ELSEVIER

Journal of Economic Dynamics and Control
20 (1996) 1307–1337

JOURNAL OF
Economic
Dynamics
& Control

Price formation in double auction markets

Timothy N. Cason*[,a], Daniel Friedman[b]

[a] Department of Economics, University of Southern California, Los Angeles, CA 90089, USA
[b] Department of Economics, University of California at Santa Cruz, Santa Cruz, CA 95064, USA

(Received August 1994; final version received October 1995)

Abstract

This paper reports 14 laboratory experiments that examine existing theories of the price formation process in the continuous double auction. The experiments feature random values and costs, and therefore a new price formation observation each period. We find that efficiency is high and rises with experience in this environment. We also find that trades with greater exchange surplus tend to occur earlier in the period, that increased trader experience reduces an anomalous intertemporal arbitrage opportunity observed previously, and that only when traders are very experienced does exchange surplus accrue disproportionately to the side of the market with a smaller number of traders.

Key words: Auctions; Experiments; Price formation; Double auction
JEL classification: D44; D82; G20

1. Introduction

No question is more central to economic theory than price formation: how do self-interested traders with private information arrive at equilibrium prices for market exchange? Since applied economists must decide when and where to assume that observed prices are at equilibrium, the price formation question is also central to economic practice.

*Corresponding author.

Financial support was provided by the National Science Foundation (SES-9223830). We acknowledge the programming support of Andrew Davis and the research assistance provided by Carl Plat, and helpful comments provided by three anonymous referees and participants at the 1994 Public Choice Society/Economic Science Association Conference. Errors are our responsibility.

SSDI 0165-1889(95)00901-7

In this paper we study empirically the price formation process in simple continuous double auction markets conducted in the laboratory. In the double auction (DA), traders freely announce bids and asks and may accept other traders' bids or asks as long as the market remains open. The DA requires no benevolent auctioneer and is the most common way of organizing major contemporary financial markets and markets for homogeneous commodities. See Friedman and Rust (1993) for a recent collection of theoretical and empirical studies and introductions to the literature on DA markets.

The DA has been studied extensively in the laboratory. Hundreds of experiments have confirmed Smith's (1962) finding that transaction prices and quantities converge quickly and reliably to competitive equilibrium in double auction markets in a wide variety of laboratory environments. By now it is folk wisdom (at least among experimental economists) that the DA market institution has remarkable powers to promote price formation.[1] Yet there still is no consensus on how and why price formation is so successful in the DA.

Our study differs from its predecessors mainly in that we designed our experiments explicitly to understand the price formation process in the light of recent theory. The vast majority of previous studies use a stationary repetitive environment, in which traders receive the same endowments each period. Here we study the double auction in a random values environment, in which traders' private values are drawn independently each period from a known uniform distribution.[2]

The random values environment is appropriate for our purposes for two reasons. First, it gives us a fresh observation of price formation each period as individual traders' incentives and the equilibrium prices change randomly; indeed, this environment in some ways overstates the degree of uncertainty traders face in contemporary markets in the field. Second, the random values environment embodies the incomplete information assumptions of recent theory, e.g., the Wilson (1987) model discussed below. By contrast, the standard stationary repetitive experiments employ 'privacy' (Smith, 1982) and so traders have no common knowledge of the distribution of other traders' values.

We also deviate from most previous DA experiments by giving each trader the capacity to buy or sell only a single unit per trading period. This allows for sharp

---

[1] Several authors have sought environments in which the DA institution would fare badly, but few have succeeded. Van Boening and Wilcox (1996) use extreme cost conditions (large avoidable costs but zero marginal costs and sunk costs) to obtain the most dramatic inefficiencies seen so far in DA markets.

[2] Random values laboratory environments are standard for studying one-sided sealed bid auctions; see Kagel (1995) for a summary. Recently Kagel (1994) has also begun to study two-sided auctions, including the DA, in a random values environment. The only other random values DA experiments we are aware of are early pilot experiments reported in our preliminary work, Cason and Friedman (1993).

theoretical tests without dubious auxiliary assumptions, such as that multiple-unit traders construct strategies separately for each unit. An important consequence of single-unit trading in our sessions with eight or ten traders is that the markets are relatively 'thin', with only zero to four transactions per period. We also vary the ratio of buyers to sellers to test the impact on price formation. Thus the environment is very challenging to traders and is closely adapted to study the price formation process.

We employ the single-unit, random values environment in 14 DA market sessions, using inexperienced, experienced, and expert traders, and various buyer/seller ratios. We consider six observable statistics: allocational efficiency, the sources of inefficiency, transaction order, the distribution of trade surplus, price change autocorrelation, and the bid/ask sequences leading to transactions. The theoretical models we evaluate make differing predictions for these statistics, which allows us to compare model performance in several dimensions.[3]

We discuss five theoretical models of the price formation process and work with three. The models differ in several respects, especially the degree of trader rationality assumed: the ZI model assumes no rationality beyond the unwillingness to trade at a loss, the BGAN model assumes some forward-looking but nonstrategic rationality, and the WGDA model assumes full strategic rationality.

Despite the intrinsic difficulty of trading in a thin, random values environment, we find that efficiency is high in every session. Efficiency rises with experience and approaches levels seen in a stationary repetitive environment. Contrary to the predictions of all three models, a bit more than half of the efficiency losses arise from trades of extramarginal units. We also find that high-surplus trades (between high-value buyers and low-cost sellers) tend to occur earlier within a trading period than low-surplus trades, and that the distribution of exchange surplus favors the side of the market with a smaller number of traders in sessions conducted with very experienced 'expert' subjects. Increased trader experience also reduces an anomalous intertemporal arbitrage opportunity observed previously.

The remainder of the paper is organized as follows. Section 2 presents the experimental design. Section 3 presents the key observable variables, the theoretical models to be tested, and the testing approach. The results are collected in Section 4. Finally, Section 5 summarizes, and offers interpretations and suggestions for future work.

[3]The single-unit trading restriction enhances the discriminative power of most statistics, including the transaction order and the distribution of surplus, and (to a lesser extent) the sources of inefficiency and the bid/ask sequences. The price change autocorrelation statistic has lower power in thin markets, but here we draw on data from a number of other independent experiments to support our conclusions.

2. Experimental design

We report 14 new DA market sessions, each with 30 or 40 trading periods. Subjects had specialized trading roles, either as single-unit buyers who could not resell or as single-unit sellers who could not repurchase. Each period t each buyer i received a specified 'resale value' v_{it} for a single indivisible unit, and similarly each seller j received a specified cost c_{jt}. If these traders transact at price P, then the exchange surplus $v_{it} - c_{jt}$ is composed of the buyer's profit $v_{it} - P$ and the seller's profit $P - c_{jt}$.

At the beginning of the session all traders were informed that all values and costs would be drawn independently each period from the discrete (rounded to the nearest penny) uniform distribution over the range [$0.00, $4.99]. Before the start of each trading period, each buyer saw her own value for that period and each seller saw his own cost, but no trader saw others' realized values or costs. The instructions in the Appendix provide additional details.[4]

Trading occurred on the multiple-unit double auction (MUDA) trading software, constrained to a single market and single unit per trader. See Plott (1991) for documentation. Buyers (sellers) are free to post bids (asks) at any time, but the market will only accept quotes that improve proposed terms of trade [i.e., higher bids or lower asks]. Every trader's screen immediately displays the current market bid and ask. Buyers (sellers) are free to accept the market ask (bid) at any time, and the transaction is executed immediately. This version of MUDA does not employ any queue of below market bids (above market asks) analogous to the specialist's order book. A transaction immediately removes both the standing market bid and ask. Traders have access to all previous transactions at all times by pressing a function key. Traders perform their record-keeping by hand on a record sheet and record-keeping is very simple for this single-unit trading capacity setting.[5]

[4] Most inexperienced sessions used the same set of values, most experienced sessions used a different set of values, and the expert sessions used a third set of values. This design feature enables us to compare results across sessions, and more importantly, to compare results across trading institutions in future research. We conducted one inexperienced session with the value set used in most experienced sessions, and one experienced session with the value set used in most inexperienced sessions. These sessions verified that our conclusions concerning experience effects were not artifacts of the specific value draws employed in the experienced treatment.

[5] A drawback of the noncomputerized record keeping and immediate execution of quotes and trades is that typographical errors are not uncommon in the early periods of inexperienced sessions. In contrast, other double auction software such as NovaNet/PLATO require traders to confirm transactions and have a smaller number of errors. We removed the obvious typographical error periods from the data prior to analysis; this accounted for 4.6 percent of the periods, most in the inexperienced sessions. Some pilots performed on similar but not identical DA software at UCSC, documented in Copeland and Friedman (1987), produced similar results to the MUDA software.

