

Understanding Market Functionality and Trading Success

James Schmitz^{*1} and David Rothschild^{†2}

¹New York University

²Microsoft Research

Abstract

PredictIt is a prediction market exchange that provides many markets linked to political issues from ongoing policy outcomes to political elections. The exchange caps accounts at \$850 in exposure for any contract with strict restrictions to one-account per person; in addition, there are high transaction and withdrawal fees. These combine to severely limit the upside to investment over the fixed costs: this ensures that traders are all retail investors. We have the individual-level data from two markets: Democratic and Republican Iowa Caucus in 2016. This data includes all orders and trades from every trader across the markets. We are able to fully reconstruct the markets and tie traders both within and between markets. We show that understanding how markets and trades works is more important to financial success than proxies for (1) confidence or funding (2) information or objectivity in trading. The work should be a call-to-action in simplifying trading for exchanges, their markets and contracts, aimed at retail investors.

KEYWORDS: prediction markets; probabilities; behavioral finance

^{*}james.schmitz@nyu.edu

[†]David@ResearchDMR.com

1 Introduction

PredictIt is a prediction market exchange that provides many markets linked to political issues from ongoing policy outcomes to political elections. The exchange was created before the 2016 election and quickly gained a robust set of users. The exchange caps traders at \$850 in exposure for any contract per Commodity Futures Trading Commission guidelines. The exchange vigorously identifies each user to ensure they only have one account. Further, all contracts are capped at 5,000 traders, but that was not an issue in the contracts we examined. In addition, the exchange charges users 5% of their wealth (gross, not net of investment) when withdrawing money, in addition to a 10% charge on all profitable trades. These restrictions certainly make this exchange unrepresentative of many other futures exchanges. It is impossible to justify institutional investors covering both the fixed cost of entering the exchange, and any given market, with these low limits and high fees. But this allows us to test something interesting: what can we learn about trading patterns in a market filled completely with retail investors?

We have the individual-level data from two markets: the Democratic and Republican Iowa Caucuses preceding the 2016 presidential election. This data includes all orders and trades from every trader across the markets. We are able to fully reconstruct the markets and tie traders between markets within and between markets. This task was both non-trivial and interesting. Exchanges save data in order to make sure they are functioning properly, but that is not necessarily the same data one would need to adequately research the individual-level trading patterns or the aggregate prices that emerge.

There is a lot of research on prediction markets doing a pretty good job in aggregating data into prices that somewhat accurately reflect the probability of upcoming events (Wolfers and Zitzewitz 2006; Rothschild 2009). Like all markets, the efficiency of the price depends on institutional investors being the marginal investor, driving the prices towards the efficient price. There is much less research on the price efficiency of markets that disincentivize institutional investors.

There are a selection of papers that examine the trading inefficiencies of prediction market traders in more open exchanges. Rothschild and Pennock (2014) examines the order books in both Intrade and Betfair in 2012 and shows how traders prefer to buy over sell, inefficiently pushing the sum of all bids in a set of mutually exclusive contracts (where one will be worth \$1 at the end of an event) much closer to \$1 than the sum of all asks. This potentially would have made it more valuable for traders looking to go long on a particular outcome to sell all other outcomes, but the paper did not have the trading data to examine this. Rothschild and Sethi (2016) has access to individual level trading data on Intrade and shows how most traders are exclusively long in one direction. But, both of these exchanges are not capped, and the papers do not address the competitive impact of these inefficiencies within the markets, between the traders.

Several papers have taken a detailed look the Iowa Electronic Market (IEM), which is interesting because it is the only other major exchange, besides PredictIt, that constrains wealth. That exchange limits traders to a \$500 initial investment in the exchange, which would make institutional investing difficult as well. Rhode and Strumpf (2008) investigates manipulation in the IEM in 2008. A key result

is that without wealth disparities it is hard for anyone to manipulate or drive the market. Another key paper (Schmitz 2011) shows the inefficiencies of traders from the perspective of a single large trader in the exchange. The research team at Iowa has also written many papers about their exchange Berg et al. (2008).

There are competing theories on why trades happen. First, various differences in risk or liquidity makes it so people with more money and patience can profit from other traders. In this model, funding is the key to better returns (Copeland and Galai 1983; Glosten and Milgrom 1985; Kyle 1985). Second, there can be heterogeneous priors or models. So people just interpret the world differently and money is made by the traders with better interpretation of the information (Harrison and Kreps 1978; Miller 1977). Neither sets of theories allow for differences in traders' understanding of the exchange, market, and specific assets being traded.

This paper is unique in that we have individual-level data and are constrained to retail investors who demonstrate a meaningful lack of understanding of the assets being traded. We use a proxy for information or objectivity to see if traders with better models of the data make more money than non-information traders. We use a proxy for funding or confidence to see if bigger, more funded traders make more money than smaller, less funded traders. This follows the two major theories of trading. Finally, we examine the efficiency of the traders in how they make trades, a proxy for how well they understand the actual workings of trading, and examine if that is better indicator of success than the other two, more traditional indicators. This is a comparison we have never seen examined in the literature.

These three variables are correlated, insofar as conditional on having one of these

attributes, the trader is slightly more likely to have any other attribute, and it is impossible for us disentangle that from our existing data. A trader who understands the working of markets well is also more likely to have money to invest and means to model information. We are not observing a trader with two of the variables, handing them the third variable, and making any claims about causation. But we have enough traders with two of the variables, and by varying the third, get a solid understanding of what correlates with better returns on investments.

We show that understanding how markets and trades works is more important in success than either proxies for objective models of information or funding. Within classes of traders who possess or do not possess a given trait, there is nothing more indicative to success in the markets than having the ability to spot the most efficient way to make a given wager. Objective models for information is also correlated with positive returns, but more funding does not necessarily correlate with higher returns. These results should be a call-to-action in simplifying trading for exchanges or contracts aimed at retail investors.

2 Data

The PredictIt team provided us with historical data for the 2016 Democratic and Republican Iowa Caucus elections. The data was organized into separate data files containing information on new orders, order modifications, and executed trades. The new order data contained unique trader IDs that allowed us to identify all of the orders sent by the same trader while maintaining each trader's anonymity. Using this

data we could study each trader’s activity individually to assess their profitability and trading style.

In addition to the data provided to us by PredictIt, we had top-of-book quote data collected every five minutes using PredictIt’s publicly available API. This data wasn’t used directly in this paper but it was still useful as it helped us evaluate the accuracy of some of our data cleaning efforts.

2.1 Basic Market Structure

It is important to understand the basic structure of the PredictIt markets and contracts. Each of these two markets included a finite set of mutually exclusive outcomes with one contract representing each candidate in the actual Iowa Caucus election. In order to wager in favor of one candidate, a trader has two choices. They could either buy Yes contracts for that candidate or buy No contracts for any other candidate. By definition, buying No contracts also provides the trader with a profitable outcome if any other candidates wins, not just the one candidate for which they would have bought a Yes contract for.

Understanding the difference between Yes contracts and No contracts and the decision-making process between trading one or the other is central to the thesis of this paper.

At market termination holders of each Yes and No contract will receive \$1 if the contract definition is satisfied and \$0 if it is not. In the Iowa Caucus markets, a Yes contract will be worth \$1 if the named candidate actually wins the caucus and \$0 otherwise. For a No contract it is the reverse: it will be worth \$0 if the named

candidate wins and \$1 if the candidate loses.

Traders are presented with a visible order book showing the demand to buy or sell each contract. All orders are priced in penny increments between \$0.01 and \$0.99.

The PredictIt exchange charges traders fees based on their trading activity. PredictIt collects a 10% fee on individual positions that are sold at a profit and no fee on positions that are sold at a loss. The exchange uses a first in, first out accounting model to identify which shares are sold and calculate their fee. Before fees these markets are essentially zero sum games with unprofitable traders losing money to profitable traders.

2.2 Data Cleaning

The Iowa Caucus data required extensive cleaning to prepare it for study. The first step was to standardize the format and column names between the two election datasets and adjust some inconsistencies in the time zones used to record the data's timestamps.

Our next step was to repair shortcomings in the data to facilitate our analysis. Our most important repair was to make adjustments to the sequence of events affecting each order so that the order's accounting is correct. For example, a common problem we found is that a new order that gets filled right away might have executions in the trade data file with timestamps that are a few hundred milliseconds before the timestamp reported in the new order data file. In this case we would adjust the timestamp of the new order to precede that of its executions. For each order we tested the output of our repair efforts to verify that the order's accounting and

sequence of events was in fact correct and was free of problems that would negatively impact our later code. Any failures were flagged for further scrutiny and knowledge gained resulted in improved code to repair the data, and eventually clean data to work with.

While performing these data repairs we found a few dozen data problems that could not be successfully resolved with our code without excessive complications. For example, there were some order modification messages that failed to modify their associated order, often because the order was filled immediately before the order could be modified by the PredictIt exchange. We also found a single overfill where an order executed 3 more shares than the order's quantity. Rather than make our data repair code more complicated to accommodate these situations, we manually assessed each occurrence and modified the data to resolve each issue in a sensible fashion.

