Asset pricing when traders produce payoff through hard computational search*

Hassan Andrabi† Peter Bossaerts‡ Konstantinos Ioannidis‡
Robert Woods‡ Nitin Yadav†

November 2026

Abstract

We study asset markets in which traders produce the payoff themselves. Each trader privately searches for solutions to a computationally hard problem, the 0–1 knapsack problem, and the asset pays the value of the best solution anyone finds during trading. Solutions are never disclosed, so prices are the only channel through which discoveries can spread, and no trader can verify that the optimum has been found. We ask whether the payoff reaches the optimum and whether prices converge to it. In 28 experimental markets, at least one participant always found the optimum. Prices converged to the resulting payoff: pooled across markets, the long-run price lies within 0.01% of the optimum, and in every market it is statistically distinct from the value of the second-best solution. In our design the payoff rule does not require the market organiser to know the optimum, thus providing a mechanism for using markets to incentivise intellectual discovery when the answer is unknown.

Keywords: asset pricing, information aggregation, endogenous payoffs, computational complexity, experimental markets, knapsack problem
JEL Codes: C92, D82, D83, G14

1 Introduction

Asset pricing theory and experiments almost always take asset payoffs as given. Dividends, liquidation values, and states of nature are determined exogenously; what is determined in the market is the price, which may or may not reflect the information traders hold about those payoffs. Even where prices affect fundamentals, as in the literature on the real effects of financial markets, the feedback runs through managers and other decision-makers who learn from prices, not through the traders who set them (Bond et al., 2012). Likewise, even where payoffs are future prices and thereby endogenous, the assumption of perfect competition and

*The research reported here was generously funded by the Leverhulme Trust. The authors would like to thank conference participants at the 2nd summer school on Cognitive Foundations of Decision Making (Ghent, 2025), the 2nd Annual Conference of the Network of the UK-Based Experimental and Behavioural Economists (Sheffield, 2025), the 15th Society for Experimental Finance Conference (Maastricht, 2025), and the 14th Society for Experimental Finance Conference (Stavanger, 2024).

†University of Melbourne, Centre for Brain, Mind and Markets.
‡University of Cambridge, Faculty of Economics.

perfect foresight takes away the ability of traders to influence them (Bossaerts et al., 2026); payoffs are endogenous only in the future.

We study a market in which the traders themselves produce the payoff. Participants trade a single asset while privately searching for solutions to a common, computationally hard problem, and the asset's terminal payoff equals the value of the best solution found by any participant during trading. Prices and payoffs are therefore determined jointly, by the same people, at the same time. Profits are made only through trading (buy low, sell high), or through deciding not to trade at all and capture the final payoff.

This setting raises two questions. First, is the payoff right: does the best solution produced during trading reach the true optimum of the problem? Second, are prices right: do transaction prices converge to the payoff that participants produce? The two questions are distinct. Prices could accurately reflect a payoff that falls short of the optimum, and participants could find the optimum without its value being reflected in prices. Because the asset pays the value of the best solution found by anyone, a single discovery is enough to secure the highest attainable payoff.

Neither outcome is guaranteed. Consider the payoff first. Producing a better solution requires cognitively costly search with uncertain returns. In standard models of costly information acquisition, traders know the quality of the information they can acquire, and hence its expected value (Grossman and Stiglitz, 1980). Our participants do not: they do not know whether a better solution remains, how likely further effort is to find one, or by how much it would raise the payoff. Indeed, determining whether a better solution exists is itself part of the computational problem they are trying to solve. Search also has a public-good component. A participant who finds a better solution raises the payoff of every asset unit in the market, not only of the units they hold. Because the payoff is set by the single best solution, search is a best-shot public good, the case in which free-riding is expected to be most severe (Hirshleifer, 1983; Harrison and Hirshleifer, 1989). It has the structure of a volunteer's dilemma: one discovery suffices, each participant would prefer someone else to bear its cost, and in the symmetric mixed-strategy equilibrium the probability that nobody succeeds rises with the number of participants (Diekmann, 1985). Our market, however, adds a private return that the volunteer's dilemma omits. A participant who finds a better solution can buy the asset below its new value before others learn of the discovery, and so profit from the price change their own discovery creates (Hirshleifer, 1971). Free-riding may therefore leave the payoff below the optimum, while trading gains push the other way; whether free-riding limits search in practice is an empirical question.

Now consider prices. Each solution gives its finder an objective lower bound on the terminal payoff and a reason to buy the asset below that value. A participant who has found the optimum resembles an insider in the markets of Plott and Sunder (1982): they hold the information that determines the payoff. Unlike those insiders, however, they cannot tell that they hold it. Participants do not know whether their own solution is the best currently available, or how much further the payoff may improve before trading ends. Above their own lower bound, valuation therefore depends on beliefs about what other participants have already found, what they may still find, and what they infer from market prices. Because solutions are never disclosed, orders

and trades are the only channel through which discoveries can reach prices, and traders must form beliefs about the beliefs of others, as in Keynes's beauty contest (Keynes, 1936; Allen et al., 2006). Prices may rationally exceed the best solution found so far if traders anticipate further discoveries, but higher-order beliefs may also keep prices away from the payoff that is eventually produced. Prices may in turn affect search: a price above a participant's own best solution may suggest that a better solution exists and prompt further search, or that someone else has already found it, inviting the free-riding described above. The market therefore combines objective lower bounds on the payoff with subjective beliefs about its remaining upside, and correct pricing requires trading to reveal the value of the best solution without revealing the solution itself.

We implement this environment with two tasks that participants performed simultaneously, in two windows open side by side on their screens. The first is the 0–1 knapsack problem (KP), a combinatorial optimisation task in which participants select the most valuable subset of items whose combined weight does not exceed a fixed capacity. The KP provides a stylised representation of intellectual discovery: valuable solutions emerge from combining components in the right way, as in invention, engineering design, scientific problem solving, and discrete resource allocation (Weitzman, 1998; Youn et al., 2015). It is also well suited to studying human search. It is easy to state but NP-hard to solve, and human performance on it declines with the computational difficulty of the instance (Murawski and Bossaerts, 2016; Bossaerts and Murawski, 2017). Feasible solutions are easy to find, but discovering the optimum requires more systematic search, and participants cannot tell whether a better combination remains: a participant who finds the optimum may not recognise it as such and may continue searching.

The second task is trading. Participants trade a single asset in a continuous double auction with an open, anonymous order book: they see all standing bids, asks, and past trades, but not who submitted them or the solutions behind them. Each participant starts the market with an endowment of cash and asset units and can buy or sell at any time during the ten minutes, but can neither borrow cash nor sell short. At the end of the market, each unit held pays the value of the best solution found by any participant, so a participant's earnings depend only on their final holdings of cash and units. Participants are not paid for the solutions they find; search pays off only through the asset. Trading therefore shapes the incentive to search: a participant's stake in a better solution is proportional to the units they hold, and every sale passes part of that stake to the buyer. The payoff rule rewards the value that participants produce, not whether anyone knows that the problem has been solved exactly: substantial value can be generated even if no one reaches the optimum.

We conducted 28 markets in five sessions with 99 participants, using 17 KP instances with 12, 18, or 24 items that vary in computational difficulty within each size. The experimenter knew the optimum of every instance; participants did not, and the optimum did not enter the payoff rule. This allows two comparisons that field data cannot provide: of the payoff participants produce with the best attainable payoff, and of transaction prices with the payoff produced. Every instance has a unique optimum, and the best inferior solution to which no further item can be added, which we call the second-best solution, lies between 1.6% and 7.5% below it. Prices that settle on the optimum can therefore be distinguished from prices that

settle on a near miss. Because the same participants trade in all markets of a session, markets within a session are not independent, and our inference accounts for this.

The payoff was right in every market. At least one participant found the optimum in all 28 markets, so the asset always paid the highest attainable value. Discovery was not left to a few: between 6 and 18 of the 19 to 20 participants in each market found the optimum independently, 71% of participants across all markets, far more than the single discovery the payoff requires. The first discovery came within the first minute in every market, and in most markets before the first trade, so that for nearly the full duration of each market the question was not whether the payoff would be produced but whether prices would reflect it. When trading began, however, only a few participants had typically found the optimum, and none of them could tell that their solution was the best possible.

Prices were right as well. Markets often opened far from the payoff, with early transactions up to a quarter above and a third below it, but prices converged. We estimate each market's long-run price from an autoregressive model of transaction prices. In 24 of the 28 markets, the 95% confidence interval includes the payoff, and in all 28 it excludes the value of the second-best solution, so the market settled on the optimum rather than on a near miss. Pooled across markets, the long-run price lies 0.01% below the payoff (95% confidence interval -0.03% to 0.01%); measured against the distance between the optimum and the second-best solution, prices settle 0.3% of the way towards the latter. Unlike in markets where traders buy above fundamental value in the expectation of reselling at a higher price (Smith et al., 1988), the remaining deviations show no systematic tendency towards overpricing. As prices converged, disagreement also disappeared: in every market the bid–ask spread narrowed, from a median of 6% of the optimum in the first fifth of trading time to about one price tick in the last fifth. The market thus priced the optimum although no participant could verify that it had been found and the solution itself was never disclosed.

We make two contributions. First, we contribute to the study of information aggregation in experimental asset markets. In those experiments the payoff is exogenous and participants receive private signals about it; the question is whether trading incorporates those signals into prices (Plott and Sunder, 1982, 1988; Forsythe and Lundholm, 1990). Markets often succeed, but aggregation is neither automatic nor complete. Prices reflect public information almost fully but less than half of traders' private information (Page and Siemroth, 2021), they can settle on values that no trader's information supports (Camerer and Weigelt, 1991), and when signals must be bought, costly acquisition can make prices less informative (Sunder, 1992; Huber et al., 2011; Corgnet et al., 2018). Our participants instead produce their information through search: each solution is a private lower bound whose quality its finder cannot judge, and the best solution produced determines the payoff itself. The market must therefore price an outcome that is still being created, using information that is entirely private. Aggregation nonetheless succeeds. Our setting also suggests why informative prices need not remove the incentive to acquire costly information (Grossman and Stiglitz, 1980): here the return to producing information comes partly from the payoff of the units one holds.

Second, we contribute to the study of markets as mechanisms for discovery. The most closely related experiments also use the KP but fix the payoff at the true optimum. In Meloso

et al. (2009), securities on individual items paid according to the optimal solution, and markets promoted discovery at least as well as a prize for the first correct solution; Bossaerts et al. (2024) show that a single asset paying the optimum produces accurate prices and effective dissemination of knowledge, whereas a complete set of securities produces noisy prices. In both designs, however, the asset pays the optimum whether or not anyone finds it, so search changes what traders know about a fixed fundamental but not the fundamental itself. We retain the single asset but change the payoff rule: the asset pays the value of the best solution participants actually produce. Settling the asset then requires only that solutions can be verified, not that anyone knows the optimum in advance. This links our design to two wider literatures. Innovation contests also reward the best solution submitted, but the prize is typically fixed by the designer and paid only to the winner (Taylor, 1995; Che and Gale, 2003; Terwiesch and Xu, 2008; Boudreau et al., 2011); in our market the reward to discovery is set by prices and shared by everyone who holds the asset. In prediction markets, traders who can influence the outcome they trade on are a concern (Wolfers and Zitzewitz, 2004; Ottaviani and Sørensen, 2007); here influencing the outcome is the purpose. We show that markets can jointly produce and price value under this payoff rule, and we discuss in the concluding section what this implies for rewarding discovery when the best attainable outcome is unknown.

The remainder of the paper is organised as follows. Section 2 describes the experimental design and procedures. Section 3 develops benchmarks for search and pricing. Section 4 reports whether the optimum was reached and whether prices converged to the resulting payoff. Section 5 discusses what these results imply for rewarding discovery when the best attainable outcome is unknown, and concludes.

2 Experimental design

2.1 The Knapsack Problem

An instance of the 0–1 Knapsack Problem consists of a set of I items, each characterised by a value v_i and a weight w_i, together with a capacity limit C. The decision variable x_i equals one if item i is included in the proposed knapsack solution and zero otherwise. The objective is to select the feasible subset of items with the highest total value:

$$\max_{\substack{x_i \in \{0,1\}, \\ i=1,2,\ldots,I}} \sum_{i=1}^{I} v_i x_i \quad \text{subject to} \quad \sum_{i=1}^{I} w_i x_i \leq C.$$

The value of a solution is the sum of the values of its selected items, while feasibility requires their combined weight not to exceed the capacity limit. Any unused capacity has no value.

The KP is simple to describe but computationally difficult to solve. It is a canonical NP-hard combinatorial optimisation problem (Sahni, 1975; Kellerer et al., 2004). An instance with I items has 2^I candidate subsets, although only a subset of these satisfies the capacity constraint. Constructing a feasible solution does not require solving the optimisation problem, but finding the exact optimum and establishing that no better feasible solution exists may require substantial search. The task is therefore well suited to the experiment: its rules are easy

to understand, while producing and recognising the optimal solution may require considerable cognitive effort.

Participants solved the KP using a drag-and-drop interface. The interface (see Figure A1) displayed all available items together with their values and weights, the capacity limit, and the current value and weight of the participant's knapsack. Items were displayed in random positions, as rectangles whose size increased with weight and whose green shading became darker with value, allowing participants to assess these attributes visually. Participants could add or remove items at any time during the market. The interface prevented them from adding an item if doing so would cause the combined weight to exceed the capacity limit. It also always displayed to participants the value of the best solution they had constructed during the market. Every feasible solution constructed through the interface was recorded, so participants did not need to submit solutions separately.

All participants faced the same KP instance within a given market, and could work on it throughout the ten-minute trading period. In each market the participants faced a KP instance they had not previously seen. Participants searched independently: they observed their own current solution and their best solution value, but not the solutions or values constructed by others. Participants were not allowed to use external aids (e.g., open a webpage with a KP solver). The construction and computational difficulty of the instances are described in Subsection 2.5.

2.2 The traded asset and payoff

In each market, participants traded a single asset, referred to as "Stock" in the experiment instructions. Let T denote the end of trading. Let S_{ntm} denote the value of participant n's best solution by time t in market m. The terminal payoff per asset unit was

$$V_m = \max_{n,t} S_{ntm}.$$

Thus, every asset unit paid the highest solution value recorded for any participant at any point during the market, regardless of who produced that solution or who held the asset at the end of trading. If no participant constructed a feasible solution, the asset paid zero.

At the beginning of each market, participants received a positive endowment of experimental cash and units of the asset. They could buy and sell units throughout the ten-minute trading period. Initial endowments varied across markets, and cash and asset holdings were reset before each new market. Markets were therefore financially separate: participants could not transfer cash or asset holdings between them.

This payoff rule makes the terminal payoff endogenous. Participant n's best solution, S_{ntm}, provides an objective lower bound on V_m. Discovering a better solution raises this lower bound and may also increase the payoff of every asset unit. Participants did not receive a separate payment for solutions they produced; any financial benefit operated through their asset holdings and trading decisions. Let O_m denote the true optimum of the KP instance used in market m. The experimenter knew O_m, but participants did not, and the asset paid V_m rather than O_m, where $V_m \leq O_m$.

Participant n's final wealth in market m was

$$W_{nm} = M_{nm} + Q_{nm}V_m,$$

where M_{nm} denotes final experimental cash and Q_{nm} denotes the number of asset units held at the end of the market. Ex post, purchasing the asset at a price below V_m increased final wealth relative to retaining the corresponding cash, while selling the asset above V_m increased final wealth relative to retaining it.

2.3 Market institution and interfaces

Trade was organised through an electronic open limit-order book, also known as a continuous double-sided auction.[1] During each market, participants could submit limit orders to buy or sell one or more units of the asset. A buy order, or bid, specified the maximum price the participant was willing to pay per unit, while a sell order, or ask, specified the minimum price the participant was willing to accept per unit.[2] Orders that did not trade immediately remained in the order book until they were executed, cancelled by the participant, or removed at the end of the market.

A transaction occurred when an incoming bid was at least as high as the lowest standing ask, or when an incoming ask was no higher than the highest standing bid. The transaction took place at the price of the standing order. Orders were ranked by price and, among orders submitted at the same price, by time of submission. Transactions were settled immediately, and any unexecuted portion of a partially filled order remained in the order book with the same limit price and retained its original time priority.

The trading platform distinguished between settled and available balances of cash and asset units. Settled balances reflected the participant's current holdings, whereas available balances excluded cash or asset units committed to standing orders. New orders must fit within the participant's available balance; the platform, therefore, prevented participants from submitting bids that exceeded their available cash or asks that exceeded their available asset holdings. Participants could therefore neither borrow experimental cash nor sell the asset short. Cancelling a standing order released the corresponding cash or asset units and restored them to the participant's available balance.

The trading interface displayed the participant's settled and available balances, an order form, the open order book, and the history of transactions within the market. Participants used the order form to specify the side, quantity, and limit price of an order, and could cancel their own standing orders directly from the order book. The resulting transaction prices provide our primary measure of pricing accuracy, while standing bids and asks are used to measure bid–ask spread. The trading and KP interfaces were open in separate browser windows side by side, allowing participants to move between solving the problem and trading the asset throughout the market. The full experimental instructions, including screenshots and detailed descriptions

[1] The software, Flex-E-Markets, is used on a subscription basis from `flexemarkets.com`, and uses price–time priority, the matching rule of major electronic exchanges such as NYSE, LSE, and Euronext.

[2] Prices were quoted in whole experimental dollars, so the minimum price increment, or tick, was one experimental dollar.

of both interfaces, are reproduced in Appendix A.

2.4 Information structure

Participants observed the complete KP instance, including the value and weight of every item and the capacity limit. They also observed their own current knapsack, their best solution value S_{ntm}, and their current cash and asset holdings. They did not observe the solutions constructed by other participants or their values, or the corresponding private lower bounds on the asset payoff. They also did not observe the highest solution value produced in the market, the terminal payoff, or the true optimum of the instance.

The open order book displayed all standing bids and asks anonymously, together with the best bid, the best ask, and the resulting bid–ask spread. Participants also observed the complete history of transaction prices within the market. They could identify their own standing orders, but not the participants who had submitted other orders or initiated past trades. Market activity was therefore the only public channel through which private discoveries could affect the information available to others, although neither orders nor transaction prices revealed the item combination underlying a solution or unambiguously identified its value.

The experimenter observed every participant's best-solution path S_{ntm}, together with all orders, cancellations, and transactions and their timestamps. We therefore observed the complete evolution of the solution values produced during each market, the realised terminal payoff V_m, and the true optimum O_m. The resulting data allow us to compare the realised payoff with the true optimum and transaction prices with the realised payoff. The order-book data also allow us to study the evolution of the bid–ask spread.

2.5 KP instances and computational complexity

The experiment used 17 distinct KP instances: five with 12 items, six with 18 items, and six with 24 items.[3] Within each instance size, the instances varied in their computational difficulty. The number of items determines the number of candidate subsets: instances with 12, 18, and 24 items have, respectively, $2^{12} = 4{,}096$, $2^{18} = 262{,}144$, and $2^{24} = 16{,}777{,}216$ possible subsets, although only a subset of these satisfies the capacity constraint. Instance size provides a coarse measure of the scale of the search problem. To capture variation in computational difficulty among instances with the same number of items, we use the Sahni-k measure introduced below.

Sahni-k measures how far the optimum lies from the solution of a simple greedy heuristic (Sahni, 1975; Meloso et al., 2009; Murawski and Bossaerts, 2016). The greedy algorithm ranks items by their value-to-weight ratio and adds them in that order whenever they fit. An instance has Sahni-$k = 0$ if this algorithm finds the optimum. Otherwise, Sahni-k is the smallest number of items that must be placed in the knapsack in advance for greedy completion of the remaining capacity to produce the optimum; any such initial selection is necessarily part of the optimal solution. Higher values mean that more items must be chosen jointly before the heuristic can

[3]The experiment also included one 10-item instance, 10-01, used in the final market of the first two sessions. It does not meet the minimum-gap criterion described below: its second-best solution lies within 0.17% of the optimum, so prices settling on the optimum cannot be distinguished from prices settling on the second-best solution. We therefore exclude its two markets from the main analysis and report them in Appendix C; including them does not change our conclusions.

do the rest, so Sahni-k provides a second measure of difficulty among instances with the same number of items. Computing the minimum value of k that guarantees the optimum requires knowing the optimum, so participants could not compute it during the market.

We required each instance to have a unique optimal solution. When selecting among candidate instances, we also imposed a minimum proportional gap between the optimal value and the value of the second-best solution, defined as the best inferior solution to which no further item can be added.[4] This restriction reduced the likelihood that payoff and price outcomes near the optimum would be observationally indistinguishable from outcomes based on a suboptimal solution. The generating parameters, candidate-pool sizes, and minimum gap were adjusted across instance sizes and target Sahni-k values because suitable instances became increasingly difficult to find at larger sizes and at the targeted values of Sahni-k. Within each instance size, we selected instances spanning different Sahni-k values so that the experiment varied both the scale and computational difficulty of the search task.

Table 1 summarises the instances used in the experiment, reporting their capacity limit, Sahni-k, optimal value, and the value of the second-best solution together with its distance from the optimum. Optimal values range from 1,313 to 1,704, so the highest attainable payoff in one market could be up to 30% higher than in another. Because the instance changed in every market and participants received no feedback on payoffs between markets, they could not use earlier prices or solutions to infer the payoff of the current market. Optimal values were also similar across instance sizes, averaging 1,523, 1,534, and 1,482 for the 12-, 18-, and 24-item instances, so the number of items carried no information about the attainable payoff. The complete item-level configurations, including the value and weight of every item and the items belonging to the unique optimal solution, are reported in Appendix B. The appendix also describes the generation and selection of the instances.

2.6 Experimental procedures

We conducted five experimental sessions, with 20 participants in four sessions and 19 participants in one session, for a total of 99 participants. Each session consisted of six ten-minute markets. We use the term "market" throughout the paper; the experimental instructions referred to each market as a "period". All participants within a session traded together in every market, and each participant attended only one session.

Each session used six different KP instances. Instance size increased across sessions: the first session used 12-item instances and the last 24-item instances, with the intermediate sessions combining adjacent sizes. Within each session, instances of different Sahni-k were spread across positions in the market sequence. Several instances were presented in more than one session, allowing us to observe outcomes for the same instance in independent participant groups. The realised assignment of instances across sessions and markets is reported in Appendix B.