Table 1
Summary of sessions

Session name	Experience level	Number of buyers	Number of sellers	Total periods
DA441	Inexperienced	4	4	30
DA442	Inexperienced	4	4	30
DA443	Inexperienced	4	4	30
DA6s44	Inexperienced	4	6	30
DA6b45	Inexperienced	6	4	30
DA6b46*	Inexperienced	6	4	30
DA6s41x	Experienced	4	6	40
DA6b42x	Experienced	6	4	40
DA443x	Experienced	4	4	40
DA6s44x	Experienced	4	6	40
DA445x	Experienced	4	4	40
DA6b46x*	Experienced	6	4	40
DA6s41xx	Expert	4	6	40
DA6b42xx	Expert	6	4	40

*Value and cost parameters for these sessions were deliberately reversed (i.e., the inexperienced session used the experienced set of draws, and the experienced session used to inexperienced set of draws).

Table 1 summarizes the 14 sessions. The six inexperienced sessions had 30 periods and the eight experienced and expert sessions had 40 periods. Each period consisted of 110 seconds of trading, which was sufficient for the typical two-to-three-unit trading volume in this design. Nine sessions employed ten traders, with unequal numbers of buyers and sellers. In the sessions with ten traders, two remained in the same role throughout; i.e., in the six-seller sessions two traders were always sellers, and in the six-buyer sessions two traders were always buyers. The remaining eight traders (and all eight traders in the equal trader number sessions) switched trading roles twice within a session. In the 30-period inexperienced sessions, roles were switched before period 9 and before period 25. In the 40-period experienced and expert sessions, roles were switched before period 11 and before period 31. This switch was common knowledge, as was the number of buyers and sellers in each session. The final two sessions employed 'experts', traders who had participated in at least two previous random-value double auction sessions, and earned greater profits than they would have earned in competitive equilibrium.

All 14 sessions were conducted at the University of Southern California. Subjects were recruited from undergraduate upper-division economics classes, so they were typically economics and business majors in their junior or senior year. No inexperienced subject had ever participated in a previous double auction session. Subjects were randomly assigned a computer and trader position and were paid \$5 when arriving to the session. Trading profits were paid at the conclusion of the session. Including instructions, sessions lasted a little less than two hours. Total earnings ranged between $10 and $36 per subject with an average of over $24.

The design provides a wealth of detailed data. We observe the entire sequence of bids and asks leading to each transaction. There are several transactions each trading period, and 30–40 trading periods each session. Across sessions, there are multiple replications using the same sequences of individual trader incentives (as explained in Footnote 4). Finally, the design includes several different levels of trader expertise and several different balances between the numbers of buyers and sellers.

3. Theoretical predictions

Detailed market data are necessary but not sufficient for the empirical study of price formation. Theory is essential to structure the data analysis. Most earlier work relies mainly on the theory of (partial) competitive equilibrium (CE), supplemented perhaps by ad hoc error terms and/or exogenous convergence terms. Such work is useful in identifying environments (or trading institutions) that promote or retard convergence to CE (cf. Friedman and Ostroy, 1995) but it sheds little light on the price formation process. For that purpose, we need theories of how traders choose actions (bids, asks, acceptances) in the DA and how these actions produce market outcomes (transactions, prices, net trades).

Fig. 1 may help sharpen the point. Panel A shows the value and cost parameters used in one particular trading period in several sessions. The most detailed prediction of CE is that there will be four transactions involving buyers B5–B8 and sellers S1–S4 at prices in the CE interval $(c_2, v_8) = (\$3.48, \$3.61)$.[6] It says nothing about the bid and ask sequences for the actual trading period shown in panel B, resulting in three transactions (between S4 and B5 at $2.76, B8 and S3 at $3.30, and B7 and S1 at $3.29).

We are aware of five theories that offer testable predictions regarding such sequences and the resulting market outcomes. Easley and Ledyard (1993) crystallizes their work of the late 1970s and early 1980s. It seeks mainly to explain convergence to CE across stationary repetitions of a DA market; see Van Boening and Wilcox (1996) for recent evidence of its robustness for this purpose. Unfortunately the theory 'does not yield [distinctive] predictions about the effects of shifts in supply and demand curves' (Easley and Ledyard,

[6] Most periods have values and costs that imply fewer transactions and a wider interval of CE prices.

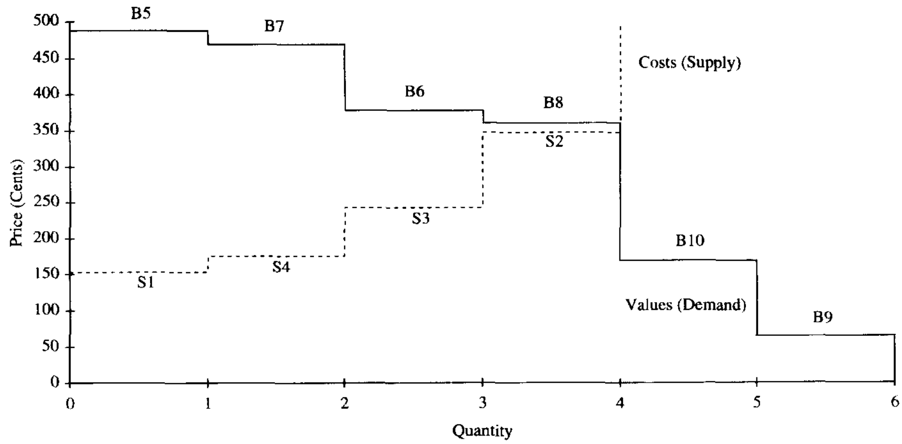

Fig. 1a. Values and costs for experienced sessions, period 34.

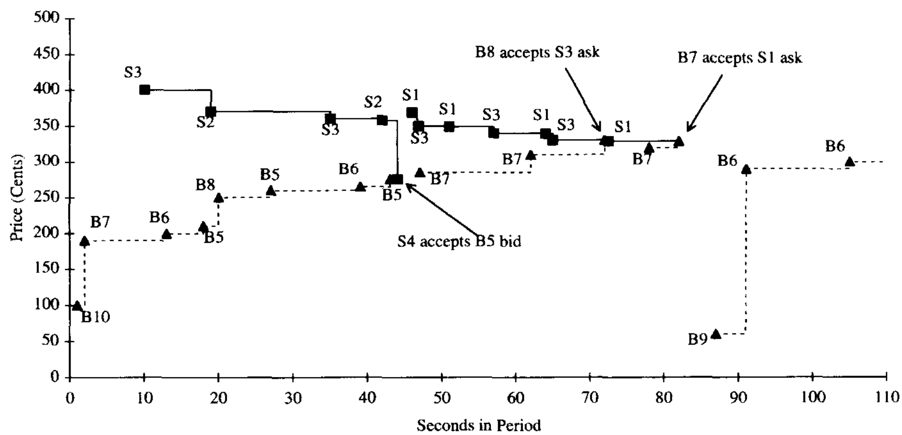

Fig. 1b. Bids, asks, and transactions for experienced session DA6b42x, period 34.

1993, p. 87; [ ]'s added here). Hence it is inapplicable to our random values environment.

The most recent model is Gjerstad (1995). It is adaptable to a random values environment, but that adaptation has not yet been fully worked out. Also, the model contains important free parameters, so proper tests require different statistical methods than the other models we examine. Full-scale tests of the Gjerstad model are beyond the scope of the present paper, although we will offer some informal comments and suggestions for such tests in the conclusion.

We present the three remaining models, called WGDA, BGAN, and ZI, in Section 3.2 after defining the key observables in Section 3.1. The basic testing approach is to use one model (ZI) as the null hypothesis and to use the other models as one-sided alternative hypotheses. This approach extends and systematizes the approach in Cason and Friedman (1993). It is appropriate conceptually because the ZI model embodies a minimal degree of rationality to which the non-strategic rationality of the BGAN model and the strategic rationality of the WGDA model can be compared. Equally important, it is appropriate empirically because the ZI model is stochastic and provides a natural reference distribution for the other two models, which are deterministic.

