WHAT MAKES MARKETS ALLOCATIONALLY EFFICIENT?*

Dhananjay K. Gode and Shyam Sunder

What determines the allocative efficiency of markets? Why are double auctions, even with untrained human traders, allocationally efficient? We provide a simple explanation for these complex phenomena by showing how externally observable rules that define a market cause high allocative efficiency when individuals remain within the confines of these rules. We also show how the oft-ignored shape of extramarginal demand and supply affects efficiency by influencing the inverse relationship between the magnitude of efficiency loss and its probability.

I. Introduction

Allocative efficiency is the ratio of the actual to the potential gains from trade, which are equal to the sum of Marshallian consumer and producer surplus. Allocative efficiency is high if the consumers who value a good the most are able to buy it from the lowest cost producers. Consequently, market designers and researchers want to know what determines the allocative efficiency of markets. Of particular interest to them is how market rules influence interactions among market participants and thereby affect efficiency. In field studies it is difficult to compute efficiency because demand and supply functions are not observable. Chamberlin’s [1948] and Smith’s [1962] experiments controlled demand and supply. Smith found that the efficiency of

*We are grateful for the financial support by the Deloitte and Touche Foundation, the Margaret and Richard M. Cyert Family Fund, the Information Networking Institute, the National Science Foundation, the Santa Fe Institute, and the Deutsche Forschungsgemeinschaft, Gottfried-Wilhelm-Leibniz-Forderpreis. Earlier versions of this paper were distributed under the title “A Comparative Analysis of Economic Institutions with Zero-Intelligence Traders.” We have benefited from comments by Mark Van Boening, Harry Butler, Daniel Friedman, John Long, Douglass North, Charles R. Plott, John Rust, Rajdeep Singh, Vernon L. Smith, Robert Wilson, and participants in presentations given at the Bonn Workshop in Mathematical Economics, Carnegie Mellon University, Cornell University, the Economic Science Association, the University of Kobe, the University of Minnesota, National Taiwan University, the Santa Fe Institute, Universitat Pompeu Fabra, Washington University, and Yale University. We have also benefited from comments by the editor Andrei Shleifer and three anonymous referees. We alone are responsible for any errors.

1. Throughout the paper we refer to “allocative efficiency” as simply “efficiency.” Allocative efficiency should not be confused with “informational efficiency” discussed in the accounting and finance literature. A market is defined to be informationally efficient with respect to a specified information set if it is not possible to devise trading schemes to profit from that information. Laboratory experiments reveal that asset markets can be informationally efficient, without being allocatively efficient (Plott and Sunder 1982, 1988; Sunder 1995).

© 1997 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology.

double auctions with human traders is close to 100 percent—the theoretical prediction assuming perfect competition, utility maximization, and Walrasian tatonnement.

Although it is comforting that theoretical predictions are robust, one is at a loss to explain how a few traders, with limited cognition, participating in a market that does not resemble Walrasian tatonnement, generate these outcomes. Is there something simple and fundamental about the interactions in a market that leads to efficient outcomes even when individuals merely follow a few cues from the market? Perhaps simple behavior, when shaped by market rules, yields the outcomes predicted by theory. Gode and Sunder [1993a] confirmed this conjecture by showing that efficiency remains close to 100 percent when human traders are replaced by traders who choose randomly subject to market rules. However, they did not explain why the efficiency is high.

We show that the high efficiency of double auctions is largely due to the rules that define them. We identify and rank a few basic rules that account for most of the efficiency. We show how successive imposition of these rules reduces the probability of inefficient exchanges among traders. As developed by North [1990, p. 3], institutions are rules of the game of a society and consist of formal and informal constraints constructed to order interpersonal relationships. Our results would help market designers understand the effect of market rules on efficiency.

Our approach differs from game theory, empirical studies, experiments with human traders, and mimicking human traders by computers. Game theory moves away from perfect competition and Walrasian tatonnement to provide valuable insights into markets. However, it is difficult even to prove that equilibrium exists, let alone solve for it, in most double auctions [Friedman 1984; Wilson 1987; Easley and Ledyard 1993]. We assume simple trader behavior for tractability, not to challenge or criticize utility maximization. We use a mathematical model instead of field data to control demand, supply, and market rules. Instead of conducting experiments or simulations [Rust, Palmer, and Miller 1993], we explain their outcomes by deriving closed-form solutions. We are not trying to accurately model human behavior; we assume simple behavior to gain insights into markets with human traders.

These simple traders have a precedent. Becker [1962] showed that, if consumers choose randomly within their budget sets, their demand curves are downward-sloping. However, Becker assumed Walrasian tatonnement, and did not analyze the

Efficiency becomes an issue whenever demand and supply intersect. Traders to the right of the intersection are the extramarginal traders, and those to the left are the intramarginal traders. Extramarginal buyers do not value the goods as much as the intramarginal buyers do, and the cost of goods to extramarginal sellers is higher than it is to intramarginal sellers. Maximum surplus is extracted, i.e., efficiency is 100 percent, if intramarginal buyers buy from intramarginal sellers (thus, no extramarginal traders trade).

We identify three causes of inefficiency: (1) traders participate in unprofitable trades; (2) traders fail to negotiate profitable trades; and (3) extramarginal traders displace intramarginal traders; i.e., the aggregate profits are not as high as they could be.

If resources are allocated by fiat or other nonmarket mechanisms, then efficiency can be arbitrarily low, even negative, depending on the shape of extramarginal demand and supply. Freedom to refuse others’ bids or asks will not increase efficiency if buyers do not know that they should not pay more than a good’s value to them and sellers do not know that they should not accept less than the good’s opportunity cost to them. Accordingly, we assume that traders are free to refuse offers and have the judgment to avoid losses.

Efficiency is still zero if a buyer, who values the good more than the seller, cannot find a seller or the two cannot agree upon a price. We show how call auctions and continuous auctions affect the probability of a buyer finding the right seller. Assuming simple trader behavior, we also show how increasing the number of rounds of bids and asks increases the probability that the buyer and the seller will find a mutually profitable price. This simple observation may explain the high efficiency of auctions that allow multiple rounds of proposals (English auction and double auctions) relative to auctions that allow only a single round (sealed-bid auction).