We put a lot of effort into data cleaning because we wanted to be able to reconstruct the full order book for each contract. Having the full order book would have given us the ability understand what trading opportunities traders saw when they looked at the PredictIt website and made their trading choices. Reconstructing the order book requires a very clean dataset because even a single missing order event can cause impossible situations such as an order book that remains in a state with executable buy and sell orders but no reported matching executions.

Unfortunately we were not able to reconstruct the order book because a portion of the order cancellation events were not recorded. We investigated with help from the PredictIt team and learned that PredictIt had a nightly process that removed

infeasible orders when traders no longer had the cash to cover an order that would potentially put traders over the \$850 contract limit. For much of our dataset these infeasible cancellations were not recorded. The missing information could not be recreated and its absence undermined our ability to reconstruct the order book.

Since the order information was flawed we developed an alternative approach to reconstruct the top-of-book prices using only the executed trade data. We know the times and prices of all of the executions and we also know the times the orders were sent. For any matching trade we can figure out which order provided liquidity because it will be the order that was placed in the order book first. Using the prices of the orders that provided liquidity we can infer the top-of-book price on one side of the order book at the time of each execution.

Using these inferred top-of-book prices we can make an assumption that these prices remain at the top-of-book until the next provide execution gives us new information about the top-of-book price. This, combined with the knowledge that the bid-ask spread must always be one penny or more, allows us to reconstruct a complete and usable set of top-of-book prices for all contracts over the life of the market.

An important question is how well this procedure approximates the top-of-book prices that would be derived from an ideal dataset. To evaluate this we can compare these prices with the five minute sampled data we collected from the PredictIt API. We performed this comparison and found that it does provide a good reconstruction of the top-of-book prices. It performs best during periods of peak trading activity, which is ideal for this research project.

2.3 Margin Linking

There was an important change to the mechanics of the PredictIt markets that took place towards the end of 2015, and that change is the introduction of something called “margin linking.” To understand what this means, first start with the PredictIt trading system deducting funds from a trader’s account whenever they make a trade to cover a possible loss in the traded contract. This is basic accounting, but now consider a trader that buys multiple No contracts that are linked together in a market such as the Iowa Caucuses that allows only one candidate to be the winner. In this case it simply isn’t possible for all of the trader’s No contracts to lose money; at most only one of the No contracts can lose money. Therefore if the PredictIt trading system were to deduct the total cost of each of the No contracts the system would be deducting more money from the trader’s account than they could lose in any possible outcome of the market.

The introduction of margin linking changed the PredictIt exchange so that it considered the maximum possible loss a trader could suffer across multiple positions in contracts in a linked market. Before margin linking the PredictIt exchange would deduct funds from a trader’s account without considering linked outcomes; after margin linking they considered linked outcomes and deducted only the maximum possible loss the trader could actually suffer. This new feature was rolled out to markets gradually over several weeks and resulted in additional tradable funds being added to some traders’ accounts. Traders were informed of this change several weeks in advance and market prices started to adjust in advance of the margin linking rollout.

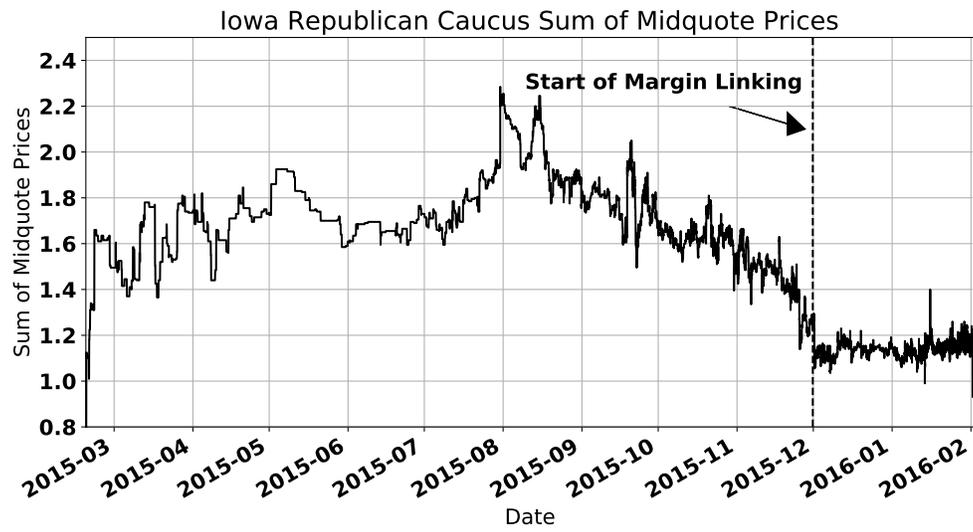


Figure 1: Iowa Republican Caucus Sum of Midquote Prices

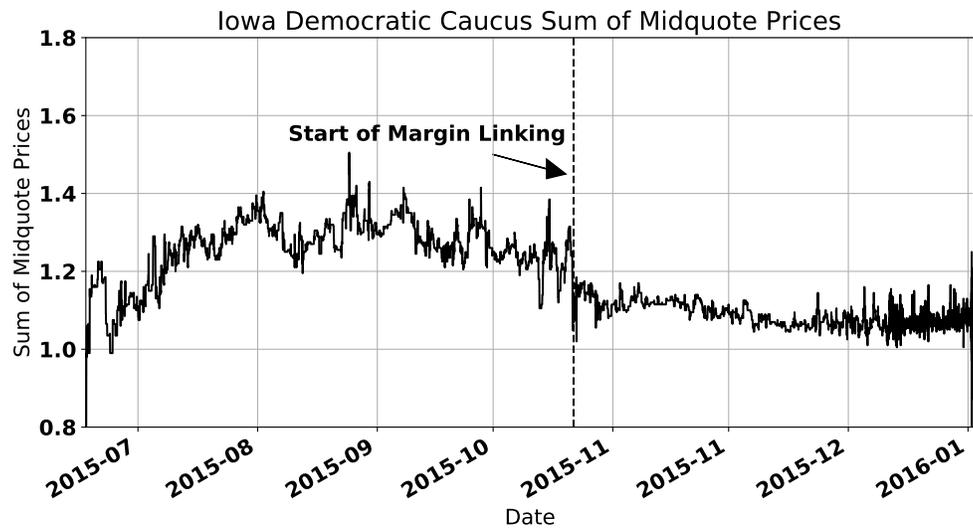


Figure 2: Iowa Democratic Caucus Sum of Midquote Prices

The impact on prices can be seen by looking at the sum of the midquotes (i.e., average of bid and ask) for all of the contracts available in a market, shown in Figures 1 and 2. Before margin linking went into effect in the Republican Caucus market, the sum of the midquote prices peaked at over \$2. Consider that in a more well functioning prediction market the sum of the midquote prices should be close to \$1. After margin linking was added on December 1, 2015, the sum decreased to the \$1.10 to \$1.20 range. Margin linking went into effect in the Democratic Caucus market on October 22, 2015 and had a similar impact on prices. In both markets traders were able to push down prices using the additional funds that were added to their accounts. This had the effect of making prices more fair for traders buying Yes contracts.

While it is tempting to start our study after margin linking, it would be impossible as many of the users already owned contracts and we cannot ignore that in describing further traders. And, the elections are not that much later than the margin linking, thus we need to consider the full time-period, including the changed rules, in this study.

2.4 General Market Activity

We can explore our cleaned market data to gain an understanding of the general market activity of this market.

The Republican and Democratic Iowa Caucus markets started in February and June of 2015, respectively. Both markets concluded on February 2nd, 2016, with the bulk of the trading activity taking place during the days and weeks leading up to

the actual election day.

As shown in Table 1, most of the trading activity in the Democratic Caucus was for candidates Sanders and Clinton. For most of the duration of this market the price of the Clinton security suggested that she was the likely winner; ultimately she did in fact win the Iowa Democratic Caucus. Figure 3 shows the security midquote prices for both candidates with the price of the Clinton security above \$0.50 most of the time.