3.1. Key observables

We will examine what we believe are the six most important observable variables:

1. *Market Efficiency* is defined as the realized gains from trade in each market period as a percentage of the potential gains.[7] Efficiency is ultimately the most important measure of market performance, but does not discriminate well among models of DA price formation because most models were *constructed to predict* high efficiency. It will be instructive to compare efficiency in our challenging environment of random values and thin markets to the more benign environments studied previously.

2. *Sources of Inefficiency*. Efficiency can fall short of 100 percent if (a) traders with extra-marginal units transact (EM-inefficiency), or (b) profitable trades are not executed (low volume or V-inefficiency). For example, in Fig. 1 a profitable trade is not executed (and trading volume is three instead of the efficient four units), so this period exhibits V-inefficiency. [It would have exhibited EM inefficiency if trader B9 or B10 had managed to transact.] We will compare the actual mix of V- and EM-inefficiency to theoretical predictions.

3. *Transaction Order*. It is natural to imagine that the first transaction each period involves the highest-value buyer and the lowest-cost seller, the second involves the second-highest-value buyer and the second-lowest-cost seller, and so forth until all gains from trade are exhausted. Departures from this 'natural' transaction order produce both the V- and EM-inefficiencies.

[7] In the Fig. 1 example period, maximum surplus is \$7.77 (the sum of the four highest buyer values minus the sum of the four seller costs). Buyer B6 (with a value of \$3.78) and seller S2 (with a cost of \$3.48) fail to trade, so \$3.78 − \$3.48 = \$0.30 of the potential surplus is not realized. Efficiency in this period is therefore 7.47/7.77 = 96.1 percent.

Hence a fine-grained analysis of efficiency will compare actual transaction order to the 'natural' order or other theoretical predictions of the trading order.

4. *Allocation of Surplus.* The transaction price determines the allocation of surplus between buyer and seller. EM-inefficient and other 'out-of-order' transactions can be expected when transaction prices outside the CE price interval skew the allocation of surplus. Some models make distinctive predictions regarding the allocation of surplus, so it is important for comparing theories empirically.

5. *Price Change Autocorrelation.* Predictable trends or reversals in transaction prices are easy to exploit in thick markets where each trader can both buy and sell. Even in our simple markets with single-unit one-way traders, predictable price changes provide an 'arbitrage opportunity' to buyers or sellers who are willing to wait for more favorable prices. The price formation theories we consider make quite different predictions regarding the possibility of such arbitrage opportunities, and the preliminary results of Cason and Friedman (1993) pointed to anomalously large reversals. Hence the price change autocorrelation deserves close empirical analysis.

6. *Bid and Ask Improvements.* Each DA transaction price arises from a two-sided auction where buyers improve (raise) bids and sellers improve (lower) asks until one of the buyers and one of the sellers reach agreement. The price formation theories differ fundamentally on how these auctions work, so the sequences of bids and asks provide a key observable for distinguishing between the models.

3.2. Predictions of the models

To conserve space, we do not offer a complete summary of the models we test. Interested readers should consult the original articles; Cason and Friedman (1993) provide some additional details. Here we just explain the testable predictions summarized in Fig. 2.

3.2.1. WGDA

The Waiting Game/ Dutch Auction (WGDA) model of Wilson (1987) regards the price formation process as a sequential equilibrium of an extensive-form game in which the private values of n single-unit buyers and m single-unit sellers are drawn from a commonly-known joint distribution. Recall that our random values environment with many trading periods closely implements the informational assumptions, and so is 'home turf' for the WGDA model. But recall also

Observables	WGDA	BGAN	ZI
1. Market Efficiency	High	High	High (estimated with simulation)
2. Sources of Market Inefficiency	Least profitable (inframarginal) trade not executed	Least profitable (inframarginal) trade not executed; slight displacement possible	Displacement by extramarginal traders
3. Transaction Order	In order of trader values	In order of trader values	Weakly in order of values (estimated with simulation)
4. Allocation of Exchange Surplus	Favors the side of the market with a smaller number of traders	Independent of the number of remaining buyers and sellers	Estimated with simulation
5. Price Change Autocorrelation	*Equal Traders:* Zero *Unequal Traders:* Slightly Positive	Positive	Approximately -0.5 (estimated with simulation)
6. Bid/Ask Improvements	By the same buyer or seller	By different buyers and sellers	Intermediate (estimated with simulation)

Fig. 2. Summary of model implications.

that the assumption of sequential equilibrium is especially demanding in the context of a double auction market. Traders are assumed to know not just the distributions of private values but also the strategies (i.e., time-, history-, and value-contingent plans to bid or ask and accept) of other players, and traders are assumed sufficiently rational to compute and use strategies that are best responses given this knowledge.

The basic idea of WGDA is that agents play a waiting game, with each buyer's (seller's) impatience arising from the possible preemption of gains by the other buyers (sellers). Eventually some buyer (or seller) finally makes a 'serious' bid (ask) – one which has a positive probability in sequential equilibrium of being accepted. If her offer is not immediately accepted, the bidder (asker) will steadily improve the offer (while other traders remain passive) until it is accepted, as in a Dutch Auction. The transactors will be the highest-value buyer and the lowest-cost seller remaining in the market because their waiting costs are largest. Hence the WGDA model predicts the 'natural' transaction order and a distinctive pattern of bid/ask improvements.

The surplus from a transaction is allocated between buyer and seller according to relative hazard rates ('impatience'), so the ratio of remaining buyers to remaining sellers determines the allocation. On average the split is equal when there are equal numbers of buyers and sellers, but otherwise it favors the shorter side of the market. Given favorable time parametrizations, all positive surplus transactions are realized, except perhaps the least valuable. No transaction will involve extramarginal traders; i.e., EM inefficiencies will not be present. Finally, arbitrage opportunities are precluded in sequential equilibrium, and with equal numbers of buyers and sellers this implies uncorrelated price changes. It implies a slight positive autocorrelation with unequal numbers, because then the buyer/seller ratio (and hence the split) becomes more extreme after each transaction, creating a small but systematic trend in the transaction price.

3.2.2. BGAN

The Bayesian Game Against Nature (BGAN) model draws on the Bertrand perspective of Friedman (1984). It centers on unobservable reservation prices for buying or for selling. A buyer seizes the market bid if she can do so at a price better than her reservation price, and she accepts the market ask whenever it exceeds her reservation price; sellers are analogous. To complete the model, the dependence of reservation prices on time and history must be specified. In Friedman (1991), this is done by means of a drastic simplification: buyers and sellers are assumed to ignore the impact that their own current bids and asks will have on subsequent offers by others. This 'Game Against Nature' assumption, together with Bayesian updating and some auxiliary assumptions similar to Wilson's (e.g., risk neutrality and proportional time parameterization), gives reservation prices as solutions of the optimal stopping problem associated with current parameter estimates for 'Nature's' bid and ask generating process. The reservation prices in turn permit numerical calculation of bids, asks, and acceptances for any set of induced value parameters, random or otherwise.

Assuming that all traders are risk-neutral (or are identically risk-averse) and have the same prior beliefs, the BGAN model predicts that transactions generally occur in the 'natural' order and include all mutually beneficial trades except perhaps the last, least profitable trade. Hence this implementation of BGAN basically agrees with the WGDA model regarding observables 1, 2, and 3.[8]

Because the BGAN model is not strategic, the predicted surplus allocation is not closely tied to the buyer/seller ratio. The model predicts a slight positive autocorrelation in price changes that is independent of the buyer/seller ratio but

[8] In BGAN traders ignore their own bids and asks in updating their beliefs, so each trader has a slightly different sample of observations. The different estimates of 'Nature' can occasionally produce reservation prices whose order differs from the value/cost order. Hence there is a slight possibility of out-of-order transactions and EM inefficiency in BGAN even with *a priori* identical two-sided traders. We will neglect that possibility here, but note that it should not be neglected in BGAN models with heterogeneous traders.

is decreasing in the number of observed bids and asks since the beginning of the period.[9]

Perhaps the most distinctive BGAN prediction concerns the sequence of bid/ask improvements. The Bertrand competition takes the form of one buyer raising another's bid until the bid exceeds the second-highest reservation price. When one buyer's reservation price is far higher than other remaining buyers, then she may eventually raise her own bid when her reservation price increases sufficiently with the passage of time. But by construction, the predominant influence is Bertrand competition, so the model predicts that successive bids typically will be made by different traders. An analogous argument leads to the same conclusion for asks.