The third source of inefficiency still remains—extramarginal traders displace intramarginal traders. If an extramarginal buyer with value $v_F$ buys instead of an intramarginal buyer with a value $v_I$, surplus is underexploited by $(v_I - v_F)$. This displace-
ment is undone if, instead of consuming the good himself, the extramarginal buyer resells it to an intramarginal buyer. However, in the real world such reselling may be limited because of transaction costs, need to consume the good, and informational asymmetry at the time of resale. Accordingly, experimental markets and game-theoretic models disallow such resale. We also do the same. Similarly, if an extramarginal seller with cost \( c_g \) sells instead of an intramarginal seller with cost \( c_i \), the loss in efficiency is \( (c_g - c_i) \). This displacement is undone if instead of producing the good himself, the extramarginal seller buys it from an intramarginal seller. We disallow such subcontracting as well. Note that retrading is different from multiple rounds of bidding and asking for a given trade, which we allow.

Given limits on retrading, the interesting question is what determines inefficiency due to displacement. Our analysis provides the following insights into the expected efficiency loss due to displacement, the product of the magnitude of efficiency loss, and its probability. The results of our analysis are summarized in Figure I.

First, the magnitude of inefficiency, \( (c_g - c_i) \) or \( (v_i - v_g) \), is determined by the shape of extramarginal demand and supply, which is often ignored. The probability of displacement depends on the way traders interact, which in turn depends on the market rules.\(^2\)

Second, we identify two rules that jointly raise the efficiency substantially by lowering the probability of displacement: (1) buyers and sellers must abide by their bids and asks;\(^3\) and (2) higher bids have priority over lower bids, and lower asks have priority over higher asks. Given binding contracts and desire to avoid losses, buyers do not bid above their valuations, and sellers do not ask below their costs. This along with price priority allows intramarginal bidders to outbid extramarginal bidders, and allows intramarginal sellers to undercut extramarginal sellers.

---

2. Rules are meaningful only if individuals follow them. Market rules are a consequence of individual rationality because they evolve out of individual choices over time. For example, the market rule that gives priority to a higher bid over a lower bid can also be interpreted as a weak form of individual rationality by the sellers. To this extent, the results of this paper have a dual interpretation as consequences of market rules or of successive refinements of individual rationality. This is further discussed in Section VII.

3. For ease of exposition, we do not consider markets in which the actual payment is different from the bid or the ask, such as second-price auctions in this paper.
$E_{\text{min}}$: Minimum expected efficiency
$E_{\text{ave}}$: Average expected efficiency assuming a uniform distribution of $\beta \sim U(0,1)$

Rules examined:
1. Freedom to accept or reject offers
2. Binding Contracts
3. Price Priority
4. Accumulation
5. Double Auction
6. Repeated Bids and Asks
7. Public Bids and Asks (with Bid Ask Improvement Rule)

(Subsection numbers are indicated in parentheses.)
(I)A) No market: Inability to avoid losses due to lack of freedom or judgment or both. ($E_{\text{min}} \rightarrow \cdot \cdot \cdot$)

Market: Freedom and judgment to refuse losses, binding contracts, and price priority.

Only one side (in our case buyers) makes offers: Sealed-bid auction

(IIB) All bids are collected before picking the highest bid. ($E_{\text{min}} = 75\%, E_{\text{ave}} = 83.3\%$)

(IIC) Not all bids are collected: $E = 1 \cdot (1 - \beta) \left[ (\alpha + (m - 1) \beta) / (\alpha + 1) \right]$ where $m$ is the number of bids collected before matching, i.e., efficiency increases as more bids are collected. This is partly why synchronized auctions are more efficient than continuous auctions.

Both sides make offers: double auctions

(IV) There is only one round, so some units may not be traded: ($E_{\text{min}} = 48.2\%, E_{\text{ave}} = 58.5\%$)

Trading continues (repeated offers) until all possible units are traded.

All bids and asks are collected before they are matched: Synchronized auctions

(VA) The current bid and ask are NOT made public. ($E_{\text{min}} = 80.8\%, E_{\text{ave}} = 89.6\%$)

(VLA) The current bid and ask are made public. ($E_{\text{min}} = 85.2\%, E_{\text{ave}} = 91.8\%$)

Bids and asks are matched as they come in: Continuous auctions

(VB) The current bid and ask are NOT made public. ($E_{\text{min}} = 74.5\%, E_{\text{ave}} = 84.5\%$)

(VLB) The current bid and ask are made public. ($E_{\text{min}} = 80.8\%, E_{\text{ave}} = 87.6\%$)

Figure I
Organization and Summary of Results

Without this basic discipline there is random allocation resulting in high probability of displacement. These two rules and the freedom to trade define the essence of a market or price system.

Third, as demand and supply functions are varied, the expected loss of efficiency has an upper bound. If extramarginal buyers value the goods much less than intramarginal buyers, the magnitude of efficiency loss is high, but its probability is low because the two basic rules prevent extramarginal buyers from bid-
ding high. If extramarginal buyers value the goods nearly as
much as intramarginal buyers, then they bid higher, which in-
creases the probability of displacement, but this is offset by the
decrease in the magnitude of loss from displacement. Similar
logic applies to sellers. Note that even though at the micro level,
individual ZI traders do not trade off profit from a proposal and
its probability of being accepted, at the market level there is
a trade-off between the magnitude of efficiency loss and its
probability.

Fourth, double auctions may be more efficient than one-sided
auctions such as sealed-bid auctions because in double auctions
more conditions must be fulfilled for an inefficient trade to occur.
Fifth, auctions that batch or accumulate bids and asks before
picking the highest bid and lowest ask, such as call auctions, may
be more efficient than continuous auctions, where a transaction
occurs as soon as a bid exceeds or equals an ask. Continuous auc-
tions are usually favored because they have faster price dis-
covery. Our analysis shows that their disadvantage may be lower
efficiency. Sixth, allowing traders to observe market data in-
creases efficiency because it allows intramarginal traders to
quickly outbid/undercut extramarginal traders.

