would receive a maximum payoff of 550 if state 1 occurred, and 50 if state 2 occurred. Type II traders would receive 75 if state 1 occurred, and 375 if state 2 occurred. Traders observed a randomly drawn (with replacement) sample of five balls, drawn from one of two hidden bingo cages, representing the different states of nature—analogous to a bull or bear market in a real context. The subjects were required to infer the state of nature from the composition of the observed sample (much like real traders try to infer the economic state of nature from variables such as the index of leading indicators). Having observed the sample and drawn an inference, subjects then traded their state-dependent claims.

The payoffs and predicted prices, given a model for predicting price, are shown in Table 1, panels A and B. Both an optimal (Bayesian; Bay) and a heuristic (representativeness; NBR) model were used to predict price and allocation outcomes. The price prediction of the representativeness model is simply the most extreme price for each possible state given the signal.

Figure 1 shows examples of the relative deviation of actual price from model predictions. Note that the behavior of the experienced traders was consistent with the theoretical results, while the behavior of the MBA students consistently deviated from theoretical predictions. That is, the deviations from theoretical predictions trended to zero for the experienced traders, but not for the MBA students. The heuristic (NBR) model predicted the students' behavior very well, which is consistent with prior studies (Camerer [1987]).

Anderson and Sunder's [1995] results are effectively an anomaly in the sense that many studies using human subjects have failed to show positive transfer of market experience. For example, Burns [1985] found that experienced wool traders did not outperform students in a simulated natural wool market. Other studies in psychology and elsewhere have suggested that experienced businesspeople and others are just as susceptible to certain errors as the most naïve students (see, for example, Kahneman, Knetsch, and Thaler [1991]).

Using predetermined draws, we replicated two sequences of draws used to initiate the trading in Anderson and Sunder [1995]. We used only inexperienced traders as subjects. For Sequence 1, Anderson and Sunder reported relatively Bayesian (economic) price outcomes. This sequence was relatively less likely to occur given the particular probability structure. However, for Sequence 2, Anderson and Sunder reported

Table 1. *Payoffs and Predicted Prices*

Panel A: Signal Sequences, Trader Types, and Market Payoffs for Realized States

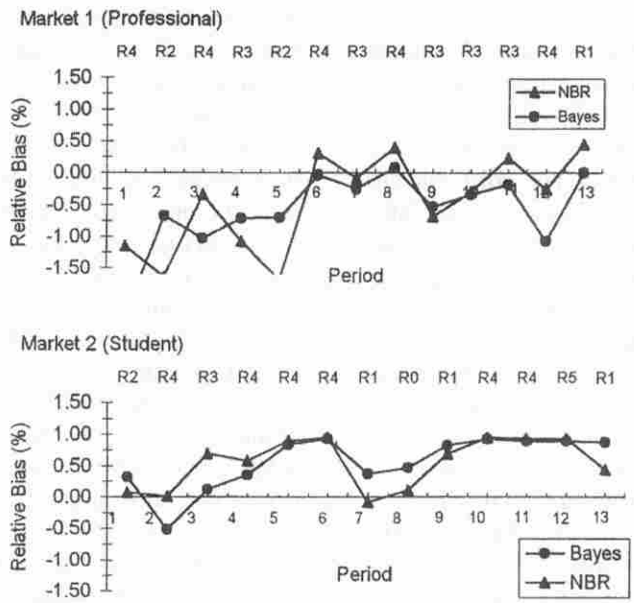
	Payoffs	
Trader Type/State	**X**	**Y**
Type I	550	50
Type II	75	375

Panel B: Predicted Market Prices, State, and Trader Type Expected To Hold Certificates after Trading Given a Signal R[#]

Signal	**R0**	**R1**	**R2**	**R3**	**R4**	**R5**
Price - Bay	360	335	288	312	423	493
Trader - Bay	II	II	II	I	I	I
E[State] - Bay	Y	Y	Y	X	X	X
Price - NBR	375	375	375	375	550	550
Trader - NBR	II	II	II	II	I	I
E[State] - NBR	Y	Y	Y	Y	X	X
Prob (S|R[#])	0.94	0.87	0.71	0.52	0.75	0.89

Note: R = red; R[#] = the # of red balls in a sample. NBR: No base rate (representatives). Bay: Bayesian.
Sequences were as shown below, with R denoting the number of red balls in a given observed sample R[#].
One: R4 R2 R4 R3 R2 R4 R3 R4 R3 R3 R4 R1 R3.
Two: R2 R4 R3 R4 R4 R4 R1 R0 R1 R4 R4 R5.
Market traders were of two types, based on differing payoffs for state X or Y. Payoffs are expressed in fictional "francs."