3.2.3. *ZI*

The final theoretical model plays a central role in our empirical analysis. Gode and Sunder (1993) introduce an algorithm of random trader behavior called 'zero-intelligence' (ZI). Traders are assumed to be rational only to the extent that they will not trade at a loss. Specifically, each bid is uniformly distributed between zero and the buyer's redemption value, and each ask is uniformly distributed between the seller's cost and the $4.99 upper bound on cost. Under the usual 'NYSE convention' for the DA, bids (or asks) which do not improve the current market bid (or ask) will not be observed. Buyers and sellers who have not yet transacted bid and ask independently in random sequence. A transaction occurs at the current market ask (or market bid) whenever a new bid exceeds (or a new ask falls below) the current market ask (or market bid). The simulation continues each period until no more trades are possible, i.e., until all remaining buyers have values below the cost of each remaining seller.[10]

[9] Price change autocorrelation is positive in the BGAN model because traders incorrectly assume that other traders' bids and asks arise as if from natural processes whose parameters are unknown but unchanging. To the extent that traders' beliefs change in response to a new bid or ask observation, their reservation prices will shift, and as a result the actual data-generating process will also shift. It can be shown that such an unanticipated shift in the data-generating processes leads to positively correlated transaction price changes. This effect dies out fairly rapidly as traders accumulate observations and therefore change their beliefs less in response to new observations. Anonymous referees correctly point out that heterogeneity across traders can lead to negative autocorrelation, and that simulations will be necessary for an adequate understanding of the model's sensitivity to initial beliefs.

[10] Note that clock time is not meaningful in such simulations. A Poisson or other random distribution of event times would imply that periods would occasionally have to run for very long periods of time. In practice the time between events is deterministic and each period takes only a few milliseconds of computer time.

These ZI simulations permit some EM-inefficiency but (by definition of the stopping rule) no V-inefficiency. It is not hard to see that, in the ZI algorithm, buyers with high values and sellers with low costs are more likely to make high bids and low asks that produce transactions. Hence the transaction order is random but has some tendency towards the 'natural' order. The allocation of surplus in ZI is far more variable than in CE and has a slight bias in the same direction as WGDA, favoring the side of the market with a smaller number of traders. The ZI assumptions imply that transaction prices are independent, and Cason and Friedman (1993, Footnote 4) show that this implies that successive price changes have a correlation coefficient of approximately -0.5. Finally, note that bid/ask sequences in ZI will be a mix of improvements by rival traders and self-improvements.

The definition of the ZI model implies a specific prediction for observable 2 – that all inefficiency must arise from extra-marginal displacement (EM-inefficiency). However, the stochastic nature of ZI implies a distribution rather than a point prediction for the other five observables. For these five observables we calculated empirical distributions using hundreds of ZI simulations based on the same sequence of value and cost draws employed in the experiment. These distributions provide the basis for Monte Carlo tests of the ZI null hypothesis. In this procedure, if the calculated observable using the laboratory data is greater than (less than) the 95th (5th) percentile of the ZI distribution, we can reject the ZI null hypothesis at the five percent significance level (for a one-sided test). Typically the WGDA and BGAN models imply observables that are outside the ZI null acceptance interval, so when the data reject ZI they may do so in the direction of one or both of the alternative models.

4. Results

We report results in six subsections, organized according to the six hypotheses summarized in Fig. 2 above. For the sake of brevity, we exclude from our presentation some alternative test specifications for the 14 current experimental sessions and also exclude the data from two pilot sessions conducted using an alternative DA software implementation (and a different set of random values). These exclusions do not materially affect our conclusions; the full results are available on request from the first author.

4.1. Market efficiency

In spite of the challenging parameter variability throughout this experiment, traders are still remarkably successful in extracting most of the available exchange surplus. Traders obtained between 85.6 and 87.5 percent of the available exchange surplus in four of the six inexperienced sessions (the outlying sessions had 81.4 and 92.1 percent efficiency), for an average inexperienced efficiency of 86.9 percent.[11] In the five experienced sessions efficiencies ranged from 91.7 to 94.2 percent with an average of 93.6 percent. Efficiencies did not improve further in the two expert sessions (with 92.9 and 94.1 percent), which suggests that overall market performance in this setting does not improve substantially after the initial inexperienced session. These efficiencies are roughly similar to efficiencies observed in early periods of many previous DA experiments using repetitively stationary environments.

Two benchmark comparisons come to mind. First, ZI efficiencies for these parameters almost always exceed 90 percent, and the observed efficiencies are significantly too low (below the 5th percentile of the ZI distribution) in five of the six inexperienced sessions. However, the observed efficiency is consistent with the ZI efficiency distribution in all but one of the eight experienced and expert sessions.

Second, recall that WGDA and BGAN both predict that all inframarginal units except the least profitable will transact, and this last inframarginal trade fails to be executed with positive probability. We have no estimates of this probability at present, but assuming that it is 1.0 provides an easily computed lower bound on efficiency. In the eight-trader sessions these lower bounds are 65 percent and 70 percent efficiency in the inexperienced and experienced sessions, respectively. In the ten-trader sessions the lower bounds are 80 percent, 77 percent, and 75 percent efficiency in the inexperienced, experienced, and expert sessions, respectively. The lower bounds increase with the number of traders because they vary closely with the number of predicted transactions in a period, and the average number of predicted transactions increases along with the number of traders. Although all observed efficiencies exceed these benchmarks, as shown in Table 2, it is worth noting that observed efficiency fails to increase with the number of traders. In the inexperienced sessions average efficiency actually falls from 88.4 percent with eight traders to 85.5 percent with ten traders.

CONCLUSION: Efficiency exceeds the lower benchmark of WGDA and BGAN in all sessions, and increases after subjects obtain trading experience. Efficiency levels are consistent with the ZI predictions for the experienced and expert sessions, but not for the inexperienced sessions.

4.2. Sources of inefficiency

Table 2 shows the decomposition of efficiency losses due to (a) transactions involving extra-marginal units (EM-inefficiency), and (b) profitable trades that

[11] Efficiency in the two inexperienced pilot sessions was slightly lower – 83 and 84 percent.

Table 2
Decomposition of efficiency losses

Experience	Traders	Average efficiency	Percent of losses for EM-inefficiency	Percent of losses for V-inefficiency
Inexperienced	8	88.4 percent	56.4 percent	43.6 percent
Inexperienced	10	85.5 percent	59.5 percent	40.5 percent
Experienced	8	92.9 percent	67.8 percent	32.2 percent
Experienced	10	93.9 percent	57.2 percent	42.8 percent
Expert	10	93.5 percent	55.8 percent	44.2 percent

are not executed (low volume or V-inefficiency).[12] Both types of inefficiency are common, although on average EM-inefficiency is more common. However, there exists substantial variation across sessions, and V-inefficiency exceeds EM-inefficiency in 4 of the 14 individual sessions.

Does the ZI model predict efficiency better with inexperienced traders than with experienced traders? We just saw that the answer was quite negative regarding the overall level of efficiency, and Table 2 shows that the evidence on the sources of inefficiency is also contrary. Although overall inefficiency decreases with experience, its sources show no clear trend across experience levels. Contrary to the suggestions of each model, greater subject experience does not appear to eliminate either type of inefficiency.

CONCLUSION: Both extramarginal trade inefficiency and low volume inefficiency are common, so the data reject the extreme predictions of all three models.

4.3. Transaction order

Recall that both the BGAN and WGDA models imply that early (later) transactions will be between high-value (low-value) buyers and low-cost

[12] Following the approach in Rust et al. (1993), we perform an 'inefficiency audit' to decompose efficiency losses. Assigning inefficiency to the two classes is straightforward if no extra-marginal units trade but trading volume falls short of the efficient level (V-inefficiency), or if trading volume equals or exceeds the efficient level and efficiency is not 100 percent (EM-inefficiency). However, if trading volume is less than the efficient level and any extra-marginal units trade, both kinds of inefficiency are present and there is no unambiguous way to allocate losses to each class. In these circumstances (which are relatively rare in the data), an extra-marginal unit has displaced an infra-marginal unit (EM-inefficiency), and another infra-marginal unit simply fails to transact (V-inefficiency). The ambiguity arises because we cannot identify which infra-marginal unit to assign to each class. Following Rust et al. (1993, App.), we assume that each of these units is equally likely to be displaced by the extra-marginal unit, so we compute the EM-inefficiency and V-inefficiency using the average value of the untraded infra-marginal units.

(high-cost) sellers. The ZI model agrees but provides a weaker prediction – that transactions between high-value buyers and low-cost sellers are more likely to occur early.