Markets “give occasion to general opulence” through partici-
pants’ “regard to their own interest” [Adam Smith 1776]. Our
analysis suggests that relentless pursuit of self-interest is not
necessary for markets to be efficient. Weak pursuit of self-interest
may be sufficient for efficient allocations in aggregate.

Section II describes the model and provides a detailed outline
for the remainder of the paper. Sections III to VI analyze the ef-
effect of seven common and important rules on efficiency. Section
VII summarizes our findings and discusses their generality.

II. MODEL

Efficiency is the joint result of demand and supply, traders’
decision rules, and market rules. Subsection II.A defines effi-
ciency; subsection II.B describes the demand and supply; subsec-
tion II.C describes trader behavior; and subsection II.D discusses
the market rules.

A. Efficiency

Allocative efficiency is an important measure of aggregate
performance. Markets are said to be allocationally efficient if
those who value the goods most are able to buy them from those who can produce them at the lowest cost. Following Smith [1962], the allocative or surplus extraction efficiency of a market is the total profit actually earned by all traders divided by the maximum total profit that could have been earned by them; the maximum total profit is the sum of Marshallian producer and consumer surplus.

B. Demand and Supply

The demand and supply are similar to those used in the experimental economics literature and have been modified so that we can derive the lower bounds on efficiency. Each buyer can buy up to one unit, which has a redemption value between zero and one. There is one intramarginal buyer (IMB) with redemption value 1 and \( n \) extramarginal buyers (EMBs) with a redemption value \( \beta \) (0 \( \leq \beta \leq 1 \)). There is one intramarginal seller (IMS) who can sell up to one unit with variable cost of zero. Figure II shows the demand and supply.

Having only one seller appears to be restrictive since it precludes competition among sellers. However, our results would be unchanged if there are many sellers because we assume simple, nonstrategic trader behavior; many sellers merely complicate the analysis. Having only one intramarginal buyer does not increase the lower bound of efficiency because if there are more intramarginal buyers, then competition among them raises bids above the redemption values of the extramarginal buyers more quickly, lowering the probability of displacement. We assume that all extramarginal buyers have the same redemption value to make it easy to (1) examine the effect of changing this value, (2) derive closed-form solutions, and (3) study the effect of changing the number of extramarginal buyers.

Given this demand and supply, the efficiency is 1 if the intramarginal buyer buys the unit, and it is \( \beta \) if an extramarginal buyer buys the unit.

C. Trader Behavior: Zero Intelligence (ZI) Traders

ZI traders bid/ask uniformly and independently over a feasible range that changes in response only to the market rules.

4. We have run simulations to confirm that lower bounds of efficiency are unchanged when the demand and supply are not flat. Other results of the paper are qualitatively unchanged if we use more general demand and supply conditions. See Gode and Sunder [1993b] for analysis of a market with multiple sellers.
Thus, they do not trade off profit from a bid/ask against its probability of being accepted; i.e., they do not optimize or even seek profits. However, they set the feasible range so that they bid below their values and ask above their costs to avoid losses in response to a market rule that requires them to pay what they bid or sell at what they ask (see note 2). We discuss the generality of this behavior in Section VII.

D. Market Rules

Of the many existing rules we pick those that are common to many important markets and show how their application transforms random allocation in a null market into a synchronized double auction. We first state the rules and then provide an outline of how they are analyzed in the remainder of the paper.

1. Voluntary trading rule: traders are free to accept or reject offers.
2. Binding contract rule: bids and asks are binding, i.e., buyers must pay what they bid; and sellers must sell at what they ask (see note 2).
3. Price priority rule: higher bids dominate lower bids, and lower asks dominate higher asks.
4. Accumulation rule: the highest bid (and the lowest ask if it is a double auction) are picked only after all bids (and asks) have been collected.

5. Double auction rule: buyers can bid as well as sellers can ask.

6. Multiple rounds rule: multiple rounds of bids and asks are allowed; i.e., if the highest bid is less than the lowest ask, then there are further rounds of bids and asks.

7. Public bids and asks with bid-ask improvement rule: a bid must be greater than previous bids, and an ask must be less than previous asks.

Subsection III.A examines a setting where traders ignore all market rules and do not have the ability to avoid losses. The resulting random allocation serves as a benchmark. Subsection III.B examines the effect of imposing three basic rules (1, 2, and 3) on traders who have the ability to avoid losses in a simple first-price sealed-bid auction. We start with this single-sided auction because we want to examine separately the effect of allowing both sides to be active. Subsection III.C examines the effect on efficiency of varying the number of bids accumulated before the highest bid is picked, i.e., the frequency with which a market is cleared. Although, for simplicity, the analysis has been conducted in the context of a sealed-bid auction, the objective is to gain insights into the relative performance of call markets (which collect all bids and asks before picking the highest bid and the lowest ask) versus continuous markets (where a transaction occurs as soon as a bid exceeds an ask). This issue is further explored in Sections V and VI.

Traditional double auctions are different from a sealed-bid auction in two aspects: (1) double auction rule: both sides are active, i.e., the buyers can bid and sellers can ask; and (2) multiple rounds rule: there are multiple rounds of bids and asks. For tractability most theoretical models have examined auctions that allow only a single round of bids and asks. Section IV isolates the effect of making both sides active while still restricting the number of rounds to be equal to one as in a double auction. Section V then shows how the efficiency increases dramatically as multiple rounds are permitted. Both these sections compare a synchronized double auction (that accumulates all bids and asks before determining whether the highest bid exceeds the lowest ask) with a continuous market (where a transaction occurs as soon as a bid
exceeds an ask). Section VI examines the effect of making the highest bid and the lowest ask public, so that traders can revise the feasible range of bids and asks in subsequent rounds.

III. Effect of Basic Market Rules (1, 2, 3, and 4)

Subsection III.A establishes a benchmark by examining the efficiency in the absence of any market rules. Subsection III.B highlights the dramatic impact of just the three basic market rules: freedom to trade, binding contracts, and price priority. Subsection III.C shows that the effect of the price priority diminishes as the number of bids accumulated before picking the highest bid (accumulation rule) decreases.

