The University of Melbourne

Finance Research Essay

The Efficient Market Hypothesis does not hold when asymmetric information arises from *computational complexity*

Shireen Tang

November 2016

The market experiment.

Acknowledgments

Greatest thanks imaginable to my advisers Prof. Peter Bossaerts and Dr. Carsten Murawski for their dedicated effort, time and guidance.

Also, special thanks to Dr. Elizabeth Bowman, Dr. Nitin Yadav, the Brain, Mind and Markets Laboratory, and my Level 12 mates Harvey, Bosco and Edison whom I've spent so many late nights working with this year.

Declaration

"This essay is the sole work of the author whose name appears on the title page. It contains no material that the author has previously submitted for assessment at the University of Melbourne or elsewhere.

To the best of the author's knowledge, the essay contains no material previously written or published by another person except where reference is made in the text of the essay."

Signed,

Abstract

When correct securities valuation involves summing bits of information, the standard (competitive, centralised) financial markets mechanism is known to be able to efficiently aggregate the individually held bits. As a result, markets price securities well, so well that even insiders (participants with strictly better information) cannot make money. Therefore, the Efficient Markets Hypothesis ("EMH") holds. Computer science would characterise this situation as a special case of "*finite sample complexity*".

Here, we ask whether EMH extends to the case where the valuation problem is of a combinatorial nature. There, it is not sufficient to just sum bits, but one has to select the right bits, and in the right combination. Computer science refers to such a situation as *"computationally complex"*. We study the case where correct valuation requires markets to solve instances of a particular computationally complex problem, the *"0-1 knapsack problem"*.

We find that valuations are incorrect (though better than those of the average participant), and insiders (those who know the correct valuation) systematically outperform. Nevertheless, prices transmit information, but in an indirect way – in interaction with trade, they "nudge" participants with incorrect solutions towards improving their valuations. Our findings have important implications – on the theory side, for the scope of the Grossman-Stiglitz paradox and hence, the case for markets to incentivise discovery; on the empirical side, for the interpretation of past tests of EMH on historical field data; and on the application side, for the promise of prediction markets.

Contents

I Introduction

The efficient market hypothesis is the hallmark of modern finance. It permeates through all walks of financial life, from how corporations make decisions, to how individuals allocate investments.

Under the strong form version of the efficient market hypothesis ("EMH"), all available information is fully reflected in prices, and insiders cannot make abnormal trading profits (Fama, 1970). It relies on perfect markets without transaction and information costs, and most importantly, assumes agents maximise utility and form rational expectations as defined by Lucas (1978).

Let us consider a problem where the market needs to correctly value traded securities, and participants face asymmetric information when searching for the optimal solution. Under EMH, the market is expected to amplify information held by insiders (those who know how to value securities correctly), such that it efficiently reflects in prices. Hence, one important consequence is that insiders cannot make consistent abnormal profits.

Experiments have consistently shown that markets value securities correctly, and hence, that EMH holds. Many studies have demonstrated that insiders failed to make profits and markets efficiently reflected all information (Plott and Sunder, 1982; 1988; Sunder, 1992; Bossaerts, Frydman and Ledyard, 2013).

However, in these studies, the correct solution is obtained by summing bits and pieces. This is described by computer science as a special case of "*finite sample complexity*" (Ehrenfeucht, Haussler, Kearns and Valiant, 1989). Finite sample complexity is a situation in which sampling improves one's estimate of the solution to the valuation problem. That is, three samples will bring us closer to the optimal solution than two. Following that, four samples increase the likelihood of reaching the solution compared to three, and so on.

The experimental success of EMH has led to the emergence of prediction markets in practice. Such markets have reduced prediction errors compared to conventional forecasting models, and have proven successful in multiple real world applications (Arrow, 2008; Plott, Court, Gillen and McKenzie, 2015; Zitzewitz and Cowgill, 2015).

At the same time, the EMH has been criticised. From an empirical perspective, multiple studies also reject the hypothesis. We continue this in the discussion. This questions whether valuation problems in the real world are truly of a finite sample complexity nature.

From a theoretical perspective, a core study by Grossman and Stiglitz (1980) stressed that perfectly efficient markets are impossible. They highlighted the paradox: if markets efficiently reflected all information, insiders have no incentive to acquire costly information, as they make no economic rents. Without trade, a competitive economy cannot be in equilibrium and markets will collapse. Laboratory experiments have confirmed the veracity of the Grossman-Stiglitz paradox in a centralised market setting; Sunder (1992) showed that participants stop buying information, or, when information is auctioned off, its sales price converges to zero.

Computer science recognises another form of complexity, namely *"computational complexity"*. In our study, it is a situation where the correct valuation depends on the solution to a combinatorial problem. Now, sampling three pieces of information is not necessarily better than two. Rather, it is the right combination of bits and pieces that will form the correct solution.

Here, we investigate EMH when correct valuation depends on solving a computationally complex problem.

We contend that a combinatorial nature better reflects problems in the field. If our understanding of market efficiency were to break down under this new framework, we may need to reconsider how we think about markets, how prices reflect information, and the mechanisms used to allocate resources and distribute wealth.

We refer to the notion of computational complexity used in computer science, defined under mathematical models of computation, such as a Turing machine (Karp, 1972; Cook, 1983). Problems are classified into difficulty classes based on algorithms and the resources required to solve an instance of the problem, such as time and memory. Although by no means a new field, recent literature has only begun to examine human decision-making under a computer science framework, on both the individual and market level (Meloso, Copic, and Bossaerts, 2009; Murawski and Bossaerts, 2016).

To examine how computational complexity affects market efficiency, we used the "*0-1 knapsack problem*". To solve an instance of the problem, one selects a subset from a set of items with given values and weights, to maximise the total value within a weight constraint. Only one combination of the correct items will yield the optimal solution.

We tested if EMH applied when correct market valuation depended on solutions to instances of the knapsack problem. There is uncertainty about the solution, which cannot simply be reduced by taking a few samples (trying a few solutions); one is never sure of the optimal solution without deriving the entire set of possibilities. As such, this is not a situation of finite sample complexity.

We tested EMH as follows. First, do participants make more money when they are better at solving the valuation problem? Second, do prices readily reveal the optimal solution?

To answer the latter, we asked if someone who is not attempting to solve the knapsack problem could merely look at prices and construct the optimal solution. Was the solution better than those entertained by the average participant?

In addition, we explored the mechanics behind information revelation. Specifically, we asked whether and how those who are entertaining wrong solutions are able to infer better solutions from prices.

To investigate this, we start from Bossaerts and Murawski (2016), who showed one type of cognitive inflexibility that explained poor performance in solving an instance of the knapsack problem. Namely, participants are reluctant to reconsider incorrect items put into the knapsack early in their search for the solution. We hypothesised that prices may improve computational performance by “nudging” individuals to reconsider the incorrect items.

The rest of the paper is organised as follows: Section II describes the methodology in terms of the participants involved, the experimental task, and the data analysed. Section III presents descriptive statistics and three key results. Section IV discusses each result in light of the EMH. Finally, Section V dissects the implications of such findings in tandem with key literature.

II Method

a. Participants

Participants were recruited from The University of Melbourne, in four experimental sessions with 18 or 20 participants each. Eligibility criteria were current students aged between 18 to 30 years with normal or corrected vision.

The final sample included a total of 78 participants (age range: 18 to 26 years, mean age: 22, standard deviation: 4, gender breakdown: 44 male, 34 female). The study was approved by The University of Melbourne Human Research Ethics Committee (Ethics ID: 1647059.1), in accordance with the World Medical Association Declaration of Helsinki. All participants provided written informed consent.

b. Experimental task

Participants attempted five instances of the "*0-1 knapsack problem*", while simultaneously trading in an online marketplace.

Nature of the problem

For each instance, participants selected a subset from a set of items with given values and weights, to maximise the total combination value subject to a total weight constraint. Such instances were similar to ones used in prior studies (Meloso et al., 2009; Murawski and Bossaerts, 2016). Participants attempted to solve the following maximisation problem in every instance, where *'i', 'w', 'v'* and *'C'* denote the item, weight, value, and capacity respectively:

$$\max \; \Sigma\, v_i w_i \;\; \text{subject to} \;\; \Sigma\, w_i x_i \;\leq\; C \;\; \text{and} \; x_i \in \{0,1\}$$

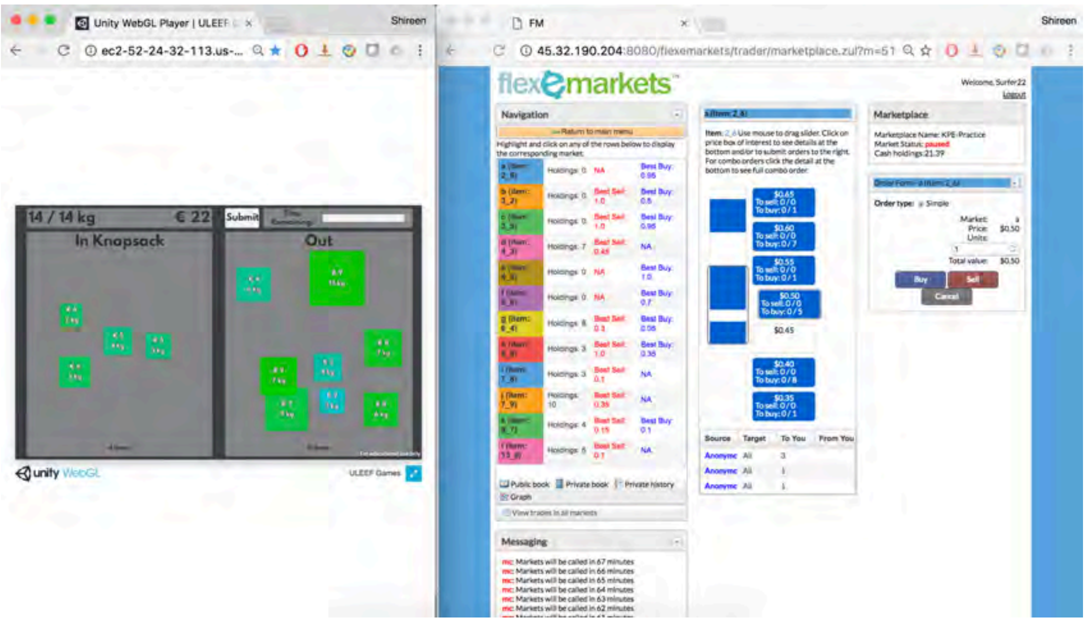
For each participant in every instance, we recorded data for every move made on each item in the problem with timestamps, and their final submitted attempt.

