How manipulable are prediction markets?*

Itzhak Rasooly†   Roberto Rozzi‡

September 18, 2024

ABSTRACT. In this paper, we conduct a large-scale field experiment to investigate the manipulability of prediction markets. The experiment involves randomly shocking prices across 817 separate markets; we then collect hourly price data to examine whether the effects of these shocks persist over time. We find that prediction markets can be manipulated: the effects of our trades are visible even 60 days after they have occurred. However, as predicted by our model, the effects of the manipulations somewhat fade over time. Markets with more traders, greater trading volume, and an external source of probability estimates are harder to manipulate.

KEYWORDS: prediction markets, field experiment
JEL CODES: C63, C93, D84, D90

†Sciences Po & Paris School of Economics.
‡University of Siena.

1. INTRODUCTION

Prediction markets — that is, markets in which traders can bet on the outcomes of various events — have proven to be powerful information aggregators. Existing research shows that the prices that arise in prediction markets can predict outcomes as well if not better than alternative forecasting methods such as expert polls (Figlewski, 1979; Roll, 1984; Pennock et al., 2001; Wolfers and Leigh, 2002; Berg et al., 2008). In addition, while alternative methods such as polling can be costly, prediction markets are largely self-financing. In part for this reason, prediction markets currently constitute the *only* source of probability estimates on many important questions. For example, although one can find prediction markets on the outcomes of various geopolitical conflicts, it is difficult to obtain public probability estimates on the outcomes of these conflicts through other means.

Perhaps due to these encouraging results, prediction markets are currently undergoing something of a renaissance. Although the prediction markets previously studied in the literature (e.g. the Iowa Electronic Markets) were often small and developed for academic purposes, recent years have seen the founding of several much larger prediction market platforms. Polymarket, founded in 2017, often handles large volumes of trade: for example, over \$380 million was traded on Polymarket in July 2024 alone. Kalshi, founded in 2021, also routinely handles tens of millions of dollars of trade on a monthly basis. Manifold Markets, also founded in 2021, has become the largest prediction market website in the world as measured by the number of markets hosted on the platform.

Despite this promise, prediction markets are hampered by long-standing concerns about manipulability. Historically, it has not been uncommon for politicians to bet on themselves in the hope of increasing the market's assessment of their electoral winning chances (Rhode and Strumpf, 2004, 2006). While this is one motivation for individuals to attempt to manipulate prediction market prices, it is not the only one: for example, concerns that terrorists would attempt to 'distract' onlookers from their true targets were cited during the cancellation of the planned Policy Analysis Market in 2003 (Hanson et al., 2006).[1] Concerns about the manipulability of prediction markets remain prominent in more recent media coverage: see, for example, FT (2023), NYT (2023) and Vox (2024).

Although it is clear that many would *like* to manipulate prediction markets, it is less clear that such attempts would be successful. Indeed, according to most existing literature (dis-

---

[1] In the words of one Nobel Laureate, '[trading] could be subject to manipulation, particularly if the market has few participants — providing a false sense of security or an equally false sense of alarm' (Stiglitz, 2003). While this concern may be valid in general, it is worth noting that the Policy Analysis Market would not have allowed traders to bet on the locations or timings of terrorist events (Hanson and Oprea, 2009).

cussed below), any manipulation attempt will be extremely shortlived in the sense that it will be rapidly 'undone' by the behavioural responses of subsequent traders. Understanding this issue also provides an indirect test of the efficiency of prediction markets (Fama, 1970). If market prices only reflect the 'fundamentals', then the effects of random trades should be transient. However, if markets are inefficient, the effect of random trades could persist.

In this paper, we study these questions using both theoretical and experimental methods. On the theoretical side, we build the first model that traces out the path of prices in a prediction market following a manipulation attempt. On the experimental side, we conduct the first large-scale field experiment on the manipulability of prediction markets.

We begin by presenting our model. In the model, agents are risk averse and optimally choose how to bet based on their beliefs about the likelihood that the event will take place. Prices are determined using the algorithm that underpins the markets that we study experimentally. Under this algorithm, prices are non-linear; and our traders understand this when they decide how much to bet. In contrast to existing literature, our model does not assume competitive equilibrium pricing. This allows us to avoid logical inconsistencies[2] and to study the full path of prices following a manipulation attempt (including 'disequilibrium' prices).

The model makes three important predictions about the impact of a manipulative trade. First, the model predicts that manipulation can systematically affect the market price — both in the short run (due to the need for behavioural adjustment), but also in the long-run (e.g., due to 'learning effects'). Second, however, the model also predicts that the effect of manipulation should be partially 'undone' by future trades: for example, if a manipulator increases the market price, then traders should believe that the price is 'too high' and take actions to correct this. The speed of price reversion is predicted to decline as the price moves back towards its original value. Third, the model suggests that the degree of reversion should systematically vary by market type. For example, markets with more traders and less scope for 'learning' from the market price should be harder to manipulate.