A. Random Allocation: Null Market

What would be the efficiency if all market rules are ignored; i.e., how much surplus would be extracted if the single available unit were allocated to a randomly picked buyer? This could occur if the price system does not exist and goods are allocated by a lottery or fiat, or in an anarchy where buyers need not pay what they bid and sellers need not sell at what they ask, or if traders lack judgment to avoid losses and bid above their valuations and ask below their costs.

Since all bidders are equally likely to get the unit, the probability of the intramarginal bidder (IMB) getting the unit (surplus extracted = 1) is $1/(n + 1)$, and the probability of an extramarginal bidder (EMB) getting the unit (surplus extracted = $\beta$) is $n/(n + 1)$. Thus, the expected efficiency ($E$) without a market is

$$E = \frac{1}{n + 1} + \frac{\beta n}{n + 1}.$$

To facilitate comparison with the efficiency of other institutions, the above equation can be rewritten as

$$E = 1 - (1 - \beta)\left(\frac{n}{n + 1}\right).$$  

An EMB displaces an IMB with probability $n/(n + 1)$, and the loss in efficiency from such displacement is $(1 - \beta)$.

Without market rules there is no trade-off between the magnitude of inefficiency $(1 - \beta)$ and its probability $(n/(n + 1))$, and thus there is no interior minimum. Figure III.A shows the effect of changing $\beta$ and $n$ on expected efficiency. As $n \to \infty$, $E \to \beta$. The lower bound of $E$ is zero, and occurs at $\beta = 0$. If $\beta$ were picked randomly and uniformly from the interval $0-1$, we could expect
the efficiency of such allocations, on average, to be 50 percent. (If there were extramarginal sellers as well, then as $n \to \infty$, $E \to -\infty$.)

B. Simple Market: Sealed-Bid Auction (Rules 1, 2, and 3)

For an institution to be a market, it must allow traders to freely accept or reject offers (rule 1) and impose the binding con-
tract rule (rule 2) and the price priority rule (rule 3). This allows high-valuation bidders to get the good by outbidding low-valuation bidders and low-cost sellers to sell the good by undercutting high-cost sellers. Thus, these rules create a price system that allows participants to express their preferences through price. These three rules are complementary because none of them has any effect without the others.

When a market replaces the random allocation rule, the probability of inefficient transaction is no longer independent of $\beta$; it decreases as $\beta$ decreases. This trade-off between the magnitude of inefficiency from an extramarginal trade and its probability leads to an interior minimum and a large increase in the lower bound of expected efficiency.

A single-unit sealed-bid auction is, perhaps, the simplest auction. It is a one-shot auction in which participants have little opportunity to learn from the actions of other traders. Even if this auction is repeated, it provides minimal feedback to traders at the end of each round. Therefore, we start with a first-price sealed-bid auction where ZI buyers bid $b_i = U(0, v_i)$, and $v_i$ is redemption value of the unit to the buyer $i$.

Event I: IMB bid $> \beta$. Probability = $(1 - \beta)$. Since EMBs cannot bid above $\beta$, IMB will win, and efficiency = 1. This efficiency is unaffected by changes in $n$, the number of EMBs.

Event II: IMB bid $\leq \beta$. Probability = $\beta$. Conditional probability IMB bid is the highest bid = $1/(n + 1)$. The expected efficiency conditional on Event II is $(1/(n + 1) + n\beta/(n + 1))$. The overall expected efficiency is, therefore,

$$E = 1(1 - \beta) + \beta \frac{1}{n + 1} + \beta^2 \frac{n}{n + 1} = 1 - (1 - \beta) \frac{\beta n}{n + 1}.$$ 

When the random allocation rule is replaced by binding contracts and price priority, efficiency increases because the probability of an inefficient trade shrinks from $n/(n + 1)$ to $\beta n/(n + 1)$ because $\beta \leq 1$. The loss from an inefficient trade $(1 - \beta)$ decreases in $\beta$, but the probability of such a trade ($\beta n/(n + 1)$) increases in $\beta$, yielding an interior minimum.

This market-level trade-off should not be confused with the individual-level trade-off faced by profit-motivated traders, who weigh the higher profit from a low bid against the lower probability of being the highest bidder. ZI traders do not make any such trade-off, yet the market-level trade-off persists.
Figure III.B plots efficiency as a function of $\beta$ and $n$. As $\beta$ increases from zero, efficiency drops until $\beta$ reaches 0.5 and then rises again. As $n$, the number of EMBs, increases, efficiency drops from 1 to a lower limit. As $n \to \infty$, $E \to (1 - \beta + \beta^2)$. The lower bound of $E$ with a random allocation rule is 0 and occurs at $\beta = 0$. Creation of the market in its most basic form increases the lower bound to 75 percent (at $\beta = 0.5$). If $\beta$ were picked randomly and uniformly from the interval $0\rightarrow1$, the average expected efficiency would be 83.3 percent, whereas the average expected efficiency without a market (and with no extramarginal sellers) is only 50 percent.

C. Varying the Number of Bids/Asks That Are Accumulated before Ranking (Rule 4)

Market designers are often interested in assessing the effect of the frequency with which a market is cleared; i.e., whether it is a call (or synchronized) or continuous auction. Continuous and synchronized auctions represent the two extremes: the first matches each bid/ask as soon as it is received, the second collects all bids and asks before matching. It is interesting to examine the middle ground by varying the number of proposals that are accumulated before matching. It is difficult to examine this in double auctions; we examine it in a sealed-bid auction.

We show that the effect of the price priority rule gradually diminishes as fewer bids are accumulated. Changing the number of bids accumulated is an alternative way of examining changes in search or waiting costs. The higher the search or waiting costs, the lower would be the number of bids accumulated before ranking, making the markets less efficient. Decreases in communication costs because of technological innovation will make markets more efficient by making the price priority more effective. This also has implications for call markets: the longer the interval between calls, the higher would be the expected efficiency.