FIGURE 1
Relative Bias Results



Note: Relative bias results from Anderson and Sunder [1995]. The pattern of changes with signals can be contrasted with results from our replication (Figure 4).

representative, or too high, price outcomes given Bayesian expectations. This sequence was relatively more likely to be witnessed by traders. We ran two laboratory market experiments for each sequence, for a total of four markets. We also compared the outcomes to those reported by Anderson and Sunder [1995].

The sequences consisted of videotaped draws of a sample of five balls, composed of different combinations of red (R) and white (W) balls. Balls were drawn with replacement from one of two hidden bingo cages, representing potential states of nature (see Figure 2). We expected that the markets with differing sequences of draws would lead to substantively different market prices.

Subjects

Subjects were either senior undergraduate or graduate students who had taken several courses in finance. None had previously participated in a laboratory market experiment, and almost all had no practical experience buying, selling, or trading securities in a "real" market. Subjects were paid from $20 to $60 for the three-hour experimental session, the same payoffs used by Anderson and Sunder [1995]. Earnings related directly to what the traders earned by buying and selling market "certificates."

Experimental Procedures

Each experimental session began with a series of exercises designed to familiarize subjects with the market procedures, and each trader was given two trading certificates per period. To initiate trading each period, subjects were shown a videotape of the sample draw of five balls used to infer cage "X" or "Y." Each market was conducted for twelve to thirteen five-minute periods. All transactions were conducted using "francs," a fictional market currency.

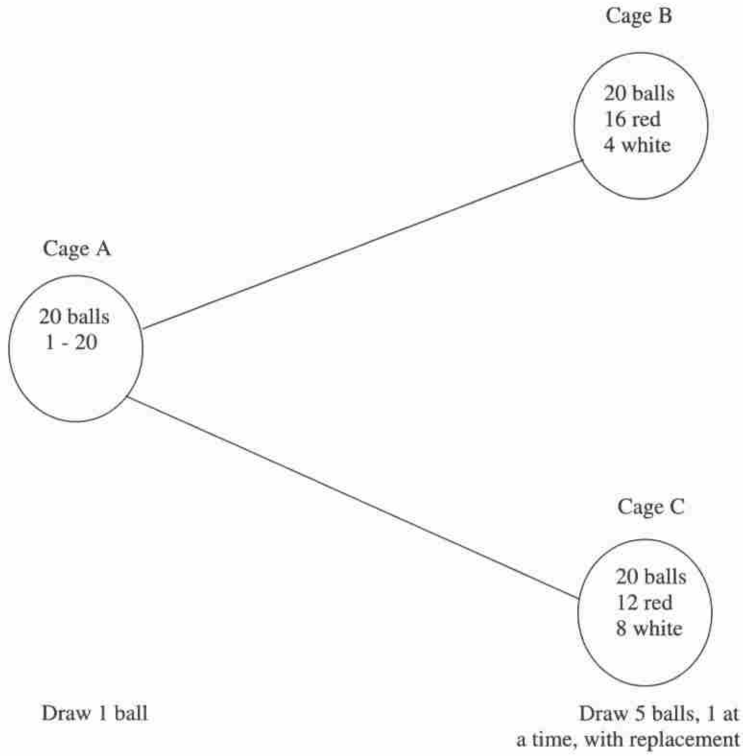
At the end of the trading periods, subjects completed a debriefing exercise and questionnaire, and were paid in cash based on their earnings. Expected price outcomes under Bayes' rule for each possible sample draw are shown in Table 1 (assuming a random sampling process). The event sequence is shown in Figure 2.