	Shares Traded	Notional Traded (\$)
Sanders	1,327,990	659,721.16
Clinton	1,223,458	568,022.08
O'Malley	44,048	23,622.64
Biden	41,566	18,150.80
Webb	5,328	2,288.70
Chafee	8	5.96

Table 1: Iowa Democratic Caucus Total Trading Activity by Candidate

In the Republican Caucus there were more candidates and there was more uncertainty about who would win. Table 2 shows that candidates Trump and Cruz had the highest trading volumes but other candidates such as Paul, Rubio and Carson also attracted a lot of trading volume. In the months leading up to market termination the prices shown in Figure 4 suggest that first Cruz and later Trump was the likely winner. Cruz ultimately won the election but prices suggested that Trump was still expected to win up until a few hours before the market concluded. The higher

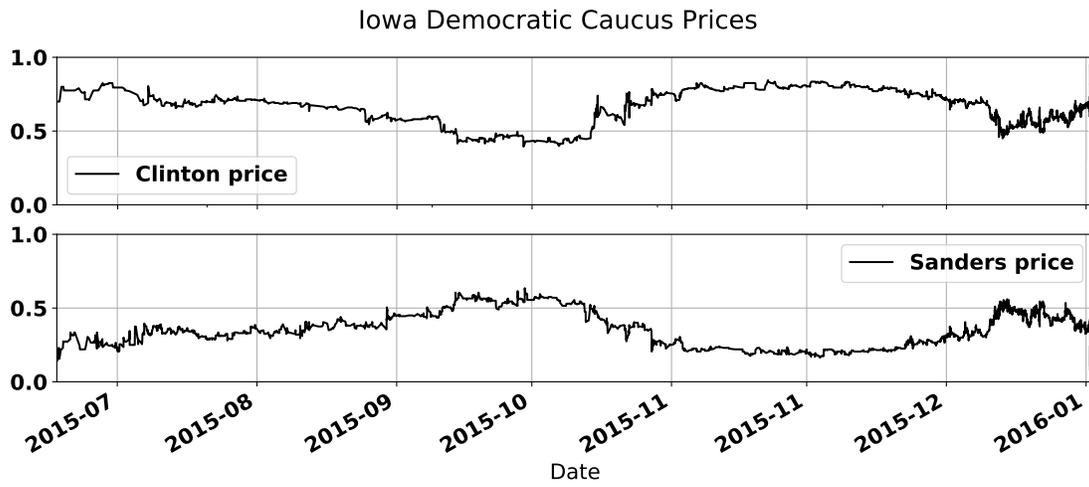


Figure 3: Iowa Democratic Caucus Prices for Clinton and Sanders

uncertainty about the outcome of this market impacted trader profitability and the results of this paper.

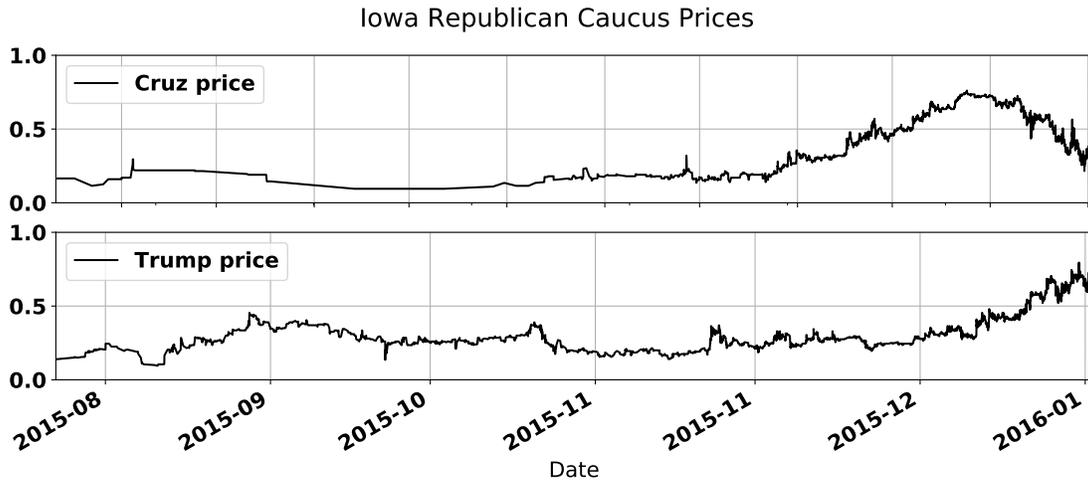


Figure 4: Iowa Republican Caucus Prices for Clinton and Sanders

	Shares Traded	Notional Traded (\$)
Trump	1,504,976	755,815.48
Cruz	1,315,862	663,317.44
Paul	834,450	421,771.84
Rubio	957,920	411,231.44
Carson	685,058	315,323.40
Huckabee	182,680	76,436.52
Bush	168,766	67,130.60
Christie	101,940	49,189.62
Santorum	78,572	39,330.02
Fiorina	69,442	32,976.98
Kasich	59,502	30,317.42
Jindal	35,196	16,366.44
Walker	17,804	10,900.46
Perry	4,302	2,182.70

Table 2: Iowa Republican Caucus Trading Activity by Candidate

Millions of dollars and shares were traded during the lifetime of both markets. Table 3 shows that the Republican market attracted more than twice the trading activity than the Democratic market and that PredictIt collected more than twice the fees.

	Shares Traded	Notional Traded (\$)	PredictIt Fees (\$)
Democrat	2,642,398	1,271,811.34	11,718.86
Republican	6,016,470	2,892,290.36	23,096.63

Table 3: Summary of Total Market Activity in Iowa Caucus Markets

The majority of the trading activity took place right before market termination. Table 4 shows the daily trading volume in shares for select dates, along with the percentage of cumulative trading volume by date. It is clear from this table that in both markets more than half of the volume took place in the last few days before market termination.

	Democratic		Republican	
	Volume	Cum. Volume (%)	Volume	Cum. Volume (%)
2015-07-01	572	0.1649	430	0.1914
2015-09-01	608	2.31	1,894	0.9952
2016-01-01	6,972	12.69	13,840	23.63
2016-01-13	16,234	17.90	46,592	30.63
2016-01-20	25,902	23.78	39,520	36.03
2016-01-27	35,760	35.10	109,450	45.91
2016-01-28	38,394	36.55	116,776	47.85
2016-01-29	60,574	38.84	194,520	51.08
2016-01-30	65,428	41.32	237,088	55.02
2016-01-31	83,936	44.50	267,734	59.47
2016-02-01	210,808	52.47	504,614	67.86
2016-02-02	1,255,814	100.00	1,933,852	100.00

Table 4: Market Activity by Date. The Volume column shows the total shares traded on each day and the Cum. Volume % column shows the percent of total volume that took place on or before that date.

2.5 Trader Behavior Summary

After all of the data was cleaned we were able to analyze trader behavior and calculate profitability metrics.

Table 5 provides a topline summary of trader activity in the two markets. There

are 3,750 and 4,452 traders who executed at least one share in the Democrat and Republican markets, respectively. Note that “Average (pre-fee) Profit” is zero by definition because before fees this market is a zero sum game. After fees the average trader suffered a small loss, with the range of profits or losses within a few thousand dollars. Also note that “Money Risked” is the total amount of money that was exposed in the market (only counting the same money once) and contributed to the zero sum game by each trader at any time during the life of the market.

	Democrat	Republican
Traders	3,750	4,452
Orders Sent	33,959	62,551
Money Risked (cumulative exposure, \$)	417,251.70	544,205.11
Average (pre-fee) Profit (\$)	0.00	0.00
Average (net-fee) Profit (\$)	-3.13	-5.19
Min (net-fee) Profit (\$)	-1,699.58	-1,701.32
Max (net-fee) Profit (\$)	1,187.54	2,882.00

Table 5: Summary of Trader Activity in the Iowa Caucus Markets

We classified traders by three key metrics: their maximum market exposure (proxy for funding or confidence), willingness to open and close contracts multiple times (proxy for information or objectivity), and if they made efficient trades (traded in the most cost-effective way to make a given wager).

The quantity of a trader’s exposure says something about how serious they are, confident they are, and how well funded they are. Larger traders are more likely

to be risk neutral and lack liquidity constraints, they they are less vulnerable than small traders to lose money on trades due to monetary objectives beyond maximizing future expected wealth. We simply characterize traders as “small” or “large.” We use \$50 in exposed risk as the boundary between the two groups, but this paper’s results are robust to adjusting that value.

Another way to classify traders is whether or not they actively traded into and out of positions on a single contract or if they followed a passive buy-and-hold approach. This is a good proxy for an information trader who is trusting their trading skills rather than trading on their personal political views or naive interpretation of the available information. We use a simple threshold: did the trader ever close a position and re-open it on a single contract? This includes traders that switched the direction (from Yes contracts to No contracts, or vice-versa) of their position and traders that closed and re-opened positions in the same direction. We classify that trader as an active trader, and the others as inactive traders, because they are more frequently adjusting their position. This is a binary classification, so there is no way to test the robustness, but we believe that any trader that ever changes their position is, at minimum, not a fully partisan trader and is providing meaningful information above the current price (or they would buy and hold). Further, with the ups and downs of the contracts we believe that it is credible to assume that a rational trader should not be long one position for the entire duration of the contest, but, of course, we will miss some traders who calculated it was not worth their time to trade beyond their initial informed trade. Thus, our classification is a precise definition insofar as most of the people we classify as active are informed, but there will be some recall issues,

with some informed traders in the inactive group.

A trader is classified as efficient if they were able to identify the most cost-effective way to make a given wager. Precisely we demand that at least 75% of their trades are efficient. As we will show later in this paper, the availability of Yes and No contracts in these markets provides traders with two trading options because Yes contracts can be replicated with No contracts and No contracts can be replicated with Yes contracts. The two options were rarely equivalent, as one was frequently more cost-effective than the other. An efficient trader is one who is able to make the most cost effective choices. This is explained in detail in subsection 3.4. Again this paper's results are robust to adjusting the 75% value.