Trading in the online marketplace

In every instance, each item from the set corresponded to a market in the online marketplace, totaling 10 or 12 markets depending on whether the instance had 10 or 12 items to pick from. Markets were organised as a continuous double-sided open book, like most electronic markets globally. Participants then traded in the online marketplace, *Flex-e-Markets*. See Figure 1 for our graphical user interface.

We recorded price data of every trade and order in the marketplace, and synchronised timestamps with each move made on items in the problem interface. Equilibrium prices of the final trade per item were used. Note that the marketplace server was unreachable during the second instance in the final session, where we were unable to record any trade data. This denial of service was beyond our control.

Figure 1: Graphical user interface for the problem task and marketplace.



Paradigm figure of each participant's screen. The knapsack problem task interface is on the left, and the online marketplace is on the right. Flex-e-Markets is organised as a double-sided online electronic market, and participants traded in markets corresponding to each item in a particular instance.

Participant payoff

Participants placed bets on items they believed to be in the optimal knapsack solution, by buying bets of a correct item in its corresponding market. Each bet was a security that traded in a particular market. A correct item was an item that belonged in the optimal knapsack solution, while an incorrect one did not. Simultaneously, they would sell bets of the incorrect items in its respective market. Short sales were not permitted. For every instance, each participant was endowed with $25 in cash holdings, and 12 random bets. The price range of a bet was bounded between $0 and $1.

Each participant's final payoff was based on trading in the online marketplace, where a $1 liquidating dividend was paid for each correct bet they held, and any change in cash holdings at the end of the trading period per instance. Earnings were cumulative across instances. Additionally, they received a fixed reward for submitting a proposed solution to each instance, and a show-up fee of $5.

Market design considerations

We designed initial allocations to induce trade, where endowments tended to be concentrated in particular bets. We expected risk-averse participants would trade bets to diversify risk across more items. Initial allocations of bets were randomised and fair, as all participants were endowed with the same number of correct bets across the five instances, such that $36 in liquidating dividends was paid per participant on average.

We addressed aggregate risk, by ensuring there were equal quantities of bets per item in the marketplace. This ensured we controlled for possible pricing prediction errors that may arise from differences in relative supplies. Additionally, we imposed a fixed reward per attempt to mitigate concerns of hedging between the problem and marketplace.

On asymmetric information held by "informed traders"

In our market experiment, all participants are endowed with the same information on a particular instance of the knapsack problem. Participants then generated asymmetric information when searching for the optimal solution to correctly value securities trading in the market. Hence, asymmetric information is the solution to a computationally complex problem. We defined a participant who held asymmetric information of the optimal solution as an *"informed trader"*, and those who do not, as an *"uninformed trader"*.

If a particular item belonged in the optimal knapsack solution, we expected the market to value corresponding bets on correct items at its fundamental worth. That is, we expected the price to equal its liquidating \$1 dividend payoff. Likewise, we expected the market to value incorrect bets at \$0, if the corresponding item was out of the optimal solution.

c. Data analysis

Our procedure tested for two definitions of market efficiency: (i) whether insiders can make abnormal profits given asymmetric information of the optimal solution, and (ii) whether prices efficiently revealed the solution for uninformed traders. Then, we analysed further for *"reluctance"* in uninformed traders – a tendency to not remove incorrect items placed into the knapsack early in the solution search.

On the first test for market efficiency

To examine if informed traders could make abnormal profits under the first definition of market efficiency, we constructed a performance metric for each participant per instance. The metric evaluated each participant's computational performance, with a score based on the distance in item space from the optimal solution. A score added 1 point if a correct item was in the proposed knapsack solution, and 1 point if an incorrect item was out, but 0 otherwise. This is then scaled out of 1.

For every instance, we compared participants' performance metrics to their respective economic payoff in cash dollars from trading in the marketplace. We estimated via a generalised linear mixed effects Gaussian model. The main regressor was the performance metric, denoted *'Score',* with fixed effects of instance difficulty '*Sahni-k*', and market experiment '*Session*'.

What followed was, whether market prices revealed information held by participants who knew the optimal solution, such that they could not make consistent abnormal profits from price differences. If the market was inefficient, participants could arbitrage mispriced securities; and make profits between an undervalued bet on a correct item and its $1 liquidating dividend, or sell an overvalued bet of an incorrect item with a price above $0.

On the second test for market efficiency

To evaluate if prices reflected the optimal solution under a second test for market efficiency, we defined a market performance metric as the distance in item space of a proposed market knapsack from the optimal solution. As prices are bound between 0 and 1, we allowed each equilibrium price to be the relative probability the market believed a particular item belonged in the optimal solution. We simulated as if the market would fill the knapsack by drawing items based on its respective price probability. We bootstrapped a proposed solution using 10,000 simulations, and defined the average as the '*market knapsack*'. If prices perfectly reflected the knapsack solution such that we draw only from optimal items, it would yield a performance metric at 1.

This method was chosen, as using prices alone would not account for the discrete nature of items in the problem. However, we also provided an alternative method for a market performance metric to confirm our findings (see appendix Figure 19).

We investigated if market prices reflected information on the optimal solution better than the average individual participant, and whether this declined as the instance became harder. Hence, we compared performance scores of the market to the average participant at every level of instance difficulty. We defined this in terms of computational complexity, by referring to the Sahni-k approximation algorithm. It first considers all combinations of a subset of items with cardinality k under "*brute force search*". Then, it fills a knapsack with items in reverse order of their value-to-weight ratio under the "*greedy algorithm*", an algorithm that simply picks the most efficient item. The optimal solution is then the knapsack with the highest value subject to the weight constraint. If k is zero, the Sahni-k coincides with the greedy algorithm.

In a complementary analysis to see if market prices were simply based on the greedy algorithm, we also compared each item's equilibrium price and respective efficiency ratio, calculated as the value-to-weight ratio of a particular item.

On information transmission

Under a third test, we extended the analysis for uninformed traders to examine how the market is able to improve participants' computational performance, and whether that related to market efficiency. We focused on two market variables that likely affected the probability of removing an incorrect item placed in the knapsack early, namely prices and trading volume (Murawski and Bossaerts, 2016). We then estimated via a generalised linear mixed effects logit model on the probability of removing an early incorrect item. The regressors were both the main and interaction effects of trading volume and prices, denoted '*Trade'* and '*Price'* respectively, with no fixed effects needed. We defined a time threshold at two minutes to specify the incorrect items placed early into a participant's proposed knapsack attempt.

In a separate analysis, to explain uncertainty in the knapsack task, we also examined the relationship between moves made on items in the problem and certain market variables. This was estimated via a generalised linear mixed effects Gaussian model. We generated a sequence of item additions "*in"* and removals "*out"* of the knapsack, represented via a "state of the knapsack" array with timestamps. From this state array, we examined moves made per item in the problem by each participant, scaled by total moves, as some participants made more moves than others.

III Results

a. Descriptive statistics

Figure 2: Summary table of average payoffs.

Average payoff	**$**	**SD ($)**
Per instance (all participants)	6.34	4.06
Per instance (for a perfect score)	8.22	3.51
Per instance (for a non-perfect score)	5.11	3.93

Figure 3: Summary table of average performance.

Average performance	**%**	**SD (%)**
Of participants in all instances	75.4	22.6
Of markets in all instances	81.7	9.5

Figure 4: Relationship between the number of participants with a perfect performance score of 1 by instance, and Sahni-k instance difficulty.

The above figure shows the number of participants who solved a particular instance at each Sahni-k level, that is, they had a performance score of 1 for that instance. Error bars reflect the standard deviation across sessions. The average proportion of participants who solved any given instance was 37.2%.

Figure 5: Summary table of the number of participants with a performance score of 1 by instance.

Number of participants that solved the instance

Sahni-k	**0**	**1**	**2**	**3**	**4**
Number of participants	47	36	37	21	5
Proportion of participants (%)	60.3	46.2	47.4	26.9	6.4

On trades in the marketplace:

Figure 6: Summary table of total trades in the marketplace by instance (and Sahni-*k* difficulty).