Assessment of Prices and Error

We use prices and relative error as the primary measures to assess the replication of market outcomes (Anderson and Sunder [1995]). We use a relative measure because the extent to which market prices can deviate from expectations within a period varies with the observed signal.[3] We also report statistical correlation results and a measure of the degree to which the market allocated goods efficiently.

Relative error (*RE*) is essentially percentage change. It is defined as the deviation of the actual price from the Bayesian expected price, divided by the difference between the Bayesian expected price and the maximum

FIGURE 2
Event Sequence



Note: The state of nature was determined in sequence. A numbered ball was first selected from cage A. Given the number of the selected ball, a second bingo cage (B or C) was then selected and a sample of five balls drawn, one at a time, with replacement. This sample was used to infer the state of nature, which was announced *after* trading for the period.

possible price (liquidation value or positive error ceiling):

$$RE = 100([(P_a - P_B)]/[(P_M - P_B)]) \quad (1)$$

where P_a is the actual market price, P_B is the Bayesian expected price conditional on the observed signal, and P_M is the maximum payout to the trader type expected to hold the asset at period end, given the observed signal. This metric equals zero if market prices conform perfectly to Bayesian expectations. Its upper bound is a function of market outcomes. A value of "1" suggests a level of error of 100%. We use the mean actual price for the last three trades of each period in this analysis (Anderson and Sunder [1995]).

We also report market allocative efficiency (AE) (Jamal and Sunder [1996]), defined as follows:

$$AE = \frac{[EV_I \mid Sample]N_I + [EV_{II} \mid Sample]N_{II}}{Max[EV_I, EV_{II}][N_I + N_{II}]} \quad (2)$$

where EV is expected value, N is the number of traders, and Max denotes the maximum of the two possible dividend values for the trader types (I and II).

Results

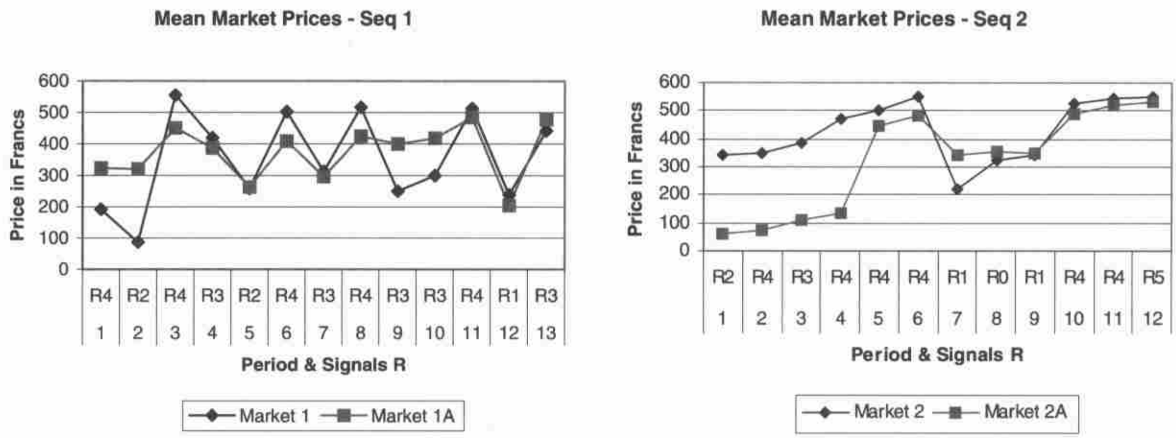
Replication, Data Consistency, Prices, and Relative Error Measures

Analysis centered around whether the observed price and relative error outcomes were internally consistent and replicated. Were the same results obtained when a sequence was repeated (e.g., Sequence 1), and did outcomes differ across the two sets of sequences? Finally, we compared our outcomes to those reported in Anderson and Sunder [1995]. Results for this analysis are based on activity from period 5 to the end of the market in each case (Friedman and Sunder [1994]).