Participants were recruited from the participant pool of the Cambridge Experimental and Behavioural Economics Group (CEBEG) laboratory. The sessions took place in person during

[4]The restriction matters. Removing a low-value item from the optimum yields a feasible solution that can be worth more than the second-best solution as defined here. Such a solution is not a plausible stopping point, because the participant holding it can see that the removed item still fits.

Instance	Capacity	Sahni-k	Optimal value	Second-best value	Distance from optimum (in value)	(in %)
Panel A: 12-item instances						
12-01	1,798	2	1,633	1,533	100	6.12
12-02	1,580	0	1,556	1,440	116	7.46
12-03	1,901	3	1,504	1,425	79	5.25
12-04	1,580	1	1,383	1,283	100	7.23
12-05	1,657	4	1,540	1,454	86	5.58
Panel B: 18-item instances						
18-01	1,172	0	1,662	1,602	60	3.61
18-02	934	2	1,533	1,474	59	3.85
18-03	1,052	1	1,605	1,548	57	3.55
18-04	1,120	6	1,563	1,503	60	3.84
18-05	1,135	4	1,493	1,434	59	3.95
18-06	1,003	8	1,350	1,302	48	3.56
Panel C: 24-item instances						
24-01	927	8	1,523	1,491	32	2.10
24-02	941	4	1,368	1,340	28	2.05
24-03	874	0	1,313	1,279	34	2.59
24-04	916	10	1,492	1,466	26	1.74
24-05	1,100	12	1,704	1,677	27	1.58
24-06	969	14	1,489	1,462	27	1.81

Notes: The table reports the 17 instances used in the main analysis, grouped by number of items. Second-best value is the value of the best inferior solution to which no further item can be added without exceeding the capacity limit. Distance from optimum is the optimal value minus the second-best value, also expressed as a percentage of the optimal value. All instances have a unique optimal solution. The 10-item instance 10-01, excluded from the main analysis, is reported in Appendix B.

Table 1: Knapsack problem instances

February and March 2024. Before participating, all participants provided informed consent. The experiment received ethical approval prior to data collection.[5] Each session lasted approximately three hours.

The experiment began with a detailed explanation of the instructions, which were read aloud by the experimenters. Participants first practised using the KP interface on its own, then practised using the market interface, and finally practised solving the KP and trading the asset simultaneously. These initial practice exercises were untimed, allowing participants to become familiar with both interfaces before facing any time constraint. Participants could ask clarifying questions throughout. The instructions and initial familiarisation phase lasted approximately 90 minutes.

Participants subsequently completed one full ten-minute practice market, in which they solved a KP and traded the asset simultaneously under the same time constraint as in the formal markets. After answering any remaining questions, we introduced a short break and then conducted the six formal markets, with a one-minute break between consecutive markets. In each market, participants were notified when five minutes and one minute remained. They received no feedback between markets about either the realised asset payoff V_m or the true optimum O_m.

At the end of each session, two markets were selected for payment: one from the first three markets and one from the final three markets. The selection was made by rolling a digital die live on a shared screen. Participants were paid the sum of their final wealth in the two selected markets, converted at a rate of 1,000 experimental dollars to one pound. A minimum total payment of £15 was guaranteed. Participants earned £39.30 on average; the minimum payment never applied.

3 Benchmarks

This section sets out the benchmarks we use to answer our two questions, and the reasons why the answer to each could be no. We do not impose a model of search or price formation. Instead, we separate two steps: the production of payoff-relevant information through participants' search, which determines the payoff, and the incorporation of that information into prices. Subsection 3.1 defines the benchmarks, Subsection 3.2 explains why the optimum may not be found even though every participant gains when it is, and Subsection 3.3 explains why prices may not reach the payoff even when the optimum is found.

3.1 Benchmarks

Discovery. Recall that S_{ntm} denotes the value of the best solution participant n has constructed by time t in market m, and let N_m denote the number of participants in that market. The discovery frontier

$$F_{tm} = \max_n S_{ntm}$$

[5] University of Cambridge ethical approval reference number: UCAM-FoE-24-01.

is the value of the best solution found by anyone at time t. Because S_{ntm} records each participant's best solution so far, F_{tm} is weakly increasing over the market period. The asset payoff is the frontier at the end of trading, $V_m = F_{Tm}$, and it cannot exceed the optimum O_m of the instance:

$$F_{tm} \leq V_m \leq O_m \quad \text{for all } t \leq T.$$

The difference $O_m - V_m \geq 0$ is the payoff shortfall, the value lost because no participant found the optimum. Our first question is whether this shortfall is zero.

A zero shortfall requires only one participant to find the optimum. Let

$$N_m^* = \#\{\, n : S_{nTm} = O_m \,\}$$

denote the number of participants who had found the optimum by the end of the market, so that $V_m = O_m$ if and only if $N_m^* \geq 1$. Two further measures describe how the optimum was reached: the share of solvers, N_m^*/N_m, and the time of first discovery,

$$\tau_m = \min\{\, t : F_{tm} = O_m \,\},$$

from which point at least one participant has found the optimum.

Prices. Let P_{km} denote the k-th transaction price in market m, $k = 1, \ldots, K_m$, in order of execution. The difference between a price and the optimum decomposes as

$$P_{km} - O_m = \underbrace{(P_{km} - V_m)}_{\text{pricing error}} + \underbrace{(V_m - O_m)}_{\text{payoff shortfall}}.$$

The pricing error answers our second question, whether prices are right; the payoff shortfall answers the first. When the optimum is found, the shortfall is zero, and the deviation of prices from the optimum equals the pricing error.

To compare markets whose payoffs differ, we express deviations as a percentage of the optimum,

$$d_{km} = 100\,\frac{P_{km} - O_m}{O_m}.$$

This rescales the difference but leaves its sign and its zero unchanged. We call the level at which d_{km} settles as trading proceeds the long-run price deviation, μ_m; Section 4 explains how we estimate it. Prices are right in market m if $\mu_m = 0$.

The second-best solution. A long-run deviation statistically indistinguishable from zero is not by itself evidence that the market priced the optimum: it may be equally indistinguishable from the payoff that a slightly worse solution would have produced. Let O'_m denote the value of the second-best solution, the best inferior solution to which no further item can be added, and let

$$\Delta_m = O_m - O'_m > 0$$

denote the distance between the two. A market that settled on the second-best solution would have a long-run deviation of $-100\,\Delta_m/O_m$. We therefore also measure the long-run deviation in units of this distance,

$$z_m = \frac{\mu_m}{100\,\Delta_m/O_m},$$

so that $z_m = 0$ corresponds to the optimum and $z_m = -1$ to the second-best solution in every market. We conclude that prices are right in market m if μ_m is statistically indistinguishable from zero and statistically distinguishable from the value implied by the second-best solution, $z_m = -1$.

3.2 Why the optimum may not be found

Finding the optimum benefits every participant who holds the asset, not only the finder, since the payoff is the value of the best solution found by anyone. A participant's earnings depend on the solutions constructed during the market only through the payoff of the units they hold, $Q_{nm}V_m$. The payoff is therefore a public good of the best-shot type, whose level is set by the single highest contribution (Hirshleifer, 1983), and search is exposed to free-riding.

A stylised version of the problem is the volunteer's dilemma (Diekmann, 1985). Suppose each of the N_m participants chooses whether to search, at a cost $\kappa > 0$ in effort and in time not spent trading, and that the optimum is found if at least one participant searches. Every participant gains $b > \kappa$ if the optimum is found, whether or not they searched themselves. Searching is then worthwhile for a participant only if no one else searches. In the symmetric equilibrium, each participant searches with probability σ, chosen so that they are indifferent between searching and not:

$$(1-\sigma)^{N_m-1} = \frac{\kappa}{b}.$$

The probability that no one searches, and the optimum is not found, is

$$(1-\sigma)^{N_m} = \left(\frac{\kappa}{b}\right)^{N_m/(N_m-1)},$$

which is strictly positive and rises with the number of participants towards κ/b. In large groups, the probability that the optimum is lost is approximately the ratio of the cost of search to its benefit, however many participants could find it. Experiments confirm that free-riding is most pronounced when the public good is of the best-shot type (Harrison and Hirshleifer, 1989). In volunteer's dilemmas, however, outcomes in which no one volunteers become less frequent as groups grow, contrary to the equilibrium prediction (Goeree et al., 2017).

Our setting differs from the stylised game in three ways that work in opposite directions. First, the benefit of a higher payoff is proportional to holdings: participants who hold more units have more to gain from discovery, while a participant who has sold all their units gains nothing. Second, the market adds a private return to search. A participant who finds a solution knows that the payoff is at least its value and can buy units from others at prices below it, a gain that accrues to the finder alone (Hirshleifer, 1971). This return weakens the free-rider problem, but it depends on prices being below the value of the solution found, and it shrinks as prices rise towards that value. Third, search is not a binary choice and does not guarantee

success: participants choose how much effort to devote to the problem, and effort raises the probability of discovery without ensuring it.

The stylised game is not meant to predict how often the optimum is found. It shows that a zero payoff shortfall is not guaranteed even when every participant would gain from discovery and the optimum is within reach. Whether it is reached is an empirical question, which we answer in Section 4.

3.3 Why prices may not reach the payoff

The natural benchmark for prices is rational expectations: in a fully revealing equilibrium, prices aggregate the information dispersed among traders and equal the payoff conditional on all of it (Grossman, 1976; Radner, 1979). Experimental markets in which some traders know the payoff and others do not often come close to this benchmark (Plott and Sunder, 1982, 1988). Here, the benchmark is demanding for three reasons.

First, the information that participants produce does not identify the payoff. A solution is a lower bound on the payoff, not a signal of it: its finder knows that the payoff is at least its value, but cannot verify whether a better solution exists or will be found. Even a participant who has found the optimum cannot tell that they do. Unlike the informed traders in Plott and Sunder (1982), the best-informed participant in our markets cannot recognise that they are the best informed. With unlimited computation, the optimum would be known to everyone once the instance was displayed, and prices should equal O_m from the start. Under realistic computational limits, prices can reach the payoff only if trading incorporates what search has produced.

Second, prices are the only public channel through which this information can spread. Participants do not observe each other's solutions or their values, so a discovery can affect others' beliefs only through the bids, asks and trades of the participant who made it.

Third, before the optimum is found, the payoff is not yet determined: it depends on search that has not yet happened. The relevant benchmark at time t is the expected payoff given the information available at t, and it can lie above the frontier F_{tm} when participants expect further discovery. Only once the optimum has been found does the benchmark become O_m. Even then, because no participant can rule out a better solution, prices could rationally settle below O_m, for example at a value close to the second-best solution if traders who found it believe it may be the best available.

Relating prices to the discovery measures from Subsection 3.1 helps interpret their path within each market. Before the first discovery of the optimum, a price above the frontier F_{tm} prices in discovery that has not yet occurred, while a price below F_{tm} is below a value that some participant has already found, and gives that participant a reason to buy. After the first discovery, a price below O_m is below a value that at least one participant has found, but that no participant can confirm is the payoff. A price above O_m exceeds every feasible payoff and is unambiguously too high, as in the bubbles of markets with known payoffs (Smith et al., 1988). Our formal tests in Section 4 concern where prices settle in the long run. These comparisons describe how they get there.

4 Results

4.1 Was the payoff right?

At least one participant found the optimal solution in every market. In all 28 markets of the main sample, $N_m^* \geq 1$ and the payoff therefore equalled the optimum, $V_m = O_m$; the same holds in the two markets that used instance 10-01 (Appendix C). The payoff shortfall was zero throughout, so the deviations of prices from the optimum analysed below are pricing errors.

Finding the optimum was not left to a few participants. Across the 28 markets, 396 of 555 participant-market observations (71%) ended with the optimal solution. The number of solvers per market ranged from 6 to 18, a share of between 30% and 90% of participants. The share was lowest in the three 24-item instances with the highest Sahni-k (24-04, 24-05 and 24-06), where 9, 6 and 8 participants found the optimum. Even in the hardest instances, then, the payoff did not rest on a single discovery.

The optimum was also found early. Measured from the opening of the market, the first discovery occurred after a median of 14 seconds and within the first minute in every market, the latest after 56 seconds.[6] Later solvers followed over the rest of the market: in the median market, half of those who eventually found the optimum had done so within 75 seconds. The payoff was therefore determined, though not known, for almost the entire trading period, although no participant could know it. Appendix C reports N_m^* and τ_m for each market.

These outcomes are far from the free-riding benchmark of Subsection 3.2. Because search does not guarantee success, the share of solvers is a lower bound on the share of participants who searched. In the symmetric equilibrium of the volunteer's dilemma with 20 participants, a search probability as high as the lowest share we observe, 30%, would require the cost of search to be below 0.1% of its benefit, $\kappa/b < 0.7^{19} \approx 0.001$. Participants searched far more than an equilibrium with any appreciable cost of search would predict. This is consistent with the private return to discovery that the market provides, and with experimental evidence that outcomes in which no one volunteers are rarer in large groups than the equilibrium implies (Goeree et al., 2017).

4.2 Estimating long-run prices

We estimate the long-run price deviation μ_m separately for each market. A single incoming bid or ask can trade against several standing orders at once; we treat each such order as one price event, with price equal to the volume-weighted average of its fills, so that the deviations d_{km} form one series per market indexed by price event k.[7] For each market we fit an autoregressive model of order p,

$$d_{km} = \alpha_m + \sum_{j=1}^{p} \rho_{jm}\, d_{k-j,m} + \varepsilon_{km},$$

[6] We measure time from the first recorded event in a market, whether a knapsack move or an order. Knapsack moves were recorded up to 13 seconds before the first order.

[7] Treating each fill as a separate observation would repeat near-identical prices and overstate the persistence of the series. Event time is also the natural clock for the series, since trades arrive irregularly.

by ordinary least squares, conditioning on the initial observations. If the model is stable, d_{km} reverts to the long-run mean

$$\mu_m = \frac{\alpha_m}{1 - \sum_{j=1}^{p} \rho_{jm}},$$

the level at which prices settle once the influence of the opening prices has died out. Conditioning on the initial observations matters here. Exact maximum likelihood treats the first prices as draws from the long-run distribution (Hamilton, 1994, ch. 5), while markets often opened far from where they settled; conditional estimation lets the opening prices inform the dynamics without forcing them to be typical of the long run. We compute the standard error of μ_m by the delta method from Newey–West standard errors (Newey and West, 1987), which allow for heteroskedasticity and for any residual autocorrelation the model leaves.[8]

The lag order is chosen market by market with a fixed rule. We fit models with $p = 1, \ldots, 5$ on a common sample, beginning at the sixth price event, so that the models are compared on the same observations. Among stable models, we take the one with the lowest Bayesian information criterion (Schwarz, 1978). If its residuals fail a heteroskedasticity-robust Breusch–Godfrey test for autocorrelation at the 5% level (Breusch, 1978; Godfrey, 1978; Wooldridge, 1991), we move to the smallest higher order that passes; if none does, we keep the original choice and rely on the Newey–West standard errors.[9]

The estimation samples contain between 19 and 368 price events (median 61), and the selected models use between one and five lags. Every selected model is stable, with largest autoregressive root below 0.95. In 6 markets, no order up to five passes the test. The conclusions do not depend on these choices: Newey–West and heteroskedasticity-robust confidence intervals have a median width ratio of 1.00, and across the five lag orders the estimate of μ_m varies by a median of 0.16 percentage points. Appendix C reports the selected order and diagnostics for each market.

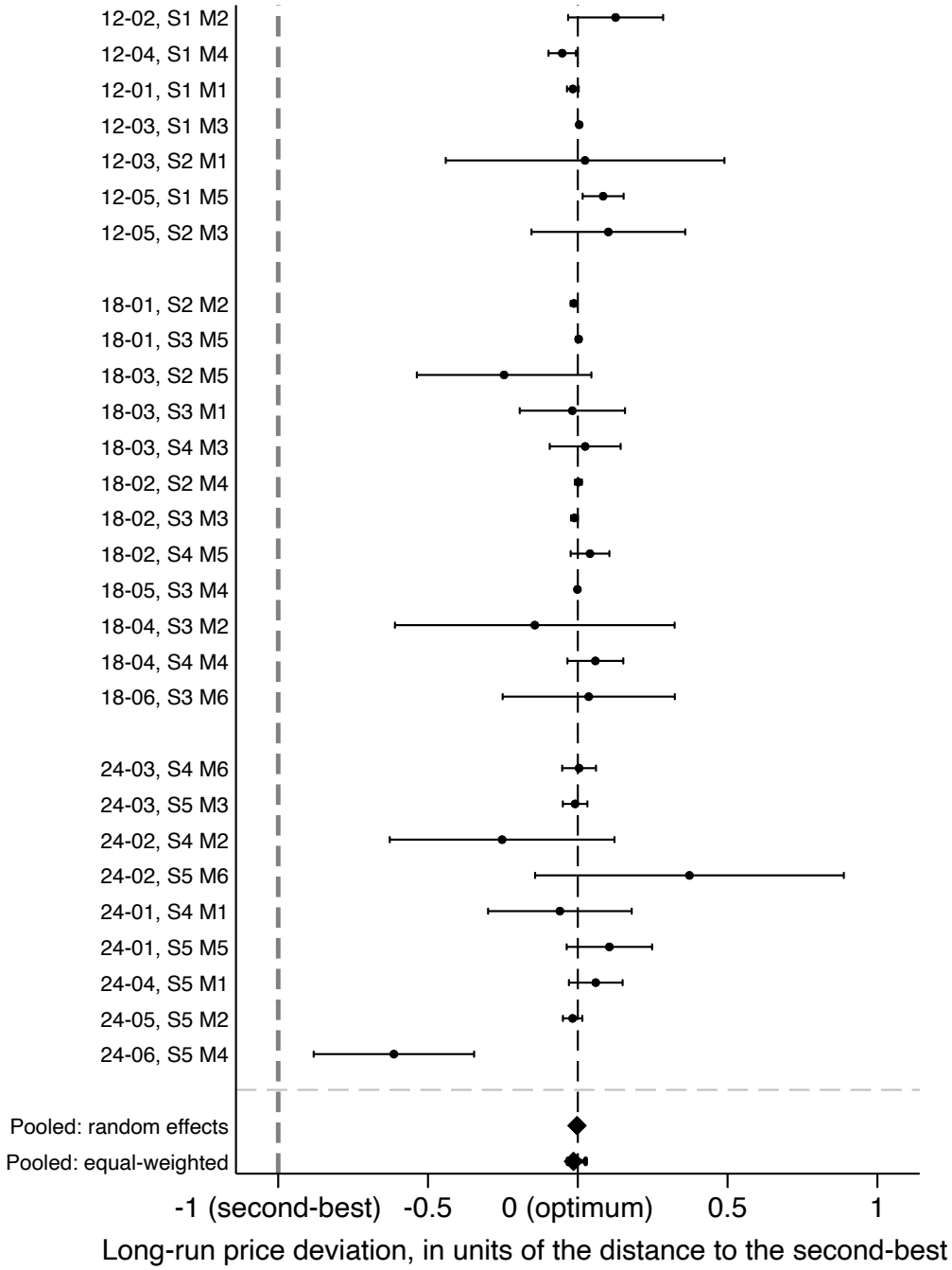
4.3 Were prices right?

Figure 1 shows the long-run price deviation of each market on the normalised scale, where 0 is the optimum and -1 the value of the second-best solution. The estimates cluster tightly around the optimum. The median long-run deviation is 0.01% of the optimum, and in 20 of the 28 markets the long-run price lies within a tenth of the distance between the optimum and the second-best solution. In 24 markets the 95% confidence interval includes the optimum. In all 28 it excludes the second-best solution, so in no market can prices be described as having settled on a near miss.

Of the four markets whose intervals exclude the optimum, three deviate from it by little. In two, prices settled slightly below the optimum, by 1% and 5% of the distance to the second-best solution (18-01 in session 2 and 12-04 in session 1); in the third, slightly above it, by 9% of that distance (12-05 in session 1), and so above every feasible payoff. The exception is the market

[8] The Newey–West lag length is $\lfloor 4(K/100)^{2/9} \rfloor$ for a market whose estimation sample contains K price events, following Newey and West (1994).

[9] We use the robust Breusch–Godfrey test rather than the Ljung–Box test (Ljung and Box, 1978) because price changes are much larger early in a market than later, and the Ljung–Box test over-rejects under such heteroskedasticity.



Notes: Long-run price deviation z_m of each market, measured in units of the distance between the optimum and the second-best solution, with 95% confidence intervals from Newey–West standard errors. On this scale, 0 is the optimum (dashed line) and -1 the value of the second-best solution (grey line). Markets are grouped by instance size and ordered by Sahni-k within size. The pooled estimates at the bottom are from a random-effects meta-analysis with markets nested in sessions and from the equal-weighted mean with a wild cluster bootstrap by session (Table 2).

Figure 1: Long-run price deviations by market

for 24-06 in session 5, the instance with the highest Sahni-k, in which prices settled 1.1% below the optimum, 62% of the way to the second-best solution. Even there, the confidence interval excludes the second-best solution. It is also one of the markets with the fewest solvers: only eight participants found the optimum (Subsection 4.1).

Table 2 pools the market estimates. Our primary estimate is from a random-effects meta-analysis in which markets are nested in sessions, since markets in the same session share participants. It weights each market by the precision of its estimate while allowing the true long-run deviation to differ across markets and sessions. Pooled across markets, the long-run price lies 0.01% below the optimum, with a 95% confidence interval from −0.03% to 0.01%. On the normalised scale, prices settle 0.3% of the way from the optimum towards the second-best solution, and the hypothesis that they settle on the second-best solution is rejected at any conventional level. The estimated variation between sessions is zero, and the variation between markets is modest, accounting for about a quarter of the variation in the estimates ($I^2 = 23\%$ on the percentage scale and 29% on the normalised scale).

The other two methods in Table 2 relax the reliance on precision weights and on the meta-analytic model. The equal-weighted mean, with inference from a wild cluster bootstrap by session (Cameron et al., 2008; Webb, 2023), gives every market the same weight, and the mean of the five session means uses only the variation between sessions. None of the methods rejects that prices settle at the optimum, and all reject that they settle on the second-best solution. The same holds for a single-level random-effects model, whose estimates are identical to those in Table 2 because the estimated variation between sessions is zero, and for a bootstrap that weights markets by their random-effects weights.