Total number of trades in the marketplace per instance

Sahni-k	0	1	2	3	4
Session B	150	190	114	202	152
Session C	138	112	112	147	149
Session D	114	91	125	148	109
Session E	107	127	141	NA	138
Average	**127**	**130**	**123**	**167**	**137**

Figure 7: Summary table of the average number of trades in a market by item type.

Average number of trades per item in each instance

Item type	Correct items	Incorrect items
Session B	13	17
Session C	13	11
Session D	11	10
Session E	10	13
Average	**12**	**13**

On item moves in the problem per participant:

Figure 8: Summary table of the average number of item moves by a participant per instance.

Average number of moves per participant by instance

Sahni-k	**0**	**1**	**2**	**3**	**4**
Session B	20	18	20	29	19
Session C	19	20	17	32	24
Session D	24	24	29	42	26
Session E	24	27	21	NA	27
Average	**22**	**22**	**22**	**34**	**24**

Figure 9: Summary table of the average number of moves per item by a participant in an instance by item type.

Average number of moves per item in each instance

Item type	**Correct items**	**Incorrect items**
Session B	2	3
Session C	3	3
Session D	3	3
Session E	3	3
Average	**3**	**3**

b. Result 1

Did informed traders make money?

To examine how computational complexity affects the first definition of market efficiency, we determined whether informed traders could make abnormal profits. We defined participants who found the optimal solution as informed traders, and expected they would trade on such asymmetric information to make profits.

Hence, we tested whether a participant's computational performance explained their economic payoff, under the following model (BIC: 2036, distribution: Gaussian):

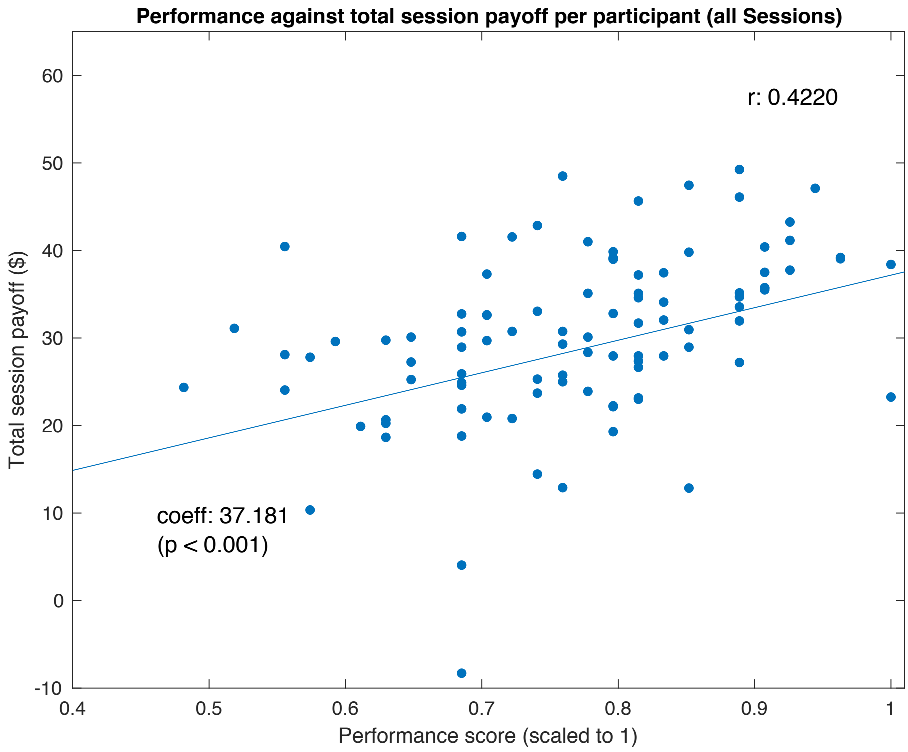
$$\text{Payoff}_{ij} = \text{Session}_i + \text{Sahni-}k_i + \beta_i \text{ Score}_{ij} + \epsilon_{ij}$$

We found a significant relationship between a participant's economic payoff against their performance score for each instance. In fact, results were even stronger when we accounted for fixed effects (coefficient: 7.011, $p < 0.001$, N: 370). We calculated a participant could make almost $3.51 more if their computational performance increased by 50%. That is, they made the correct decision on 50% more items to put in or remove out of the knapsack attempt. We also noted an insignificant constant term, which suggested that a participant makes zero payoffs if they guessed all items incorrectly. This result implied the market was not efficient.

The average payoff for participants who perfectly solved an instance was $8.22, compared to $5.11 for those who did not. Moreover, the maximum payoffs in each instance were only by participants who had found the optimal solution. The maximum payoff was $21.05 for the easiest instance with Sahni-*k* 0, and $12.50, $13.70, $16.25, and $11.95 for Sahni-*k* 1 to 4 of increasing difficulty respectively.

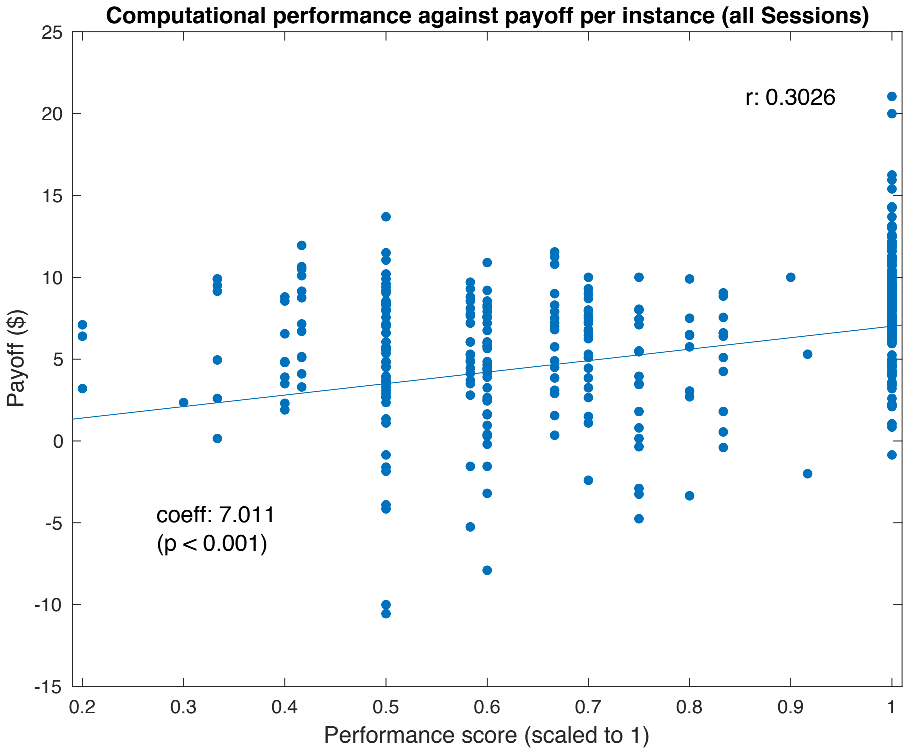
Furthermore, the average payoff was \$8.41 for Sahni-k 0, and \$5.13, \$5.03, \$5.93 and \$7.08 for Sahni-k 1 to 4 of increasing difficulty respectively. We noted the distribution of payoffs is significantly consistent between instances based on pairwise t-tests, except when Sahni-k was 0 or 4 (see appendix Figures 16 and 18 for significance tests).

Figure 10: Relationship between computational performance, and total session payoff from trading in the marketplace (by participant).



The distribution of each participant's total payoff evaluated against their respective performance score, aggregated across all five instances in that session. The Pearson correlation 'r' between a participant's computational performance score across all five instances and their total session payoff was 42.20%.

Figure 11: Relationship between computational performance, and economic payoff from trading in the marketplace (by instance).



The distribution of each participant's payoff per instance evaluated against his or her respective performance score for that instance. The Pearson correlation between computational performance and economic payoff of individual instances was 30.26%.

b. Result 2

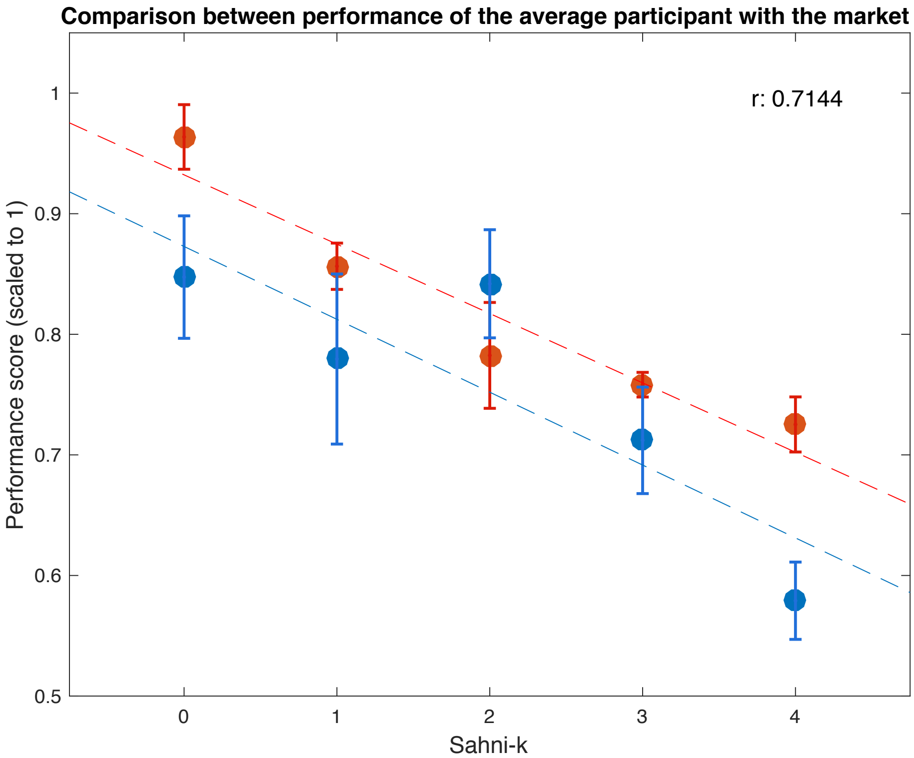
Did prices efficiently reflect the optimal solution?