To test these predictions, we conduct the first-ever large-scale field experiment on the manipulability of prediction markets. Unusually for a field experiment, our experiment takes place at the market level: in total, we have $n = 817$ different prediction markets in our sample. Our experiment involves randomly placing *yes bets* and *no bets* on different markets; we also include 'control' markets in which we do not intervene. The yes bets are designed to increase the price by 5 percentage points; likewise, the no bets decrease the price by 5 percentage

---

[2] Specifically, competitive equilibrium assumes a continuum of traders, none of whom can move the market price (Aumann, 1964); in contrast, the whole point of our model is to study the adjustment of prices after the price is pushed in a particular direction by a trade.

points. To investigate whether the effects of our bets persist, we then collect hourly price data over a 30 day period (leading to ∼600,000 price observations in total). To understand important heterogeneities, we collect rich data on each market's type including the total volume and number of traders at the point at which our bet takes place.

Our experiment yields three main sets of results. First, we show that the effects of our manipulations are visible in the data even 60 days after they have occurred: market prices in the 'yes' group are higher than they would otherwise have been, and market prices in the 'no' group are correspondingly lower. Second, however, as predicted by our theoretical model, the behavioural responses of subsequent traders generate some reversion of prices towards their original values. As in the model, reversion is relatively quick in the week following our bets but slows as time progresses. Third, the degree of reversion varies across markets in just the way that is suggested by our model. In particular, markets which are duplicated on a different prediction platform (which provides an 'external' source of probability estimates), markets with more traders, and markets with more 'activity' (measured, e.g., by the number of comments left by traders) are more difficult to manipulate.

Taken together, our results imply that prediction markets *can* be manipulated — although the extent to which any manipulation persists depends on market characteristics in systematic ways. This result is suggested not just by theory (e.g. due to learning effects), but also appears in every sub-sample of the experimental data that we investigate. The success of our manipulations may be due to their relatively small size: while a shock of 5 percentage points is large enough to be statistically detectable, it appears to have been small enough to result in plausible looking market probabilities. In contrast, very large shocks to market prices may be insufficiently 'persuasive' to persistently move market prices.

Our paper builds on several literatures that study the manipulability of prediction markets. An important precursor to our work is Camerer (1998), who studies the impact of making random bets on particular horses in the context of pari-mutuel racetrack betting.[3] Since Camerer (1998)'s setting is not a prediction market, it differs from ours in multiple ways: for example, Camerer (1998) studies a very different population of traders operating in an environment with a different pricing rule and a substantially more rapid resolution of uncertainty.[4] In addition, unlike in our experiment, Camerer (1998) cancels his bets around 12 minutes after they have been made. For these reasons, it is perhaps unsurprising that we

---

[3] See also Brown and Yang (2017) for a related experiment on 'anchoring'.

[4] In addition, since horse races are repeated, it is plausible Camerer (1998)'s bettors broadly agreed on what the prices of different horses ought to be. As our model makes clear, this makes manipulating the market price substantially more difficult and may help explain the contrast to our results and those that Camerer (1998) obtains.

obtain very different results: while our trades have persistent effects even 60 days after they are made, the effect of Camerer (1998)'s bets appear to vanish within around 20 minutes.

Somewhat closer to our study, a small literature attempts to trace out the impact of historical manipulation attempts within prediction markets. For example, Rhode and Strumpf (2004, 2006) study attempts to manipulate US prediction markets in the early 20th century; while Hansen et al. (2004) examines an attempt to manipulate prediction markets during the 1999 Berlin state elections.[5] Such studies clearly establish that many have attempted to manipulate prediction markets. However, it is difficult to assess whether such attempts were successful using this approach since it is difficult to estimate the path that prices would have taken in the absence of the manipulation attempt. Our approach entirely side-steps this issue through use of a randomised field experiment.

The manipulability of prediction markets has also been studied by a series of laboratory experiments: see Hanson et al. (2006); Oprea et al. (2008); Veiga and Vorsatz (2009, 2010); Jian and Sami (2012); Deck et al. (2013); Buckley and O'Brien (2017) and Choo et al. (2022).[6] However, the prediction markets created in the laboratory are typically quite different from those observed in the field.[7] Moreover, due to the logistical difficulty of constructing markets in the laboratory, it is typically only feasible to run at most a handful of markets in the study. This makes identifying the impact of a manipulation difficult (especially given that observations within a market cannot be treated as independent) and makes the kind of heterogeneity analysis that we conduct impossible. For these reasons, it is useful to complement the existing laboratory evidence with evidence from the field.

Perhaps most relevant for our purposes, Rhode and Strumpf (2006) report the results of an experiment on the Iowa Electronic Markets that involves making 15 bets in total on 2 (inter-related) markets. Since Rhode and Strumpf (2006) were interested in manipulating a particular market (on the 2000 US Presidential Election), their experiment is very different from the large-scale and across-market field experiment whose results we report here. For this reason, our study is better powered to detect the long-run effects of manipulative trades.