Mean market prices are shown in Figure 3. Panel A of Figure 3 depicts prices for Markets 1 and 1A, where the traders observed Sequence 1. Panel B depicts prices for Markets 2 and 2A, where the traders observed Sequence 2. First, we assert that the markets pass the "interocular trauma test (eyeball)" of Savage (Friedman and Sunder [1994, p. 92]). Visually, subjects in a given set of markets appear to be reacting to "draws" in similar ways.

This conclusion is borne out by analysis of the market prices. Prices in Markets 1 and 1A are not statistically different from each other (p = 0.83), and are

FIGURE 3
Mean Market Transaction Prices for Each Period of Markets 1 and 1A (Sequence 1 Markets) and Markets 2 and 2A (Sequence 2 Markets)



Panel A Panel B

Note: R(#) is the number of red balls in the observed sample of five balls.

highly correlated ($r = 0.78$; $p = 0.01$). Furthermore, for Markets 1 and 1A, the prices move in the same direction as those reported in Anderson and Sunder [1995], and do not differ statistically ($p > 0.15$).

Similar results hold for Markets 2 and 2A. Market prices across the two markets do not differ ($p = 0.60$), and prices are highly correlated ($r = 0.83$; $p = 0.01$). As with Markets 1 and 1A, the outcomes observed do not differ statistically from those reported in Anderson and Sunder [1995] ($p > 0.28$).

Prices from the two sequences (Markets 1 and 1A versus Markets 2 and 2A) were significantly different from each other ($p = 0.059$). In these markets, the price outcomes of inexperienced traders mimic those of the more experienced traders, as well as the inexperienced traders of Anderson and Sunder [1995]. The results suggest that the sequences of draws used to initiate market activity matters.

Prices in Markets 2 and 2A tend to stabilize at levels well above Bayesian expectations, consistent with the behavior reported in Anderson and Sunder [1995] and other studies (e.g., Camerer [1987]). Prices are relatively more variable in Markets 1 and 1A than in Markets 2 and 2A.

Outcomes using the relative error measure (REM) are depicted by market in Figure 4. The REM results are similar to those obtained for prices: Subjects appear to respond to signals in a consistent fashion based on correlation analysis. For Markets 1 and 1A (panel A), $r = 0.88$ ($p = 0.00$); for Markets 2 and 2A (panel B), $r = 0.67$ ($p = 0.07$).

The REM outcomes in Markets 1 and 1A are not different from each other ($p = 0.83$), and are consistent with the Anderson and Sunder [1995] results ($p > 0.13$). There is also no difference between Markets 2 and 2A ($p = 0.79$), and outcomes are again consistent with Anderson and Sunder ($p > 0.75$). The REM outcomes for Markets 1 and 1A do differ from those of Markets 2 and 2A ($p = 0.03$).

Recall that the level of error in these graphs would tend to zero to the extent that trading behavior is consistent with Bayesian expectations. For Markets 1 and 1A, the level of the REM outcomes often approaches zero and is quite variable. In contrast, the level of the REM outcomes of Markets 2 and 2A is relatively stable and approaches "1" by the end of each market. Taken together, these price and relative error results suggest that 1) different behavior is occurring with the different sequences of draws, 2) behavior is replicated internally, and 3) behavior is consistent with the Anderson and Sunder [1995] results. This suggests that the use of predetermined draws of the samples used to generate market activity does not by itself significantly alter market prices.