We then investigated how complexity affected the market under a second definition of market efficiency. Specifically, we focused on how it related to Sahni-k instance difficulty. To determine whether prices efficiently reflected all information, even that held by informed traders who had the optimal solution; we examined market performance. Can uninformed traders gauge from prices if a particular item belonged in the optimal solution? If so, does the market perform better than the average participant?

The average performance of the market knapsack across all sessions was 96.4% for Sahni-k 0, compared to the average participant performance at 87.1%. Following that, the average performance was 85.6%, 78.3%, 75.8% and 72.5% for Sahni-k 1 to 4 respectively, compared to the average participant performance at 81.1%, 86.7%, 74.6% and 57.8%.

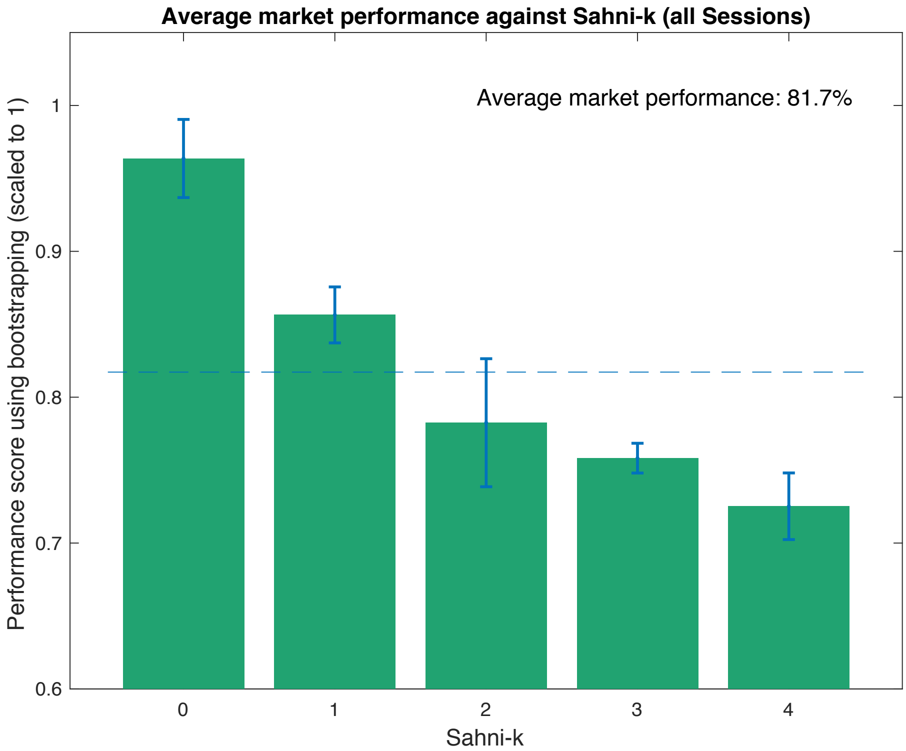
We used an alternative approach to evaluate market performance and found similar conclusions. Market performance also exhibited a monotonically declining relationship as an instance became more computationally difficult, in terms of Sahni-k (see appendix Figure 19). Additionally, we briefly examined the correlation between an item's price and its respective efficiency ratio, and found little relationship (see appendix Figures 20 to 24).

Figure 12: Comparison between the average participant performance and market performance, evaluated against Sahni-k instance difficulty.



The above compares the average market performance (red) against the average individual participant performance (blue), at each level of Sahni-k instance difficulty. Error bars indicate the standard deviation of the average score across all sessions. The performance score was calculated as distance in item space of the proposed knapsack attempt from the optimal solution. The Pearson correlation between the average participant and market performance was 71.44%.

Figure 13: Average market performance using price probabilities for bootstrapping, evaluated against Sahni-k instance difficulty.



The above figure shows the performance score of the 'market knapsack' averaged across all sessions at each level of Sahni-k instance difficulty. The 'market knapsack' was constructed using the price probability of each item, and simulated 10,000 times through bootstrapping. The score was then calculated as the distance in item space from the optimal solution.

d. Result 3

How does the market transmit information?

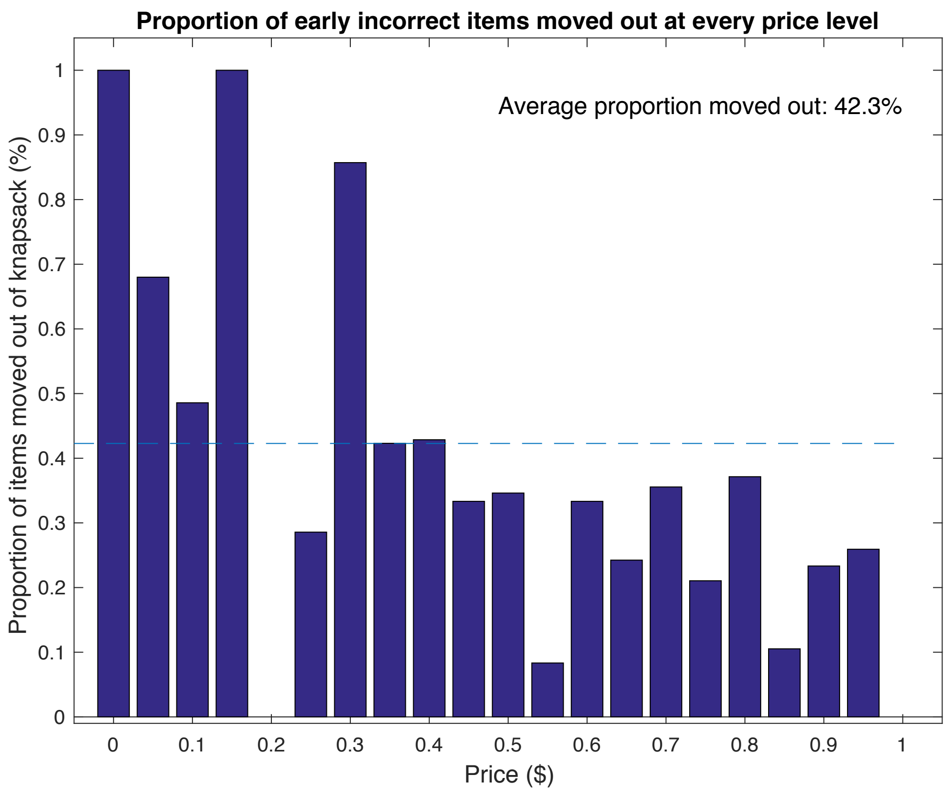
We extended the analysis to understand how the market could improve computational performance for uninformed traders. If prices efficiently reflected whether a particular item was in the optimal solution or not, we expected participants to respond to price signals and trade on such information. Hence, does the average participant perform better as a group compared to when searching for the solution individually?

We tested under the following model (BIC: 2144, distribution: logit):

$$\text{Moveout}_{ij} = \alpha_i + \beta_1 \text{Trades}_{ij} + \beta_2 \text{Price}_{ij} + \beta_3 \text{Trades}_{ij} \text{ x } \text{Price}_{ij} + \epsilon_{ij}$$

Results showed that the probability of removing an incorrect item placed into the knapsack early increased with low prices (coefficient: -3.491, $p < 0.001$, N: 479), but more so for highly traded items with low prices (coefficient: 13.725, p: 0.027, N: 479). However, there was no significant relationship between simply trading volume and the probability of removing items at the 5% level (p: 0.068, N: 479). The marginal effect of a 10% change in prices is a 10.5% increase in the probability of removing an incorrect item included early in the search.

Figure 14: Proportion of items moved out at every price level, given they were incorrectly placed into the knapsack early in the solution search.



The above shows the proportion of total early incorrect items moved out of the knapsack, at every price level. An early incorrect item is defined as an item that does not belong in the optimal solution, placed into the proposed knapsack attempt within a two-minute time threshold from the start of trading.

IV Discussion

Experiments have shown that, in a world where asymmetric information reflects finite sample complexity, the EMH obtains. Specifically, informed traders are unable to make abnormal profits, and their information can directly be read off market prices. Markets correctly value securities given the available information.

Here, we tested EMH when correct valuation required participants to solve a combinatorial problem. Information heterogeneity emerges because different participants follow different strategies when searching the solution space. Correct valuation is computationally complex. We find that EMH does not hold, because those that manage to solve the valuation problem correctly make more money, and prices do not reveal the correct valuation. Nevertheless, prices – in interaction with trade – transmit information, but in an indirect way: they "nudge" participants to reconsider parts of their valuation model that are mistaken.

a. On Result 1

Informed traders make money.