Table 6 shows that in both markets most traders are small and inactive. With activity being a binary variable, it is impossible for us to consider more comparably sized groups (we will later show that the vast majority of traders are also inefficient). Trading theory assumes that small and inactive traders would lose money to large and active traders. As large active traders, without liquidity constraints and using optimized models, would take money from small and inactive traders who are potentially liquidity constrained (due to non-expected return maximizing trades) and using inefficient or biased ways to analyze available information (which are simply less precise than the models for active traders).

	Democratic		Republican	
	Large	Small	Large	Small
Active	295	252	413	275
Inactive	764	2,439	980	2,784

Table 6: Number of traders by volume and activity in both markets.

In both markets the majority of traders exclusively traded Yes contracts. Across both markets, just 17 percent traded only No contracts and an additional 22 percent traded both No and Yes contracts. The remaining 61 percent traded only Yes contracts. Thus, with so many traders only trading Yes contracts and with the structural benefits of No contracts, it is not surprising that there are often-times better trading opportunities by transacting No contracts instead of Yes contracts.

The next section walks through the methods of evaluating if traders chose the most efficient option for any given trade and if they could be classified as efficient or not.

3 Dominant Contracts

3.1 Hypothetical Market

Consider a hypothetical market within PredictIt with n mutually exclusive outcomes. The market provides Yes and No contracts for traders to trade. Each possible outcome has a Yes and No contract with a payoff that is linked to whether or not that outcome is the one that is realized at market termination. A Yes contract is worth

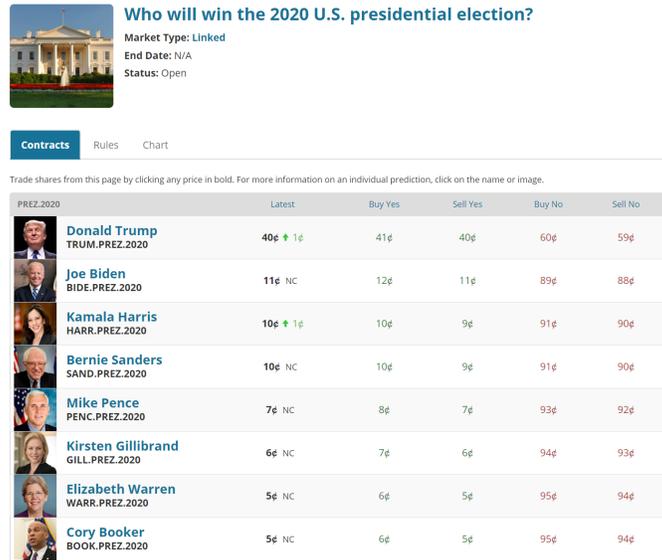


Figure 5: Sample market from PredictIt

\$1 if and only if its outcome is realized and \$0 otherwise. A No contract is worth \$1 if and only if its outcome is not realized and \$0 if it is. Before market termination the price of any Yes or No contract is always greater than (or equal to) \$0 and less than (or equal to) \$1. Figure 5 is image of an actual market in PredictIt where you can see a series of mutually exclusive possible outcomes and a set of Yes and No prices.

We are going to start with two simplifying assumptions that we will later relax: (1) no bid/ask spread and (2) no transaction costs. There are Yes contracts Y_i and No contracts N_i for all i in $i = 1 \dots n$. Since there is no bid/ask spread, $Y_i = 1 - N_i$ and $N_i = 1 - Y_i$. Since there are no transaction costs, the profit from a Yes contract is $1 - Y_i$ and the profit from a No contract is $1 - N_i$. In the real market, PredictIt charges a 10% fee on gains that reduce the profits to $(1 - 10\%)(1 - Y_i)$ or $(1 - 10\%)(1 - N_i)$.

And as Figure 5 shows there is a spread between the cost to buy and sell a given contract. Since prices are in penny increments, the spread between the best bid and best ask prices must be a minimum of 1 penny wide.

In this market the sum of the Yes contract prices $\sum_{i=1}^n Y_i$ may or not be equal to \$1.

This market also features margin linking, where the exchange considers the mutual exclusivity of the outcomes and the fact that the maximum possible loss a trader can suffer may be less than the total cost of all of the contracts traded. This possibility is clearly seen when a trader buys multiple No contracts. Because only one outcome $i = 1..n$ will be realized, only one No contract can lose money; the other No contracts will make money. The amount of money the exchange deducts from each trader's account is equal to the maximum possible loss the trader can suffer for any outcome given all of the contracts in their possession.

Consider a trader who thinks that the Yes contract linked to outcome a is underpriced. The trader can make an investment in 1 share of Yes contract a for a price of Y_a . The trader's profit if outcome a is realized is $1 - Y_a$. The trader will suffer a loss of Y_a if outcome a is not realized. Alternatively, the trader can replicate the payoff of $1 - Y_a$ if a is realized by trading c contracts of every No contract except for the one linked to outcome a , where

$$c = \frac{1 - Y_a}{\sum_{i=1, i \neq a}^n 1 - N_i} \quad (1)$$

With this set of No contracts the payoff if a is realized is

$$\sum_{i=1, i \neq a}^n c(1 - N_i) \quad (2)$$

It is easy to show that this is equal to $1 - Y_a$ if outcome a is realized.

Because of margin linking, the trader's cost for trading this set of No contracts is always equal to the maximum possible loss. The trader's loss if some other outcome b is realized is

$$cN_b - \sum_{i=1, i \neq a, b}^n c(1 - N_i) \quad (3)$$

The No contract for outcome b suffers a loss but the others in the traded set earn a profit of $1 - N_i$ per share. It can be shown that the net result for outcome b and all the other outcomes $i = 1..n, i \neq a$ are equal to each other. We can see this by setting two loss equations equal to each other and simplifying (shown in Appendix A).

Now, when is it a better choice to buy this set of No contracts instead of a Yes contract? The optimal choice is the one that can be purchased for the lowest cost but with the same payoff if outcome a is realized.

With the first simplifying assumption, we apply the pricing relationship of $Y_i = 1 - N_i$ to convert all of the variables to prices of Yes contracts. Thus, Yes contracts are more expensive for the same return if:

$$Y_a > c(1 - Y_b) - \sum_{i=1, i \neq a, b}^n cY_i \quad (4)$$

with

$$c = \frac{1 - Y_a}{\sum_{i=1, i \neq a}^n Y_i} \quad (5)$$

Which simplifies to:

$$\sum_{i=1}^n Y_i > 1 \quad (6)$$

Buying a replicating set of No contracts is a better choice when the sum of the Yes contract prices is greater than \$1. Our empirical observations of this market in Figures 1 and 2 shows that most of the time this was the case. The impact this had on the trader's transactions is explored in the next section. As this proof suggests, there were many trades for Yes contracts that were more expensive than an equivalent trade executed with No contracts.

Now consider a trader that makes an investment of 1 share of No contract a for a price of N_a . The trader's profit if a is not realized is $1 - N_a$. The trader will suffer a loss of N_a if outcome a is realized. Alternatively, the trader can replicate the payoff of $1 - N_a$ if a is not realized by trading d contracts of every Yes contract except for the one linked to outcome a , where

$$d = \frac{1 - N_a}{1 - \sum_{i=1, i \neq a}^n Y_i} \quad (7)$$

With this set of Yes contracts the payoff if another outcome b is realized is

$$d(1 - Y_b) - d \sum_{i=1, i \neq a, b}^n Y_i \quad (8)$$

This is equal to the payoff $1 - N_a$.

The trader's loss if outcome a is realized is equal to the sum of the Yes contracts purchased. This is the maximum possible loss.

$$d \sum_{i=1, i \neq a}^n Y_i \tag{9}$$

The formula for d in equation 7 contains a potential problem when replicating a No contract for outcome a . The numerator will be negative or zero when

$$1 \leq \sum_{i=1, i \neq a}^n Y_i \tag{10}$$

This situation will occur in the trading history because $\sum_{i=1}^n Y_i$ was frequently larger than \$1 and there were some candidates who were unlikely to win, with Yes contract prices that were close to \$0. Therefore the sum of Yes contract prices excluding just one contract $\sum_{i=1, i \neq a}^n Y_i$ was also sometimes greater than or equal to \$1. This will cause d to be undefined or negative.

If d is undefined replicating the payoff of the No contract with other contracts is not possible. If d is negative then the formula is directing the trader to buy a negative number of Yes contracts. This can be done by instead buying $-d$ No contracts. This actually does work and can correctly replicate the payoff of the No contract. Interestingly, the cost of that replicating set of No contracts will also be negative. This means that when this trading opportunity presents itself it is always prudent of the trader to trade the replicating set of No contracts instead of the original No contract.

Another potential problem with d in equation 7 is if the numerator is close to zero. When that happens $|d|$ will become very large and in reality the replicating set of contracts may become untradable given PredictIt's trading limits on individual contracts. In the dataset this situation occurs only a handful of times but it must be articulated for thoroughness. To manage this issue we constrain valid values of d to $|d| < 10$. This is a conservative limit that will help us later when we want to make statements comparing the actual trades that took place and the alternative replicating set of contracts that the trader could have traded instead.