Market Efficiency

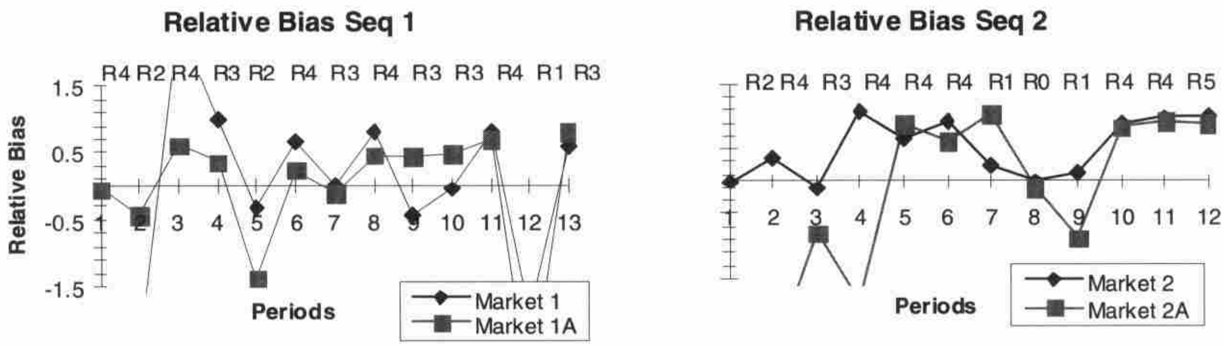
Figure 5 shows market allocative efficiency measures. Both sets of markets (panels A and B) are efficient at a level consistent with the prior literature despite the fact that the market prices are not consistent with competitive outcomes (Smith [1994]). Ex post, the "right" traders appear to hold the certificates traded in the market, which implies that subjects did not misperceive the underlying distribution and make wrong state choices.

Conclusion

One of our primary objectives in this article was to demonstrate the degree of control analysts could obtain

FIGURE 4

Relative or Percentage Bias ([Actual Price–Predicted Price] / [Period's Payoff–Predicted Price]) Each Period, Markets 1 and 1A (Sequence 1 Markets; Panel A) and Markets 2 and 2A (Sequence 2 Markets; Panel B)



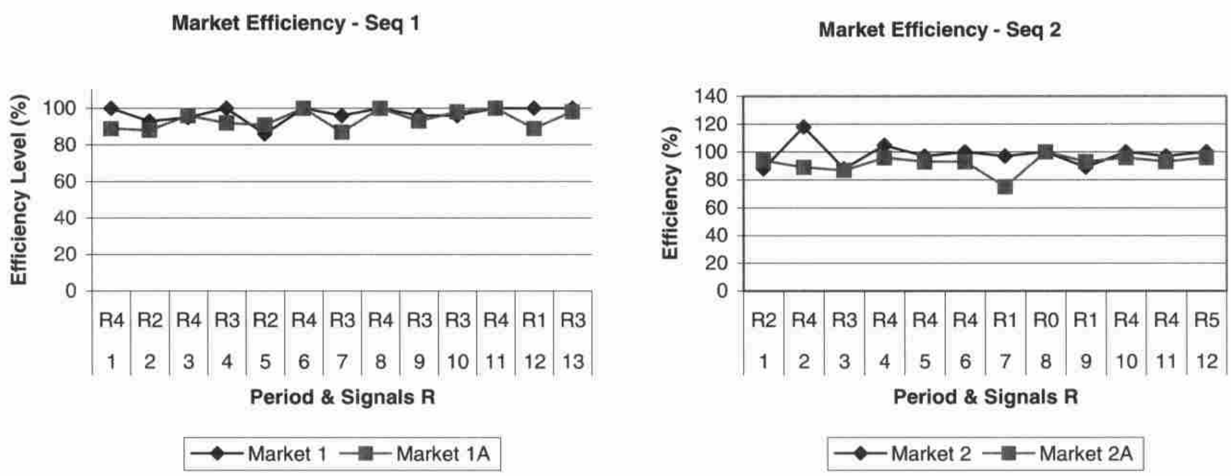
Panel A

Panel B

Note: R[#] = red balls in the sample of five balls. NBR: No base rate; representativeness model Bayes: Bayesian model.

FIGURE 5

Market Allocative Efficiency by Periods for Markets 1 and 1A and Markets 2 and 2A



Panel A

Panel B

Note: Measure reflects the extent to which predicted traders held the certificates at the end of trading. R(#) = the number of red balls in a sample of five balls.

over data extraction by using experimental techniques. We demonstrated that 1) behavior from an existing study—Anderson and Sunder [1995]—could be replicated, and 2) market behavior was not significantly affected by the use of predetermined (videotaped) signals. Our analysis suggests that price and allocation results are consistent. That is, within a given market environment with particular characteristics, results are similar, and across differing market environments, results differ in a predictable fashion.