Let $m$ ($1 \leq m \leq n + 1$) be the number of bids accumulated before matching. As shown in the Appendix, the expected efficiency $E$ is given by

$$E = 1 - (1 - \beta)\frac{(n + 1 - m) + (m - 1)\beta}{n + 1}.$$  \hspace{1cm} (3)

Thus, the probability of an extramarginal bidder getting the unit is $(n + 1 - m) + (m - 1)\beta/(n + 1)$. The coefficient of $m$ is $-\frac{(1 - \beta)}{(n + 1)}$. Each additional bid accumulated decreases the probability of the extramarginal buyer getting the unit by
(1 - \beta)/(n + 1) and therefore increases the expected efficiency by
(1 - \beta)\beta/(n + 1). When the good is awarded to the first bidder
(m = 1), expression (3) reduces to expression (1), which is for ran-
dom allocation. When all bids are accumulated (m = n + 1), ex-
pression (3) reduces to expression (2), which is for a sealed-bid
auction. In this sense, m is the price priority parameter: when
m = 1, there is no price priority; when m = n + 1, price priority
is fully implemented.

Sealed-bid auctions are one-sided auctions; the next section
examines the effect of allowing both buyers and sellers to play an
active role.

IV. DOUBLE AUCTIONS (RULE 5) WITH ONLY ONE ROUND

Many markets are two-sided auctions: buyers submit bids,
and sellers submit asks. Many of these auctions are also con-
ducted in multiple rounds, i.e., if the highest bid does not equal
or exceed the lowest ask in the first round, then a second round
of bids and asks is solicited. It is important to separate the effect
of two-sided actions and multiple rounds. After watching double
auctions, one wonders whether a large part of the efficiency is
simply because double auctions allow multiple rounds of bids and
asks, which ensures that eventually there is a trade.

To address this issue, this section examines a double auction
with only one round. All bids and asks are collected, and if the
highest bid exceeds the lowest ask, then there is a transaction.
Otherwise, the market terminates with the unit left untraded;
i.e., we examine a synchronized double auction that is stopped
after the first round. The analysis highlights the inefficiency
caused by traders failing to conclude mutually profitable trades.
As shown in the Appendix, the expected efficiency is

\[
E = 0.5 - \frac{n\beta(1 - 2\beta)}{2(n + 2)} .
\]

Figure IV plots this function. The minimum, for all values of
n, occurs at \beta = 1/3 and is given by 0.5 - n/(54(n + 2)), which
is approximately equal to 48.2 percent as n \to \infty. The average,
assuming \beta \sim U(0,1), is 58.5 percent. In contrast, as shown in
expression (7) in subsection VI.A, if the auction is continued until
all possible units are traded (and the highest bid and the lowest
ask at the end of each round are made public), then the minimum
expected efficiency is 85 percent, and the average expected effi-

ciency is 92 percent. In some cases only half of the total surplus is extracted in the first round; if the double auction does not allow multiple rounds, its efficiency with ZI traders will be reduced significantly. The same results would be obtained in a sealed-bid auction in which the seller randomly chooses a reservation price below which he would not sell, instead of passively accepting the highest bid as in the previous section. The efficiency in our two-sided auction is lower because there is a finite probability that the unit will not be traded because of unsuccessful “bargaining” between the seller and the buyers.

V. DOUBLE AUCTIONS WITH MULTIPLE ROUNDS (RULE 6)

We study two types of double auctions: synchronized and continuous. Subsection V.A examines the former, and subsection V.B examines the latter.

A. Synchronized Double Auction

A synchronized double auction is a sequence of call auctions. First, all bids and asks are collected. The bids and asks are simultaneous; i.e., traders do not know others’ bids and asks when they
submit their own. If the highest bid is greater than or equal to the lowest ask, then the highest bidder and the lowest asker transact.\footnote{The transaction price, depending on the clearing rule, can lie anywhere between the bid and the ask. Since we do not study price patterns in this paper, we leave this rule unspecified.} Otherwise, the market moves into the next round. The highest bid and the lowest ask are carried over from one round to the next as current bid and current ask, respectively. The current bid/ask is updated in a later round only by a higher bid/lower ask. A transaction clears all bids and asks. This process continues until the only available unit is sold.

In this version of the synchronized market, bids and asks are not made public. Since ZI traders do not remember their own bids or asks, they do not change the range of their bids or asks from one round to the next.

We are able to derive the expected efficiency of these auctions only\footnote{We have conducted simulations to analyze synchronized and continuous markets when the number of extremarginal traders is finite. The results remain qualitatively unchanged except that in all cases as the number of extremarginal traders decreases, the efficiency monotonically increases, and the relative advantage of batching in synchronized DA diminishes.} for \( n \to \infty \). As shown in the Appendix, the expected efficiency \( E \) of this market is given by

\[
E = 1 - (1 - \beta) \frac{\beta^n}{1 - \beta(1 - \beta)}.
\]

The probability of an inefficient transaction is \( \beta^n/(1 - \beta (1 - \beta)) \). In comparison, the probability of an inefficient transaction in a sealed-bid auction as \( n \to \infty \) is \( \beta \) (see expression (2) in subsection III.B). It can easily be shown that \( \beta > \beta^n/(1 - \beta (1 - \beta)) \) for all values of \( \beta \) except \( \beta = 0 \), where both are zero. Thus, with ZI traders the synchronized double auction is always more efficient than the sealed-bid auction.

The graph of expression (5) is labeled “Synch. Without data” in Figure V. The minimum expected efficiency of this double auction is about 80.8 percent (at \( \beta = 0.639 \)), which is higher than 75 percent (at \( \beta = 0.5 \)) for the sealed-bid auction. The average expected efficiency assuming a uniform distribution of \( \beta \) is 89.6 percent.

The efficiency is higher in the double auction because two conditions must be satisfied for an inefficient transaction to occur: (1) IMB bid remains less than \( \beta \), and (2) IMS ask falls below \( \beta \). In contrast, in the sealed-bid auction only the first condition needs to be fulfilled for an inefficient trade to occur.
B. Continuous Double Auction: Not Waiting for All Proposals

In contrast to the synchronized auction, which operates in a batch mode, many double auctions are continuous. They do not accumulate bids and asks before matching them. A transaction occurs as soon as a bid matches or exceeds an ask, without waiting for other bids/asks. Not waiting for bids and asks leads to faster price discovery; we examine how it affects efficiency.