To summarize, if d is defined and $|d| < 10$ the payoff of a No contract might be replicated by either a set of Yes contracts or a set of No contracts. When is it a better choice to buy the replicating set of contracts instead of the No contract? As before it depends on which choice has the lowest cost for the same payoff. The No contract is more expensive for the same return if:

$$N_a > d \sum_{i=1, i \neq a}^n Y_i \tag{11}$$

with

$$d = \frac{1 - N_a}{1 - \sum_{i=1, i \neq a}^n Y_i} \tag{12}$$

This is easily reduced to:

$$1 > \sum_{i=1}^n Y_i \tag{13}$$

Buying the replicating set of contracts is a better choice when the sum of the

Yes contract prices is less than \$1 (which makes sense with our previous finding on when it is better to buy No versus Yes contracts). Our empirical observations of this market in Figures 1 and 2 shows that this was rarely the case.

Using this simplified model we evaluate if the Yes contract trades are strictly dominated by a replicating set of No contracts. To test this empirically we use the prevailing midquote at the time of each Yes contract buy trade and compared it to the maximum possible loss of a replicating set of No contracts, priced using the prevailing midquotes of the No contracts. We found that for the trades that took place after margin linking was initiated in the Democratic and Republican caucus markets the replicating set of No contracts dominated the Yes contract trades 96.1% and 99.7% of the time, respectively. This means that the majority of the time the No contracts were the cheaper trade. We limited this to only the trades after margin linking was initiated because the earlier trades are less interesting because of the pricing issues shown in Figures 1 and 2.

Scatter plots of the data are shown in figure 6. A black diagonal line is included to clearly show the break-even point where the cost of the Yes contract and the cost of the replicating set of No contracts are equal. Trades that appear below the line are trades that are dominated by the set of No contracts.

We also analyzed the No contract buy trades in both markets and compared them to the replicating set of Yes or No contracts. We found that after margin linking was initiated in the Democratic and Republican Caucus markets, the No contract trades were replicatable (meaning d is defined and $|d| < 10$) and dominated by the replicating set of Yes contracts only 2.9% and 0.2% of the time, respectively, and by

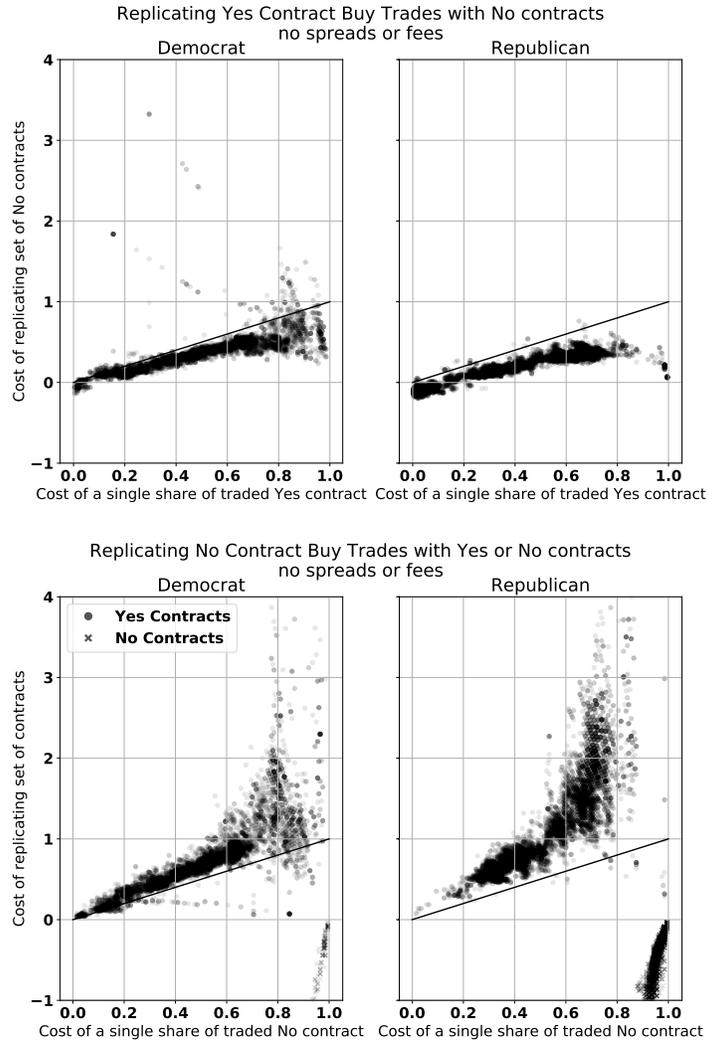


Figure 6: Top: Buy Yes trades plotted against the optimal replicating set of No contracts. Bottom: Buy No trades plotted against the optimal replicating set of Yes contracts. All contract trades are priced with the prevailing midquote at the time of the trade. Contract returns do not include PredictIt's 10% fee.

the replicating set of No contracts 1.3% and 41.3%, respectively. These percentages mean that again the majority of the time the No contracts were the cheaper trade. In the Republican market the No contract trades were frequently dominated by a replicating set of No contracts, suggesting that No contracts were still the better option but the traders should have traded different No contracts than the one that actually executed.

Here we are considering only the trades that took place after margin linking was initiated because before margin linking market prices were much more distorted and the results would be less interesting. Additionally, all current PredictIt markets use margin linking, making our results applicable to what a current observer would find in a market today.

3.2 Including Market Spreads

The equations for c and d change slightly if the hypothetical market's prices include bid-ask spreads. If the trader crosses the spread to purchase Yes or No contracts linked to outcome a , they are buying contracts at Y_a^{ask} or N_a^{ask} . These trades could each be replicated by crossing the spread and buying a set of No or Yes contracts, respectively. The formulas for c and d become:

$$c = \frac{1 - Y_a^{ask}}{\sum_{i=1, i \neq a}^n 1 - N_i^{ask}} \quad (14)$$

$$d = \frac{1 - N_a^{ask}}{1 - \sum_{i=1, i \neq a}^n Y_i^{ask}} \quad (15)$$

If the numerator in equation 15 is negative then one must instead calculate a replicating set of No contracts with this formula:

$$d' = \frac{1 - N_a^{ask}}{\sum_{i=1, i \neq a}^n Y_i^{bid} - 1} \quad (16)$$

The numerator in equation 16 may again be negative. If this is the case then there is no replicating set of contracts for that trade.

Our approach becomes more difficult in the case of providing liquidity. If a trader purchases a Yes or No contract by providing liquidity, they had to place an order in an order book and wait for another trader to sell contracts at that price. A trader could also do this for a replicating set of contracts, but it would be uncertain if the trader would be able to execute all of their orders at those prices because the trade is dependent on the arrival of a willing counter-party. For the purpose of our analysis, and our argument of No contracts generally dominating Yes contracts, we will avoid this complexity by instead comparing the price trader's provide for a trade with the price of the replicating set of contracts if the trader were to instead cross the spread and execute their trades immediately. This is a conservative assumption that results in a higher cost for the trader than the reality of potentially trading at a better price by providing liquidity, but this avoids the uncertainty of whether or not the trader would be able to get filled providing liquidity in the set of contracts at the desired prices.

With this assumption, the formulas for c and d become:

$$c = \frac{1 - Y_a^{bid}}{\sum_{i=1, i \neq a}^n 1 - N_i^{ask}} \quad (17)$$

$$d = \frac{1 - N_a^{bid}}{1 - \sum_{i=1, i \neq a}^n Y_i^{ask}} \quad (18)$$

If the numerator in equation 18 is negative then one must instead calculate a replicating set of No contracts with this formula:

$$d' = \frac{1 - N_a^{bid}}{\sum_{i=1, i \neq a}^n Y_i^{bid} - 1} \quad (19)$$

The numerator in equation 19 may again be negative. If this is the case then again there is no replicating set of contracts for that trade.

The empirical results changed by a small amount after our analysis included the bid-ask spreads of the contract prices. We analyzed the post-margin linking trades that took liquidity separately from the trades that provided liquidity. For the trades that took liquidity, the Yes contract trades were dominated by the replicating set of No contracts 87.8% and 96.7% of the time in the Democratic and Republican markets, respectively. For the provide trades, the percentages are 66.9% and 94.8%.

Our analysis of the No contract trades continued to show that these trades were rarely dominated by a replicating set of Yes contracts. For the Democratic market, after margin linking went into effect the No contract trades that took liquidity and provided liquidity were replicated and dominated by a set of Yes contracts 1.7% and 0.9% of the time and by a set of No contracts 1.3% and 0.6% of the time, respectively. For the Republican market, after margin linking went into effect the

No contract trades were replicated and dominated by the set of Yes contracts 0% of the time. The No contract trades were replicated and dominated by the replicating set of No contracts when taking and providing liquidity 39.2% and 24.1% of the time, respectively. Again, the No contracts were still the better option but frequently the traders should have traded different No contracts than the one that actually executed.