Our first result showed a significant link between computational performance and economic payoff, specifically when participants solved the instance perfectly.

Because prices were not efficient, there was scope for informed traders to arbitrage profits from price differences between undervalued bets on a correct item and its $1 liquidating dividend, and sell overvalued bets on an incorrect item with a price above $0. Hence, participants could trade on any informational advantages they generated from finding the optimal solution. Our results confirmed that informed traders with a better performance score made significant profits compared to those uninformed. If markets were efficient, then participants with the solution cannot abnormally profit, as such asymmetric information should fully reflect in prices under the efficient market hypothesis. We confirmed the market was not efficient in our next result.

Our results diverge from findings by Sunder and Plott, where insiders were generally unable to make abnormal profits. The authors showed that profits to insiders ultimately disappear, information aggregation holds, and markets are competitive and efficient in two seminal papers (1982; 1988). Additionally, Sunder (1992) showed that insiders would not pay for information, confirming the Grossman-Stiglitz paradox. As those who are better at solving combinatorial problems make more money in our setting, one can envisage that participants would want to pay for information on optimal solutions (e.g., an indication of which items are not in the optimal knapsack). Consequently, the Grossman-Stiglitz paradox is unlikely to apply to the case of computational complexity. Future experiments should test this, in analogy with Sunder (1992).

b. On Result 2

Prices are better than the average participant and become noisier as instances become more difficult.

If markets satisfied EMH, we would expect the market to perform at least as well as the average individual participant, as uninformed traders could readily gauge information of the optimal solution from prices. We confirmed that market prices revealed information that could be used to beat the average participant. However, prices became less informative as instance difficulty increased. We measured instance difficulty using a metric, Sahni-k, which has been found to predict individual solution scores (Murawski and Bossaerts, 2016).

Prices performed best when Sahni-k was 0, or when the instance was computationally easy such that using the greedy algorithm was sufficient. The market solved the instance quickly, with prices reaching equilibrium early into trading time across all sessions. However, when the instance was difficult at Sahni-k 4, the average market performance through prices was much lower at 72.5%. Therefore, market efficiency only held well for the simplest instance of the combinatorial problem, and our result indicated it declined as instances became computationally complex.

When evaluated in terms of price informativeness, the EMH obtains only for the simplest combinatorial instances. Computational complexity quickly affected informativeness, monotonically declining in Sahni-k.

Our results showed that individual level findings in Bossaerts and Murawski (2016) extend to the market level. There too, individual scores were found to decline with Sahni-k instance difficulty. Additionally, we confirmed this monotonic relationship in an alternative method and found similar results (see appendix Figure 19).

Although market prices were unable to efficiently reflect the optimal solution, we note the computational power of a collective market is still stronger than any one individual. From a previous study in *Science* (Meloso et al., 2009), more participants reported the correct solution under a market system compared to when they searched individually. In a secondary result, we confirmed this when we compared to findings in Murawski and Bossaerts (2016). Participant success rates in our market experiment were at an average of 87.1%, 81.1%, 86.7%, 74.6% and 57.8% for Sahni-k 0 to 4 respectively, compared to relatively lower individual success rates at 74.4%, 44.0%, 48.0%, 36.0% and 2.7% found in Murawski and Bossaerts (2016).

One explanation for why the market does better than the average individual is that the market explores more possibilities collectively. This was a key result from Murawski and Bossaerts (2016), where they found that individuals showed great heterogeneity in their search for the optimal solution with little overlap.

With 20 participants in our market, one critique is if there was sufficient liquidity, or the appropriate market dynamics that were comparable to the real world. However, experimental studies by Sunder, Plott and Smith (1978; 1982; 1988; 1992) used far fewer subjects and still successfully achieved market dynamics that allowed them to investigate market efficiency. In fact, Meloso et al. (2008) indicated that 15 subjects were sufficient to recreate the market dynamics necessary in an experiment.

Given the nature of combinatorial problems, we expected the market to value items in combination with all other items, and not in isolation. Hence, we expected that item prices would have little relationship with its relative efficiency ratio (see appendix Figures 20 to 24). Furthermore, we found that prices for the same items varied considerably across all sessions, suggesting that prices are uninformative and the market is generally unclear of the optimal solution. This general result held unless an instance was Sahni-k 0, which reduced to the greedy algorithm, such that correct items were also the most efficient.

From past research by Murawski and Bossaerts (2016), participants appeared to start with the greedy algorithm as a general heuristic, which perhaps explains why both individual participants and the market do well when Sahni-k is 0. However, better participants tend to be algorithmically flexible, and are able to solve instances using multiple approaches. This continues our next test on how participants improved their computational performance through the market despite such inefficiency.

c. On Result 3

The market "nudges" to overcome cognitive inflexibility.

In recent findings, Bossaerts and Murawski (2016) showed that participants tend to be reluctant to reconsider incorrect items placed into the knapsack early, known as path dependence in search, where they may not remember why they added a particular item early on. This explained why some were better at solving instances than others, such as being able to simultaneously replace items when reconsidering their attempt. Murawski and Bossaerts attributed this cognitive inflexibility to limited episodic memory.

We investigated the feedback from prices to problem solving and found that prices, in interaction with trade, helped improve participants' computational performance. It increased the probability of removing an incorrect item in the knapsack considered early in the search, with an effect for low priced items amplified by trade, but none for trading volume itself.

We examined this result in the context of our previous findings. Markets are not efficient when information is combinatorial, but may still perform better than individuals, by improving participants' computational performance collectively. Our findings indicated the market hardly revealed information in complex instances, yet participants still paid attention to crucial items that helped them improve valuations.

As such, markets transmit information in unconventional ways. Rather than directly revealing information, as in traditional accounts of EMH and the theory behind it (Grossman and Stiglitz, 1980), prices, in interaction with volume, "nudge" participants into improving their solutions.

From an individual standpoint, there are incentives for participants to consider items with low prices, especially when drawn to their attention by trading volume activity. Because participants can never be sure of the solution, low prices provide an opportunity to speculate and buy bets cheaply, in the hope of arbitraging a profit between a potential \$1 liquidating dividend and its undervalued trading price. Conversely, there was no incentive to buy an expensive \$1 item despite being confident it belonged in the optimal solution.

We hypothesise this as one possible reason why “nudging” occurs. Markets provide incentives to consider not only more items, but also items that are generally unfavourable. Hence, markets widen the search space for the solution. To test this, future experiments should consider allowing for short sales. We predict that participants will short sell expensive securities and generate trading activity from doing so. This interaction between prices and volume will likely also “nudge” uninformed traders to consider placing certain items into their proposed knapsack, thereby also replacing incorrect ones due to the weight constraint.

V Key implications

a. On the Grossman-Stiglitz paradox

From a theoretical standpoint, the Grossman-Stiglitz paradox does not apply in our experiment as insiders can make trading profits, and prices are not fully informative of the solution. Nevertheless, prices are still able to transmit information indirectly, in interaction with trading volume activity. These findings indicate that the scope of the paradox does not extend to markets with computational complexity.

What are the implications for intellectual discovery? Under the assumptions of Grossman and Stiglitz, if intellectual discovery leads to information aggregation of ideas in the market, then prices will reveal the outcome such that no one can profit. There would be no incentive for anyone to innovate and acquire costly information. However, in a previous *Science* paper on the knapsack problem, Meloso et al. (2009) showed this was not correct. Our paper adds to their findings.

We contend that intellectual discovery is better described in a combinatorial framework. Innovation is a process that combines only the right pieces of information in the correct way. This is not necessarily the same as aggregating bits of information together and expecting a solution. Murawski and Bossaerts (2016) found that most U.S. patents for inventions between 1790 and 2010 were novel combinations of existing technologies.

In fact, markets have been proven to improve the search for solutions to computationally complex problems. This implies that there is incentive to innovate, and in analogy to our experiment, participants will try out new ideas.

Meloso et al. (2009) showed that markets provided incentives to solve complex combinatorial problems that performed at least as well under a "*prize system*", if not collectively better. The prize system only rewards the first individual to solve the problem, characteristic of the patent system today. However, their research suggests that we can promote collective innovation and information sharing through market mechanisms, rather than design policies for individual protection, which may suffocate the creation of new ideas (Nuvolari, 2004). With such markets, there is no need to provide separate compensation through prizes or patents (Meloso et al., 2009).

Furthermore, in empirical findings from the British Industrial Revolution, Nuvolari (2004) emphasised the limits to the patent system. The author found it was collective invention that led to the burst of new technologies during the period. Our experiment showed that market participants performed better as a collective group with prompts from interacting market variables, compared to individual search for the solution alone. This also reverberates with results from Meloso et al. (2009) who found their market performed collectively better than the patent system that rewarded only the winner. They both critique the patent system, where ownership is likely fragmented, and costly market inefficiencies arise from monopoly rights.

b. On empirical studies of EMH

Currently, there is little consensus on whether the EMH holds in markets. Historical field data seems to both prove and disprove the theory. We revisit this debate and offer a new interpretation from our findings.