	Percent of optimum		Normalised		
	Estimate	p: optimum	Estimate	p: optimum	p: second-best
Random effects	−0.009 [−0.028, 0.010]	0.362	−0.003 [−0.008, 0.003]	0.348	< 0.001
Equal-weighted	−0.001 [−0.095, 0.182]	0.994	−0.014 [−0.033, 0.026]	0.229	< 0.001
Session means	0.005 [−0.130, 0.139]	0.931	−0.013 [−0.044, 0.017]	0.288	< 0.001

Notes: Pooled long-run price deviations across the 28 markets of the main sample, with 95% confidence intervals in brackets, in percent of the optimum (μ_m) and in units of the distance to the second-best solution (z_m). p-values test whether the pooled deviation equals the optimum (0) and the second-best solution (−1). Random effects: meta-analysis with markets nested in sessions, estimated by restricted maximum likelihood. Equal-weighted: mean across markets, with inference from a wild cluster bootstrap by session (Webb weights). Session means: t test on the five session means.

Table 2: Pooled long-run price deviations

Robustness. The conclusions do not depend on how the long-run deviation is estimated. With heteroskedasticity-robust instead of Newey–West standard errors, the same 24 intervals include the optimum and all 28 exclude the second-best solution. Dropping the first 20% of price events in each market, to remove most of the opening phase, all 28 intervals still exclude the second-best solution, and 20 include the optimum; with fewer observations from the volatile

opening, estimates are more precise and small deviations become detectable. Adding two lags to the selected model, 23 intervals include the optimum and 27 exclude the second-best solution, the exception being the market for 24-06 discussed above, whose interval now just reaches it. Across these alternatives, the long-run deviation changes by a median of 0.14 percentage points. In the two markets for instance 10-01, whose second-best solution lies within 0.17% of the optimum, both intervals include the optimum and neither excludes the second-best solution, as the design anticipated (Appendix C).

4.4 How prices got there

Markets often opened far from the payoff. Across the first five price events of each market, prices ranged from 31.9% below the optimum to 25.7% above it, and the first trade was below the optimum in 19 of the 28 markets. Nearly all trading took place after at least one participant had found the optimum: in 22 markets the first discovery came before the first trade, and only 14 of 3,158 price events preceded it. When trading began, however, the optimum was held by only a few participants, a median of three per market and never more than nine, and in 13 markets by at most two. Most traders therefore began without the optimum, and those who held it could not tell that they did. Prices did not need to anticipate a discovery that had yet to happen, but they had to come to reflect one that a small minority held and no one could recognise as the best possible.

Prices approached the optimum from both sides. After the first discovery, the median market traded above the optimum, and thus above every feasible payoff, in 30% of its price events and below it in 44%, with the remainder at the optimum itself. Neither persistent overpricing nor persistent underpricing emerged: the deviations shrank around the optimum rather than settling on one side of it. Measured by the last time a price lay further from the optimum than the second-best solution, prices settled within that distance after a median of 104 seconds, or the first sixth of price events.

The order book converged along with prices. We measure the bid–ask spread as the difference between the lowest standing ask and the highest standing bid. Over the first two minutes of trading, the median spread was 6.0% of the optimum; over the last two minutes it was 0.07%, a single price tick. The spread narrowed in every market. Appendix D shows the best bid, best ask and transaction prices in all 30 markets, together with the time of first discovery. In most markets, discovery came first and the quotes closed in on the optimum within two or three minutes. In several, prices first gathered around the value of the second-best solution before moving up to the optimum, and in the market for instance 24-06 they remained between the two until the end of trading.

5 Discussion and conclusion

We studied markets in which the traders produce the asset's payoff. In every market, at least one participant found the optimal solution, so the payoff was the highest attainable. Prices converged to that payoff and were statistically distinguishable from the value of the second-best solution in every market. The market thus priced a value that its participants had produced

but could not recognise as the best possible.

Two features of the results help explain why this worked. First, discovery was early and, by the end of each market, redundant: the optimum was found within the first minute in every market and eventually by most participants, so the payoff did not hinge on a single finder and free-riding did not bind. Second, prices aggregated information that no one could verify. Unlike the informed traders in Plott and Sunder (1982), participants who had found the optimum could not know that they were the best informed, and each knew only a lower bound of the payoff. Prices nonetheless reached the value they had found, approaching it from both sides rather than resting at a lower bound.

These results bear most directly on how to reward intellectual discovery when the answer is unknown. Meloso et al. (2009) show that markets for claims on the items in the optimal solution can outperform a prize for the first to find it, and Bossaerts et al. (2024) show that a single security paying the optimal value prices accurately and spreads knowledge about the solution more effectively than a complete set of item-level claims. In both designs, however, claims are settled against the optimal solution, which the organiser must know in advance. As a proposal for organising discovery, this is circular: an organiser who knew the solution would have no need for a market to find it. Meloso et al. (2009, p. 1338) acknowledge the point, describing payment on the very best solution as an idealisation and noting that in practice discoverers would be rewarded for the best solution attained so far. Our design implements exactly that. It keeps the single security of Bossaerts et al. (2024) but pays the value of the best solution found during trading, so the organiser needs only to verify that a candidate is feasible and to measure its value, which for problems such as ours is easy even when finding the optimum is hard. In our markets, this weaker requirement came at no cost: the optimum was found in every market, and prices valued it correctly. Participants can thus be rewarded according to the best outcome collectively produced, while prices communicate its value without anyone disclosing the solution.

The design has a historical precedent. After James Watt's steam engine patent expired in 1800, Cornish mines were financed under the cost-book system, in which ventures were divided into transferable shares and adventurers often held shares in several mines, giving them a stake in improvements across the district. From 1811, *Lean's Engine Reporter* published the design and performance of engines at different mines, and performance rose sharply after the slow progress of the patent period (Nuvolari, 2004). Historical data cannot reveal whether these improvements exhausted what was attainable, because the optimum is unobservable; the laboratory makes that comparison possible. Our setting is also the harder test. A Cornish share paid according to the average performance of engines, which rose as the *Reporter* spread the best designs (Bossaerts et al., 2024), whereas our asset paid the value of the single best solution, and the market revealed no solution or part of one, only the prices at which participants traded.

Because the optimum was always found, and usually by at least one participant before trading began, our data say little about how prices behave ahead of discovery or when the payoff falls short of the optimum. Harder instances, shorter markets, variation in group size, and public disclosure of solutions would test the mechanism where it is most likely to fail.

Asset pricing usually takes payoffs as given and asks whether prices reflect them. When

traders produce the payoff themselves, a market must both secure the best attainable value and price it; in our experiment, it did both.

References

Allen, F., Morris, S., and Shin, H. S. (2006). Beauty contests and iterated expectations in asset markets. *The Review of Financial Studies*, 19(3):719–752.

Bond, P., Edmans, A., and Goldstein, I. (2012). The real effects of financial markets. *Annual Review Financial Economics*, 4(1):339–360.

Bossaerts, P., Bowman, E., Fattinger, F., Huang, H., Lee, M., Murawski, C., Suthakar, A., Tang, S., and Yadav, N. (2024). Resource allocation, computational complexity, and market design. *Journal of Behavioral and Experimental Finance*, 42:100906.

Bossaerts, P., Fattinger, F., Van Den Bogaerde, F., and Yang, W. (2026). Asset pricing in a world of imperfect foresight. *Management Science*.

Bossaerts, P. and Murawski, C. (2017). Computational complexity and human decision-making. *Trends in Cognitive Sciences*, 21(12):917–929.

Boudreau, K. J., Lacetera, N., and Lakhani, K. R. (2011). Incentives and problem uncertainty in innovation contests: An empirical analysis. *Management Science*, 57(5):843–863.

Breusch, T. S. (1978). Testing for autocorrelation in dynamic linear models. *Australian Economic Papers*, 17(31):334.

Camerer, C. and Weigelt, K. (1991). Information mirages in experimental asset markets. *Journal of Business*, 64(4):463–493.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, 90(3):414–427.

Che, Y.-K. and Gale, I. (2003). Optimal design of research contests. *The American Economic Review*, 93(3):646–671.

Corgnet, B., Deck, C., DeSantis, M., and Porter, D. (2018). Information (non) aggregation in markets with costly signal acquisition. *Journal of Economic Behavior & Organization*, 154:286–320.

Diekmann, A. (1985). Volunteer's dilemma. *Journal of Conflict Resolution*, 29(4):605–610.

Forsythe, R. and Lundholm, R. (1990). Information aggregation in an experimental market. *Econometrica: Journal of the Econometric Society*, 58(2):309–347.

Godfrey, L. G. (1978). Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. *Econometrica: Journal of the Econometric Society*, pages 1293–1301.

Goeree, J. K., Holt, C. A., and Smith, A. M. (2017). An experimental examination of the volunteer's dilemma. *Games and Economic Behavior*, 102:303–315.

Grossman, S. (1976). On the efficiency of competitive stock markets where trades have diverse information. *The Journal of Finance*, 31(2):573–585.

Grossman, S. J. and Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American Economic Review*, 70(3):393–408.

Hamilton, J. D. (1994). *Time series analysis*. Princeton, Princeton University Press.

Harrison, G. W. and Hirshleifer, J. (1989). An experimental evaluation of weakest link/best shot models of public goods. *Journal of Political Economy*, 97(1):201–225.

Hirshleifer, J. (1971). The private and social value of information and the reward to inventive activity. *The American Economic Review*, 61(4):561–574.

Hirshleifer, J. (1983). From weakest-link to best-shot: The voluntary provision of public goods. *Public Choice*, 41(3):371–386.

Huber, J., Angerer, M., and Kirchler, M. (2011). Experimental asset markets with endogenous choice of costly asymmetric information. *Experimental Economics*, 14(2):223–240.

Kellerer, H., Pferschy, U., and Pisinger, D. (2004). *Knapsack problems*. Springer.

Keynes, J. M. (1936). *The general theory of employment, interest and money*. Macmillan.

Ljung, G. M. and Box, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303.

Meloso, D., Copic, J., and Bossaerts, P. (2009). Promoting intellectual discovery: Patents versus markets. *Science*, 323(5919):1335–1339.

Murawski, C. and Bossaerts, P. (2016). How humans solve complex problems: The case of the knapsack problem. *Scientific Reports*, 6(1):34851.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Newey, W. K. and West, K. D. (1994). Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies*, 61(4):631–653.

Nuvolari, A. (2004). Collective invention during the British Industrial Revolution: The case of the Cornish pumping engine. *Cambridge Journal of Economics*, 28(3):347–363.

Ottaviani, M. and Sørensen, P. N. (2007). Outcome manipulation in corporate prediction markets. *Journal of the European Economic Association*, 5(2-3):554–563.

Page, L. and Siemroth, C. (2021). How much information is incorporated into financial asset prices? experimental evidence. *The Review of Financial Studies*, 34(9):4412–4449.

Plott, C. R. and Sunder, S. (1982). Efficiency of experimental security markets with insider information: An application of rational-expectations models. *Journal of Political Economy*, 90(4):663–698.

Plott, C. R. and Sunder, S. (1988). Rational expectations and the aggregation of diverse information in laboratory security markets. *Econometrica: Journal of the Econometric Society*, pages 1085–1118.

Radner, R. (1979). Rational expectations equilibrium: Generic existence and the information revealed by prices. *Econometrica: Journal of the Econometric Society*, pages 655–678.

Sahni, S. (1975). Approximate algorithms for the 0/1 knapsack problem. *Journal of the Association for Computing Machinery*, 22(1):115–124.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, pages 461–464.

Smith, V. L., Suchanek, G. L., and Williams, A. W. (1988). Bubbles, crashes, and endogenous expectations in experimental spot asset markets. *Econometrica: Journal of the Econometric Society*, 56(5):1119–1151.

Sunder, S. (1992). Market for information: Experimental evidence. *Econometrica: Journal of the Econometric Society*, 60(3):667–695.

Taylor, C. R. (1995). Digging for golden carrots: An analysis of research tournaments. *The American Economic Review*, pages 872–890.

Terwiesch, C. and Xu, Y. (2008). Innovation contests, open innovation, and multiagent problem solving. *Management Science*, 54(9):1529–1543.

Webb, M. D. (2023). Reworking wild bootstrap-based inference for clustered errors. *Canadian Journal of Economics*, 56(3):839–858.

Weitzman, M. L. (1998). Recombinant growth. *The Quarterly Journal of Economics*, 113(2):331–360.

Wolfers, J. and Zitzewitz, E. (2004). Prediction markets. *Journal of Economic Perspectives*, 18(2):107–126.

Wooldridge, J. M. (1991). On the application of robust, regression-based diagnostics to models of conditional means and conditional variances. *Journal of Econometrics*, 47(1):5–46.

Yadav, N., Murawski, C., Sardina, S., and Bossaerts, P. (2020). Is hardness inherent in computational problems? Performance of human and electronic computers on random instances of the 0-1 knapsack problem. In *ECAI 2020*, pages 498–505. IOS Press.

Youn, H., Strumsky, D., Bettencourt, L. M., and Lobo, J. (2015). Invention as a combinatorial process: Evidence from US patents. *Journal of the Royal Society Interface*, 12(106):20150272.

Appendix A Experiment instructions

Introduction

You have been invited to participate in a research trading experiment. The purpose of this research is to better understand how different structures of financial markets impact productivity and effort in solving the Knapsack Problem.

You will ***earn real cash*** based on your performance today. It is important – both for your benefit, as your reward depends on your performance, and for us, as we aim for robust results – that you ***fully understand*** the experiment. As such, we encourage you to be proactive in asking any questions you may have.

The **schedule** for today is as follows:

1. **Collect forms**: we will collect your Plain Language Statement, Consent, and Demographic information forms.
2. **Instructions**: we will read through these instructions and allow you to ask questions.
3. **Experiment platform**: we will introduce you to the Knapsack Problem Interface and the online trading platform, Flex-E-Markets.
4. **Practice period**: we will conduct a full practice period of the experiment and allow you to ask any final questions.
5. **Experiment**: we will conduct six (6) formal periods of the experiment.
6. **Payment**: we will randomly select two (2) periods for payment out of the six (6) formal periods of the experiment and pay you the sum of the values of your holdings at the end of the two (2) selected periods converted to GBP.

Overview

The experiment will be administered over **six (6)** independent periods. In each period, you will be given a starting endowment of **experimental dollars** and units of an **experimental asset** called '**Stock**'. You will be able to trade experimental dollars and Stocks with other participants through an online marketplace called **Flex-E-Markets**.

Additionally, you will have access to a **Knapsack Problem**. Solving the Knapsack problem will offer information on the potential value of Stocks. This information may assist your decisions to buy or sell units of Stock in the online marketplace.

Your **performance** in each period will be the **sum** of:

1. the **experimental dollar**s you have remaining at the end of the period; and
2. the **payoff** from the **Stocks** you hold at the end of the period (to be explained later in these instructions).

Each period will last for **10 minutes**. The experimenters will notify you at halftime as well as one minute before the period ends.

At the end of the experiment, your **take home earnings** will be determined by your performance in **two (2)** randomly selected periods. This means it is in your best interest to **perform well in all periods**.

What is the Knapsack Problem?

The **Knapsack Problem** involves your filling a large bag, called a **Knapsack**, with a selection of treasures which we will call '**Items**'. Each **Item** will have a **Value** (in Dollars) and a **Weight** (in kilograms). The more Items you add to your Knapsack, the more valuable it will be. However, a Knapsack has a **Capacity Limit** (in kilograms) such that you will be **unable** to include selections of Items whose combined weight **exceeds** the **Capacity Limit**.

The goal of the Knapsack Problem is to **maximise** the **combined value** of Items in the Knapsack, **without** the **combined weight** of the Items going **over the Capacity Limit**.

In each of the six periods there will be a unique Knapsack Problem. However, within a period the **same** Knapsack Problem is given to **all** participants.

You will access these Knapsack Problems through an **online interface**. We will **demonstrate** this interface to you now and give you time to **practice** before progressing. If you need a reminder of anything we demonstrate, a full breakdown of the Knapsack Problem interface can be found in **Appendix A.1**.

What is a Stock?

A **Stock** is an experimental asset that can be bought or sold in an online marketplace using experimental dollars. At the **end of the period**, each **unit of Stock** will yield a **payoff** that is determined as follows: we track the Knapsacks found by each participant during the period, and then take the value of the highest value Knapsack found across **all participants**. The payoff of **each unit of Stock** will be this **highest value**.

For example, imagine that someone discovers a Knapsack worth \$1,000, and this is the highest value discovered across all participants throughout the period. In this case, each unit of Stock will pay \$1,000 at the end of the period. If instead the highest value discovered was \$1,700, each unit of Stock will pay \$1,700.

Each unit of Stock **pays** the **same for everyone**, as it pays the highest-value Knapsack found across **all participants**, not what you found.

So, if the highest-value Knapsack you found was \$1,000, but another participant found a Knapsack fitting \$1,200 and this was the highest value anyone ever discovered during the period, each unit of Stock will pay \$1,200 at the end of the period for all participants. Also, each unit of Stock will pay this highest value regardless of what you paid to purchase it.

This means your earnings will **increase** when you **buy** Stocks at a price **lower** than the payoff at the **end of the period**, and **sell** Stocks at a price **higher** than the payoff at the **end of the period**.

Attempting to find the highest-valued solution to the Knapsack Problem will provide you with additional information on the paying value of Stocks, but it may also change the paying value by increasing the highest-valued Knapsack found. If no one attempts to solve the Knapsack Problem, each unit of Stock will pay \$0.

You should use the Knapsack Interface to get an idea of how high the maximum value of the Knapsack can possibly be. This will allow you to estimate what the final payoff of Stock might be. To be considered for the highest-value Knapsack found, the Knapsack solution **must** be **entered through the Knapsack interface**.

Importantly, you are **not** paid based on the value of the Knapsack you discover. You are paid based on the money you make in the market from buying and selling units of the Stock.

At the end of each period, your holdings will be cleared, and a new period will commence with new holdings. This means that each period is **independent** and has no relation with the periods before or after.

What is the Online Marketplace?

You will have access to an online Marketplace, called Flex-E-Markets. Flex-E-Markets is set up with the same market structure as LSE, NYSE, and other major exchanges. The Marketplace enables you to buy and sell units of Stock from other participants using experimental dollars. The trading price of Stocks will be set by the participants in the experiment.

We will **demonstrate** the online marketplace to you now and give you time to **practice** before progressing. If you need a reminder of anything we demonstrate, a full breakdown of the Flex-E-Markets interface can be found in **Appendix A.2**.

Take Home Payment

At the end of the experiment, **two (2) periods** will be selected as payment periods. One payment period will be randomly selected from periods 1, 2, and 3. The other payment period will be randomly selected from periods 4, 5, and 6.

There is nothing fundamentally different between periods 1, 2, 3, and periods 4, 5, 6, except that you will be starting with different endowments of cash and the Stock. Paying for one period in each set of periods ensures that everyone's pay is fair and does not depend on the initial endowments.

Your payment for this experiment will be determined by the sum of your performance in the two randomly selected periods, converted to British pounds at a ratio of **1,000:1**.

Your performance in each Period will be the **sum** of

1. **Experimental dollars**: the cash remaining at the end of the period; and,
2. **Stock payoffs**: the **payoffs** from **units of Stock** held at the **end of the period**.

For example, if you end up with \$10,000 experimental dollars and ten (10) units of Stock, and the highest-value knapsack found had a value of \$2,000 then your payment for that period - if it was to be drawn as one of the payment periods - will be:

$$\frac{[(10 \text{ Stocks} * \$2,000 \text{knapsack value}) + \$10,000 \text{ Dollars}]}{1,000} = £30.00 \text{ GBP}$$

The **minimum** payment for the complete experiment is **£15**. You will not receive less than £15, even if your calculated performance corresponds to a lower payment. We expect the average payment to be between £30 and £50.

Important

Please do not **resize the windows, refresh the page** or **close any tabs** during the experiment.

You are **not** allowed to access any webpages or programs other than the Knapsack Problem and Flex-E-Markets. Please do **not** take any **photos**, use your **phones**, or **look at the screens** of other participants.

Failure to comply will lead to **exclusion from the experiment** and **forfeiture of payment**.

Thank you!

Thank you for participating in our research experiment. We hope you will enjoy it.

A.1 Instructions appendix: Knapsack Interface

To assist you in solving the Knapsack Problem, you will have access to a drag-and-drop interface which displays:

- all available Items with their corresponding Weight and Value, and if they are currently in your Knapsack or not;
- the Weight of your current Knapsack;
- the Value of your current Knapsack;
- the Capacity Limit of the Knapsack; and,
- the highest-valued Knapsack you have discovered in this period.

An annotated diagram of the Interface is shown below.

The Interface will prevent you from going over the Capacity Limit of the Knapsack, so if you attempt to add an Item that would put you over the Capacity Limit, the Interface will stop that Item from being added. You will have to remove Items (one or more) from the Knapsack to free up Capacity.

You should use the Interface to test different selections of Items inside the Knapsack and identify the highest-value selection.

Accessing the Knapsack Problem

You should use **Mozilla Firefox** (web browser) to access the **Knapsack Problem**. You will have a separate piece of paper which has your username and password to sign in, as well as the URL.

Once signed in, you will see a message instructing you to "Please wait for the next period to commence". When the experimenter starts the period, you will be automatically redirected to the Knapsack Problem.