3.3 Including Market Spreads and Fees

Now we relax both assumptions and consider the real PredictIt market that has bid-ask spreads and charges a 10% fee on all profits for any trade.¹ Now, because of the fee, the trader's payoff for any profitable No contract trade is reduced by an amount that is proportional to $1 - N_i^{ask}$. As a result, the number of shares to buy for each No contract to replicate the payoff of a Yes contract is no longer a constant value c . The number of shares of each contract needs to be increased or decreased slightly for the trader to trade optimally. The optimal trade will minimize the maximum possible loss for any No contract, minimizing the amount of money deducted from the trader's account by PredictIt to cover the margin linking.

It is helpful to illustrate this with a concrete example. Consider a market with contract prices as shown in the top of Table 7 and a trader that wants to use No contracts to replicate the returns of a single Yes contract for candidate A. If the trader bought 0.942 shares of No contracts for candidates B and C using (14), the

¹Recall that the PredictIt exchange uses a first in, first out accounting model to calculate fees. Fees are based on profitable trading of individual shares, not the net profitability over a day or some other time period. If Trader A, in Market Z, sells a contract for a \$5 profit, then loses \$15 in the same market the next hour, they would be charged \$0.50 on successful profit, even though they later lost \$10 that day.

net loss if candidates B or C wins will be slightly different from each other because of the fee charged for the one profitable No contract. Because of margin linking, the PredictIt exchange will use the maximum possible loss to determine how much money to deduct from the trader's account. In this case the amount is 0.348 to cover the loss if candidate C wins.²

²Note that if there was a fourth candidate D and the trader went long on candidate A by buying Yes for A, then they lose, but if they went long on candidate A by selling No for B and C, they still wins with a victory by D.

				Market Prices			
				Yes		No	
				Contract		Contract	
				Bid	Ask	Bid	Ask
Candidate A	0.34	0.35	0.65	0.66	0.40	0.41	
Candidate B	0.59	0.60	0.89	0.90	0.942	-0.292	
Candidate C	0.10	0.11	0.942	-0.292	0.902	-0.348	

Not adjusted for fees				Adjusted for fees			
-----------------------	--	--	--	-------------------	--	--	--

		Shares of	Payoff if X wins				Shares of	Payoff if X wins	
		No contracts	w/o fee	w/ fee			No contracts	w/o fee	w/ fee
A	0.650	0.585			A	0.650	0.585		
B	0.942	-0.292	-0.301			B	0.949	-0.299	-0.308
C	0.942	-0.292	-0.348			C	0.902	-0.252	-0.308

Table 7: Contract payoffs for a replicating set of No contracts with the number of No contracts adjusted or not adjusted to consider fees. Observe on the left that when the number of No contracts purchased for candidates B and C is the same the loss if candidates B or C win is the same before fees but is different after fees. The number of No contracts purchased can be slightly adjusted to make the losses equal if candidates B or C win, as shown on the right.

This isn't optimal, however. By slightly increasing the number of No contracts for candidate B and decreasing the number of No contracts for candidate C the trader can decrease the maximum possible loss to 0.308, resulting in less money being deducted from their account through margin linking. This is illustrated in the

lower right of Table 7. When considering fees it is necessary to adjust the quantities to achieve the minimum cost replicating set of contracts.

To calculate the formula for the number of shares c_i for each No contract N_i , start with the same payoff equation from the previous section but reduce each payoff to account for the fee f :

$$(1 - f)(1 - Y_a^{ask}) = (1 - f) \sum_{i=1, i \neq a}^n c_i * (1 - N_i^{ask}) \quad (20)$$

Our goal is to find a formula for the number of each No contract to buy, c_i , for $i = 1..n, i \neq a$.

If outcome a is not realized then some other outcome b has occurred and contract Y_b is worth \$1 and N_b is worth \$0. The trader's loss will be:

$$c_b N_b^{ask} - (1 - f) \sum_{i=1, i \neq a, b}^n c_i (1 - N_i^{ask}) \quad (21)$$

This loss must be equal to the loss of some other potential outcome x .

$$c_b N_b^{ask} - (1 - f) \sum_{i=1, i \neq a, b}^n c_i (1 - N_i^{ask}) = c_x N_x^{ask} - (1 - f) \sum_{i=1, i \neq a, x}^n c_i (1 - N_i^{ask}) \quad (22)$$

Solving for c_x gives:

$$c_x = \frac{(1 - Y_a^{ask}) \prod_{k=1, k \neq a, x}^n g(N_k^{ask}, f)}{\sum_{i=1, j \neq a}^n (1 - N_i^{ask}) \prod_{w=1, w \neq a, i}^n g(N_w^{ask}, f)} \quad (23)$$

with

$$g(p, f) = p + (1 - f)(1 - p) = (1 - f) + fp \quad (24)$$

Observe that when the fee $f \rightarrow 0$, $g(p, f) \rightarrow 1$ and c_x reduces to 14. Also notice what function g represents. This is the amount of money returned to the trader's account when a No contract purchased for N^{ask} closes out for \$1, net of fees. This amount varies by a small amount based on the value of N^{ask} .

We can also derive a formula for d_x that will enable a trader to replicate the payoff of a single No contract with a set of Yes contracts.

Consider the payoff of a single No contract for outcome a . If some other outcome b is realized, it will have a positive payoff that will be reduced by the fee f . The replicating set of Yes contracts must have the same payoff for outcome b . That leads to the equality

$$(1 - f)(1 - N_a^{ask}) = (1 - f)d_b(1 - Y_b^{ask}) - \sum_{i=1, i \neq a, b}^n d_i Y_i^{ask} \quad (25)$$

The payoff of the set of Yes contracts for all outcomes other than a must be equal to each other. Therefore, we also know that

$$(1 - f)d_b(1 - Y_b^{ask}) - \sum_{i=1, i \neq a, b}^n d_i Y_i^{ask} = (1 - f)d_x(1 - Y_x^{ask}) - \sum_{i=1, i \neq a, x}^n d_i Y_i^{ask} \quad (26)$$

Solving for d_x gives:

$$d_x = \frac{(1-f)(1-N_a^{ask}) \prod_{k=1, k \neq a, x}^n g(Y_k^{ask}, f)}{(1-f)(1-Y_x^{ask}) \prod_{k=1, k \neq a, x}^n g(Y_k^{ask}, f) - \sum_{i=1, i \neq a, x}^n Y_i^{ask} \prod_{w=1, w \neq a, i}^n g(Y_w^{ask}, f)} \quad (27)$$

The d_x values may be negative, indicating that the No contract cannot be replicated by Yes contracts. To replicate with other No contracts instead, the following formula can be used:

$$d'_x = \frac{(1-f)(1-N_a^{ask}) \prod_{k=1, k \neq a, x}^n g(N_k^{ask}, f)}{-N_x^{ask} \prod_{k=1, k \neq a, x}^n g(N_k^{ask}, f) + (1-f) \sum_{i=1, i \neq a, x}^n (1-N_i^{ask}) \prod_{w=1, w \neq a, i}^n g(N_w^{ask}, f)} \quad (28)$$

If the d'_x values are also negative then the No contract cannot be replicated with either Yes or No contracts.

Our empirical results changed when our equations considered PredictIt's 10% fee on each profitable trade. For Democratic market, the Buy Yes trades were replicated and dominated by the replicating set of No contracts 87.6% and 63.9% of the time for the take and provide trades. For the Republican market, the percentages are 89.6% and 83.2%. Scatter plots of the data are available in Figure 7.

When comparing the No contract buy trades with the replicating set of Yes contracts, the No contract was rarely dominated by either Yes or No contracts. In the Democratic market the No contract trades were replicated and dominated by a set of Yes contracts 1.7% and 0.9% of the time and by a set of No contracts 0.3% and 0.1% of the time when taking and providing liquidity. In the Republican market, the No contract trades were dominated by a set of Yes contracts 0% of the time. The

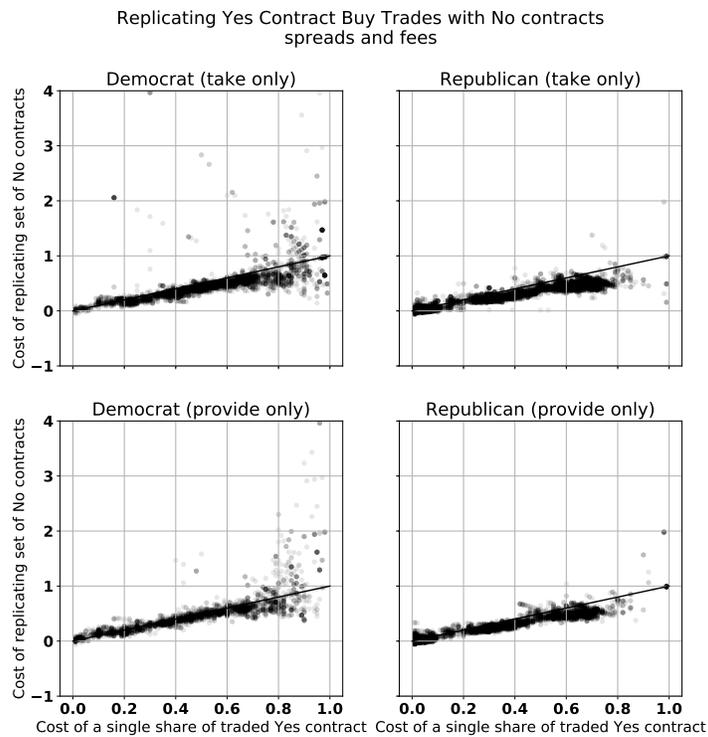


Figure 7: Buy Yes trades that took liquidity plotted against the optimal replicating set of No contracts. All trades are priced using the contract's bid-ask spread. Contract returns include PredictIt's 10% fee.