To test these implications we estimate the two rank correlation coefficients, between (1) transaction order and buyer value and (2) transaction order and seller cost. Each theoretical model leads us to expect that buyer values are negatively correlated with transaction order – high-value buyers should be first in transaction order and low-value buyers should be last in transaction order. The models similarly predict that seller costs are positively correlated with transaction order. Indeed, WGDA and BGAN both make the very strong prediction that the buyer value (seller cost) and transaction order rank correlation coefficient is -1.0 (1.0) because traders should transact exactly in the order of their values.

Table 3 presents the correlation estimates by session type. [The table separates the sessions with six buyers from those with six sellers because the correlation estimate is always closer to zero for the side of the market with six traders.] The estimates have the predicted signs in all but the inexperienced four buyer–four seller case. However, as indicated by the p-values in parentheses, the estimates are often not significantly different from zero – particularly for the side of the market with six traders. Again simulations of ZI traders for each row provide a convenient null hypothesis. Note that in these simulations, as in the data, the correlation is closer to zero for the side of the market with six traders. Note also that every estimated rank correlation lies well within the 5th and 95th percentiles of the ZI simulations, and in many instances the estimate falls remarkably near the midpoint of these percentiles. We conclude from this test that the rank correlation estimates are entirely consistent with those implied by the ZI model, and are much closer to zero than the extreme predictions of the WGDA and BGAN models. For the example period shown in Fig. 1, the first transaction between B5 and S4 has high exchange surplus, as does the third transaction between B7 and S1. However, the second transaction between B8 and S3 has substantially lower exchange surplus and should not precede the third transaction between B7 and S1 according to BGAN and WGDA. In this period and most others, the ordering of trades is correlated with but not perfectly predicted by the value and cost ranking.[13]

CONCLUSION: High-surplus trades (between high-value buyers and low-cost sellers) tend to occur earlier in the period than low-surplus trades, but the

[13] A regression of gains from trade on the number of seconds left in the period provides additional evidence that the earliest trades tend to occur between low-cost sellers and high-value buyers. The estimated coefficient on the number of seconds left is positive and significant for every dataset.

Table 3
Spearman rank correlation estimates between transaction order and trader values

			Observed		ZI simulation			
					Buyer value		Seller value	
Experience	Traders	Obs.	Buyer value	Seller value	5th percentile	95th percentile	5th percentile	95th percentile
Inexperienced	4 Buyers 4 Sellers	168	− 0.257 (< 0.01)	− 0.017 (0.822)	− 0.46	− 0.07	− 0.11	0.24
Inexperienced	4 Buyers 6 Sellers	59	− 0.256 (0.051)	0.044 (0.741)	− 0.52	− 0.17	− 0.15	0.18
Inexperienced	6 Buyers 4 Sellers	142	− 0.085 (0.313)	0.169 (0.045)	− 0.29	0.05	0.08	0.38
Experienced	4 Buyers 4 Sellers	134	− 0.140 (0.108)	0.105 (0.226)	− 0.36	− 0.02	− 0.02	0.28
Experienced	4 Buyers 6 Sellers	181	− 0.192 (< 0.01)	0.063 (0.398)	− 0.38	− 0.09	− 0.08	0.21
Experienced	6 Buyers 4 Sellers	166	− 0.173 (0.026)	0.204 (0.009)	− 0.31	− 0.03	0.05	0.31
Expert	4 Buyers 6 Sellers	78	− 0.075 (0.516)	0.127 (0.269)	− 0.33	− 0.04	− 0.10	0.24
Expert	6 Buyers 4 Sellers	87	− 0.125 (0.247)	0.230 (0.032)	− 0.20	0.07	0.06	0.33

One-tailed p-values in parentheses for the null hypothesis that the estimate equals zero. Each row of the ZI simulation is based on 100 markets, using the same value and cost draws as the experiment sessions.

correlation estimates are much more consistent with the weaker predictions of the ZI model than with the strong implications of both the WGDA and BGAN models.

4.4. Allocation of surplus

A striking implication of the WGDA model that was not testable using earlier experiments is that the side of the market with a smaller number of traders will receive a larger share of the exchange surplus. Therefore, WGDA implies that the fraction of total exchange surplus received by sellers is highly (and positively) correlated with the fraction of remaining traders that are buyers after the transaction. Intuitively, if buyers substantially outnumber sellers, sellers should have more bargaining power and thus extract a larger percentage of the exchange surplus. The nonstrategic nature of the BGAN model suggests no strong correlation between these variables.[14]

Our first test of this hypothesis simply examines the correlation of the fraction of total surplus received by sellers with the fraction of remaining traders that are buyers. Note that in our six buyer–four seller sessions the fraction of remaining traders that are buyers on transaction t, which we denote $BF_t =$ (number of remaining buyers)/(number of remaining traders), is $BF_1 = 5/8 = 0.625$, $BF_2 = 4/6 = 0.667$, $BF_3 = 3/4 = 0.75$, and $BF_4 = 2/2 = 1.0$. For the four buyer–six seller sessions, this fraction is $BF_1 = 3/8 = 0.375$, $BF_2 = 2/6 = 0.333$, $BF_3 = 1/4 = 0.25$, and $BF_4 = 0/2 = 0.0$. Of course, the fraction remains constant at 1.0 in sessions with equal numbers of buyers and sellers, so the eight-trader sessions are not useful for this test. For the three experience levels – inexperienced, experienced, and expert – the estimated correlation coefficients are 0.03, 0.06, and 0.16, respectively. Only the expert estimate is significantly different from zero (p-value $= 0.02$).[15] Thus the evidence, though weak, suggests that with increased experience the side of the market with a smaller number of remaining traders receives more than its share of the exchange surplus, as predicted by the WGDA model.

[14]We should mention that BGAN has a second-order effect going in the same direction as the positive WGDA correlation, arising from differences in order-statistic distributions with unequal numbers of buyers and sellers. The Bertrand competition of BGAN tends to put the bid (ask) at the second-highest (second-lowest) reservation price, creating a slight asymmetry when the number of buyers differs from the number of sellers. As we already noted in Section 3.2.3, the ZI model also has 'a slight bias' in the same direction (summarized in Footnote 15).

[15]This correlation coefficient estimated for the ZI simulation ranges from slightly negative to slightly positive, with a median and mean of about 0.03. Only the expert estimate of 0.16 can reject the ZI null hypothesis.

Our second test relates price changes to changes in the fraction of remaining traders that are buyers, given that higher prices imply that more of the exchange surplus goes to sellers. The WGDA model implies that the price change $D_t = P_t - P_{t-1}$ on transaction t should be positive when the fraction of traders that are buyers is increasing (i.e., $BF_t - BF_{t-1} > 0$), and should be negative when the fraction of buyer traders is falling, so generally the price change should be positively correlated with the buyer fraction change. Furthermore, the WGDA model states that the time series of price changes exhibits no autocorrelation except as induced by the change in the fraction of remaining traders that are buyers. By contrast, BGAN implies a positive autocorrelation in the price change time series but that the buyer fraction change on a transaction has no major independent effect. We test these implications with the following regression:

$$D_t = b_0 + b_1 D_{t-1} + b_2[BF_t - BF_{t-1}] + e_t\,. \tag{1}$$

The b_1 coefficient is an estimate of the price change autocorrelation, controlling for the changing ratio of buyers and sellers captured by the b_2 coefficient. The predicted coefficient signs under WGDA are $b_1 = 0 < b_2$, while BGAN predicts $b_1 > 0 = b_2$. The ZI model predicts $b_1 < 0$, since it implies that price changes are negatively autocorrelated.

Table 4 presents estimation results for Eq. (1), separated by experience level. [The residuals follow an AR(1) process, requiring a GLS transformation.] Because the model predictions refer to within-period price changes, only those periods with at least three transactions generate observations for this estimation. Consequently, the sample sizes are small compared to the other results presented here. Nevertheless, this table provides several notable results. First, note that the data support the ZI prediction for the inexperienced sessions, but this price change autocorrelation rises with experience. [Section 4.5 below provides corroborating evidence for this conclusion.] Second, note that the fraction of buyer traders coefficient (b_2) is only statistically significant with expert subjects. Furthermore, with expert subjects there appears to be mixed support for both WGDA and BGAN because $b_1 > 0$ and $b_2 > 0$.