There are several ways of modeling a continuously clearing double auction. Modeling such markets with ZI traders is difficult because they have identical strategies and therefore theoretically have the same speed. In our version of a continuous auction, it is assumed that every trader bids/asks before any trader bids/asks again, and so on with the traders bidding/asking in a random order within a round. If a new bid exceeds the current bid, which starts at zero, it becomes the current bid. Similarly, if a new ask is less than the current ask, which starts at one, it becomes the current ask. As soon as the current bid and current ask cross, there is a transaction. Otherwise, another trader bids/asks. If all traders have bid/asked and the current bid and current ask do not cross, then the market moves into the next round. This process is continued until the only available unit is sold.

In this section we assume that bids and asks are not made public, and therefore the probability distribution of bids and asks of ZI traders does not change from one round to the next. As before, we examine the efficiency as \( n \to \infty \). As shown in the Appendix,

\[
E = 1 - (1 - \beta) \frac{\beta + \beta^2}{2(1 - \beta(1 - \beta))}.
\]

As compared with the probability of an inefficient transaction in a synchronized auction \( (\beta^2(1 - \beta(1 - \beta))) \), from expression (5), the probability of an inefficient transaction in a continuous auction is strictly higher except at \( \beta = 0 \). The efficiency of a continuous auction is lower because it allows an EMB to buy from the IMS before a higher IMB bid can be received.

The efficiency is plotted as the bottom curve in Figure V, labeled “Cont. Without Data.” Minimum expected efficiency is about 74.5 percent at \( \beta = 0.56 \). The average expected efficiency assuming a uniform distribution of \( \beta \) is 84.5 percent. Note that

\[ \text{7. For a discussion of problems in modeling continuous double auctions, see Gode and Sunder [1992].} \]
the apparently small change in market rules (from synchronized to continuous auction) lowers the minimum expected efficiency by almost 6 percent. Thus, similar to subsection III.C, accumulating bids and asks before ranking makes price priority effective because an extramarginal trader is unable to displace an intramarginal trader merely because he submitted his bid or ask before the intramarginal trader.
VI. Effect of Making Bids and Asks Public (Rule 7)

Allowing traders to observe the outstanding bid or ask and to revise their bids and asks in response also raises market efficiency because the extramarginal traders are outbid or undersold sooner. We assume that the ZI traders adjust the feasible range in response to information about the current bid and current ask.

A. Synchronized Auction When Bids and Asks Are Public

We modify the synchronized auction so that the bids and asks are made public. Instead of bidding \( b_i \sim U(0,v_i) \), ZI buyers now bid \( b_i \sim U(\min(v_i, \text{current bid}), v_i) \), where \( v_i \) is the reservation value to the buyers. The ZI seller asks \( a_i \sim U(0,\text{current ask}) \).

There is no change in expected efficiency for cells 1, 3, and 4 (see Figure VI in the Appendix); in cell 2 the second-round IMB bid will be greater than \( \beta \), raising the efficiency in this cell to 1. Thus,

\[
E = 1 - (1 - \beta)^2.
\]

The probability of an inefficient transaction is \( \beta^2 \), which is less than the probability of inefficient transaction when traders do not observe market data that is \( \beta^2/(1 - \beta(1 - \beta)) \) (see expression (5)). Knowing bids and asks increases efficiency for all values of \( \beta \) because the IMB outbids the EMBs earlier. The minimum expected efficiency is raised from 80.8 percent to 85.2 percent at \( \beta = 2/3 \), an increase of 4.4 percent. This is shown as the top curve in Figure V, labeled “Synch. With Data.” The average expected efficiency assuming a uniform distribution of \( \beta \) is 91.8 percent.

Comparison of expressions (2) and (7) reveals how two-sided competition increases efficiency. In a sealed-bid auction an inefficient trade occurs when the IMB bid \( \leq \beta \); whereas in a double auction it occurs only if the IMB bid \( \leq \beta \) and the IMS ask \( \leq \beta \), which is less likely. Comparing expression (4) (as \( n \to \infty \)) with expression (7), the expected efficiency of a synchronized auction with public bids, we get the percentage of surplus exploited in the first round as

\[
E = \frac{0.5 - 0.5\beta^2 + \beta^3}{1 - \beta^2(1 - \beta)} = 0.5 \left( 1 + \frac{\beta^3}{1 - \beta^2(1 - \beta)} \right)
\]

This shows that for many values of \( \beta \) only 50 percent of the surplus is extracted in the first round.
B. Continuous Auction When Bids and Asks Are Public

We test whether the efficiency of continuous auction also increases if bids and asks are public. A ZI buyer now bids \( b_i \sim U(\min(v_i, \text{current bid}), v_i) \), where \( v_i \) is the redemption value of the unit to the buyer. The ZI seller asks \( a_i \sim U(0, \text{current ask}) \).

There is no change in the calculation of expected efficiency for cells 1, 3, and 4 (see Figure VI in the Appendix). In cell 2 the second-round IMB bid will be greater than \( \beta \), and so in this cell efficiency would be 1. Thus,

\[
E = 1 - (1 - \beta)/((\beta + \beta^2)/2).
\]

(9)

The probability of an inefficient trade is \((\beta + \beta^2)/2\), which is less than the probability of an inefficient trade in a continuous auction without public data, but is greater than the probability of an inefficient trade in a synchronized auction with public data.

The curve in Figure V labeled “Cont. With Data” plots expression (9). The availability of data improves the efficiency of the continuous auction for all values of \( \beta \) except at 0 and 1 and raises the minimum from about 74.5 percent (at \( \beta = 0.56 \)) to 80.75 percent (at \( \beta = 1/\sqrt{3} \)), a gain of about 5 percent. The average expected efficiency increases from 84.5 percent to 87.6 percent.

VII. CONCLUSIONS AND GENERALITY OF RESULTS

The results of the paper are summarized in Figure I. The paper suggests that for all their complexity, some important market phenomena have simple explanations. We show that simple and externally observable rules that define markets may be responsible for most of the high allocational efficiency of markets. As long as the traders stay within the confines of market rules, allocational efficiency remains high even if human traders are replaced by Zero Intelligence traders. Behavior and unobservable motivations of individuals apparently have only a small effect on allocational efficiency.