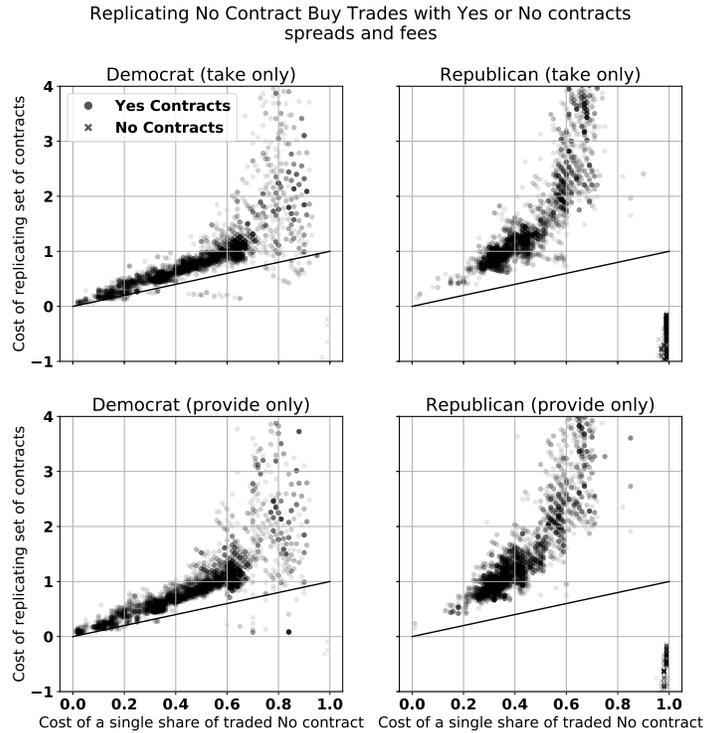


Figure 8: Buy No trades that took liquidity plotted against the optimal replicating set of Yes or No contracts. All trades are priced using the contract’s bid-ask spread. Contract returns include PredictIt’s 10% fee.

No contract trades were dominated by a set of No contracts 12.8% and 6.7% of the time when taking and providing liquidity. Charts of the data are shown in Figure 8.

3.4 Evaluating Trader Efficiency

We can employ our concept of dominated contracts to evaluate all of the executed trades, both before and after margin linking was implemented, and determine if they are or are not dominated by a replicating set of contracts. We are including

trades executed in the pre-margin linking period because some traders only executed trades during that time and without including those trades we would have no way of classifying those traders. Trades that are replicatable and are dominated by the replicating set of contracts are deemed inefficient trades. Trades that are either not replicatable or are not dominated by the replicating set of contracts are deemed efficient trades. Note that the way we have defined efficiency, there is no plausible reason for someone to desire the inefficient contract.

We can then look at each traders' executed volume and determine what percent was executed efficiently. As noted above, we classify traders that executed at least 75% of their volume efficiently as efficient traders; the remaining are classified as inefficient traders. Our results are robust to changes to the cutoff percentage.

4 Results

Using the metrics described in sections 2.5 and 3.4 we categorized all of the traders in each market according to the size, activity, and efficiency metrics, labeling each as small or large, inactive or active, and inefficient or efficient. We organized similar traders into subgroups and analyzed the trading behavior and profitability of each subgroup separately.

We found that the largest group of traders in both markets were inefficient, small, and inactive. Table 8 shows that nearly half of all traders are in this category. Further, about 70% of all traders are inefficient and roughly the same amount of are small, but about 85% are inactive.

	% of Traders		Exposure (\$)		Profit (\$)		ROI	
	D	R	D	R	D	R	D	R
Efficient Small Active	2.21	0.87	19.20	19.78	0.62	4.65	3.27	23.49
Efficient Small Inactive	19.33	16.13	13.16	12.62	-2.12	0.511	-16.15	4.05
Efficient Large Active	3.23	2.07	463.66	519.13	24.87	-1.28	5.36	-0.24
Efficient Large Inactive	7.44	6.22	298.55	338.21	37.16	36.70	12.45	10.85
Inefficient Small Active	4.51	5.30	17.66	18.11	-4.21	0.976	-23.82	5.39
Inefficient Small Inactive	45.71	46.41	12.81	12.94	-4.04	-1.41	-31.52	-10.88
Inefficient Large Active	4.64	7.21	455.57	448.94	9.46	-25.56	2.08	-5.69
Inefficient Large Inactive	12.93	15.79	335.07	309.84	-36.33	-32.45	-10.84	-10.47

Table 8: Percentage of traders by volume, activity, and efficiency in both markets. Profit: net of fees, per trader. Exposure: max total exposure, per trader. ROI: profit over Exposure, per trader. There were 3,750 traders in the Democratic market and 4,452 in the Republican market.

Profits are dispersed, but two things are clear: efficient traders were generally profitable, and while there are mixed directions, there are larger absolute numbers for groups with higher average exposures (which makes sense, as they are risking more money).

Dividing profits by Exposure provides the Return on Investment or ROI. This helps scale trader profits by the capital risk to better compare the success of the different approaches to trading.

We decompose gross profits into three categories: spread profit, bias profit, and

position profit.

The spread profit is the accumulation of small price gains from trading at a better price by providing liquidity instead of taking liquidity. We calculated this as the difference between the executed price and the contract midquote at the time of each execution. Observe that in general the traders who were efficient or active were better at earning spread profits.

The bias profit is the accumulation of gains from the price distortion associated with a market's set of contract prices adding up to more than \$1. To calculate this we first normalized the contract midquote of each executed contract by the sum of all the midquotes. The bias profit was the difference between the normalized midquote and the actual midquote. As expected, efficient traders had positive bias profits.

The position profit is all of the remaining gross profits after spread and bias profits are removed. This is the profit from favorable price movements of the contracts. Traders that were better at identifying the right candidates at the right time were better at accumulating position profits.

PredictIt's 10% fee is subtracted from the gross profit to yield the net profit. We show all how all of our categories did by profit type in Table 9.

	Net Profit		Spread		Bias		Position		Fee	
	D	R	D	R	D	R	D	R	D	R
Efficient Large Active	5.4	-0.3	7.3	8.1	4.4	5.9	-2.6	-9.7	-3.8	-4.6
Efficient Small Active	3.3	23.5	3.5	0.7	5.0	0.7	-2.1	27.0	-3.1	-4.9
Efficient Large Inactive	12.5	10.9	-0.7	-1.7	2.9	6.7	13.6	9.9	-3.3	-4.0
Efficient Small Inactive	-16.2	4.1	-1.0	-3.9	4.6	1.5	-17.4	10.4	-2.3	-3.9
Inefficient Large Active	2.1	-5.7	-0.2	1.8	-0.9	0.3	5.9	-2.9	-2.7	-4.8
Inefficient Small Active	-23.8	5.4	0.5	-0.3	-3.2	-2.8	-18.9	13.6	-2.2	-5.1
Inefficient Large Inactive	-10.8	-10.5	-1.7	-1.7	-2.1	-3.2	-4.6	-1.7	-2.4	-3.9
Inefficient Small Inactive	-31.5	-10.9	-2.3	-2.9	-5.2	-9.4	-22.0	5.6	-1.9	-4.2

Table 9: Spread, Bias, Position, Fee are all components of Net Profit and are divided by Exposure, per trader, for comparison purposes. There were 3,750 traders in the Democratic market and 4,452 in the Republican market.

Efficient traders earned bias profits by definition. They are frequently making the most efficient trades, and in markets that frequently have inefficiencies, there are many opportunities for earning bias profits. Generally active traders earned spread profits. The active traders know they could accumulate small profits by providing liquidity instead of taking liquidity. Position profits are much more mixed, even for the efficient traders. But, this is not the easiest view to digest, so we reconstitute Table 9 into Table 10 which illustrates the how each categories does compared to the categories with just one variable shifted. For instance, on the top row we show Efficient traders do, versus Inefficient, when they are “Large Active”, “Small Active”,

“Large Inactive”, and “Small Inactive”. So, Efficient Large Active have 12.8 and 11.9 percentage points higher ROI for Spread + Bias than Inefficient Large Active. But, Efficient Large Active have 8.5 and 6.8 percentage points lower ROI for Position than Inefficient Large Active. Overall, Efficient Large Active have 3.3 and 5.4 percentage points higher ROI than Inefficient Large Active.