We also estimated Eq. (1) using data from 900 simulated ZI markets pooling across groups of sessions as in Table 4. In the typical regression the intercept estimate is zero and the lagged price change coefficient (b_1) is negative and about -0.5. This of course arises from the negative price change autocorrelation implied by ZI, and the b_1 estimates for experienced and expert sessions shown in Table 4 are both greater than the 95th percentile of the distribution of b_1 estimates from the ZI simulation estimates. The estimate for b_2 is typically positive for the ZI simulation, but it is significantly greater than zero for less than one-quarter of the regressions. The b_2 estimates shown in Table 4 are all below the 95th percentile of the ZI simulation estimates, so the data do not reject the

Table 4

GLS estimates of price change equation (1)

Experience level	Obs.	Intercept (b_0)	D_{t-1} (b_1)	$[BF_t - BF_{t-1}]$ (b_2)	Adj. R-squared	DW of transformed residuals
Inexperienced	38	5.02	−0.619***	86.8	0.29	1.790
		(17.10)	(0.124)	(103.4)		
Experienced	68	−3.06	−0.126	22.8	−0.02	2.010
		(5.25)	(0.166)	(43.5)		
Expert	31	−4.22	0.511**	189.6***	0.23	2.035
		(7.36)	(0.216)	(73.5)		

Standard errors in parentheses. * denotes significantly different from zero at ten percent level; ** denotes significantly different from zero at five percent level; *** denotes significantly different from zero at one percent level.

ZI null hypothesis for this parameter. The data support the WGDA prediction for b_2 with increased experience, while ZI of course predicts no experience impact. On balance, therefore, we conclude that WGDA deserves some credit for predicting the distribution of exchange surplus for expert subjects.

CONCLUSION: There is no evidence that remaining buyer and seller numbers influence the exchange surplus distribution for inexperienced subjects, but the correlation predicted by WGDA is significant for expert subjects.

4.5. Price change autocorrelation

Recall that the WGDA model precludes intertemporal arbitrage opportunities, and so for equal numbers of buyers and sellers it implies that transaction prices follow a martingale:

$$\mathrm{E}[P_t | P_{t-j} \text{ for all } j > 0] = P_{t-1} . \tag{2}$$

For the statistical tests we shall test the hypothesis that prices follow a second-order martingale:

$$P_t = P_{t-1} + u_t ,$$

where

$$\mathrm{E}[u_t] = 0 \text{ and } \mathrm{cov}[u_t, u_{t-s}] = 0 \text{ for all } s \neq 0 . \tag{3}$$

We test the null hypothesis of the covariance restriction of (3) with $s = 1$, where $u_t = P_t - P_{t-1}$.[16] Preliminary work, reported in Cason and Friedman (1993), identified a startling anomaly in five of the most relevant existing data sets.[17] Typically autocorrelation estimates were significantly negative and near -0.5 for inexperienced subjects, as implied by the ZI model. This result has been subsequently replicated by others using different datasets and parameters – for example, see the discussion of Kagel (1994) and Gjerstad (1995) below. With experienced subjects the estimates reported in Cason and Friedman rose to

[16] See Fama (1970) for a review of martingale tests on securities prices.

[17] The previous data we analyzed were originally reported in Cox and Oaxaca (1990), Friedman (1993), Smith and Williams (1982), and Williams and Smith (1984), as well as three previously unreported experiments. These experiments were designed for other purposes, and are thus inferior to the new sessions reported here for evaluating these price formation hypotheses. See Cason and Friedman (1993) for details.

about -0.25, leading us to conjecture that 'subject inexperience may be one source of the negative serial correlation in price changes' (Cason and Friedman, 1993, p. 268).

The present experiment confirms this conjecture. The autocorrelation estimates for each of the fourteen sessions are reported in Fig. 3, identified by experience level. Note that the MUDA institution replicates the previously-reported significantly negative autocorrelation for the inexperienced sessions. With increasing experience, however, these estimates rise significantly. Pooled across sessions, the estimated correlations are -0.51, -0.26, and $+0.33$ for the inexperienced, experienced, and expert sessions, respectively.[18] The negative pooled estimates are significantly different from zero at the five percent level. One of the six inexperienced and three of the six experienced session estimates reject the ZI null hypothesis because they exceed the 95th percentile of the ZI simulation distribution. Both of the expert sessions strongly reject the ZI null. The expert sessions estimate should be viewed with some caution because of its small sample size ($N = 31$), which also probably explains why this estimate is significantly different from zero at only the ten percent level. Nevertheless, it does suggest some support for the BGAN prediction of positive correlation with increasing experience.

The increase in price change autocorrelation with experience is statistically significant. Treating each session as an independent observation, we employ a nonparametric Mann–Whitney test of the null hypothesis that the autocorrelation does not increase with experience. Using the six inexperienced and six experienced sessions, the test statistic is $U = 6$, which is significant at the five percent level. Pooling the two expert sessions with the six experienced sessions leaves the test statistic unchanged at $U = 6$, which is significant at the one percent level for this greater sample size.

Kagel (1994) and Gjerstad (1995) also document this pattern of increased price change autocorrelation with increased experience. Kagel reports two inexperienced and three experienced sessions also using random values and random costs. However, his experiments employ 12, 14, or 16 subjects each, leading to greater transaction volume and therefore a larger number of observations for each price change correlation estimate. Averaged across sessions, his autocorrelation estimates increase from -0.46 for inexperienced subjects to -0.30 for experienced subjects. Gjerstad calculates autocorrelation estimates for five inexperienced and seven experienced sessions previously reported in Smith and Williams (1983), Ketcham et al. (1984), and Smith and Williams (1990). Averaged across sessions, these autocorrelation estimates increase from -0.42

[18] The autocorrelation estimates are -0.20 and -0.37 for the two pilot sessions, which are closer to the experienced range even though the pilot sessions employed inexperienced subjects.

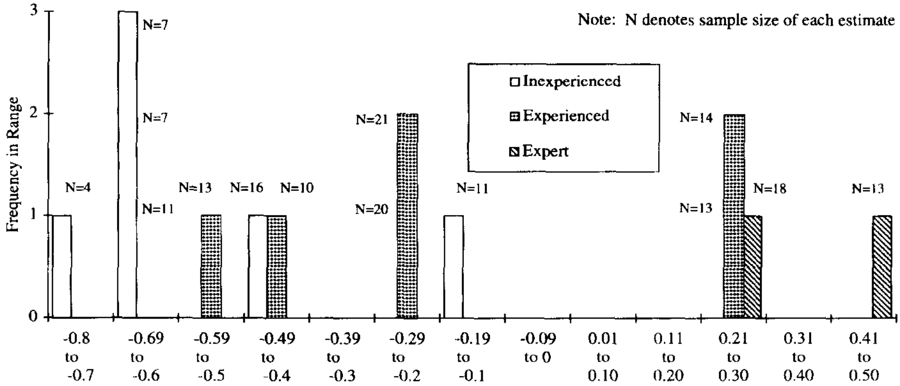

Fig. 3. Distribution of price change autocorrelation estimates across sessions.

for inexperienced subjects to -0.26 for experienced subjects.[19] Taken by themselves, neither the datasets reported in Kagel (1994) nor those in Gjerstad (1995) can statistically reject the hypothesis that autocorrelation does not increase with experience. However, pooling these additional 17 sessions with the sessions reported in the present paper in a meta-analysis strongly rejects the null hypothesis of no experience effect at better than the five percent level [Mann–Whitney $U = 54$, with 13 inexperienced and 16 experienced sessions].

CONCLUSION: Price change autocorrelation estimates for the inexperienced sessions are negative and reject the WGDA and BGAN models in favor of the ZI model. However, the autocorrelation estimates increase with experience, and the expert sessions appear more consistent with the BGAN and WGDA models and are clearly inconsistent with the ZI model.

4.6. Bid and ask improvements

The bid/ask sequences are particularly useful for differentiating between the WGDA and BGAN perspectives of trader competition. Recall that WGDA implies that successive improvements on a bid (ask) by a given buyer (seller) culminating in an acceptance by a seller (buyer) will be common, and successive

[19] Gjerstad (1995) also reports price change autocorrelation estimates for four 'Minnesota Returns Markets', all conducted over a 26-day period. The number of transactions in each market ranged between 2338 and 3145, and his price change autocorrelation estimates range between -0.42 and -0.31 for the four sessions. However, it is difficult to assign the level of experience to these markets since they were not conducted in the self-contained sessions usually employed in laboratory studies.

improvements by different buyers (sellers) will be rare. BGAN implies the opposite – namely, that successive bid (ask) improvements by different buyers (sellers) will be common. For a sharp test of the observable data we use the following empirical definitions:

'*Not BGAN' Definition*: A transaction is deemed inconsistent with BGAN if any outstanding bid (or ask) since the previous transaction is improved by the same trader currently holding the outstanding bid (or ask).