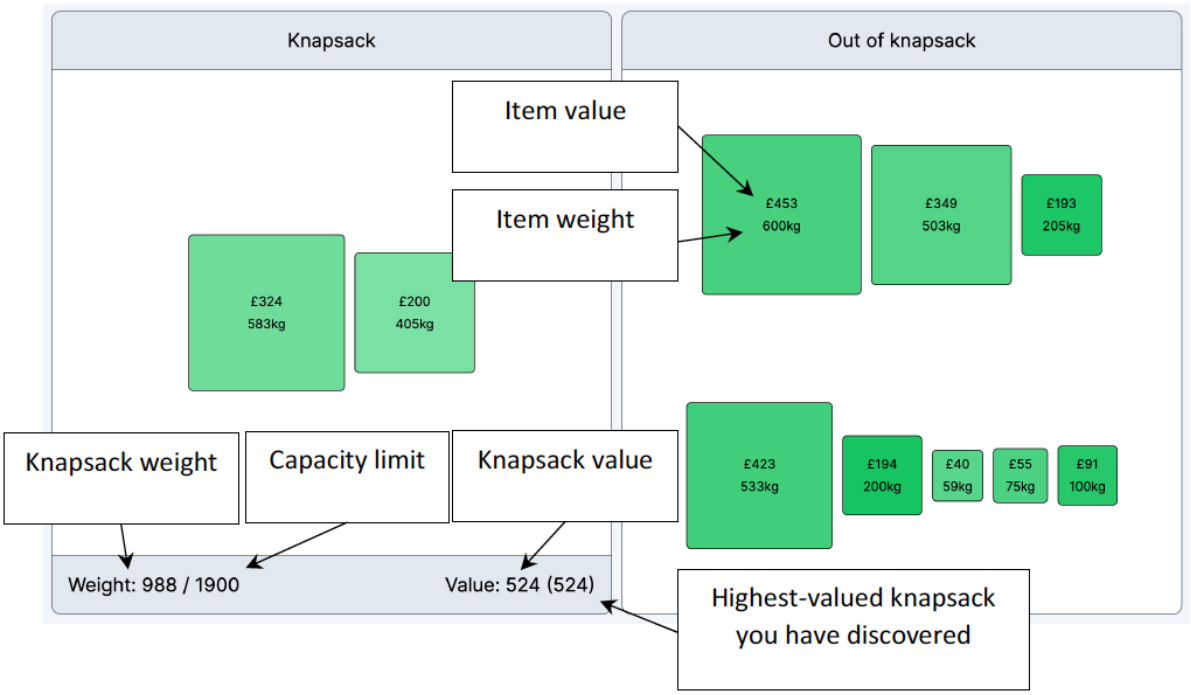


Figure A1: Knapsack Interface

A.2 Instructions appendix: Flex-e-Markets Interface

All trading will take place through an electronic and online marketplace, **Flex-E-Markets**.

Limit orders

In **Flex-E-Markets** you submit **limit orders**, which are orders to buy or sell at a price you determine, or, if possible, at any better price.

Transactions take place from the moment a buy order with a higher price matches or surpasses a sell order with a lower price and are settled instantly. The trade occurs at the price at which the **standing** (earlier) order was placed. Orders remain in the market until (i) they are traded; (ii) you **cancel** the order; or (iii) the period ends.

Interface

The Flex-E-Markets interface features three sections, your **Balance**, the **Order Form**, and the **Order Book**. Each section is depicted in the diagram below. Your Balance is shown at the top left. The Order Form appears in the bottom left. The Order Book and Trade History appear on the right.

Balance

The **Balance** section allows you to see:

- Your current **balance of cash**; and,
- Your current **balance of Stocks**.

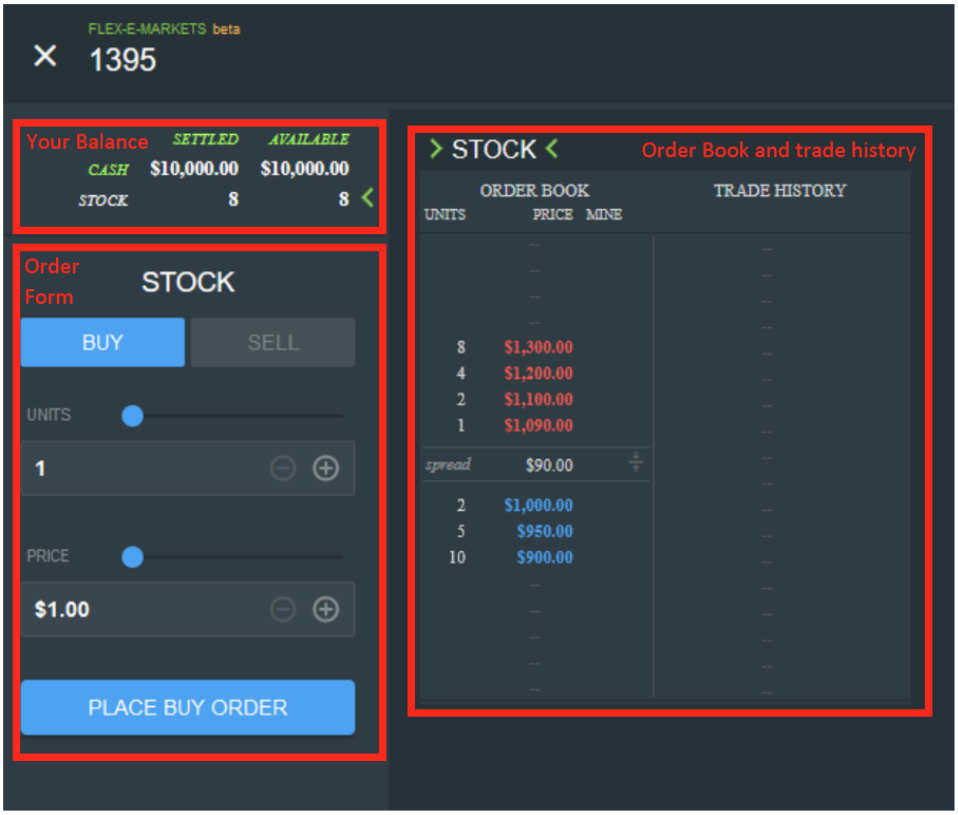


Figure A2: Trading (Flex-e-Markets) Interface

These balances are broken into **Settled** and **Available**.

Settled Balances are what you currently own. This will be your starting endowment, plus any cash you receive for selling Stocks, minus the Stocks you sold, minus any cash you paid to purchase Stocks, plus the Stocks you purchased.

Available Balances are equal to your **Settled Balances** minus anything you currently have submitted to the market through standing orders (orders that have been placed, but not traded or cancelled). For example, if your Settled Balance is $10,000 and you have one standing order to buy 1 unit of Stock for $1,000, your remaining Available Balance will be $9,000. Similar, if your Settled Balance of Stocks is 10, but you have two (2) standing orders to sell 1 unit each, your Available Balance of Stocks will be 8. Standing orders will remain in the market until they are traded (by another order) or cancelled by the participant. Cancelling an order will return the Available quantities.

When a Period ends, you are paid based on your Settled balances, if you have any standing orders they will be cancelled and returned to you.

Order Form

The **Order Form** allows you to **place buy and sell orders**. To do this you must specify:

- **Order Side**: whether you wish to 'BUY' (displayed in blue) or 'SELL' (displayed in red) units;
- **Order Units**: the number of **Units** you wish to buy or sell; and,

- **Limit Price**: the price which you are willing to pay in the case of a buy order, or require to receive in the case of a sell order for a single unit of the experimental asset.

After specifying your order using the Order Form, you should click the 'PLACE ORDER' button to submit the order to the market.

In the example below the participant is looking to BUY 1 Unit of Stock for $1,000.

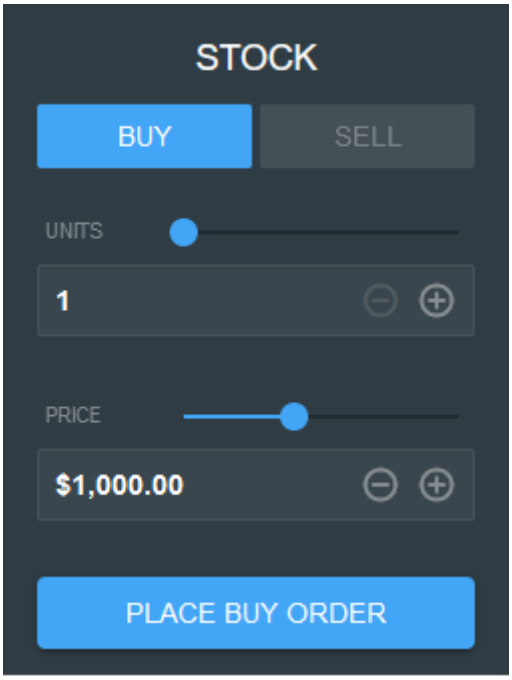


Figure A3: Order form

The Flex-E-Market's interface will prevent you from placing orders that you cannot pay for or do not have units of Stock to supply. It does this using your Available balance. For example if you had two units of Stock Settled, but two orders in the market to Sell one unit each, you would not be able to place any further sell orders.

Order Book

The **Order Book** allows you to see all current orders in the marketplace. **Orders to buy** units are shown in **blue** and **orders to sell** units are shown in **red**. Orders that belong to you will be accompanied by an '**x**' in the '**Mine**' column. The 'x' allows you to easily identify your current standing orders. Clicking the 'x' will **Cancel** the corresponding order.

The Order Book also displays the history of trades in the current period. The history of trades shows prices at which units of Stock have been bought and sold in the past.

In the example below there are three (3) standing orders to sell a unit of Stock for $1,090. One (1) of those units belongs to you, and you can cancel it by clicking the 'x'. There are two (2) standing orders to buy a unit of Stock for $1,000. Neither of these orders were placed by you.

The interface also displays the **spread**, which is the difference between the highest price that a buyer is willing to pay for a unit of Stock, and the lowest price that a seller is willing to accept.

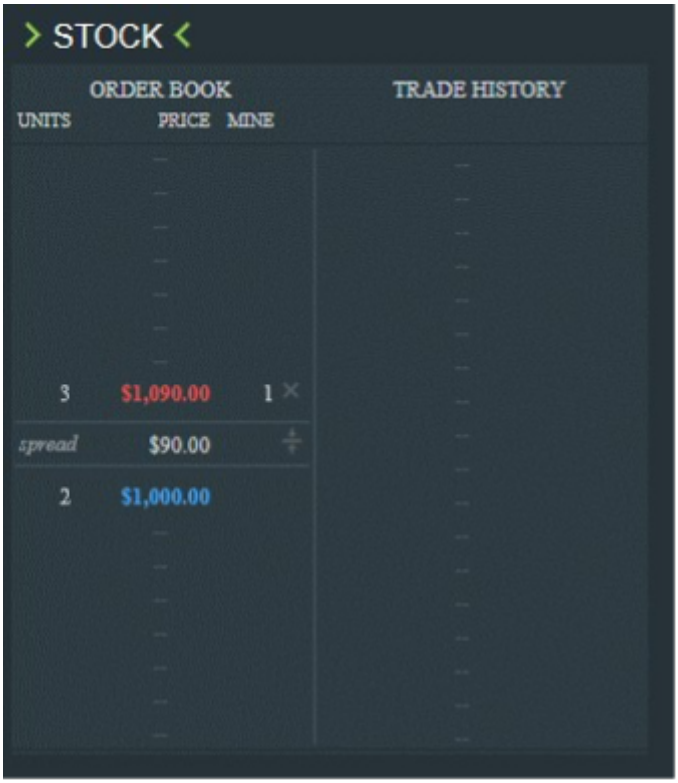


Figure A4: Order book

Accessing the Online Marketplace

You should use **Google Chrome** (web browser) to access the **Flex-E-Markets** online market. You will have a separate piece of paper which has your account, email, and password to sign in, as well as the URL.

To sign in to Flex-E-Markets, go to the URL provided and press 'Sign In' at the top right of the screen. Enter your Account, Email, and Password as provided on your paper and press the Sign in button.

Once signed in, you will see a list of Marketplaces. Please select the Marketplace `KME_2024_1` with `ID 1395`. The Marketplace will be closed. When the experimenter starts the period, the interface will automatically open the Online Marketplace.

Appendix B Construction and configuration of KP instances

B.1 Generation and selection procedure

We generated candidate instances with the aim of obtaining variation in both problem size and Sahni-k. Within each problem size, we sought instances spanning a broad range of Sahni-k values, while requiring a unique optimum and sufficient separation between the optimum and inferior terminal solutions.

For each candidate instance, item weights and value-to-weight ratios were sampled independently from uniform distributions, and item values were constructed from these two quantities. Capacity was initially set to one-half of the sum of all item weights. This normalised capacity lies near the region in which previous evidence finds that random KP instances become especially difficult for both computational solvers and human decision-makers (Yadav et al., 2020). Finally, item values, weights, and capacity were rescaled by a common factor to place the optimal value within a comparable target range across instances.

For each candidate instance, we used a branch-and-bound solver to identify the optimal solution and the highest-valued inferior terminal solution. The solver bounded the value attainable from each partial solution by filling the remaining capacity fractionally with items in decreasing order of their value-to-weight ratio. A feasible solution was classified as terminal if no additional item could be added without exceeding the capacity limit. We retained only instances with a unique optimum. For each such instance, we computed Sahni-k by considering every subset of k items in the optimal solution and greedily completing the remaining capacity, beginning with $k = 0$ and increasing k until the greedy completion reproduced the optimum.

Candidate instances were grouped by their target Sahni-k, and only those with a unique optimum were eligible for selection. We also required a minimum proportional gap between the optimal value and the highest-valued inferior terminal solution.[10] The size of this minimum gap and the extent of the candidate search were adjusted across problem sizes and target Sahni-k values, as suitable instances became increasingly rare in larger problems and at the targeted values of Sahni-k. One instance was then selected from the eligible candidates for each target combination of problem size and Sahni-k.

The selected 12-item instances had Sahni-k values from 0 to 4, the 18-item instances had values of 0, 1, 2, 4, 6, and 8, and the 24-item instances had values of 0, 4, 8, 10, 12, and 14. These values provide broad variation in Sahni-k within and across problem sizes. Sahni-k need not, however, be globally monotonic in difficulty when it approaches the number of available items.[11]

B.2 Configuration of instances

[10] We refer to this solution in the main text as the second-best solution.

[11] For intuition, in a 12-item problem, identifying nine items to include can sometimes be reframed as identifying three items to exclude. Very high values of Sahni-k therefore need not imply proportionally greater search difficulty.

Item	1	2	3	4	5	6	7	8	9	10
Value	300	350	400	450	47	20	8	70	5	5
Weight	205	252	352	447	114	50	28	251	19	20
Optimal	✓		✓	✓					✓	✓

Notes: The instance has 10 items, capacity 1,044, and Sahni-k = 3. The unique optimal solution has value 1,160. The highest-valued inferior terminal solution has value 1,158, corresponding to 99.83% of the optimal value.

Table B1: Instance 10-01

Item	1	2	3	4	5	6	7	8	9	10	11	12
Value	570	413	157	335	315	20	252	58	12	22	59	17
Weight	477	419	193	428	438	42	550	244	66	156	427	151
Optimal	✓	✓		✓	✓							

Notes: The instance has 12 items, capacity 1,798, and Sahni-k = 2. The unique optimal solution has value 1,633. The highest-valued inferior terminal solution has value 1,533, corresponding to 93.88% of the optimal value.

Table B2: Instance 12-01

Item	1	2	3	4	5	6	7	8	9	10	11	12
Value	250	257	573	476	237	33	141	47	37	112	11	28
Weight	212	265	598	500	284	41	237	131	135	430	47	276
Optimal	✓	✓	✓	✓								

Notes: The instance has 12 items, capacity 1,580, and Sahni-k = 0. The unique optimal solution has value 1,556. The highest-valued inferior terminal solution has value 1,440, corresponding to 92.54% of the optimal value.

Table B3: Instance 12-02

Item	1	2	3	4	5	6	7	8	9	10	11	12
Value	489	293	34	305	234	183	20	239	50	169	45	50
Weight	410	312	40	449	375	321	45	543	129	438	230	505
Optimal	✓	✓		✓	✓	✓						

Notes: The instance has 12 items, capacity 1,901, and Sahni-k = 3. The unique optimal solution has value 1,504. The highest-valued inferior terminal solution has value 1,425, corresponding to 94.75% of the optimal value.

Table B4: Instance 12-03

Item	1	2	3	4	5	6	7	8	9	10	11	12
Value	221	169	424	391	29	178	183	315	39	89	10	12
Weight	203	178	462	473	39	241	280	545	136	413	65	120
Optimal	✓	✓	✓	✓		✓						

Notes: The instance has 12 items, capacity 1,580, and Sahni-k = 1. The unique optimal solution has value 1,383. The highest-valued inferior terminal solution has value 1,283, corresponding to 92.77% of the optimal value.

Table B5: Instance 12-04

Item	1	2	3	4	5	6	7	8	9	10	11	12
Value	88	211	368	37	400	561	296	35	59	51	38	100
Weight	83	203	364	40	433	644	450	101	174	151	150	514
Optimal		✓	✓		✓	✓						

Notes: The instance has 12 items, capacity 1,657, and Sahni-k = 4. The unique optimal solution has value 1,540. The highest-valued inferior terminal solution has value 1,454, corresponding to 94.42% of the optimal value.

Table B6: Instance 12-05

Item	1	2	3	4	5	6	7	8	9	10	11	12
Value	304	204	250	74	242	40	104	63	180	201	120	114
Weight	161	121	150	48	160	28	77	48	163	212	157	160
Optimal	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		

Item	13	14	15	16	17	18
Value	184	50	48	45	21	22
Weight	259	120	132	138	75	129
Optimal						

Notes: The instance has 18 items, capacity 1,172, and Sahni-k = 0. The unique optimal solution has value 1,662. The highest-valued inferior terminal solution has value 1,602, corresponding to 96.39% of the optimal value.

Table B7: Instance 18-01

Item	1	2	3	4	5	6	7	8	9	10	11	12
Value	169	273	312	241	172	49	164	20	199	19	36	3
Weight	89	144	175	136	100	33	117	16	163	21	67	6
Optimal	✓	✓	✓	✓	✓		✓		✓			✓

Item	13	14	15	16	17	18
Value	76	73	48	43	16	5
Weight	168	177	129	117	128	75
Optimal						

Notes: The instance has 18 items, capacity 934, and Sahni-k = 2. The unique optimal solution has value 1,533. The highest-valued inferior terminal solution has value 1,474, corresponding to 96.15% of the optimal value.

Table B8: Instance 18-02

Item	1	2	3	4	5	6	7	8	9	10	11	12
Value	151	178	255	121	242	243	228	5	187	187	153	59
Weight	78	103	156	79	160	165	158	4	150	170	154	65
Optimal	✓	✓	✓	✓	✓	✓	✓		✓			

Item	13	14	15	16	17	18
Value	65	83	35	54	60	14
Weight	88	132	77	121	149	89
Optimal						

Notes: The instance has 18 items, capacity 1,052, and Sahni-k = 1. The unique optimal solution has value 1,605. The highest-valued inferior terminal solution has value 1,548, corresponding to 96.45% of the optimal value.

Table B9: Instance 18-03

Item	1	2	3	4	5	6	7	8	9	10	11	12
Value	151	145	59	148	246	124	310	244	191	33	39	53
Weight	79	84	37	94	157	80	223	196	194	69	86	138
Optimal	✓	✓		✓	✓	✓	✓	✓	✓			

Item	13	14	15	16	17	18
Value	43	4	51	43	29	14
Weight	128	12	204	202	166	82
Optimal		✓				

Notes: The instance has 18 items, capacity 1,120, and Sahni-k = 6. The unique optimal solution has value 1,563. The highest-valued inferior terminal solution has value 1,503, corresponding to 96.16% of the optimal value.

Table B10: Instance 18-04

Item	1	2	3	4	5	6	7	8	9	10	11	12
Value	220	78	167	158	7	233	206	216	113	215	114	163
Weight	110	40	95	100	5	184	179	202	113	222	138	260
Optimal	✓	✓	✓	✓		✓	✓	✓		✓		

Item	13	14	15	16	17	18
Value	87	80	57	20	8	6
Weight	145	166	154	82	33	35
Optimal						

Notes: The instance has 18 items, capacity 1,135, and Sahni-k = 4. The unique optimal solution has value 1,493. The highest-valued inferior terminal solution has value 1,434, corresponding to 96.05% of the optimal value.

Table B11: Instance 18-05

Item	1	2	3	4	5	6	7	8	9	10	11	12
Value	185	36	102	283	10	92	105	84	111	200	178	81
Weight	97	19	57	167	6	69	80	68	92	181	180	129
Optimal	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	

Item	13	14	15	16	17	18
Value	68	92	75	40	30	23
Weight	111	208	199	118	122	96
Optimal						

Notes: The instance has 18 items, capacity 1,003, and Sahni-k = 8. The unique optimal solution has value 1,350. The highest-valued inferior terminal solution has value 1,302, corresponding to 96.44% of the optimal value.

Table B12: Instance 18-06

Item	1	2	3	4	5	6	7	8	9	10	11	12
Value	184	116	36	91	230	170	216	141	118	110	146	88
Weight	94	60	19	50	134	100	129	85	73	81	114	108
Optimal	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	

Item	13	14	15	16	17	18	19	20	21	22	23	24
Value	39	16	27	37	56	19	37	3	13	1	13	2
Weight	51	26	46	71	141	72	144	13	64	5	124	40
Optimal										✓		

Notes: The instance has 24 items, capacity 927, and Sahni-k = 8. The unique optimal solution has value 1,523. The highest-valued inferior terminal solution has value 1,491, corresponding to 97.90% of the optimal value.

Table B13: Instance 24-01

Item	1	2	3	4	5	6	7	8	9	10	11	12
Value	184	130	160	161	155	147	92	16	89	69	123	53
Weight	94	67	92	105	102	102	66	12	77	62	111	56
Optimal	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓

Item	13	14	15	16	17	18	19	20	21	22	23	24
Value	5	60	45	51	37	28	19	27	20	18	7	4
Weight	6	80	66	92	98	86	62	90	85	114	60	87
Optimal	✓											

Notes: The instance has 24 items, capacity 941, and Sahni-k = 4. The unique optimal solution has value 1,368. The highest-valued inferior terminal solution has value 1,340, corresponding to 97.95% of the optimal value.

Table B14: Instance 24-02

Item	1	2	3	4	5	6	7	8	9	10	11	12
Value	173	145	155	108	93	168	23	35	82	85	96	47
Weight	98	86	97	68	59	109	15	23	56	63	73	36
Optimal	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Item	13	14	15	16	17	18	19	20	21	22	23	24
Value	103	62	72	42	24	29	16	14	14	6	17	6
Weight	81	74	115	74	55	80	64	68	85	39	120	98
Optimal	✓											

Notes: The instance has 24 items, capacity 874, and Sahni-k = 0. The unique optimal solution has value 1,313. The highest-valued inferior terminal solution has value 1,279, corresponding to 97.41% of the optimal value.

Table B15: Instance 24-03

Item	1	2	3	4	5	6	7	8	9	10	11	12
Value	20	22	180	114	177	175	154	182	154	108	123	125
Weight	9	10	93	59	94	102	92	112	95	73	96	98
Optimal			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Item	13	14	15	16	17	18	19	20	21	22	23	24
Value	123	84	5	83	73	7	25	35	38	27	11	6
Weight	112	98	6	101	90	9	35	49	82	105	85	116
Optimal												

Notes: The instance has 24 items, capacity 916, and Sahni-k = 10. The unique optimal solution has value 1,492. The highest-valued inferior terminal solution has value 1,466, corresponding to 98.26% of the optimal value.