Efficiency is lower if (1) traders indulge in unprofitable trades, (2) traders fail to negotiate profitable trades, and (3) extramarginal traders displace intramarginal traders. Voluntary exchanges among traders who have the judgment to avoid losses eliminate the first source of inefficiency. Multiple rounds of bids and asks, i.e., multiple opportunities to negotiate, reduce ineffi-
ciency due to the second source. Section IV shows that the expected efficiency with ZI traders is only 50 percent if the double auction is ended after the first round of bidding. Further rounds raise this lower bound to 81 percent (see Figure IV) by reducing the probability of no trade.

The expected loss of efficiency when intramarginal traders are displaced is a product of the magnitude of inefficiency from displacement and its probability. The magnitude of inefficiency depends only on the shape of extramarginal demand and supply, which are often ignored, but the probability of displacement depends, in addition, on the market rules.

Without a price system there is random allocation so that the probability of displacement converges to one as the number of extramarginal buyers increases. Efficiency approaches zero as the redemption values of extramarginal buyers approach zero. (It could be negative if extramarginal sellers are also present.) Imposition of the binding contract rule and the price priority rule creates a price system in its most basic form. This system discriminates against extramarginal bidders because their redemption values are low, and their lower bids are given lower priority. Increasing the extramarginal buyers' redemption values increases the probability of displacement but lowers the loss of efficiency when displacement does occur. This market-level trade-off raises the lower bound on expected efficiency from zero to 75 percent. This market-level trade-off exists even if individuals do not trade off profit from a bid against the probability of it being accepted. The price priority rule can also be interpreted as sellers having the rationality to accept the highest bid and the buyers having the rationality to accept the lowest ask. As mentioned in footnote 2, this illustrates a fascinating duality between human thinking and institutional "codification" of that thinking which trains and protects less-rational traders—if the trading rules are "smart," the traders need not be.8

The double auction rule (allowing sellers to ask as well as buyers to bid) increases efficiency to 81 percent in a synchronized double auction because now an inefficient trade requires both the intramarginal ask and the intramarginal bid to be lower than the extramarginal bid. In a sealed-bid auction, inefficient trade

8. We thank an anonymous referee for elucidating this point.
requires only the latter. The accumulation rule (accumulating bids and asks before matching) makes price priority more effective because extramarginal buyers cannot get the unit merely by bidding before the intramarginal buyer. For example, in a sealed-bid auction without accumulation, i.e., if the unit is sold to the first bidder, the probability of displacement is \( n/(n + 1) \), which is the same as that without a market. The probability of displacement decreases as more bids are accumulated before ranking. A continuous auction's efficiency is lower than that of a synchronized auction because of the same effect (see Figure V). Making bids and asks public also increases efficiency because the extramarginal traders are outbid by intramarginal traders sooner. The efficiency of both synchronized and continuous markets increases by about 5 percent with public bids and asks (see Figure V).

We now discuss the generality of our results. Our simulations show that the results are qualitatively unchanged with more general demand and supply functions. The basic intuition and results are also qualitatively unchanged if instead of choosing bids and asks uniformly over the entire feasible range, all traders become more or less “greedy.” The transactions will take fewer or more rounds of bidding. The results would be different if the structure of trading strategies is not the same for all traders. If intramarginal buyers became more greedy, and extramarginal buyers and intramarginal sellers remained unchanged, then the chances of extramarginal trades will increase, thereby lowering efficiency.

The ZI traders do not bid and ask more aggressively if the market is to be terminated after fewer rounds. Intelligent traders may partly adjust for fewer rounds of bids and asks by bidding higher and asking lower, which will reduce the probability of no-trade and increase the efficiency. So our results may overstate the role of multiple rounds of bids and asks.

The effect of random behavior of noise traders is also discussed in the finance literature [De Long et al. 1990]. Its focus is on how the random actions of noise traders introduce excessive volatility in stock prices, and how correlation in noise traders’ behavior across stocks introduces risk that rational traders cannot diversify away. However, the focus of this paper is on markets for only one commodity. So we do not examine the issue of correlated behavior across markets. The prices in markets with ZI traders are much more volatile than experimental markets with
human traders. However, the allocational efficiency is still quite high. Allocational efficiency of double auctions with human traders is significantly lower if some traders have market power [Holt, Langan, and Villamil 1996]. Since ZI traders are nonstrategic, it is not possible to analyze the consequences of market power in our framework.

We analyze centralized markets where all bids and asks are funneled into a central clearinghouse for matching. In contrast, in “open outcry” markets a buyer and seller trade directly with each other; i.e., the matching is “decentralized” [Osborne and Rubinstein 1990]. In such markets the expected efficiency with ZI traders will be lower than even the continuous double auction and the sealed-bid auction because in a decentralized market the intramarginal seller, in effect, is the first one to enter the market. The probability that the intramarginal seller will get to see the intramarginal buyers bid before trading with an extramarginal buyer decreases from 1 to 0 as \( n \), the number of extramarginal traders, increases from 0 to infinity. In contrast, in a continuous centralized market, the probability that the intramarginal buyer submits his bid before the intramarginal seller enters the market is 0.5 regardless of the number of extramarginal buyers. Our analysis leads us to conjecture that the decentralized matching reduces efficiency but permits a larger volume of trade. The efficiency of decentralized matching will be higher than completely random allocation because in the former a seller has a better chance of trading with the IMB than with the EMBs she meets along the way. Random allocation, in contrast, affords equal chance of trading with all buyers.

Understanding how externally observable market rules can cause a price system to efficiently aggregate unobservable individual preferences [Hayek 1945], and why simple individual behavior can generate highly efficient allocations in markets [Simon 1981], is a step toward unraveling the mystery of the invisible hand.
Appendix

Derivation of Expression 3

$m$ bids are accumulated before ranking.