'*Not WGDA' Definition*: A transaction is deemed inconsistent with WGDA if it is an accepted bid (ask) and any bid (ask) improvements since the previous transaction are made by different buyers (sellers).[20]

The bids, asks, and transactions from an example period shown in Fig. 1b are useful to illustrate these definitions. Twelve bids and asks precede the first transaction, but no trader improves her own outstanding bid or ask. Therefore, the bid and ask pattern for this transaction is consistent with BGAN and inconsistent with WGDA. Before the second transaction sellers S1 and S3 continually replace each others' market asks (and the transaction is an accepted ask), so the second transaction is inconsistent with WGDA. Furthermore, buyer B7 improves his own standing market bid, so this transaction is also inconsistent with BGAN. Finally, the third transaction is an accepted ask and no ask improvements are made by other sellers (i.e., S1 always holds the market ask), so this transaction is consistent with WGDA.

Table 5 presents the Not BGAN and Not WGDA frequencies for each type of session.[21] The Not BGAN frequency ranges between 40 and 54 percent, and the Not WGDA frequency ranges between 76 and 84 percent. The frequencies tend to be lowest in the inexperienced sessions, but the experience effect does not seem significant.

[20] Two qualifications are in order. First, the Not BGAN definition is too strict in that the model does permit occasional improvements by the trader currently holding the outstanding bid (or ask). We are comfortable with the strict definition because the results here still favor BGAN. Second, the WGDA implication applies only to 'serious' bids and asks – those that have a positive probability of acceptance in equilibrium – but traders beliefs are not observable. We ran the tests described below on our data excluding bids and asks outside the range [$1.00, $4.00]. We chose this range because it includes more than 99 percent of all transactions, and traders undoubtedly learn after the first few periods that transactions at more extreme prices are highly unlikely. The results, available on request, do not seem sensitive to the nonserious bid and ask exclusion rule, so we are comfortable with the simple definition of Not WGDA.

[21] In the ZI model of random bid and ask submission, an increase in the number of traders would increase the Not WGDA frequency and reduce the Not BGAN frequency, so we report ten-trader and eight-trader sessions separately in Table 5. The results, however, indicate no significant pattern across the number of traders.

Table 5
Frequency of Not BGAN and Not WGDA transactions

			Observed		ZI simulation			
					Not BGAN		Not WGDA	
Experience level	Traders	Obs.	Not BGAN	Not WGDA	5th percentile	95th percentile	5th percentile	95th percentile
Inexperienced	8	168	80 (48%)	127 (76%)	49%	65%	47%	66%
Inexperienced	10	192	76 (40%)	155 (81%)	43%	61%	51%	69%
Experienced	8	134	72 (54%)	104 (78%)	51%	66%	47%	66%
Experienced	10	347	180 (52%)	290 (84%)	46%	62%	53%	68%
Expert	10	165	86 (52%)	136 (82%)	47%	62%	55%	70%

Each row of the ZI simulation is based on 200 markets for the ten-trader sessions and 100 markets for the eight-trader sessions, using the same value and cost draws as the experiment sessions.

The last four columns of the table indicate that the Not BGAN and Not WGDA frequencies are not representative of ZI behavior. Actual Not BGAN frequencies mostly are near the 5th percentile of the ZI simulations; indeed, for the inexperienced sessions we can narrowly reject ZI in the direction of BGAN at the five percent level on this criterion. Actual Not WGDA frequencies always exceed the 95th percentile of the ZI simulations, so on this criterion we can easily reject the WGDA and the ZI models in the direction of BGAN. That is, in every session we have too many Not WGDA transactions to be consistent with ZI (and WGDA) and in inexperienced sessions we have too few Not BGAN transactions to be consistent with ZI (and WGDA). The experienced sessions go in the same direction but less strongly favor BGAN.

CONCLUSION: Although none of the theoretical models fully accounts for observed bid and ask patterns, the patterns are better described by the BGAN model than by WGDA or ZI.

Before closing the results section, perhaps we should mention some additional exploratory analyses conducted to identify factors associated with greater deviations from the CE price interval in this random values environment. An ANOVA decomposition analysis finds that CE price deviations are greatest in the periods with high potential trading surplus, on the first transaction each period, in the initial periods each session, and for inexperienced subjects. Details are available from the first author.

5. Discussion

One message of our results is very comforting. Even with no auctioneer and even with small numbers of traders with diverse private information, the participants in our double auction market experiment were successful at discovering prices at which they could extract most of the exchange surplus. Trading efficiency exceeded the lower theoretical benchmark in all sessions, so overall the markets were quite efficient in this very challenging environment.

Another message is less comforting. The available theoretical models of the price formation process leave much to be desired. Each model has some success, of course. BGAN partisans can point to its clear victory in explaining the bid/ask sequence. Despite definitions that stack the deck in favor of WGDA and despite the advantage of ZI as the null hypothesis, counts of 'Not WGDA' behavior always lead to strong rejection of the null in favor of BGAN, while counts of 'Not BGAN' behavior are usually well below those generated by the ZI null hypothesis (or WGDA). The rather imprecise but positive coefficient estimates for price change autocorrelation in expert sessions also lend some support to the BGAN model. WGDA partisans can take some comfort in the same finding, as well as from the small but significant effects in expert sessions of unequal numbers of buyers and sellers. ZI fans can point to its rather accurate predictions of the transaction order as well as its generally good predictions of outcomes (especially autocorrelation coefficients) in sessions with inexperienced traders. Perhaps more importantly, ZI now seems the only natural source for a null hypothesis in assessing the performance of any more complicated model.

Despite these successes, it seems safe to say at this point that none of these three models adequately explains price formation in double auction markets. Each model has conspicuous failures, the flip side of the successes of other models. The productive way to look at these mixed results is as clues for new theories. The pattern of success and failure of the current models should point the way for constructing new models. Given that the most 'rational' of available models did not perform especially well on its 'home turf', perhaps the most appropriate next step is to explore new models that incorporate significant bounded rationality.

One promising candidate is Gjerstad (1995). Its basic idea is that each trader regards the double auction as a Nature process, but unlike in BGAN, each trader forms beliefs about Nature based on a window (length specified exogenously) of most recent observations. The model is stochastic because the timing of bids and asks is random, but the bid and ask prices are deterministic best responses to current beliefs. The model reduces to a form of ZI that allows low volume inefficiency when the window length is 0, and resembles an impatient BGAN when the window length is long.

The empirical approach we have used here will not be adequate for testing Gjerstad (1995) or other models with free parameters. The appropriate way to proceed seems to be to fit some parameters and to assign an exogenous distribution to others, and to compare the resulting distributions of observables to the data. [The BGAN and even the WGDA models, not to mention other hybrid models, can also be endowed with free parameters for trader-specific risk aversion, etc.] We hope to pursue that approach in the future.

More experiments also are in order. Some aspects of price formation (such as price change autocorrelation) are better explored in thicker markets; for example, with traders who have a two-to-six-unit trading capacity. This multiple-unit environment may also be useful for testing alternative explanations for the negative price change autocorrelation observed with inexperienced subjects. For example, negative autocorrelation may arise mainly from the price variance introduced by a few weak double auction bargainers ('rabbits'), who tend to transact at unfavorable prices;[22] and multiple-unit capacity experiments might be useful for identifying such individuals. Finally, it might be useful to

[22] We thank John Dickhaut for this suggestion.

investigate price formation in other market institutions. Comparisons across simple market institutions may give new insight into the price formation process in the double auction.

Appendix: Instructions

A.1. General

This is an experiment in the economics of market decision making. The National Science Foundation and other research organizations have provided funds for the conduct of this experiment. The instructions are simple, and if you follow them carefully and make good decisions, you can earn a CONSIDERABLE AMOUNT OF MONEY which will be PAID TO YOU IN CASH at the end of the experiment.

In this experiment we are going to create a market for a fictitious good. As a BUYER or SELLER in this market, you can use your computer terminal to purchase or sell a unit of the good. Remember that the information on your computer screen is PRIVATE. To insure the best results for yourself and complete data for the experimenters, DO NOT TALK with other market participants while trade is in progress, and DO NOT DISCUSS your information with others at any point during the experiment.

Each time period for buying and selling is called a TRADING DAY and will actually last two minutes (120 seconds) or less. The start of the Trading Day will be announced by a beep at your terminal. The digital clock on your screen will tell you how much time is left in the Trading Day. At the end of the Trading Day, your profits for that day will be computed as explained below. There will be an intermission of 20 seconds, and a new Trading Day will then begin. You will receive a new unit of the good or new cash holdings for each new day. However, your profits accumulate from day to day. There will be about 30–35 Trading Days in this experiment.