Table B16: Instance 24-04

Item	1	2	3	4	5	6	7	8	9	10	11	12
Value	16	73	52	211	162	234	58	129	271	62	191	187
Weight	7	34	26	111	86	125	33	75	160	40	152	170
Optimal		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Item	13	14	15	16	17	18	19	20	21	22	23	24
Value	82	74	134	80	111	80	61	25	13	13	1	4
Weight	89	83	154	93	168	169	166	85	59	59	6	39
Optimal		✓										

Notes: The instance has 24 items, capacity 1,100, and Sahni-k = 12. The unique optimal solution has value 1,704. The highest-valued inferior terminal solution has value 1,677, corresponding to 98.42% of the optimal value.

Table B17: Instance 24-05

Item	1	2	3	4	5	6	7	8	9	10	11	12
Value	9	60	72	134	67	89	59	216	91	49	189	110
Weight	4	28	34	66	36	49	35	144	63	35	135	80
Optimal		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Item	13	14	15	16	17	18	19	20	21	22	23	24
Value	154	58	141	156	52	99	59	64	13	55	32	28
Weight	114	43	106	161	54	108	66	138	29	126	116	157
Optimal	✓	✓	✓									

Notes: The instance has 24 items, capacity 969, and Sahni-k = 14. The unique optimal solution has value 1,489. The highest-valued inferior terminal solution has value 1,462, corresponding to 98.19% of the optimal value.

Table B18: Instance 24-06

B.3 Allocation of instances across sessions

Session	Market	Instance	Items	Capacity	Sahni-k
1	1	12-01	12	1,798	2
1	2	12-02	12	1,580	0
1	3	12-03	12	1,901	3
1	4	12-04	12	1,580	1
1	5	12-05	12	1,657	4
1	6	10-01	10	1,044	3
2	1	12-03	12	1,901	3
2	2	18-01	18	1,172	0
2	3	12-05	12	1,657	4
2	4	18-02	18	934	2
2	5	18-03	18	1,052	1
2	6	10-01	10	1,044	3
3	1	18-03	18	1,052	1
3	2	18-04	18	1,120	6
3	3	18-02	18	934	2
3	4	18-05	18	1,135	4
3	5	18-01	18	1,172	0
3	6	18-06	18	1,003	8
4	1	24-01	24	927	8
4	2	24-02	24	941	4
4	3	18-03	18	1,052	1
4	4	18-04	18	1,120	6
4	5	18-02	18	934	2
4	6	24-03	24	874	0
5	1	24-04	24	916	10
5	2	24-05	24	1,100	12
5	3	24-03	24	874	0
5	4	24-06	24	969	14
5	5	24-01	24	927	8
5	6	24-02	24	941	4

Notes: Each row identifies the KP instance assigned to one experimental market. Instance identifiers combine the number of items with a unique instance number. The same identifier is used whenever an instance appears in more than one session. Markets are numbered in the order in which they occurred within each session.

Table B19: Assignment of KP instances across sessions and markets

Appendix C Market-level estimates

This appendix reports the market-level results behind Section 4. Table C1 gives, for each market, the discovery measures of Subsection 4.1 and the long-run price deviation estimated as in Subsection 4.2, together with the selected autoregressive order and the residual diagnostic used to select it. Panel A covers the 28 markets of the main sample, in the order of Figure 1. Panel B reports the two markets that used instance 10-01, which is excluded from the main analysis because its second-best solution lies within 0.17% of the optimum. In both markets the optimum was found and the confidence interval includes it, but, as the design anticipated, neither interval excludes the second-best solution.

Table C2 reports the long-run price deviation of each main-sample market under the three alternative specifications summarised in Subsection 4.3: heteroskedasticity-robust instead of Newey–West standard errors, re-estimation without the first 20% of price events, and two additional lags. Across all four specifications, the confidence interval excludes the second-best solution in every market except one: the market for instance 24-06 in session 5, whose interval reaches it by a small margin when two lags are added.

Instance	Session	Market	N_m	N_m^*	τ_m (s)	Events	p	BG p-value	μ_m [95% CI]	z_m
Panel A: Main sample										
12-02	1	2	20	15	3.9	246	5	0.303	0.941 [−0.241, 2.123]	0.126
12-04	1	4	20	13	8.2	368	2	0.442	−0.376 [−0.707, −0.045]	−0.052
12-01	1	1	20	16	3.8	303	4	0.753	−0.101 [−0.222, 0.019]	−0.017
12-03	1	3	20	15	4.3	368	1	0.132	0.025 [−0.011, 0.061]	0.005
12-03	2	1	19	13	24.3	59	2	0.134	0.128 [−2.315, 2.571]	0.024
12-05	1	5	20	15	5.7	249	4	0.195	0.472 [0.090, 0.854]	0.085
12-05	2	3	19	15	4.6	19	2	0.755	0.569 [−0.863, 2.002]	0.102
18-01	2	2	19	14	41.0	39	5	0.000	−0.046 [−0.085, −0.007]	−0.013
18-01	3	5	20	17	10.3	41	2	0.268	0.010 [−0.021, 0.042]	0.003
18-03	2	5	19	13	12.6	77	4	0.475	−0.874 [−1.908, 0.161]	−0.246
18-03	3	1	20	12	22.1	57	2	0.681	−0.065 [−0.689, 0.559]	−0.018
18-03	4	3	20	15	10.2	164	4	0.679	0.088 [−0.333, 0.508]	0.025
18-02	2	4	19	16	8.0	27	1	0.371	0.008 [−0.037, 0.054]	0.002
18-02	3	3	20	18	14.2	31	2	0.435	−0.043 [−0.087, 0.001]	−0.011
18-02	4	5	20	17	12.3	65	2	0.096	0.157 [−0.092, 0.405]	0.041
18-05	3	4	20	16	8.3	38	5	0.146	−0.005 [−0.023, 0.013]	−0.001
18-04	3	2	20	13	8.1	53	5	0.000	−0.551 [−2.343, 1.241]	−0.144
18-04	4	4	20	15	18.8	125	4	0.697	0.225 [−0.133, 0.583]	0.059
18-06	3	6	20	15	18.8	42	1	0.447	0.130 [−0.892, 1.151]	0.036
24-03	4	6	20	18	21.2	88	4	0.214	0.011 [−0.135, 0.157]	0.004
24-03	5	3	20	16	20.7	33	5	0.012	−0.023 [−0.129, 0.083]	−0.009
24-02	4	2	20	13	25.8	158	4	0.000	−0.517 [−1.285, 0.251]	−0.253
24-02	5	6	20	16	14.3	42	1	0.413	0.763 [−0.292, 1.818]	0.373
24-01	4	1	20	14	56.0	107	4	0.001	−0.126 [−0.629, 0.378]	−0.060
24-01	5	5	20	13	22.1	37	5	0.114	0.222 [−0.078, 0.521]	0.105
24-04	5	1	20	9	13.4	63	5	0.000	0.105 [−0.052, 0.261]	0.060
24-05	5	2	20	6	31.5	49	4	0.120	−0.028 [−0.079, 0.024]	−0.017
24-06	5	4	20	8	46.7	70	1	0.121	−1.113 [−1.598, −0.628]	−0.614
Panel B: Instance 10-01 (excluded from the main analysis)										
10-01	1	6	20	13	17.8	340	2	0.521	1.601 [−1.426, 4.627]	9.283
10-01	2	6	19	10	10.6	85	4	0.594	−0.724 [−6.766, 5.318]	−4.197

Notes: N_m: participants in the session. N_m^*: participants who found the optimal solution by the end of the market. τ_m: seconds from the opening of the market (the first order-book event or knapsack move) to the first discovery of the optimum. Events: price events used in estimation. p: selected autoregressive order. BG p-value: heteroskedasticity-robust Breusch–Godfrey test of the selected model's residuals. μ_m: long-run price deviation, in percent of the optimum, with 95% confidence interval from Newey–West standard errors. z_m: long-run deviation in units of the distance between the optimum and the second-best solution (0 = optimum, −1 = second-best). Markets are ordered as in Figure 1.

Table C1: Discovery and long-run prices by market

Instance	Session	Market	Main μ_m [95% CI]	Robust SE [95% CI]	First 20% dropped μ_m [95% CI]	AR($p+2$) μ_m [95% CI]
12-02	1	2	0.941 [−0.241, 2.123]	[−0.109, 1.991]	0.007 [−0.050, 0.063]	0.717 [−0.095, 1.528]
12-04	1	4	−0.376 [−0.707, −0.045][a]	[−0.702, −0.049][a]	−0.018 [−0.032, −0.004][a]	−0.346 [−0.709, 0.017]
12-01	1	1	−0.101 [−0.222, 0.019]	[−0.221, 0.018]	−0.120 [−0.273, 0.032]	−0.101 [−0.209, 0.006]
12-03	1	3	0.025 [−0.011, 0.061]	[−0.010, 0.060]	0.000 [−0.020, 0.021]	0.025 [−0.011, 0.061]
12-03	2	1	0.128 [−2.315, 2.571]	[−2.434, 2.690]	−0.007 [−0.702, 0.687]	0.093 [−1.981, 2.166]
12-05	1	5	0.472 [0.090, 0.854][a]	[0.092, 0.852][a]	0.264 [−0.070, 0.598]	0.459 [0.124, 0.793][a]
12-05	2	3	0.569 [−0.863, 2.002]	[−0.864, 2.003]	−0.001 [−0.143, 0.141]	0.648 [−1.166, 2.462]
18-01	2	2	−0.046 [−0.085, −0.007][a]	[−0.091, −0.001][a]	−0.041 [−0.076, −0.006][a]	−0.041 [−0.080, −0.003][a]
18-01	3	5	0.010 [−0.021, 0.042]	[−0.019, 0.039]	−0.002 [−0.053, 0.049]	0.011 [−0.027, 0.049]
18-03	2	5	−0.874 [−1.908, 0.161]	[−1.977, 0.230]	−0.460 [−1.277, 0.358]	−0.088 [−1.934, 1.759]
18-03	3	1	−0.065 [−0.689, 0.559]	[−0.929, 0.799]	−0.077 [−0.219, 0.065]	−0.001 [−0.555, 0.553]
18-03	4	3	0.088 [−0.333, 0.508]	[−0.313, 0.489]	0.011 [0.003, 0.019][a]	0.233 [−0.438, 0.904]
18-02	2	4	0.008 [−0.037, 0.054]	[−0.040, 0.056]	0.021 [−0.002, 0.044]	0.009 [−0.043, 0.061]
18-02	3	3	−0.043 [−0.087, 0.001]	[−0.130, 0.044]	−0.044 [−0.064, −0.024][a]	−0.043 [−0.079, −0.006][a]
18-02	4	5	0.157 [−0.092, 0.405]	[−0.146, 0.460]	0.017 [−0.242, 0.276]	0.105 [−0.136, 0.346]
18-05	3	4	−0.005 [−0.023, 0.013]	[−0.021, 0.010]	−0.009 [−0.025, 0.007]	−0.008 [−0.026, 0.009]
18-04	3	2	−0.551 [−2.343, 1.241]	[−2.480, 1.378]	−0.389 [−0.913, 0.134]	−0.606 [−1.252, 0.041]
18-04	4	4	0.225 [−0.133, 0.583]	[−0.064, 0.514]	0.084 [−0.079, 0.248]	0.327 [−0.221, 0.875]
18-06	3	6	0.130 [−0.892, 1.151]	[−1.665, 1.925]	−0.112 [−0.349, 0.124]	0.052 [−0.970, 1.074]
24-03	4	6	0.011 [−0.135, 0.157]	[−0.182, 0.204]	0.042 [−0.034, 0.119]	−0.017 [−0.158, 0.123]
24-03	5	3	−0.023 [−0.129, 0.083]	[−0.126, 0.079]	0.038 [0.022, 0.054][a]	0.035 [0.018, 0.052][a]
24-02	4	2	−0.517 [−1.285, 0.251]	[−1.206, 0.172]	−0.160 [−0.483, 0.163]	−0.604 [−1.024, −0.183][a]
24-02	5	6	0.763 [−0.292, 1.818]	[−0.426, 1.952]	0.035 [0.014, 0.056][a]	0.477 [−0.399, 1.354]
24-01	4	1	−0.126 [−0.629, 0.378]	[−0.524, 0.272]	0.026 [−0.190, 0.243]	0.011 [−0.421, 0.442]
24-01	5	5	0.222 [−0.078, 0.521]	[−0.064, 0.508]	−0.018 [−0.045, 0.009]	0.048 [−0.048, 0.144]
24-04	5	1	0.105 [−0.052, 0.261]	[−0.034, 0.243]	0.107 [0.049, 0.164][a]	0.141 [−0.041, 0.323]
24-05	5	2	−0.028 [−0.079, 0.024]	[−0.076, 0.021]	−0.038 [−0.087, 0.011]	−0.040 [−0.097, 0.018]
24-06	5	4	−1.113 [−1.598, −0.628][a]	[−1.649, −0.577][a]	−0.952 [−1.312, −0.593][a]	−0.874 [−1.822, 0.074][b]

Notes: Long-run price deviation μ_m, in percent of the optimum, with 95% confidence intervals, for the 28 markets of the main sample. Main: selected AR(p) model, Newey–West standard errors. Robust SE: same estimate, heteroskedasticity-robust standard errors. First 20% dropped: model reselected and estimated without the first 20% of price events. AR(p+2): two lags added to the selected model, Newey–West standard errors. [a] Interval excludes the optimum. [b] Interval does not exclude the value of the second-best solution.

Table C2: Long-run price deviations under alternative specifications

Appendix D Market-level bid and ask paths

Figures D1–D30 show the best bid, the best ask and transaction prices in each market, by session and market. Prices are measured in units of the distance between the optimum and the second-best solution, so that 0 is the optimum (dashed horizontal line) and −1 the second-best solution (dotted line) in every market. Bids, asks and transaction prices more than two such distances from the optimum are drawn at the edge of the plot. Time is measured from the opening of the market on a logarithmic scale, and the dashed vertical line marks the first time a participant found the optimum. Gaps in the bid or ask line indicate periods in which that side of the order book was empty. For the two markets that used instance 10-01 (session 1, market 6, and session 2, market 6), the second-best solution lies only two experimental dollars below the optimum, so a deviation of one price tick corresponds to half the distance between the two. Sharp moves in the final seconds reflect orders cancelled or executed as the market closed.