Our results suggest that the discrepancy is because past studies fail to differentiate the problems that markets try to solve. We reconcile the two, and hypothesise that EMH holds only where valuation depends merely on finite sample complexity. Whereas, in cases that dispute the EMH, valuation may really depend on solving a computationally complex problem.

We acknowledge that past empirical literature supports the hypothesis well. In a study on U.S. mutual funds, Fama and French (2012) showed that managers are unable to make abnormal returns after accounting for fund expenses. Fama (1991; 1998) also surveys a vast body of event studies, and ample evidence points to markets as efficient. He suggests that anomalies are either chance results or due to methodology errors.

On the other hand, a stream of critique on the EMH comes from behavioural finance literature. Because investors are subject to cognitive biases, rational expectations fail to hold and market prices are unlikely to be truly efficient. In empirical literature, DeBondt and Thaler (1985) proxy for overreaction through price reversals in winning and losing stocks, and suggested that investors do not always react rationally to new information. Pouget and Biais (2005) also observed overconfidence in an experimental market, and found it reduced trading performance, whereas self-monitoring enhanced performance. Other cognitive biases explored by literature include loss aversion, psychological accounting, and herding (Tversky and Kahneman, 1979; Gervais and Odean, 2001; Barber and Odean, 2001; Spyrou, 2013).

Current models assume we can aggregate a very particular type of asymmetric information. In some cases, information in the real world can be aggregated; but we recognise that information is sometimes better described in a combinatorial way. In these cases, it should not be surprising that EMH does not hold.

We identified a case where EMH may not hold, that is, when the market solves for a problem with combinatorial information, which is far more computationally complex. We thought prices directly reflected information, yet found they revealed little in complex instances. Rather, information transmission is through a "nudge" from the interaction of prices and trading volume. In our results, prices acted as an intervention for traders to reconsider and remove incorrect items put into the knapsack early, when they would be reluctant to do so otherwise. This only begins to change how we may think about information transmission in markets for the future.

c. On the promise of prediction markets

Prediction markets are now making their debut on several fronts. However, will they always work? Although prediction markets perform well when information aggregation is sufficient, our results reveal they may fail if valuation is the solution to a computationally complex problem.

Research shows that prediction markets provide an efficient revelation mechanism when information is dispersed among individuals (Arrow, 2008; Plott et al., 2015). In the field, there is a growing focus on the use of prediction markets to solve problems; beginning with the Iowa Electronic Markets used in the 1988 U.S. presidential election. In *Science*, Arrow (2008) indicated that such markets reduced prediction errors compared to conventional forecasting models. From an application stance, Plott et al. (2015) showed markets predicted box office revenues through two information aggregation mechanisms, and Zitzewitz and Cowgill (2015) cite success in corporations when used for sales forecasts.

Such prediction markets rely on prices to transmit information under the EMH. Markets are generically efficient as the underlying information is of finite sample complexity (Ehrenfeucht et al., 1989). The market can sample to merely add pieces of individually held information, which quickly converges to the solution. We know this as information aggregation (Grossman and Stiglitz, 1980; Hellwig, 1980; Diamond and Verrecchia, 1981). Asparahouva and Bossaerts (2015) provide evidence of this type of information aggregation even in a decentralised market, but noted that participants did not seem to be aware that trade took place at efficient prices.

In reality, markets may never be efficient in indicating the optimal solution. We question if it is too simple to assume a world under finite sample complexity, such that market efficiency generically holds and information aggregation is sufficient to value securities. We acknowledge there is often insufficient computational power in time available to take the number of samples required to solve the valuation problem.

Moreover, we note our controlled experiment had little aggregate information to begin with. That is, participants in the market started with the same information set, and as a whole, the market knew little about the optimal solution. Even if we assume finite sample complexity, can the market still value securities when there is little asymmetric information to aggregate?

In a concluding remark, we question the promise of prediction markets. Imagine markets in lithium, sodium, magnesium, sulfur and other chemical elements. These are potential components to the future of battery technology, which involves combining chemical compounds in precise ways, such as lithium-air, lithium-sulfur or sodium-ion (Pol and Venere, 2015). Which market should we invest in, if only a select combination will survive as the best battery technology? Can markets solve such a computationally complex problem?

We know that prediction markets will struggle to work in this case. Given the implications of our results, markets will fail to predict the lasting battery technology, because EMH does not hold and information aggregation is insufficient to solve the combinatorial problem. However, our results also reassure that markets will still give us incentives to search for the right solution. We caution against relying on only prediction markets for answers. After all, promises were not made to be broken.

VI References

Asparouhova E., and Bossaerts P. Experiments on Percolation of Information in Dark Markets. *The Economic Journal*, in press.

Arrow, K. J., Forsythe, R., Gorham, M., Hahn, R., Hanson, R., Ledyard, J. O., and others (2008). The promise of prediction markets. *Science*, New York, Washington, 320(5878), 877.

Barber, B. M., and Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *Quarterly Journal of Economics*, 261–292.

Barberis, N., and Thaler, R. (2003). A survey of behavioral finance. *Handbook of the Economics of Finance*, 1, 1053–1128.

Biais, B., Hilton, D., Mazurier, K., and Pouget, S. (2005). Judgmental overconfidence, self-monitoring, and trading performance in an experimental financial market. *The Review of Economic Studies*, 72(2), 287–312.

Bondt, W. F., and Thaler, R. (1985). Does the stock market overreact? *The Journal of Finance*, 40(3), 793–805.

Bossaerts, P., Frydman, C., and Ledyard, J. (2013). The speed of information revelation and eventual price quality in markets with insiders: comparing two theories. *Review of Finance*, 49.

Campbell, J. Y., Lo, A. W. C., MacKinlay, A. C., et al. (1997). The econometrics of financial markets, (2). *Princeton University Press*, Princeton, NJ.

Cook, S. A. (1983). An overview of computational complexity. *Communications of the ACM*, 26(6), 400–408.

Cowgill, B., and Zitzewitz, E. (2015). Corporate prediction markets: Evidence from Google, Ford, and Firm X. *The Review of Economic Studies*, 82(4), 1309–1341.

Diamond, D. W., and Verrecchia, R. E. (1981). Information aggregation in a noisy rational expectations economy. *Journal of Financial Economics*, 9(3), 221–235.

Ehrenfeucht, A., Haussler, D., Kearns, M., and Valiant, L. (1989). A general lower bound on the number of examples needed for learning. *Information and Computation*, 82(3), 247–261.

Fama, E. F. (1991). Efficient capital markets: II. The Journal of Finance, 46(5), 1575–1617.

Fama, E. F. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, 49(3), 283–306.

Fama, E. F., and French, K. R. (2012). Size, value, and momentum in international stock returns. *Journal of Financial Economics*, 105(3), 457–472.

Gervais, S., and Odean, T. (2001). Learning to be overconfident. *Review of Financial Studies*, 14(1), 1–27.

Gillen, B. J., McKenzie, J., Plott, C. R., et al. (2015). Two information aggregation mechanisms for predicting the opening weekend box office revenues of films: box office prophecy and guess of guesses. *California Institute of Technology paper.*

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

Hellwig, M. F. (1980). On the aggregation of information in competitive markets. *Journal of Economic Theory*, 22(3), 477–498.

Jegadeesh, N., and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1), 65–91.

Kahneman, D., and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, 263–291.

Karp, R. M. (1972). Reducibility among combinatorial problems: in complexity of computer computations. *Springer*, 85–103.

Lucas Jr., R. E. (1978). Asset prices in an exchange economy. *Econometrica: Journal of the Econometric Society*, 1429–1445.

Malkiel, B. G. (2003). The efficient market hypothesis and its critics. *The Journal of Economic Perspectives,* 17(1), 59–82.

Malkiel, B. G., and Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383–417.

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.

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

Plott, C. R., and Sunder, S. (1982). Efficiency of experimental security markets with insider information: An application of rational-expectations models. *The Journal of Political Economy*, 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*, 1085–1118.

Pol, V., and Venere, E. (2015). Sodium-ion batteries are potential power technology of future, *Purdue University News.* Available at: <http://www.purdue.edu/newsroom/releases/2015/Q3/sodium-ion-batteries-are-potential-power-technology-of-future.html> [Accessed: November 4, 2016]

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, 99–118.

Spyrou, S. (2013). Herding in financial markets: a review of the literature. *Review of Behavioral Finance*, 5(2), 175–194.

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

Wolinsky, A. (1990). Information revelation in a market with pairwise meetings. *Econometrica: Journal of the Econometric Society*, 1–23.

VII Appendix

a. Average participant payoffs per instance

Figure 15: Relationship between the average economic payoff from trading and Sahni-k instance difficulty (all solutions).