A.2. Sources of profits

Each buyer can purchase a single unit of the good each day from any seller. The PERSONAL RECORD SHEETS I am now passing out display the personal VALUE of the unit for the days in which you will be a buyer. [You will be a buyer in some days and will be a seller in other days.] Each buyer's value is assigned randomly each day as described below. A buyer with value v who purchases a unit at price p earns PROFIT $= v - p$ that day. For example, if a buyer has a VALUE of \$2.86, she would earn a profit of \$2.86 − \$1.94 = \$0.92 if she purchases a unit at price \$1.94. Enter these figures on the practice side of your record sheet now (example day P1) to see how this profit accounting works.

Note that she would earn a negative profit (lose money) if she paid a price above her value \$2.86. Her profit for that day would be zero (\$0.00) if she doesn't purchase a unit.

Each seller can sell a single unit of the good each day to any buyer. The PERSONAL RECORD SHEETS also display the personal COST of the unit for the days in which you will be a seller. Each seller's cost is assigned randomly each day as described below. A seller with cost c who sells a unit at price p earns PROFIT $= p - c$ that day. For example, a seller with a COST of \$1.51 would earn a profit of \$1.94 – \$1.51 = \$0.43 if he sells his unit at price \$1.94. Enter these figures on the practice side of your record sheet now (example day P2) to see how this profit accounting works. Note that he would earn a negative profit (lose money) if he sel ls at a price below his cost \$1.51. His profit for that day would be zero (\$0.00) if he doesn't sell his unit.

Each day each buyer's value and each seller's cost is determined by a computerized random drawing of the 500 numbers (to the nearest penny) between \$0.00 and \$4.99. Each of numbers \$0.00, \$0.01,…,\$4.98, \$4.99 is equally likely on each draw. The value for each buyer and the cost for each seller is drawn *independently* each day. This means that the value for each buyer and the cost for each seller do not depend on the actions of any participant in the experiment, nor on the numbers drawn by the other buyers or the other sellers that day, nor on the numbers drawn on other days.

After the last trading day, we will verify your total trading profits for all days in the experiment, and you then will be paid this amount in cash (US currency). Note that buyer's initial cash and seller's unsold units do *not* contribute to your take-home pay in this experiment. At a minimum, you will keep the \$5.00 you have already received for arriving to the experiment today. You will also receive your additional earnings from trading.

Let met emphasize that you cannot profitably buy or sell more than one unit per day, because you have a one-unit trading capacity on your record sheet. After you trade your unit you should calculate your profits on your record sheet and STOP your market participation. If you should trade more than one unit, you can only lose money because we will calculate your profits based on a value of 0 (or a cost of \$5 in the case of two units sold). Don't make that mistake.

A.3. How to buy or sell

Before we begin the experiment, each of you will proceed at your own pace through a brief (about 20-minute) tutorial that explains how the market software works. A couple of features of this program will not be used in the experiment, so you need not pay close attention to them in the tutorial. In particular, we will only operate one market today, so you don't need to worry about the multiple-market discussion in the tutorial; furthermore, you will only be allowed to trade up to one unit per day, so you can ignore the tutorial's references to multiple-unit quantities.

Throughout the experiment you will be earning profits in dollars (as indicated in your PERSONAL RECORD SHEETS), but all bids, asks, and prices on the computer screens will be in CENTS. Therefore, do not use a decimal point when typing in the numbers.

⟨RUN TUTORIAL NOW⟩

To summarize:

1. The F1 key is used to submit BIDS to the market.
2. The F2 key is used to submit ASKS to the market.
3. The control key is used to accept BIDS (with ctrl + F1) or accept ASKS (with ctrl + F2).
4. The ALT key is used to cancel BIDS or ASKS, and you may use this feature at any time.
5. For the days in which you are a buyer, you may only submit BIDS (with F1) or accept ASKS (with ctrl + F2).
6. For the days in which you are a seller, you may only submit ASKS (with F2) or accept BIDS (with ctrl + F1).

All trading profits after the initial practice day 0 are yours to keep, so make your decisions carefully. You may ask questions at any time by raising your hand. Are there any questions now?

References

Cason, Timothy and Daniel Friedman, 1993, An empirical analysis of price formation in double auction markets, in: D. Friedman and J. Rust, eds., The double auction market (Addison-Wesley, Reading, MA) 253–283.

Copeland, Thomas E. and Daniel Friedman, 1987, The effect of sequential information arrival on asset prices, Journal of Finance 42, 763–797.

Cox, James C. and Ronald L. Oaxaca, 1990, Using laboratory market experiments to evaluate econometric estimators of structural models, Working paper (University of Arizona, Tucson, AZ).

Easley, David and John O. Ledyard, 1993, Theories of price formation and exchange in double oral auctions, in: D. Friedman and J. Rust, eds., The double auction market (Addison-Wesley, Reading, MA) 63–97.

Fama, Eugene, 1970, Efficient capital markets: A review of theory and empirical work, Journal of Finance 25, 383–423.

Friedman, Daniel, 1984, On the efficiency of double auction markets, American Economic Review 74, 60–72.

Friedman, Daniel, 1991, A simple testable model of double auction markets, Journal of Economic Behavior and Organization 15, 47–70.

Friedman, Daniel, 1993, How trading institutions affect financial market performance: Some laboratory evidence, Economic Inquiry 31, 410–435.

Friedman, Daniel and Joseph Ostroy, 1995, Competitivity in auction markets: An experimental and theoretical investigation, Economic Journal 105, 22–53.

Friedman, Daniel and John Rust, eds., 1993, The double auction market (Addison-Wesley, Reading, MA).

Gjerstad, Steven, 1995, Price formation in double auctions, Ph.D. thesis (University of Minnesota, Minneapolis, MN).

Gode, Dharanjay K. and Shyam Sunder, 1993, Allocative efficiency of markets with zero intelligence (ZI) traders: Market as a partial substitute for individual rationality, Journal of Political Economy 101, 119–137.

Kagel, John H., 1994, Double auction markets with stochastic supply and demand schedules: Call markets and continuous auction trading mechanisms, Manuscript (University of Pittsburgh Department of Economics, Pittsburgh, PA).

Kagel, John H., 1995, Auctions: A survey of experimental research, in: J. Kagel and A. Roth, eds., Handbook of experimental economics (Princeton University Press, Princeton, NJ) 501–585.

Ketcham, Jon, Vernon L. Smith, and Arlington Williams, 1984, A comparison of posted-offer and double-auction pricing institutions, Review of Economic Studies 51, 595–614.

Plott, Charles, 1991, A computerized laboratory market system and research support systems for the multiple unit double auction, Social science working paper 783 (California Institute of Technology, Pasadena, CA).

Rust, John, Richard Palmer, and John Miller, 1993, Behavior of trading automata in a computerized double auction market, in: D. Friedman and J. Rust, eds., The double auction market (Addison-Wesley, Reading, MA) 155–198.

Smith, Vernon, 1962, An experimental study of competitive market behavior, Journal of Political Economy 70, 111–137.

Smith, Vernon, 1982, Microeconomic systems as an experimental science, American Economic Review 72, 923–955.

Smith, Vernon and Arlington Williams, 1982, The effects of rent asymmetries in experimental auction markets, Journal of Economic Behavior and Organization 3, 99–116.

Smith, Vernon and Arlington Williams, 1983, An experimental study of alternative rules for competitive market exchange, in: R. Englebrecht-Wiggins, M. Shubik, and R. Stark, eds., Auctions, bidding and contracting (New York University Press, New York, NY) 307–334.

Smith, Vernon and Arlington Williams, 1990, The boundaries of competitive price theory: Convergence, expectation and transactions costs, in: L. Green and J. Kagel, eds., Advances in behavioral economics, Vol. 2 (Ablex Publishing, New York, NY) 3–35.

Van Boening, Mark and Nathaniel Wilcox, 1996, Avoidable cost: Ride a double auction roller coaster, American Economic Review 86, forthcoming.

Williams, Arlington and Vernon L. Smith, 1984, Cyclical double-auction markets with and without speculators, Journal of Business 57, 1–33.

Wilson, Robert, 1987, On equilibria of bid–ask markets, in: G. Feiwel, ed., Arrow and the ascent of modern economic theory (MacMillan Press., Houndmills) 375–414.