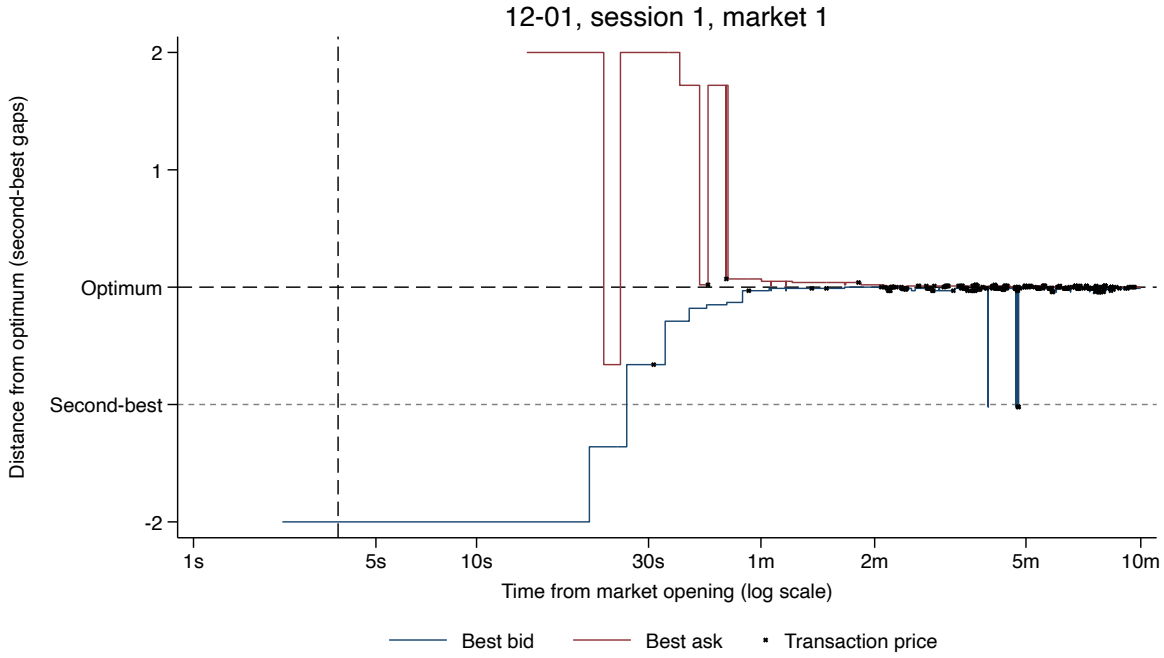


Figure D1: Time series plot for Session 1, market 1

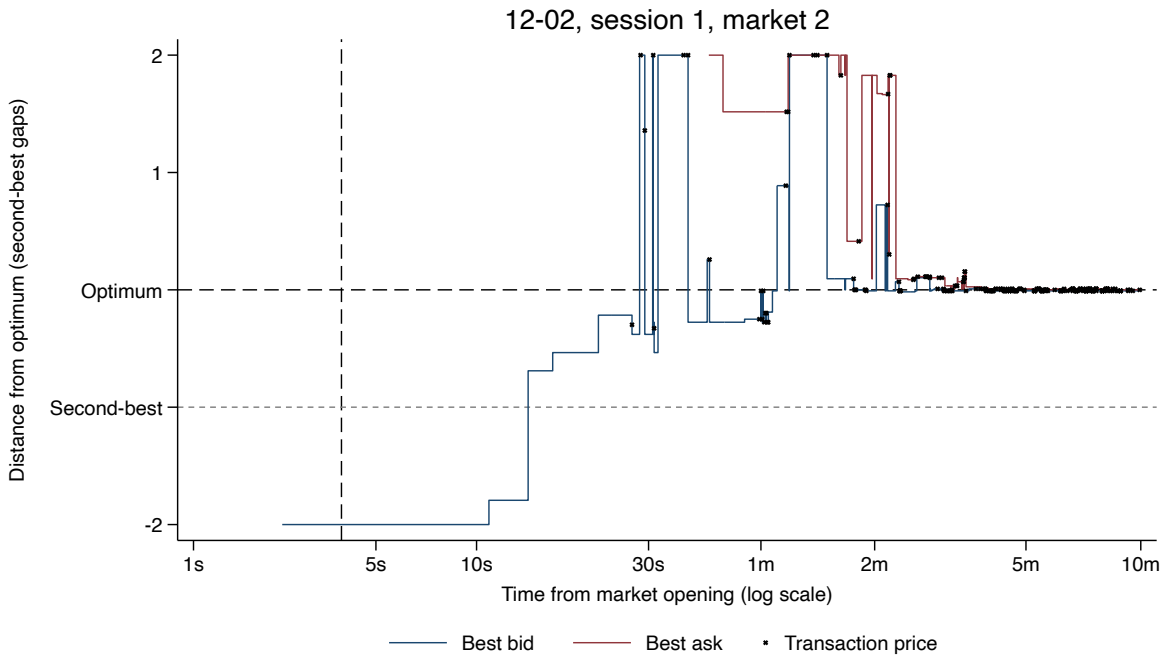


Figure D2: Time series plot for Session 1, market 2

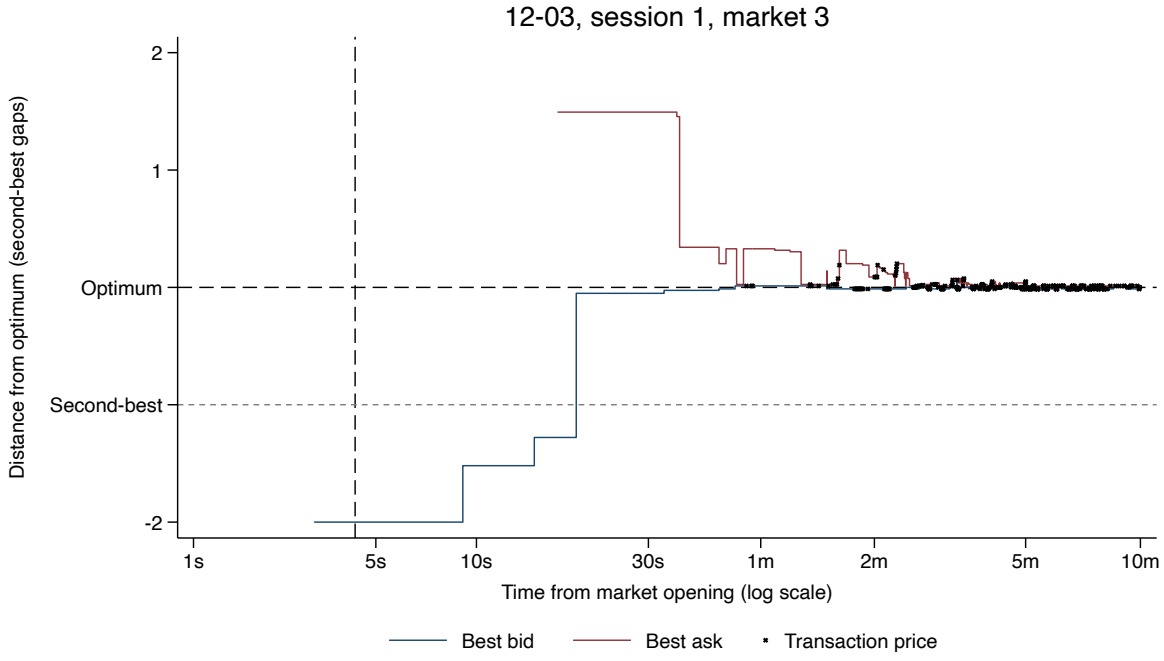


Figure D3: Time series plot for Session 1, market 3

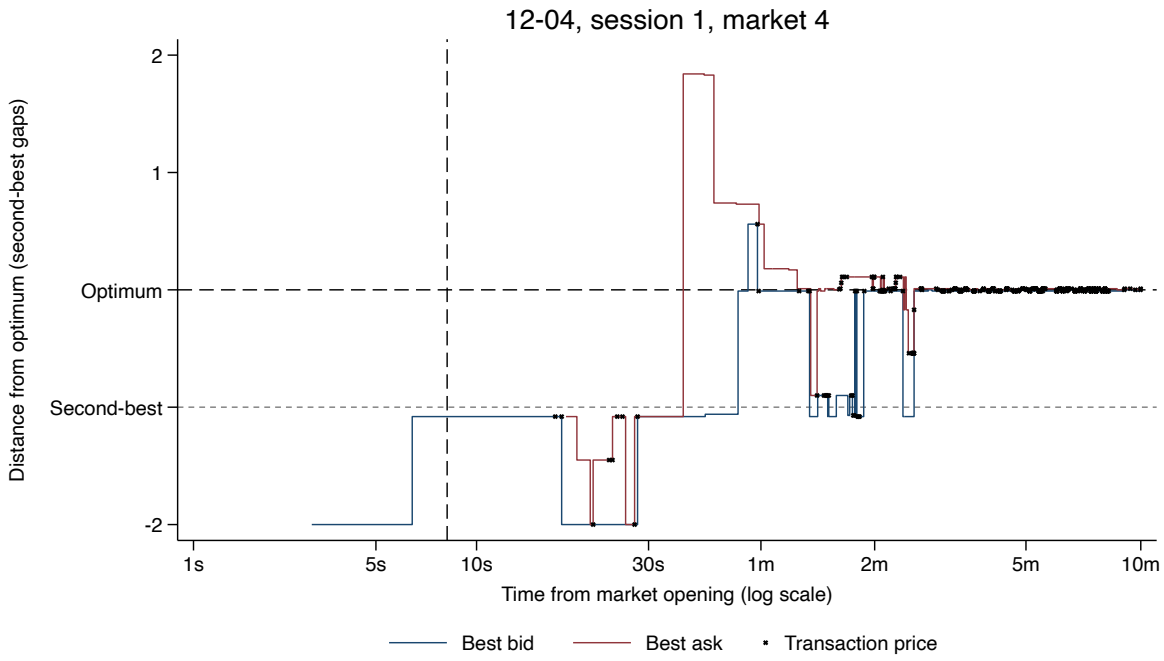


Figure D4: Time series plot for Session 1, market 4

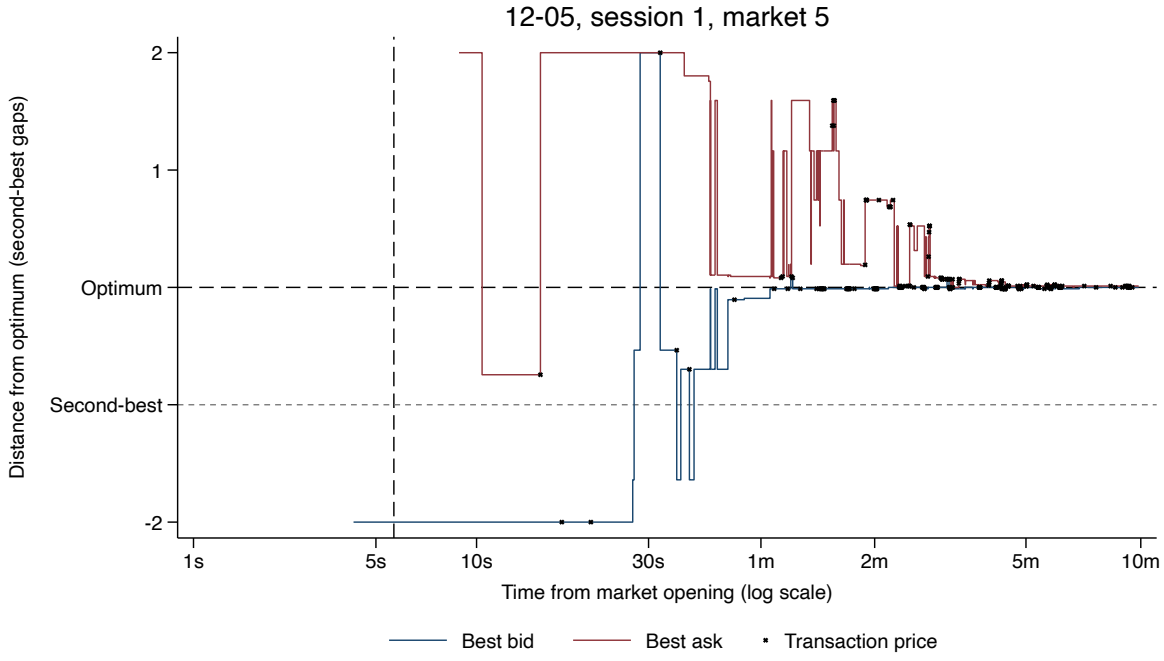


Figure D5: Time series plot for Session 1, market 5

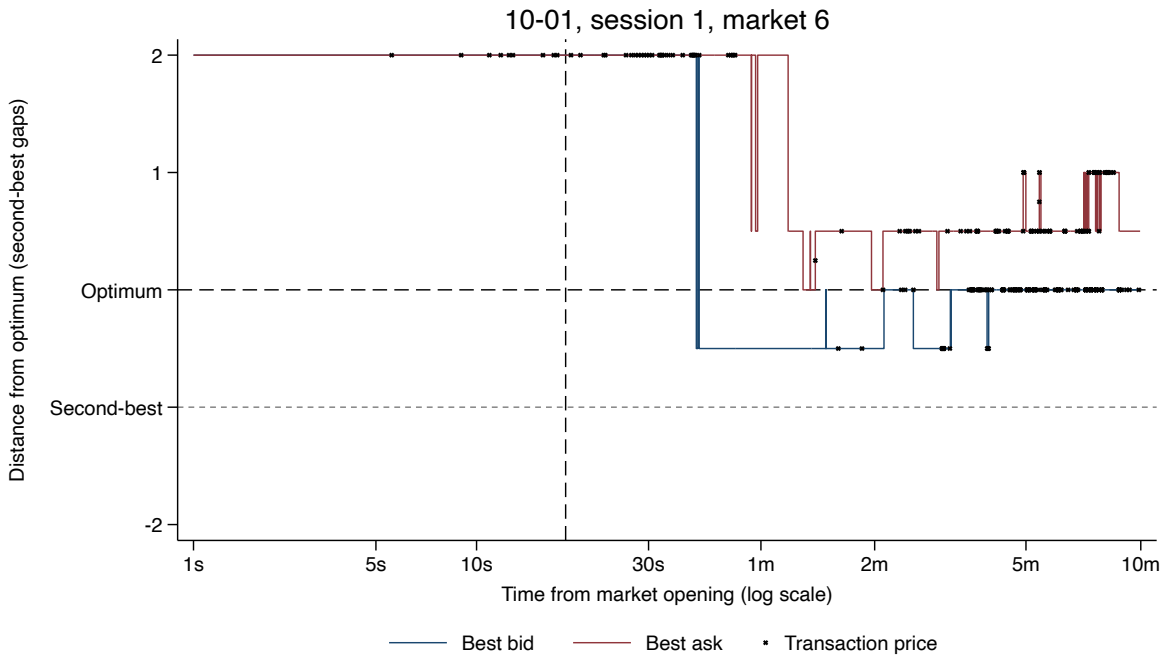


Figure D6: Time series plot for Session 1, market 6

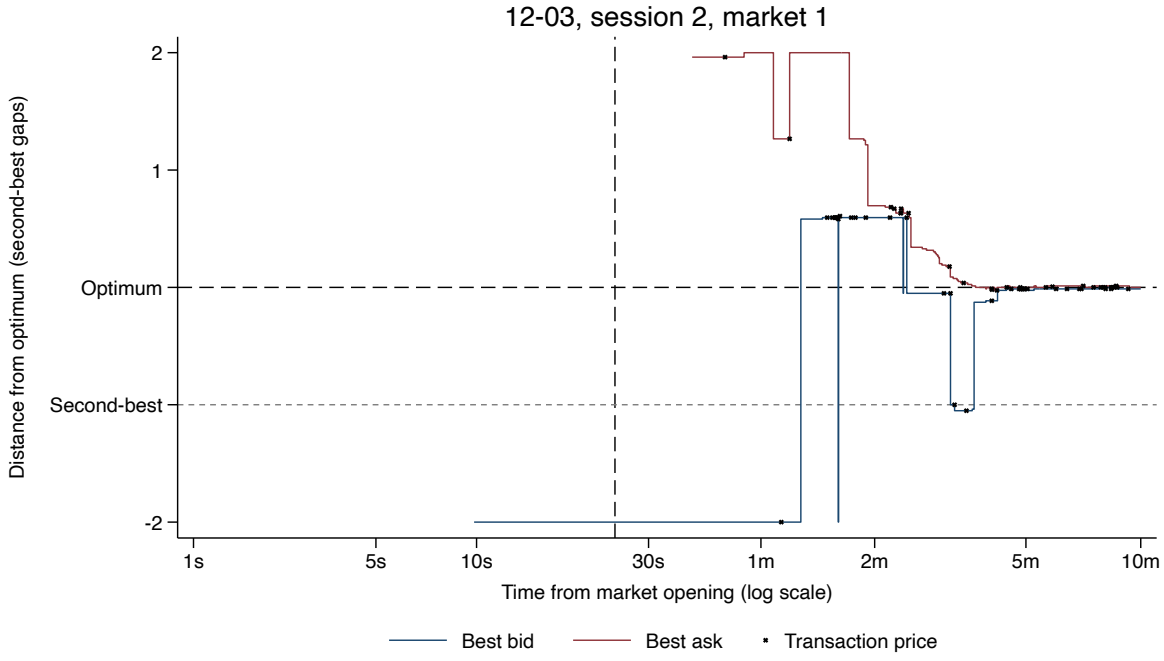


Figure D7: Time series plot for Session 2, market 1

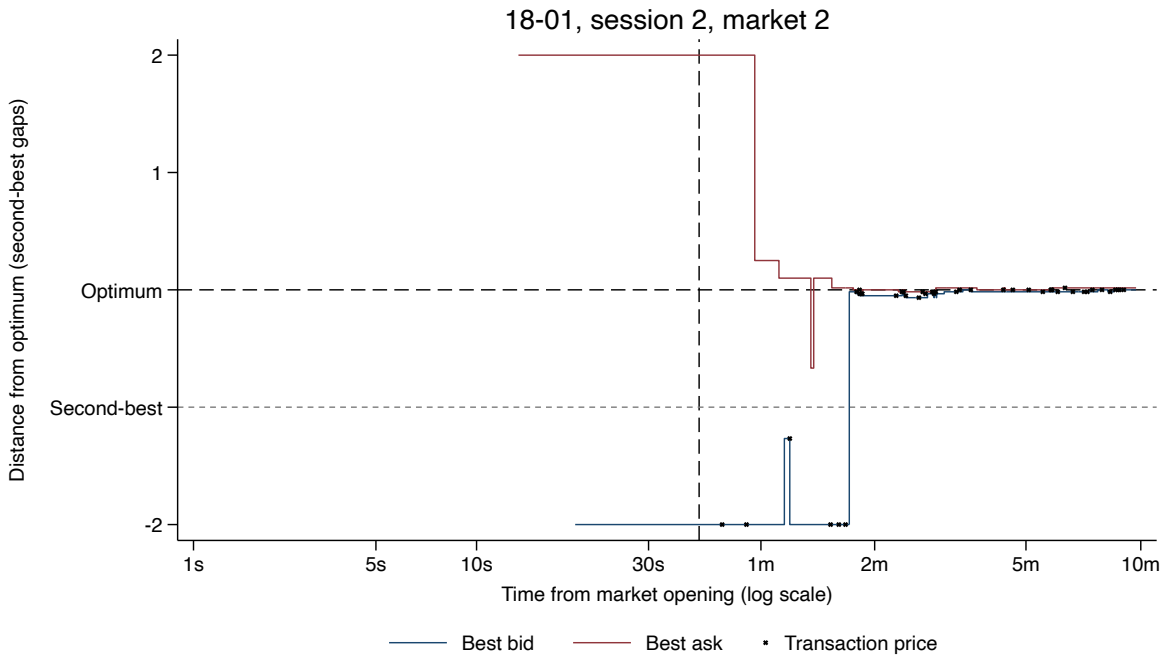


Figure D8: Time series plot for Session 2, market 2

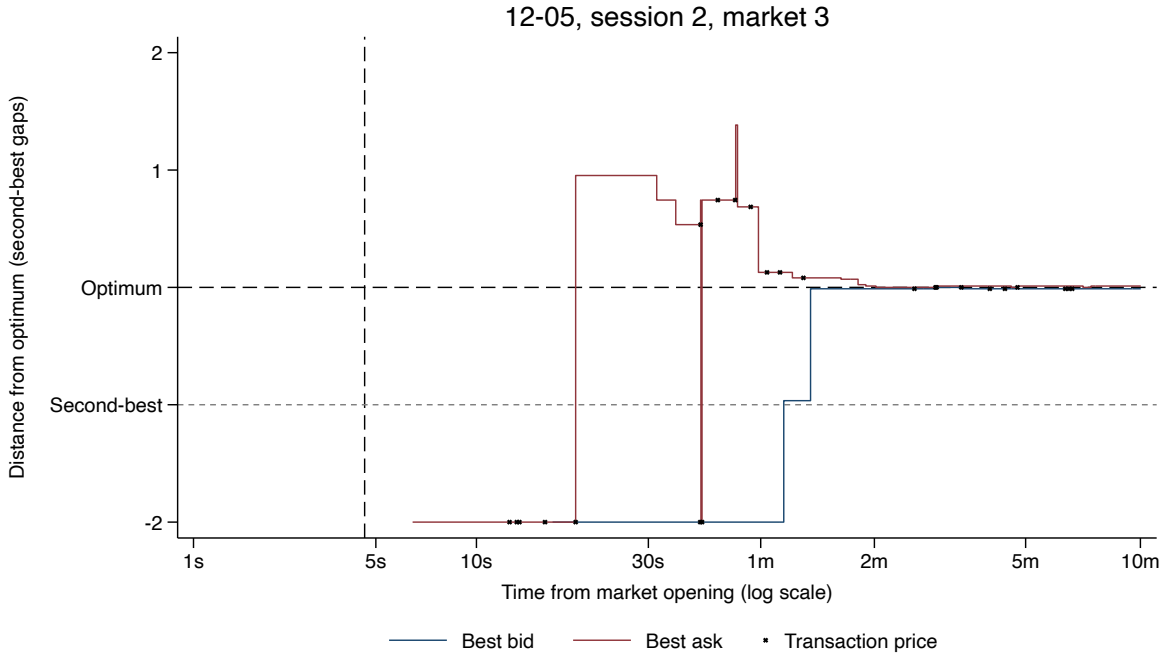


Figure D9: Time series plot for Session 2, market 3

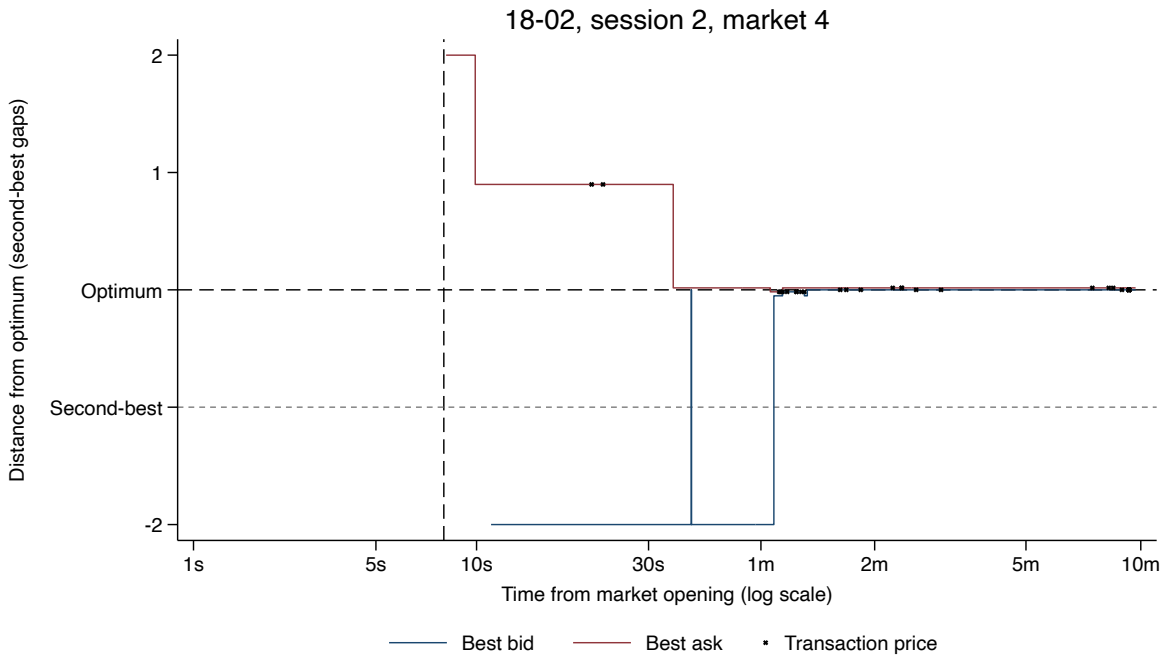


Figure D10: Time series plot for Session 2, market 4