- Intramarginal bid is among the $m$ bids
  - Probability of IMB bid $> \beta$
    - Probability $= \frac{\binom{m}{n} \cdot \beta}{\binom{m}{n} + 1}$
  - Probability $= 1 - \frac{m}{n + 1}$, efficiency $= \beta$

- Intramarginal bids is not among the $m$ bids
  - Probability $= (1 - \beta)$, efficiency $= 1$

- IMB bid $\leq \beta$
  - Probability $= \beta$

- IMB bid is the highest among the $m$ bids
  - Probability $= 1/m$, efficiency $= 1$

- IMB bid is not the highest
  - Probability $= 1 - 1/m$, efficiency $= \beta$

$$E = \frac{m}{n + 1} \left( (1 - \beta) + \beta \left( \frac{1}{m} + (1 - \frac{1}{m}) \beta \right) \right)$$

$$= 1 - (1 - \beta) \cdot \frac{(n + 1 - m) + (m - 1) \beta}{n + 1}.$$ 

Derivation of Expression 4

As before, we study the market outcome with ZI traders under the demand and supply specified in Figure II and as $n \to \infty$. There is no trade in the first round if the maximum bid is less than the IMS ask. Let $r$ be the IMS ask in the first round.

Event I: $1 > r > \beta$.

Since the EMB bid cannot be greater than $\beta$, a transaction can occur if and only if the IMB bid $\geq r$ (probability $= (1 - r)$, surplus extracted $= 1$); there is no transaction if the IMB bid $< r$ (probability $= r$, surplus extracted $= 0$). Therefore, the expected efficiency conditional on event I is $1 - r$.

Event II: $0 < r \leq \beta$.

Case A: The IMB bid $> \beta$, and it wins (probability $(1 - \beta)$, surplus extracted $= 1$).

Case B: The IMB bid $\leq \beta$ (probability $= \beta$), and a transaction occurs if the IMB or an EMB bid $\geq r$ (probability $= (1 - (r/\beta)^{n+1})$). The probability of the IMB winning is $1/(n + 1)$, and the probability of an EMB winning is $n/(n + 1)$. The expected effi-
iciency conditional on case B is \((1 - (r/\beta)^n + 1)/(1 + \beta n)/(n + 1)\).

The expected efficiency conditional on Event II = \((1 - \beta) + \beta (1 - (r/\beta)^n + 1)/(1 + \beta n)/(n + 1)\). Since \(r\) is a random variable, \(r \sim U(0,1)\), we can eliminate the conditioning on \(r\) by taking expectations:

\[
\text{Expected Efficiency} = \int_{0}^{1} (1 - r)dr + (1 - \beta)\int_{0}^{\beta} dr + \frac{\beta(1 + n\beta)}{n + 1} \int_{0}^{\beta} (1 - (r/\beta)^{n+1})dr
\]

\[
= 0.5 - \frac{n\beta(1 - 2\beta)}{2(n + 2)}.
\]

**Derivation of Expression 5**

We examine the efficiency as \(n \to \infty\). As \(n\) increases, the highest extramarginal bid converges to \(\beta\). Figure VI shows the probabilities of the four possible combinations of intramarginal bid and intramarginal ask. \(E\) denotes expected efficiency.

Cell 1: IMS ask \(\leq \beta\), and IMB bid \(\leq \beta\) (probability = \(\beta^3\)). An EMB has the highest bid of \(\beta\), which is \(\geq\) IMS ask. Therefore, a transaction occurs, and the surplus extracted is \(\beta\).

Cell 2: IMS ask > \(\beta\), and IMB bid \(\leq \beta\) (probability = \(\beta(1 - \beta)\)). No transaction can occur until the bids/ask move into one of the other three cells in a later round. The other rounds are identical to the first round since the budget-constrained ZI traders do not observe bids or asks and hence do not change the range of their bids or asks; the surplus extracted is \(E\).

Cell 3: IMS ask > \(\beta\), and IMB bid > \(\beta\) (probability = \((1 - \beta)^3\)). Since EMBs bid \(\leq \beta\), they are out of the market, and the IMB gets the unit in the first round or later; the surplus extracted is 1.

Cell 4: IMS ask \(\leq \beta\), and IMB bid > \(\beta\) (probability = \(\beta(1 - \beta)\)). Transaction occurs in the first round; the surplus extracted is 1.

Thus,

\[
E = \beta^3 + \beta(1 - \beta)E + (1 - \beta)^3 + \beta(1 - \beta) = 1 - \frac{(1 - \beta)^3}{1 - \beta(1 - \beta)}.
\]
Derivation of Expression 6

As before, we examine the efficiency as $n \to \infty$. As $n$ increases, more EMBs bid before the IMB bids or the IMS asks. Also the current bid, before the IMB bids or the IMS asks, approaches $\beta$. Let $E$ denote the expected efficiency. Consider Figure VI and the following events:

Event 1: IMB bid arrives before IMS ask (probability = 0.5).
   Cells 1 and 2: The IMB bid $\leq \beta$.
   Cell 1: The IMS follows with an ask $< \beta$ (probability = $\beta^2$). An EMB will trade, and efficiency = $\beta$.
   Cell 2: The IMS follows with an ask $> \beta$ (probability = $\beta(1 - \beta)$). No transaction occurs until a subsequent round, and efficiency = $E$.

Cells 3 and 4: The IMB bid $> \beta$ (probability = $(1 - \beta)$), and the IMB would eventually trade. Efficiency = 1.

Expected efficiency conditional on Event 1 equals $\beta^3 + \beta (1 - \beta)E + (1 - \beta)$. 
Event 2: IMB bid arrives after IMS ask.

Cells 1 and 4: The IMS ask ≤ β, (probability = β). An EMB will trade, and efficiency = β.

Cells 2 and 3: The IMS ask > β.

Cell 2: IMB follows with a bid < β (probability = β (1 − β)). No transaction occurs until a subsequent round, and efficiency = E.

Cell 3: IMB follows with a bid > β (probability = (1 − β)²). EMBs are outbid, the IMB would eventually trade, and efficiency = 1.

Expected efficiency conditional on Event 2 is β² + β(1 − β) E + (1 − β)².

Thus,

\[ E = 0.5(1 - \beta) + \beta^2 + \beta(1 - \beta)E + 0.5(\beta^2 + (1 - \beta)^2 + \beta(1 - \beta)E). \]

\[ = 1 − (1 - \beta)^{-1} \frac{\beta + \beta^2}{2(1 - \beta(1 - \beta))}. \]

---

**References**