The results also support the conclusion that the use of predetermined draws does not substantially alter observed market behavior. Our data do not support the idea that using predetermined draws causes subjects to suspect deception, or to be confused about the nature of the underlying distribution from which draws are made. Our results suggest that predetermined draws can be used to control the nature and extent of experience given experimental subjects, and can thus improve the replicability and consistency of results in laboratory markets. This outcome substantially broadens the possible scenarios for obtaining data in many finance and economic settings.

Acknowledgments

The authors thank the Eli Broad Graduate School of Management and the Department of Accounting for financial support. Any errors are their own.

Notes

1. Davis and Holt [1993] point out that some researchers even avoid the use of computer algorithms to maintain "realism" in their experiments.
2. Evidence suggests that deception, even when consciously used, does not adversely affect experimental results. Bonetti [1998] found no adverse effects in economic contexts. Korn [1997] reviews similar findings in the psychology literature.
3. The dependence of price deviation from expectation on signals is demonstrated by the following example (using market parameters given in Table 1). Given an R1 signal (one red ball in a sample of five balls) in the markets, conditional on Bayesian expectations and prior results suggesting that naïve subjects tend to trade at higher than optimal prices, the positive error ceiling is 40 (375–335). However, given an R3 signal, the positive error ceiling is 238 (550–312). Thus, the absolute magnitude of price deviation may differ substantially as signals change. Furthermore, the same absolute change in price, given different signals, may suggest very different behavior on the part of the trader(s). Ceiling effects are an issue.

References

Anderson, M.J., and S. Sunder. "Professional Traders as Intuitive Bayesians." *Organizational Behavior and Human Decision Processes*, 64, (1995), pp. 185–202.

Becker, G.S. "Irrational Behavior and Economic Theory." *Journal of Political Economy*, 70, (1962), pp. 1–13.

Bonetti, S. "Experimental Economics and Deception." *Journal of Economic Psychology*, 19, (1998), pp. 377–395.

Burns, J. "Experience and Decision Making: A Comparison of Students and Businessmen in a Simulated Progressive Auction." In *Research in Experimental Economics*. Greenwich: JAI Press, 1985.

Camerer, C. "Do Biases in Probability Judgment Matter in Markets? Experimental Evidence." *American Economic Review*, 5, (1987), pp. 981–997.

Davis, D., and C.A. Holt. *Experimental Economics*. Princeton: Princeton University Press, 1993.

Friedman, D., and S. Sunder. *Experimental Methods: A Primer for Economists*. New York: Cambridge University Press, 1994.

Hirshleifer, J. *Price Theory and Applications*. Englewood Cliffs, N.J.: Prentice Hall, 1976.

Jamal, K., and S. Sunder. "Bayesian Equilibrium in Double Auctions Populated by Biased Heuristic Traders." *Journal of Economic Behavior and Organization*, 31, (1996), pp. 273–291.

Kahneman, D., J.L. Knetsch, and R.H. Thaler. "The Endowment Effect, Loss Aversion, and Status Quo Bias." *Journal of Economic Perspectives*, 5, (1991), pp. 193–206.

Keppel, G. *Design and Analysis: A Researcher's Handbook*. Englewood Cliffs, NJ: Prentice Hall, 1982.

Korn, J.H. *Illusions of Reality: A History of Deception in Social Psychology*. Albany, NY: State University of New York Press, 1997.

Orne, M.T. "Communication by the Total Experimental Situation: Why It is Important, How It is Evaluated, and Its Importance for the Ecological Validity of Findings." In P. Pliner, L. Krames, and T. Alloway, eds., *Communication and Affect: Language and Thought*. New York: Academic Press, (1972), pp. 159–191.

Smith, Adam. *The Wealth of Nations*, ed. R.H. Campbell, A.S. Skinner and W.B. Todd. Oxford: Clarendon Press, 1976.

Smith, V.L. "Economics in the Laboratory." *Journal of Economic Perspectives*, 8, (1994), pp. 113–131.