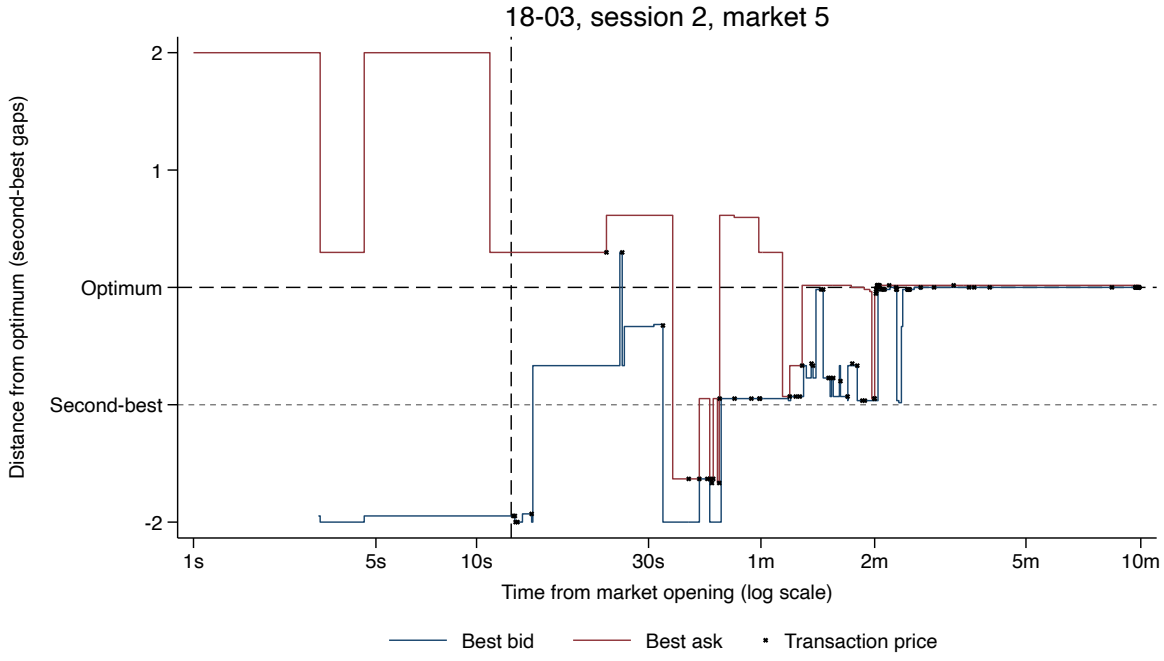


Figure D11: Time series plot for Session 2, market 5

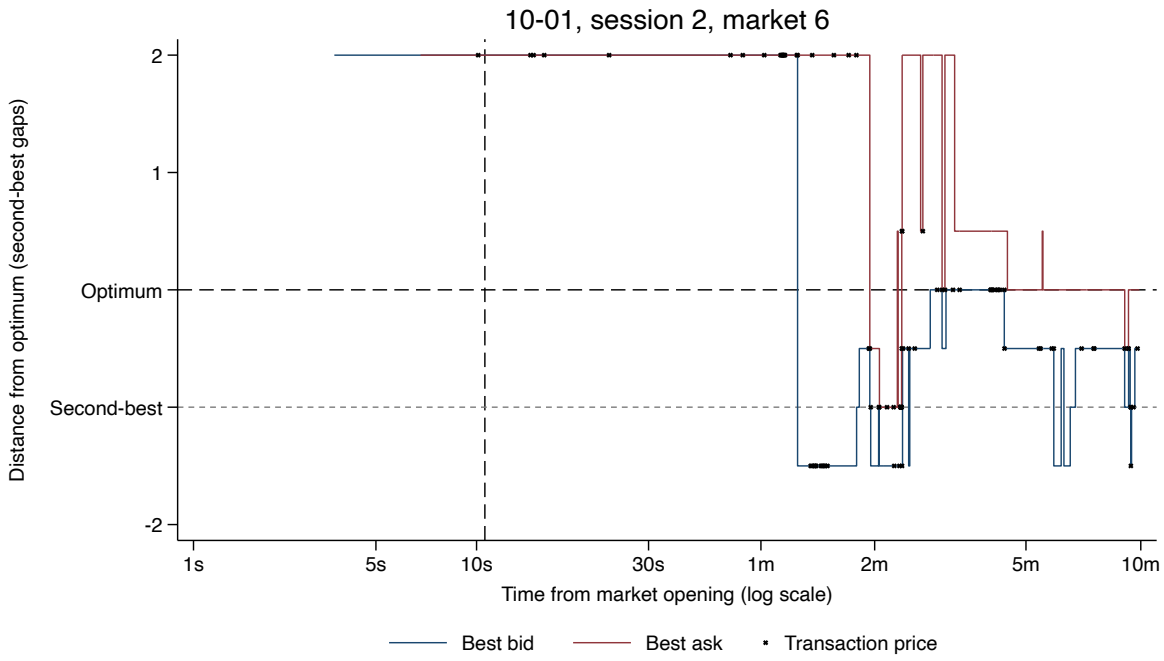


Figure D12: Time series plot for Session 2, market 6

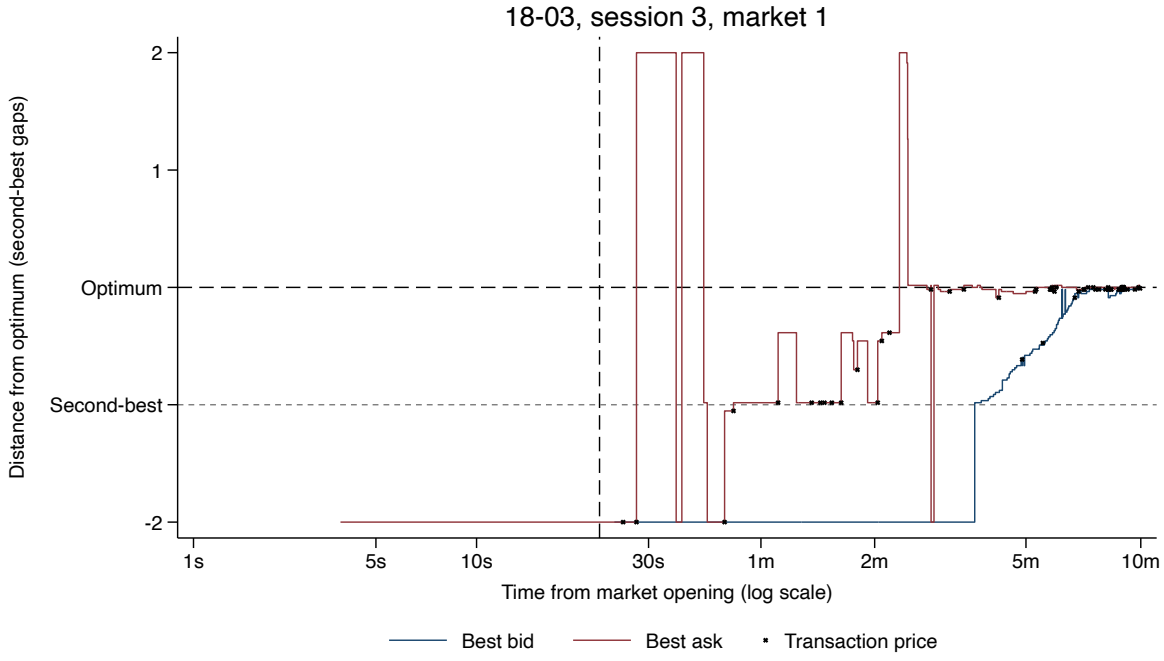


Figure D13: Time series plot for Session 3, market 1

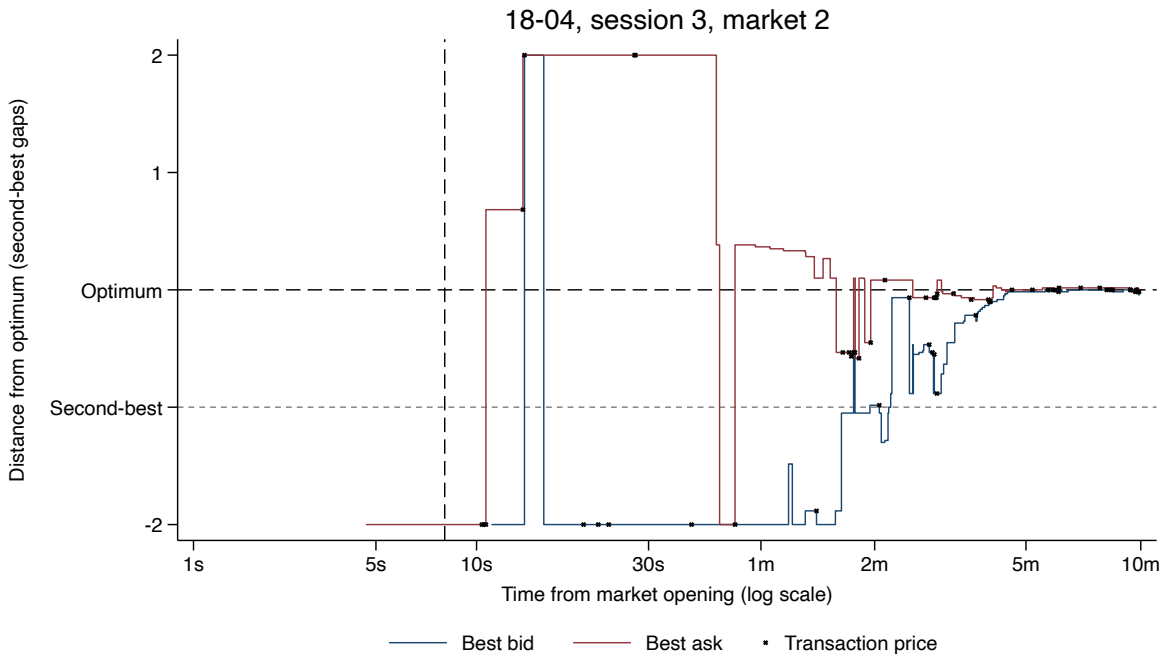


Figure D14: Time series plot for Session 3, market 2

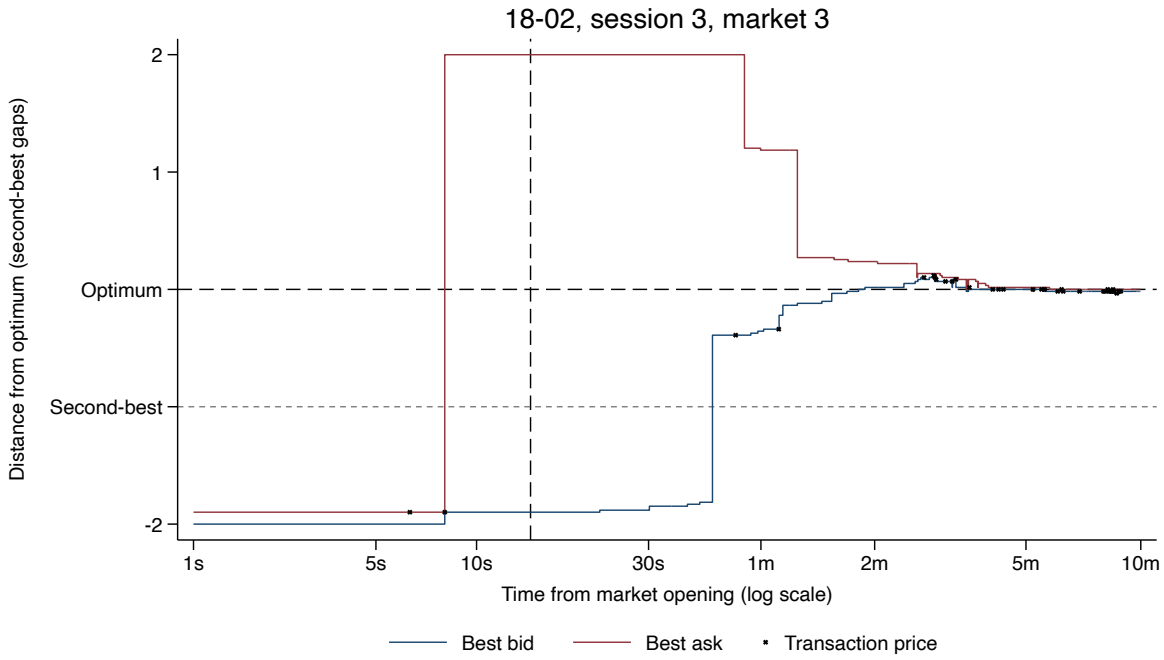


Figure D15: Time series plot for Session 3, market 3

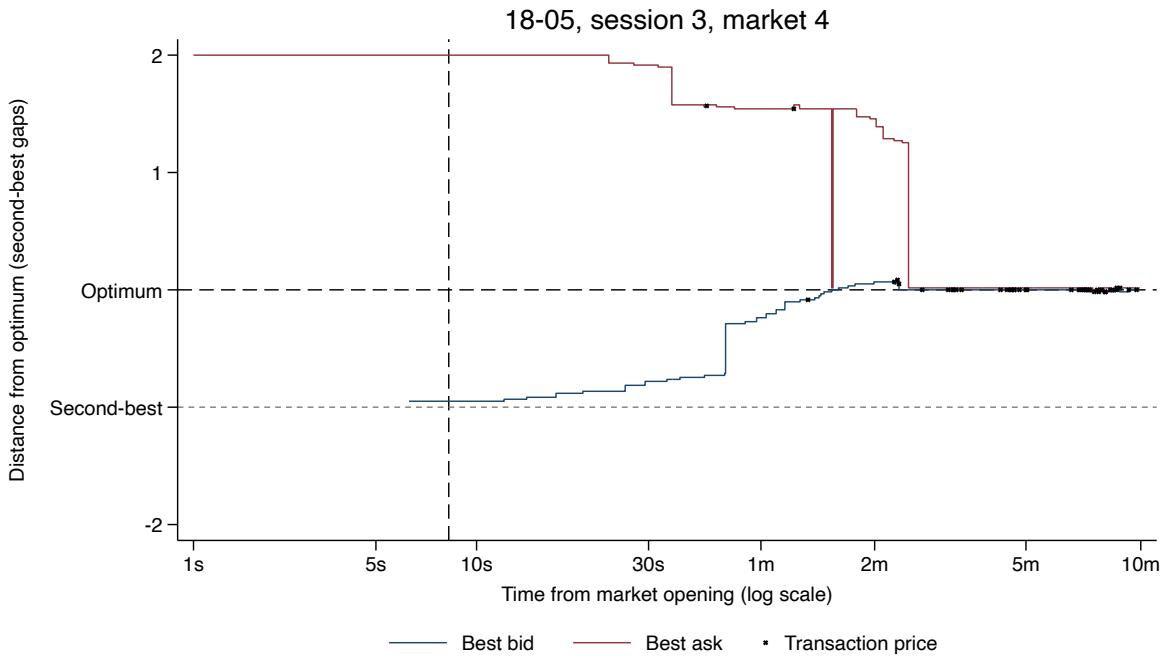


Figure D16: Time series plot for Session 3, market 4

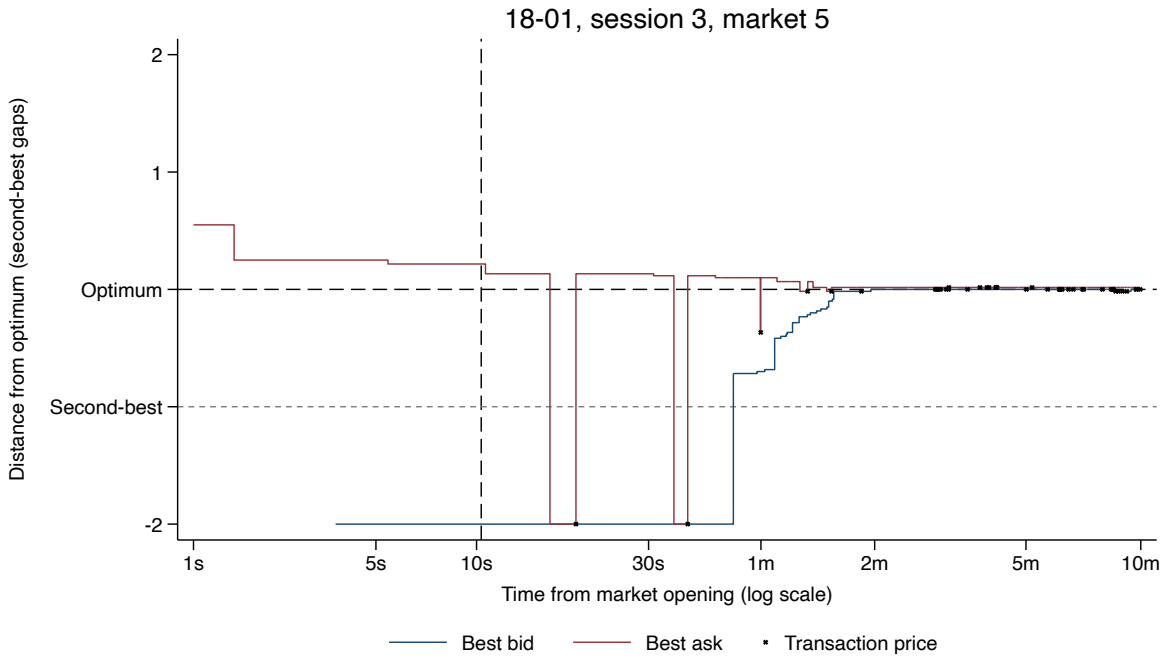


Figure D17: Time series plot for Session 3, market 5

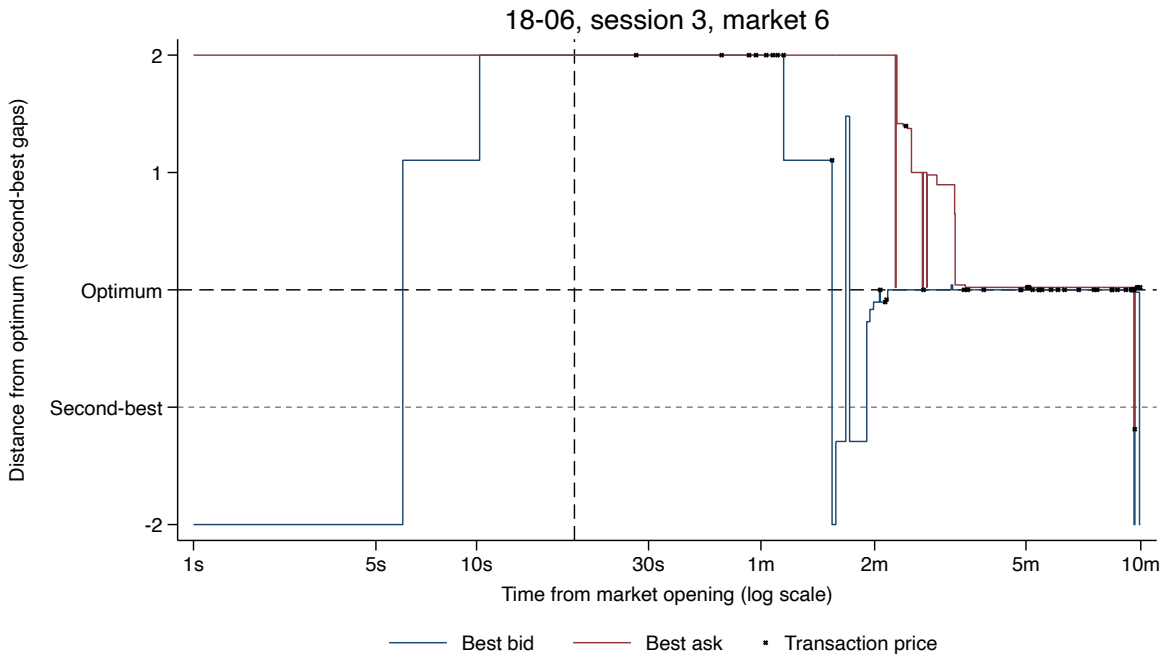


Figure D18: Time series plot for Session 3, market 6

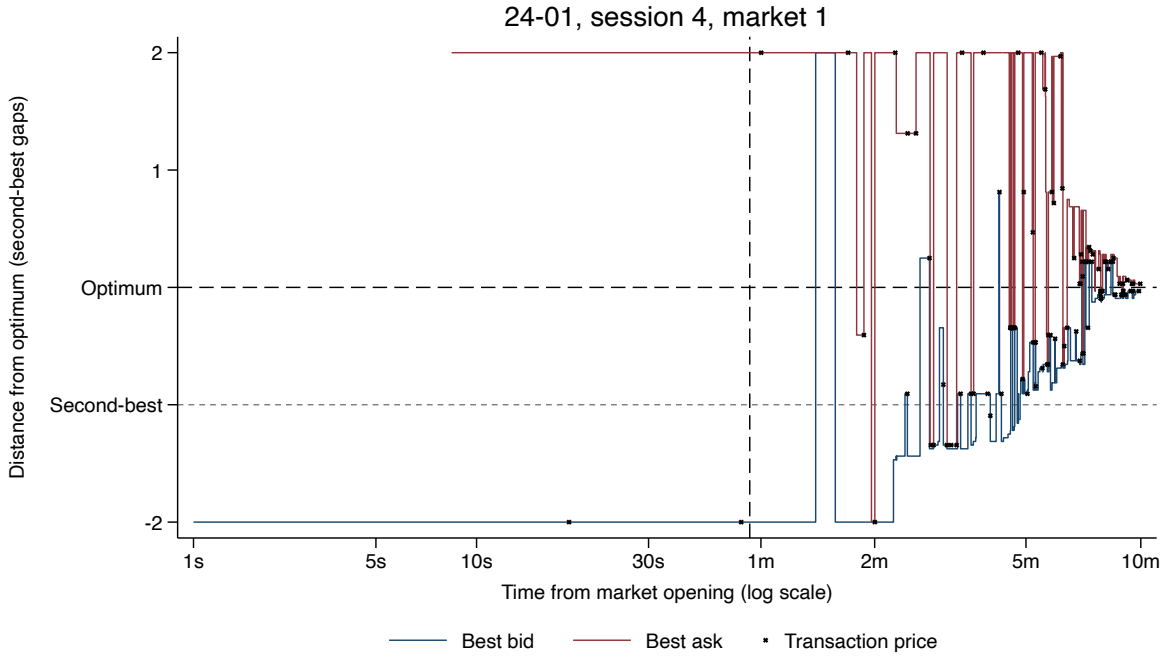


Figure D19: Time series plot for Session 4, market 1

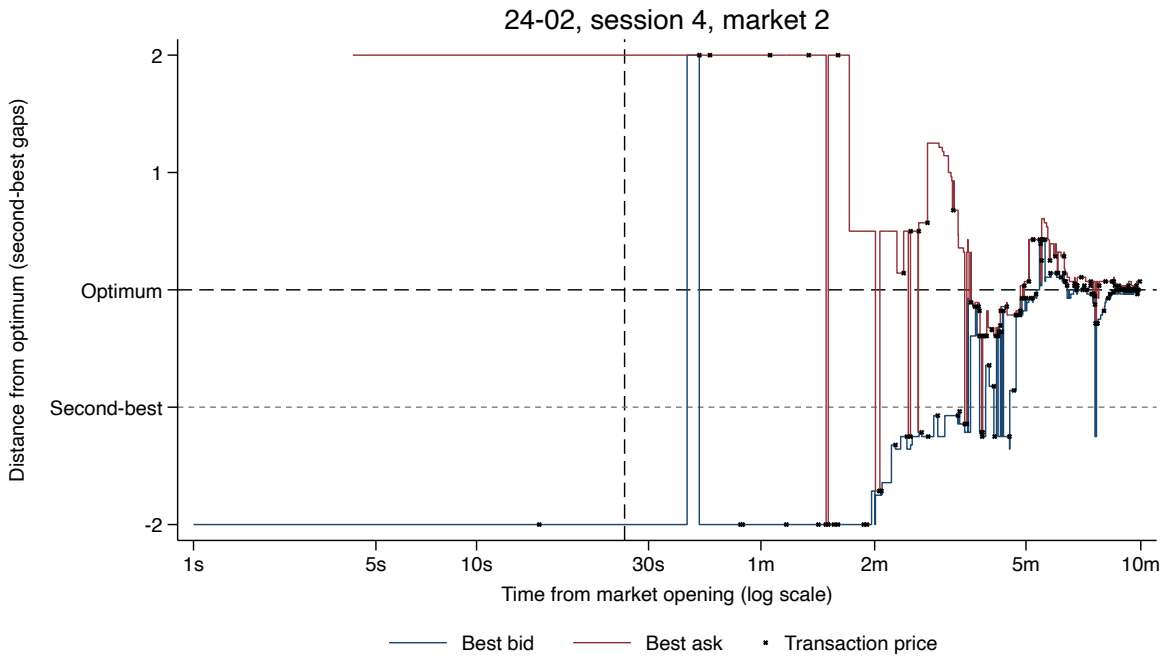


Figure D20: Time series plot for Session 4, market 2