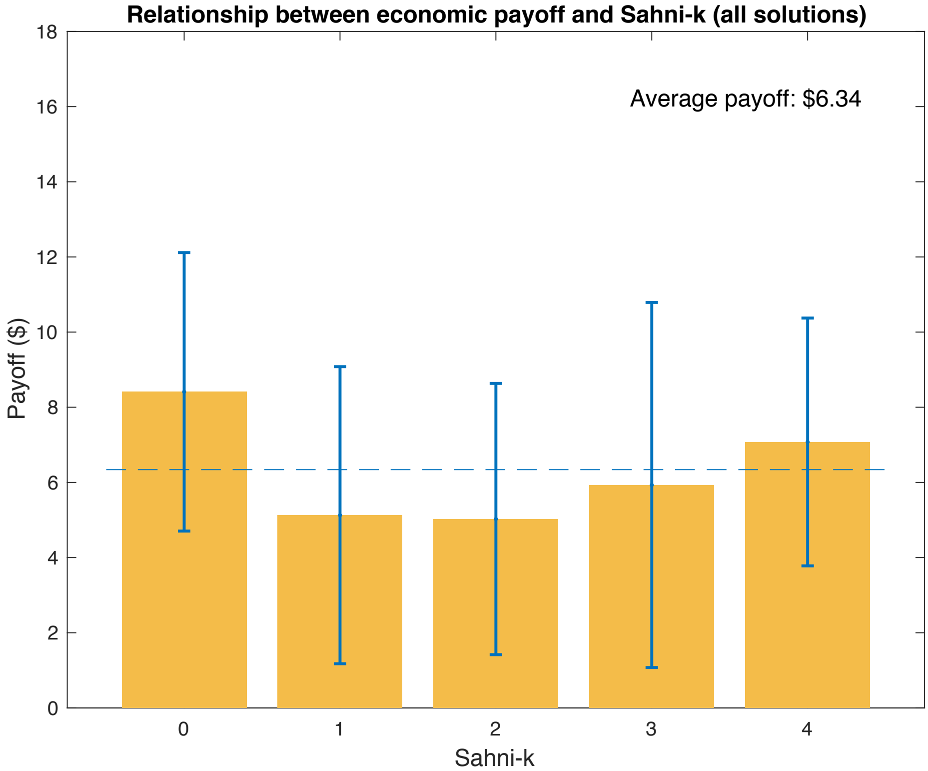
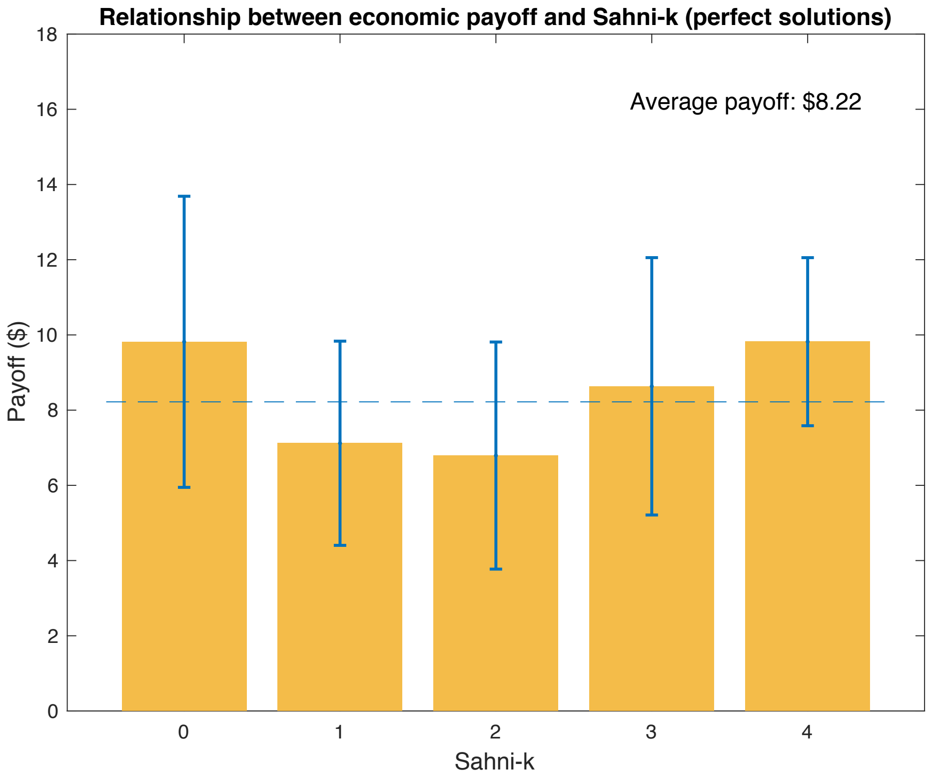


Figure 16: Pairwise t-tests between Sahni-k by instance (all solutions).

	0	**1**	**2**	**3**	**4**
0	–	***	***	***	***
1	–	–	0	0	***
2	–	–	–	0	***
3	–	–	–	–	***
4	–	–	–	–	–

**** Denotes a significant t-test between the specified Sahni-k pair with p-value < 0.001. If significant, it indicates the average payoff for that instance was significantly different from the average payoff of the other instance pair. The above table shows that average payoffs were significantly different if instance difficulty was Sahni-k 0 or 4, but not between other pairs of instance difficulty.*

Figure 17: Relationship between the average economic payoff from trading and Sahni-k instance difficulty (perfect solutions).



The average economic payoff at every level of Sahni-k instance difficulty, given the optimal solution was found. This is the same as a perfect score of 1. The average payoff for a perfect score across all levels of instance difficulty was $8.22.

Figure 18: Pairwise t-tests between Sahni-k by instance (perfect solutions).

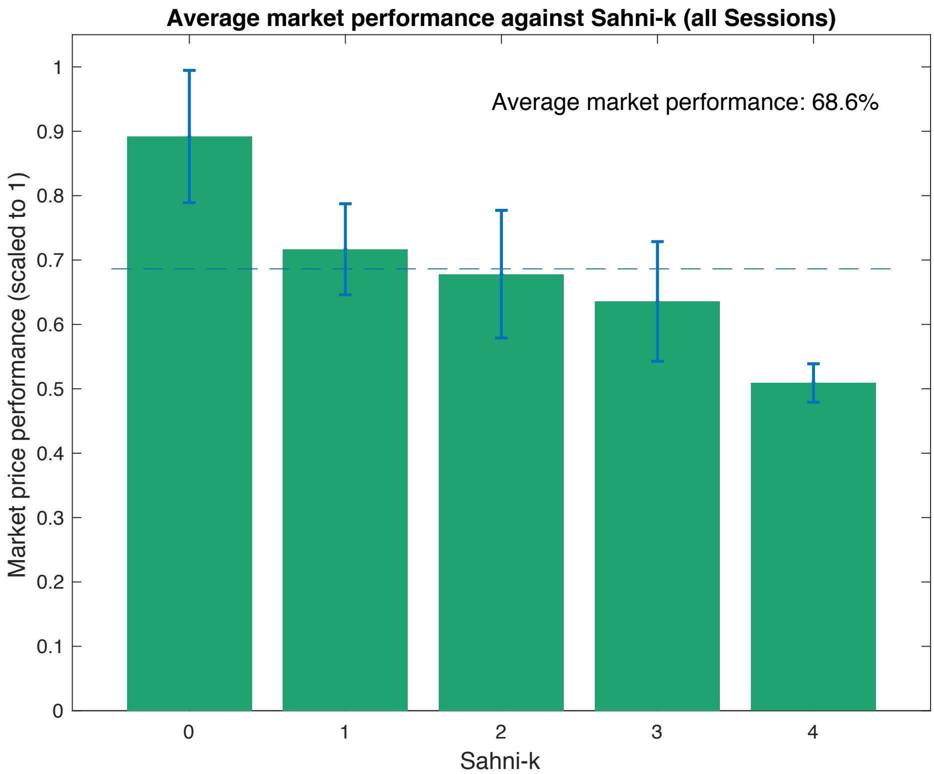
	0	**1**	**2**	**3**	**4**
0	–	***	***	***	***
1	–	–	0	0	***
2	–	–	–	0	***
3	–	–	–	–	***
4	–	–	–	–	–

**** Denotes a significant t-test between the specified Sahni-k pair with p-value $<$ 0.001. If significant, it indicates the average payoff for that instance was significantly different from the average payoff of the other instance pair. The above table shows that average payoffs were significantly different if instance difficulty was Sahni-k 0 or 4, but not between other pairs of instance difficulty.*

b. Alternative method for Result 2

We defined market performance using equilibrium prices, such that a correct item will yield a performance score of the price itself, and an incorrect item will yield a performance score of one minus its price. Then, we summed across all possible items and scale to 1. Hence, for market prices to perfectly reflect the optimal knapsack solution, it would yield a performance metric at 1.

Figure 19: Relationship between average market price performance and Sahni-*k* instance difficulty, using prices (all Sessions).



We found the average performance in a second market performance metric across all sessions was 89.2% for Sahni-k of 0, indicating that market prices made the correct decision of whether a particular item was correctly in or correctly removed from the optimal knapsack 89.2% of the time on average. Following that, average market performance was at 71.7%, 67.8%, 63.6% and 50.9% for Sahni-k 1 to 4 respectively.

c. On efficiency ratios

Relationship between item prices and their respective efficiency ratios, and the price variation across all sessions.

Figure 20: Price variation per item ordered by relative efficiency ratio for Sahni-*k* 0 (by instance).