[Democratic , Republican] in terms of difference in percent ROI

Efficient - Inefficient	Large Active	Small Active	Large Inactive	Small Inactive
Spread + Bias	12.8 , 11.9	11.2 , 4.5	6.0 , 9.9	7.4 , 9.9
Position	-8.5 , -6.8	16.8 , 13.4	18.2 , 11.6	4.6 , 4.8
ROI	3.3 , 5.4	20.5 , 18.1	23.3 , 21.4	15.3 , 16.0
Active - Inactive	Efficient Large	Efficient Small	Inefficient Large	Inefficient Small
Spread + Bias	9.5 , 9.0	4.9 , 3.8	2.7 , 7.0	4.8 , 12.6
Position	-16.2 , -19.6	15.3 , 16.6	10.5 , -1.2	3.1 , 8.0
ROI	-7.1 , -11.2	19.5 , 19.4	12.9 , 4.8	7.7 , 16.3
Large - Small	Efficient Active	Efficient Inactive	Inefficient Active	Inefficient Inactive
Spread + Bias	3.2 , 12.6	-1.4 , 7.4	1.6 , 2.7	3.7 , 7.4
Position	-0.5 , -36.7	31.0 , -0.5	24.8 , -16.5	17.4 , -7.3
ROI	2.1 , -23.8	28.7 , 6.8	25.9 , -11.1	20.7 , 0.4

Table 10: Differences in ROI holding other two variables constant. There were 3,750 traders in the Democratic market and 4,452 in the Republican market.

Efficient dominates inefficient conditional on the other two variables. Only in Position ROI for Large Active traders are efficient traders not getting a better return

than inefficient traders (and their overall ROI is still higher).

Active does very well against inactive conditional on the other two variables. While active traders are not as impressive versus inactive as efficient versus inefficient, they are consistent. Only in the Efficient Large category are they actually getting a lower ROI. Which means efficient large inactive traders do better than efficient large active traders.

Large has relatively mixed results relative to small conditional on the other two variables. Being a large trader does not tell us much about their return.

We run a regression to test what we see in the tables: what is the return on investment with dummy variables for size, activity, and efficiency. We ran this separately for each market. Simple ordinary least squares: $ROI = \text{constant} + \text{Efficiency} + \text{Size} + \text{Activity}$. Results in Table 11.

	Democratic	Republican
constant	-0.35 ** (0.01)	-0.11 ** (0.02)
Efficient	0.14** (0.02)	0.15** (0.03)
Large	0.20** (0.02)	0.01 (0.03)
Active	0.13** (0.03)	0.10* (0.04)

Table 11: Regression coefficients on ROI: * Significant at 5%, ** Significant at 1%.

Efficiency is the only variable that is significant in both markets. Active is slightly smaller and not fully significant in the Republican market, but also consistent. Size is big and positive in the Democratic market, but near zero and non-significant in the Republican market. It is tempting to try to deconstruct the differences between

the two markets, but the biggest distinction is the number of competitive candidates. In future work we hope to replicate this over hundreds of markets and deconstruct the correlation of market-level characteristics and our three variables, but until then we using the two markets as adding power to our findings across various types of markets, rather than a test between them.

There is a high degree of classification overlap for the traders that traded in both markets. We found that there were 2,042 traders that traded in both markets; this is 54% of the traders in the Democratic market and 46% of the traders in the Republican market. For those 2,042 traders, 65.6% of them had the same efficiency classification in both markets. Table 12 shows the cross tabulation of the efficiency classification in both markets. We also found that 80.0% had the same size classification and 75.7% had the same activity level classification.

	Republican Classification	
	Efficient	Inefficient
Democrat classification		
Efficient	242	437
Inefficient	265	1,098

Table 12: Efficiency classification cross tabulation in the Democrat and Republican markets for the 2,042 traders that traded in both markets.

An important question we need to answer is if the 65.6% overlap of efficiency classification is statistically significant. If the traders who were classified as efficient knew nothing about trading but were classified as efficient because they randomly

made trades that happened to be efficient, their efficient classification would be nothing more than an ex-post evaluation and they would not be more likely to trade efficiently in other markets. We wish to show that this is not the case.

To test the significance we can do a simulation to determine the distribution of random classifications. For the 2,042 overlapping traders, 33% are classified as efficient in the Democratic market and 25% are classified as efficient in the Republican market. Both statistics are the same for all of the traders in their respective markets. If we randomly assign classifications to the traders' activity in each market in these same proportions, we find that 58.3% of the traders have same classification with a standard deviation of 1.1%. Using this distribution we find that the z-score of our 65.6% measurement is 6.7. Clearly this has statistical significance and our results show that trading efficiency is a real trader characteristic for many people. But, there are some traders who are efficient in one market and not the other. These traders are going to be more profitable where they are efficient and we have no way of knowing if it is luck that occasionally makes an unknowledgeable person pick the efficient trade or if it is laziness that an efficient person chooses the less efficient option in a given market.

5 Discussion

There are a lot of inefficient trades. The vast majority of trades of Yes contracts could have been done with No contracts. If there were institutional traders or enough retail traders that figured this out, the advantage of trading efficiently would have

shrunk, but it did not.

Considering that large traders have more wealth on the line, it is remarkable that they did not get better returns with larger volume. Theory would dictate that these traders are not constrained by liquidity, so they can trade efficiently based on their information. Further, they may even have more information than other traders. But, this is not found in our results.

Considering they objectively apply their information, it is reasonable that active traders did better than inactive traders (most of the time). These traders have better models of the available information, which likely also includes being non-partisan in how they trade.

But, most important: understand how markets work. Efficient traders dominate inefficient traders, holding everything else constant. While we solved for efficiency with a reasonably complex model that would have been costly for any retail trader to employ, simple heuristics would have sufficed. For instance, a trader could simply have traded only No contracts whenever the Yes contracts for a set of mutually exclusive outcomes were priced above \$1.

We suggest that exchanges like PredictIt focus on lowering the barriers to efficiency for retail traders. This could be a wizard on the PredictIt confirmation page where they can tell users if there is another strictly dominant trade that they should consider instead of their current trade. Or it execute the more efficient trade for the trader, without even telling the trader. For instance: if it a trader wants to buy Trump for President in 2020 for \$0.45, but Harris to not be President in 2020 is selling for \$0.40, the exchange simply buy the Harris contract on the back-end,

but display the Trump for President to the trader. The exchange could split the savings with either the exchange, buyer, or seller. If both Harris and Trump lose, the exchange could profit, or again share the winnings with either the buyer or seller.

If the exchange made efficient trades easier or automatic, it would eliminate much of the easy profits of the efficient traders. This transfer of wealth from the unknowledgeable about trading to knowledgeable keeps some people in the market, but gets them spending resources on the marginal zero-sum gains, rather than adding information. In theory, this should force some of these more capable traders into information trading (while some, with lower information, may leave the exchange). This could also allow some traders with a lot of information, but little understanding of trading, to make money and move into the market. In net, the trade-offs could lead to less trading, but should lead to higher information in the market.

Critically, this could be an important testing ground to narrowing the advantages that institutional traders have over retail traders in other exchanges. Many retail investors actually know things about companies or the economy, but do not know how to make the trades that could supply this information into the market.

References

- Berg, Joyce, Robert Forsythe, Forrest Nelson and Thomas Rietz. 2008. “Results from a dozen years of election futures markets research.” *Handbook of experimental economics results* 1:742–751.
- Copeland, Thomas E and Dan Galai. 1983. “Information effects on the bid-ask spread.” *the Journal of Finance* 38(5):1457–1469.
- Glosten, Lawrence R and Paul R Milgrom. 1985. “Bid, ask and transaction prices in a specialist market with heterogeneously informed traders.” *Journal of financial economics* 14(1):71–100.
- Harrison, J Michael and David M Kreps. 1978. “Speculative investor behavior in a stock market with heterogeneous expectations.” *The Quarterly Journal of Economics* 92(2):323–336.
- Kyle, Albert S. 1985. “Continuous auctions and insider trading.” *Econometrica: Journal of the Econometric Society* pp. 1315–1335.
- Miller, Edward M. 1977. “Risk, uncertainty, and divergence of opinion.” *Journal of Finance* 32(4):1151–1168.
- Rhode, Paul W. and Koleman S. Strumpf. 2008. “Manipulating political stock markets: A field experiment and a century of observational data.” Working Paper, University of Arizona.

- Rothschild, David. 2009. “Forecasting elections comparing prediction markets, polls, and their biases.” *Public Opinion Quarterly* 73(5):895–916.
- Rothschild, David and David M Pennock. 2014. “The extent of price misalignment in prediction markets.” *Algorithmic Finance* 3(1):3–20.
- Rothschild, David M and Rajiv Sethi. 2016. “TRADING STRATEGIES AND MARKET MICROSTRUCTURE: EVIDENCE FROM A PREDICTION MARKET.” *Journal of Prediction Markets* 10(1).
- Schmitz, James E. 2011. “Algorithmic trading in the Iowa electronic markets.” *Algorithmic Finance* 1(2):157–181.
- Wolfers, Justin and Eric Zitzewitz. 2006. “Interpreting prediction market prices as probabilities.” Working paper No. 12200, National Bureau of Economic Research.