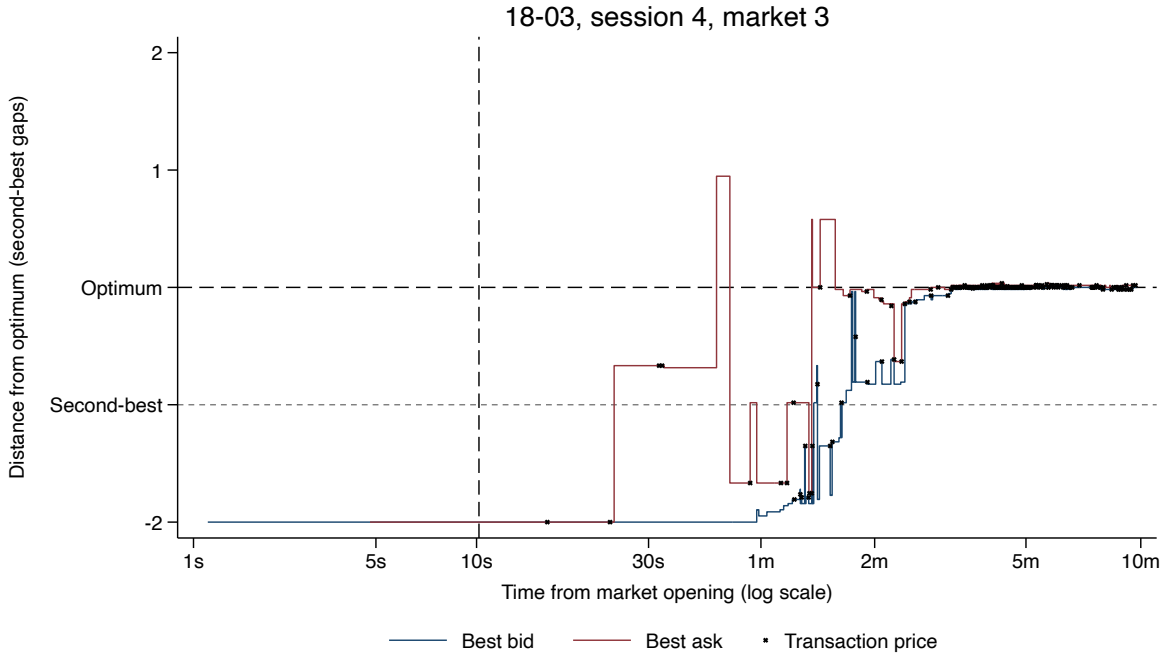


Figure D21: Time series plot for Session 4, market 3

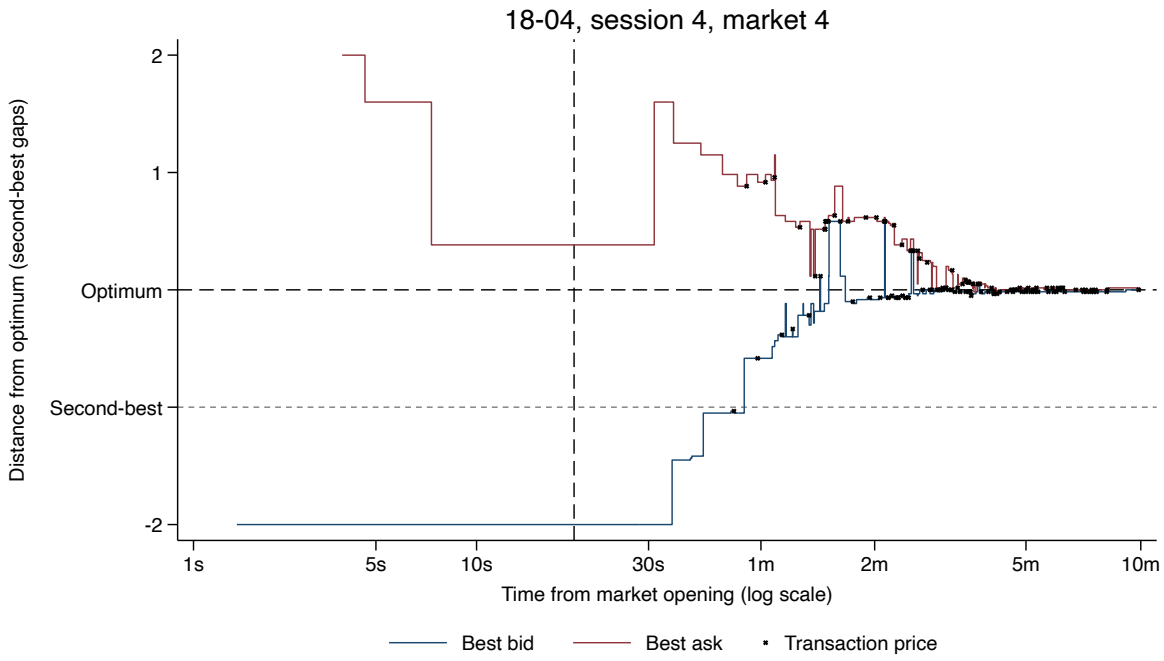


Figure D22: Time series plot for Session 4, market 4

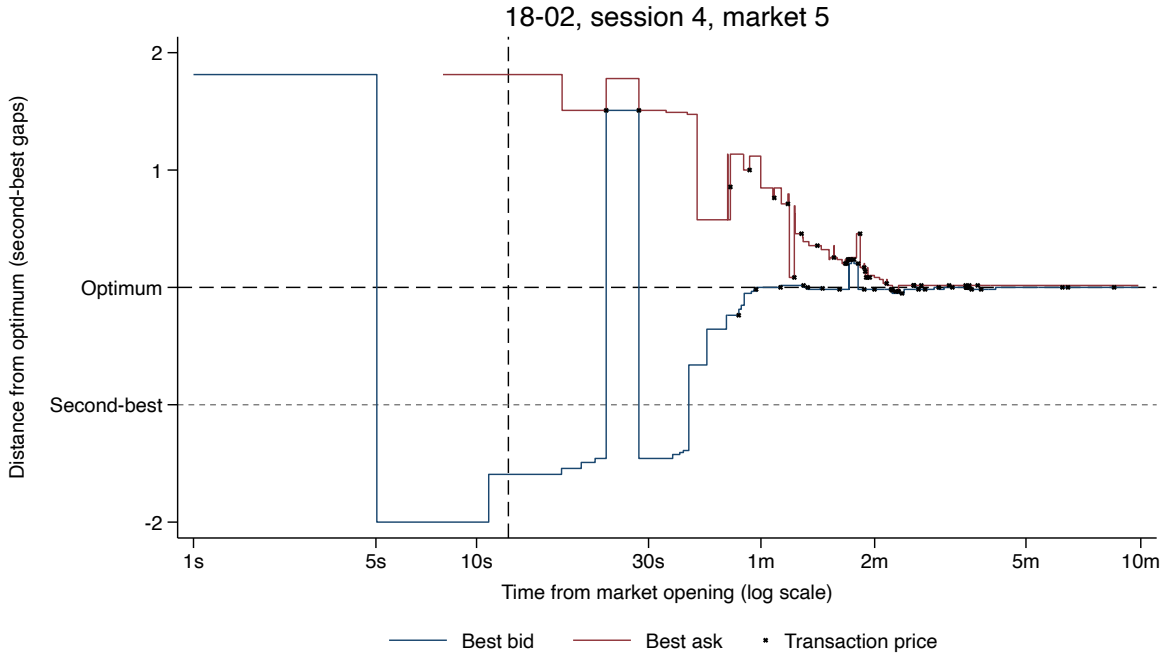


Figure D23: Time series plot for Session 4, market 5

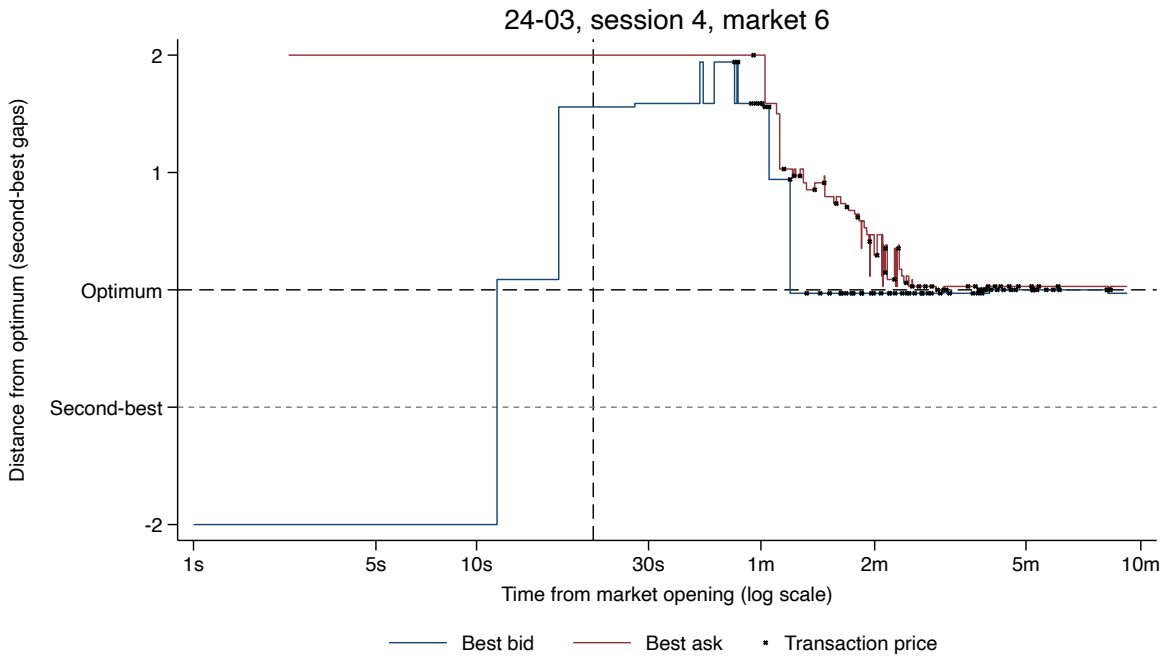


Figure D24: Time series plot for Session 4, market 6

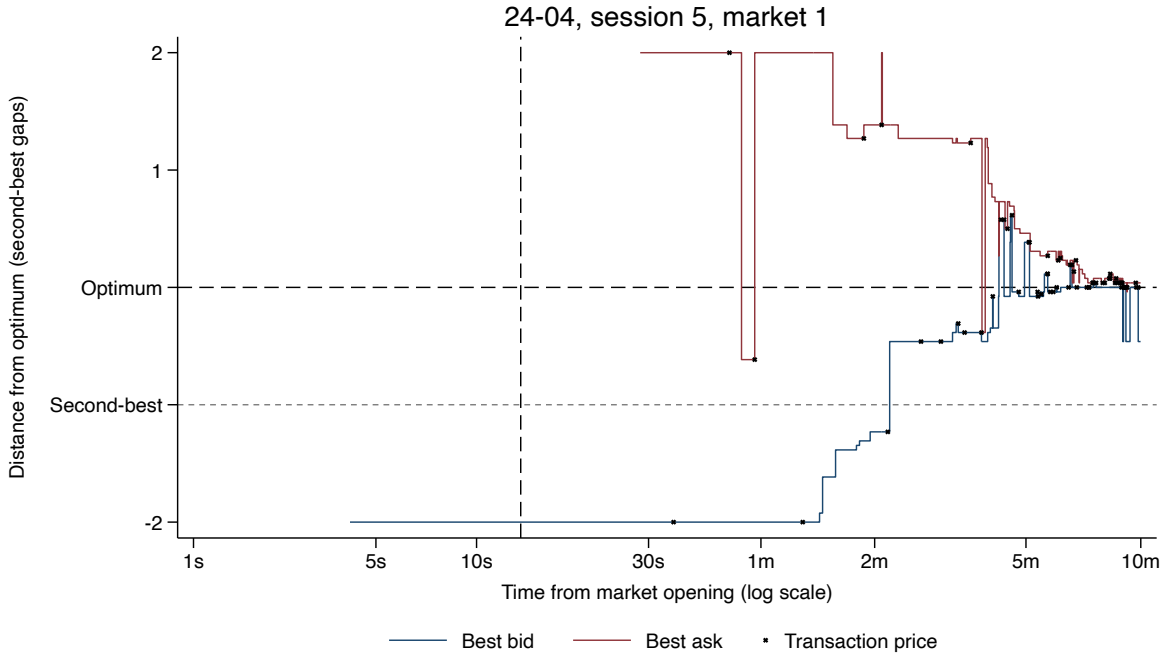


Figure D25: Time series plot for Session 5, market 1

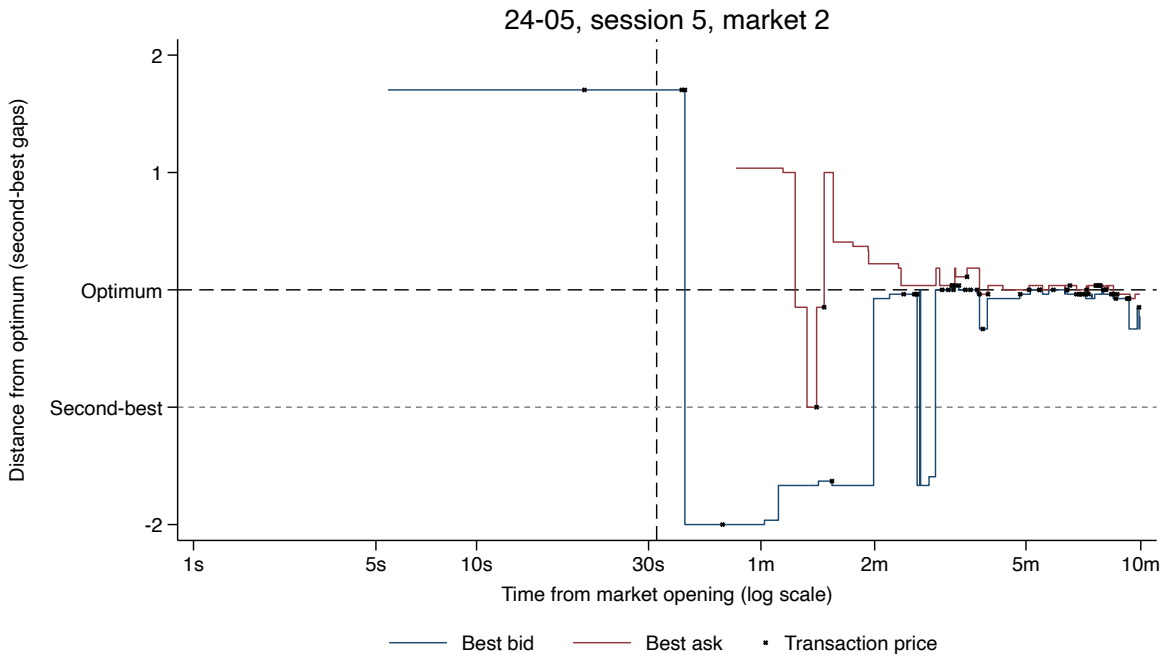


Figure D26: Time series plot for Session 5, market 2

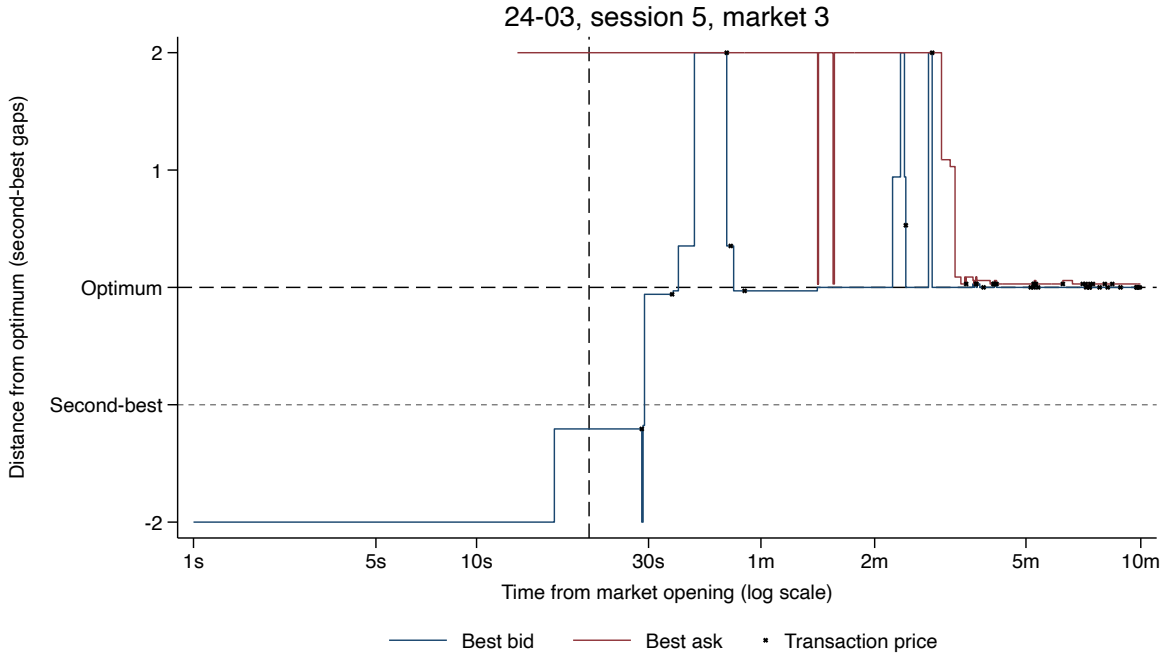


Figure D27: Time series plot for Session 5, market 3

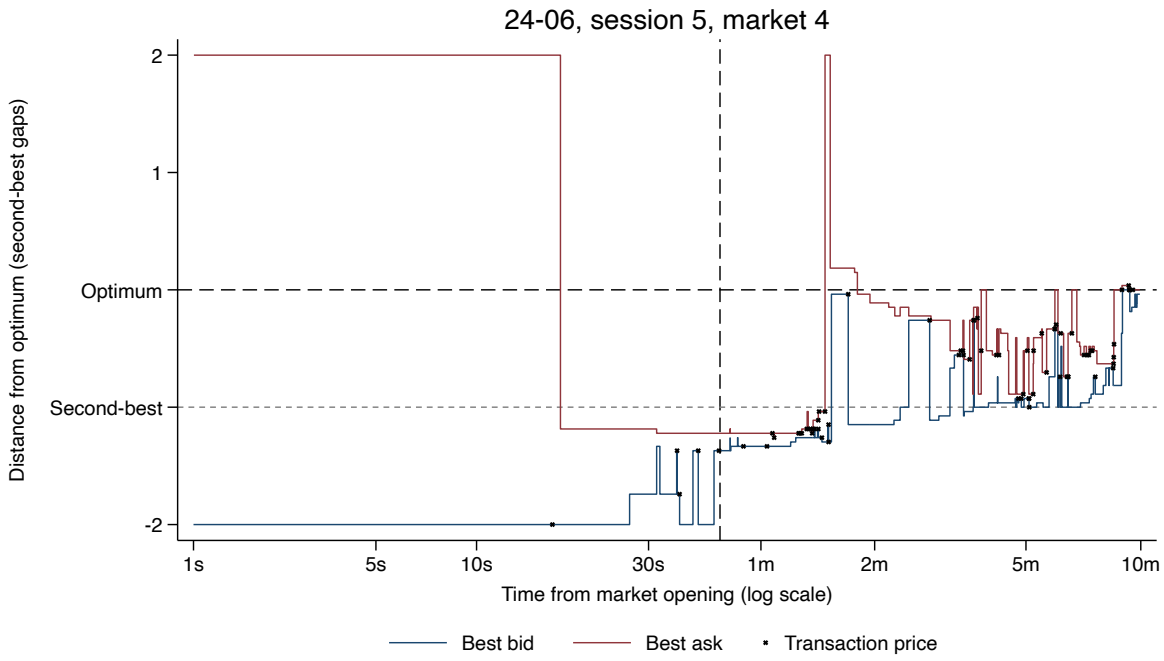


Figure D28: Time series plot for Session 5, market 4

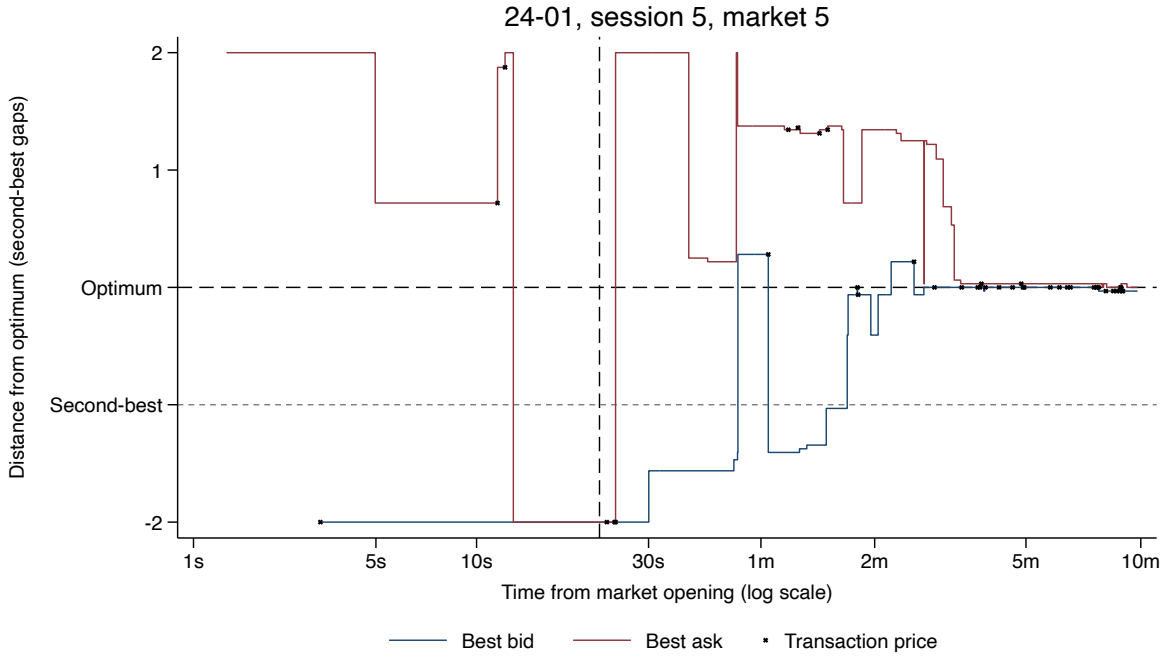


Figure D29: Time series plot for Session 5, market 5

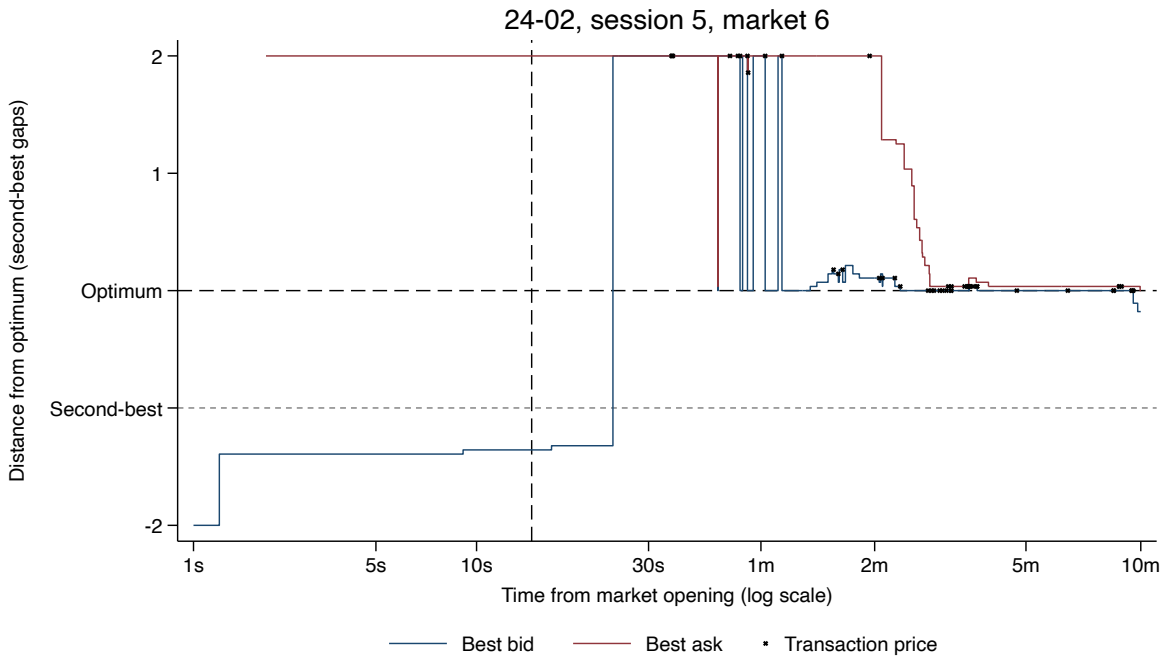


Figure D30: Time series plot for Session 5, market 6