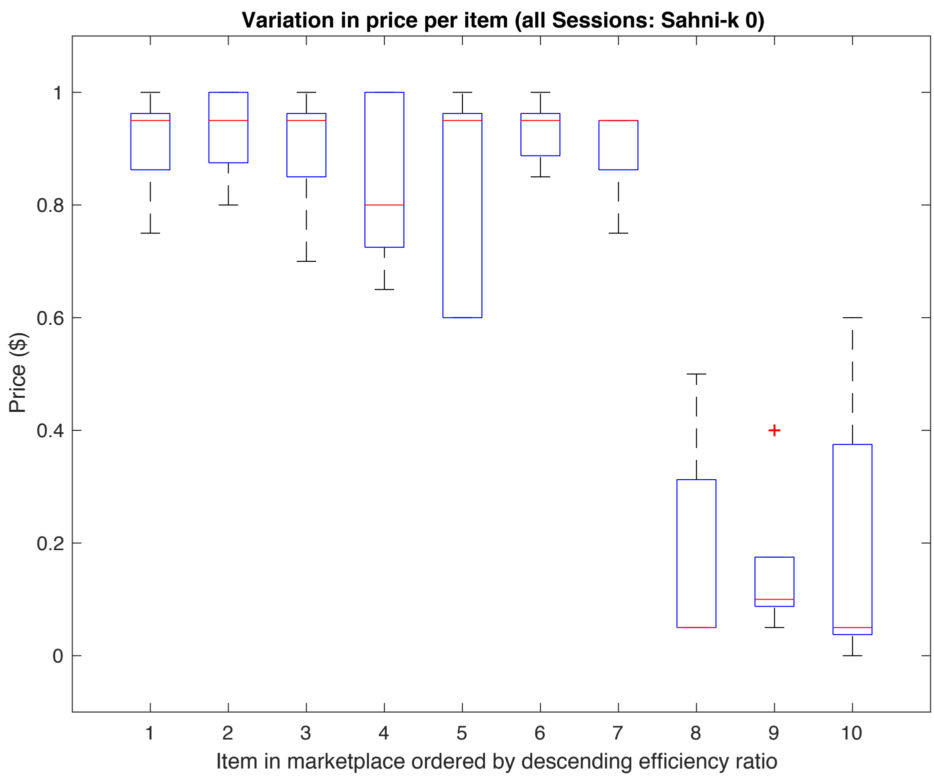


Figure 21: Price variation per item ordered by relative efficiency ratio for Sahni-k 1 (by instance).

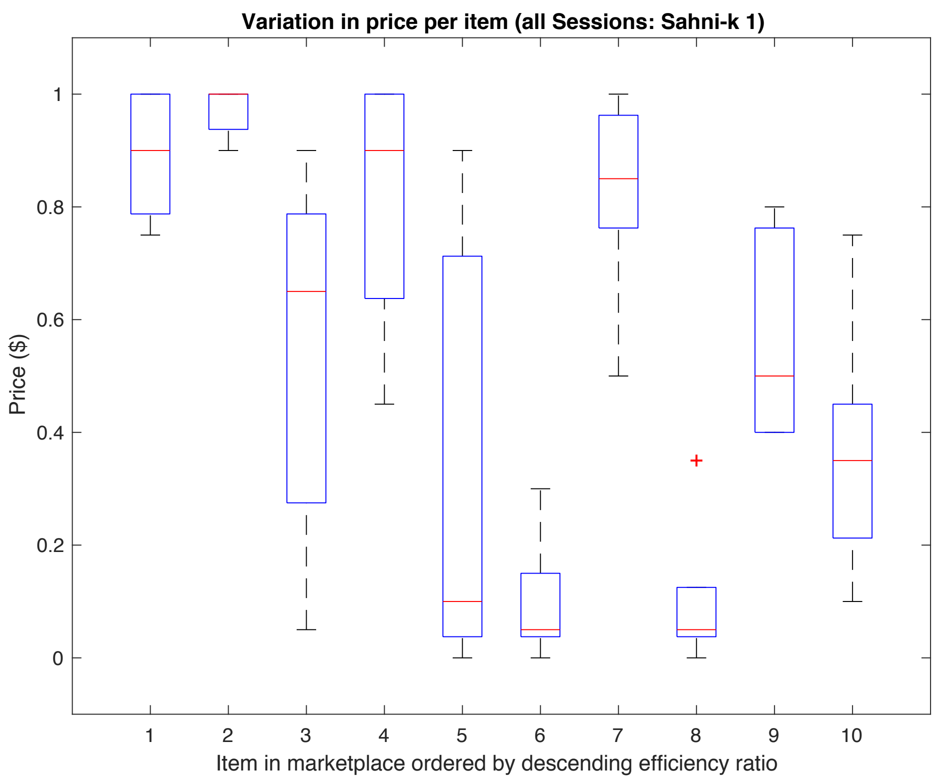


Figure 22: Price variation per item ordered by relative efficiency ratio for Sahni-k 2 (by instance).

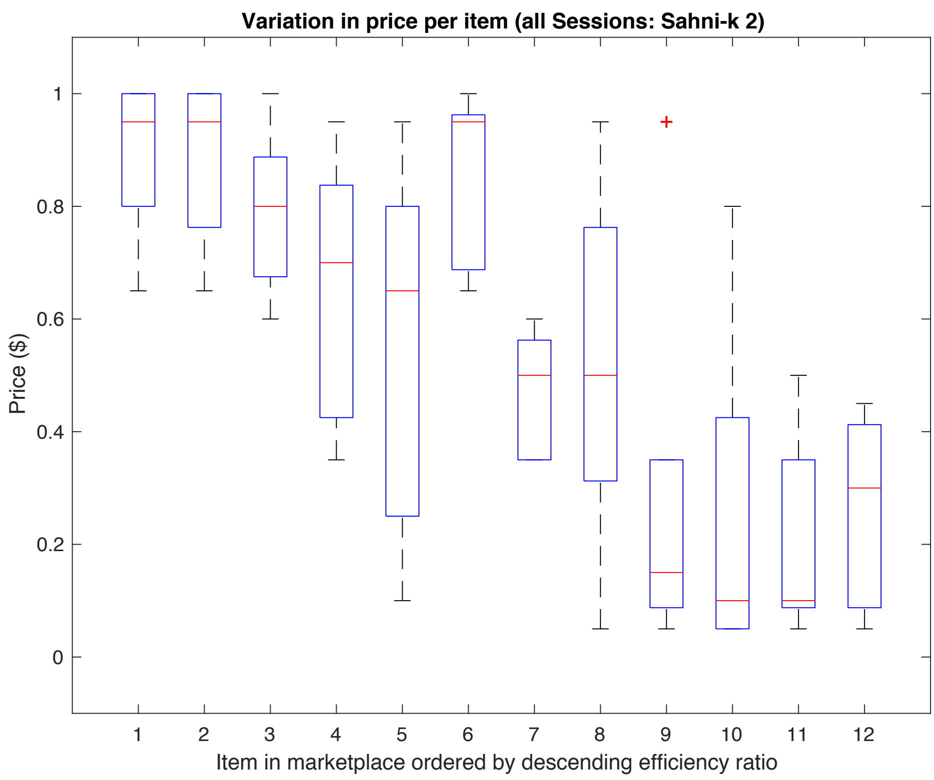


Figure 23: Price variation per item ordered by relative efficiency ratio for Sahni-k 3 (by instance).

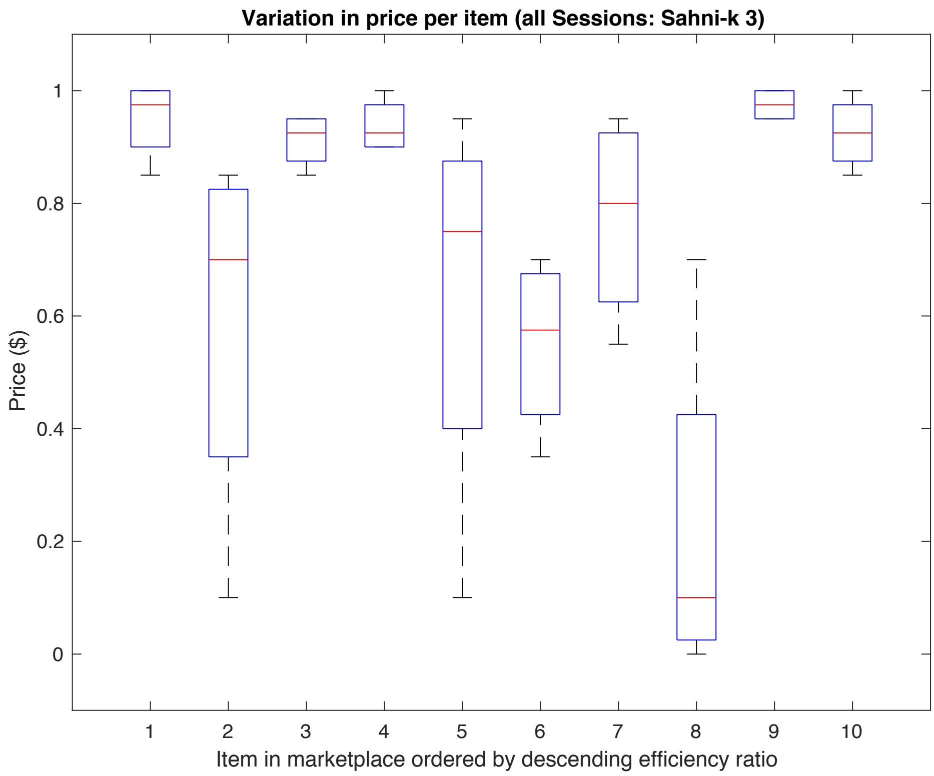


Figure 24: Price variation per item ordered by relative efficiency ratio for Sahni-k 4 (by instance).