Finally, although our paper contributes to the empirical literature on prediction markets, we also contribute to the theoretical literature. In particular, we adapt 'price theory' style models of prediction markets (Gjerstad, 2005; Manski, 2006; Wolfers and Zitzewitz, 2006)

---

[5] See also Rothschild and Sethi (2016) for an examination of an apparent manipulation attempt during the 2012 US President Election.

[6] For laboratory evidence on prediction markets that does not study the question of manipulability, see also Filippin and Mantovani (2023) and Mantovani and Filippin (2024).

[7] For instance, they are populated by a different group of traders (typically undergraduate students) who form beliefs based on quite stylised information about abstract topics (e.g. the number of balls in an urn).

so that they can cast light on price dynamics following a manipulation attempt.[8] To do this, we relax the assumption that agents are price takers and that prices are determined by competitive equilibrium; instead, agents in our model understand the extent to which their trades will alter the market price. Our model is, to our knowledge, the first attempt to theoretically trace out the path of prices following a price manipulation attempt.

The remainder of this article is structured as follows. Section 2 presents our theoretical model; Section 3 describes the platform that we study experimentally; and 4 presents our experimental design. Our empirical results are contained within Section 5. Section 6 concludes with a discussion of the open questions raised by this research.

2. Manipulation in theory

In this section, we develop some theoretical expectations regarding the impact of a manipulation attempt. As discussed earlier, our model builds on seminal work by Manski (2006), Gjerstad (2005) and Wolfers and Zitzewitz (2006). In contrast to these papers, prices in our model are determined by a common pricing algorithm instead of competitive equilibrium, which allows us to study the adjustment path of prices following a shock. In addition, our agents understand that their trades impact the market price and take this into account when choosing how to trade.

Market. We consider a single market in which traders can bet on whether an event will or will not take place. Traders can buy and sell *yes shares*, each of which pays out 1 currency unit if and only if the event takes place. Likewise, traders can buy and sell *no shares*, each of which pays out 1 currency unit if and only if the event does not take place. As pointed out by Gjerstad (2005), one can assume without loss of generality that traders buy either yes shares or no shares (but not both).

Traders. There are a finite number of traders, indexed by $i \in \{0, 1, ..., m\}$. Each trader believes that the event will occur with probability $\pi_i \in [0, 1]$; thus, the prior beliefs of traders can differ. For simplicity, we assume that each trader has the same initial wealth $w > 0$ and vNM utility function $u \colon \mathbb{R}^+ \mapsto \mathbb{R}$. Let $q_y^i \geq 0$ and $q_n^i \geq 0$ denote trader i's purchase of yes and no shares respectively (as discussed before, either $q_y^i = 0$ or $q_n^i = 0$). We assume that

---

[8] For models in the 'rational expectations' tradition (from which we depart), see also Allen and Gale (1992); Kumar and Seppi (1992); Hanson and Oprea (2009); Ottaviani and Sørensen (2007).

each trader maximises expected utility, i.e.

$$\mathbb{E}[u(q_y^i, q_n^i)] = \begin{cases} \pi_i u(w + q_y^i - C(q_y^i)) + (1 - \pi_i)u(w - C(q_y^i)) & \text{if } q_y^i \geq 0 \\ \pi_i u(w - C(q_n^i)) + (1 - \pi_i)u(w + q_n^i - C(q_n^i)) & \text{if } q_n^i \geq 0 \end{cases} \tag{1}$$

where $C(q_y^i)$ and $C(q_n^i)$ are the costs of purchasing yes and no shares. We assume that u is twice differentiable with $u'(w_s) > 0$ and $u''(w_s) < 0$ for all $w_s > 0$, where w_s is the agent's wealth in a generic state. To prevent corner solutions, we assume that $\lim_{w_s \to 0^+} u'(w_s) = \infty$. Finally, we assume that u exhibits decreasing absolute risk aversion, i.e. that $-u''(w_s)/u'(w_s)$ is strictly decreasing in w_s for all $w_s > 0$. Although this assumption is unnecessarily strong, it simplifies the analysis considerably; see the discussion below.

Pricing. We assume that costs are determined by the constant product rule; this pricing algorithm underpins the prediction market that we study experimentally (see Section 3). Under this rule, prices are set by an automated market maker (AMM) that holds reserves of yes and no shares, denoted (y, n). To illustrate the mechanics of the rule, suppose that a trader wishes to spend x currency units to purchase yes shares (analogous comments apply to purchases of no shares). The trader first transfers x to the AMM, which converts this into x yes shares and x no shares; thus, the AMM's reserves become $(y + x, n + x)$. The AMM then transfers the trader the number of shares q that satisfies

$$(y + x - q)(n + x) = yn \tag{2}$$

That is, it supplies the number of yes shares necessary to restore the product of its reserves to its previous value yn. This rule implicitly defines the number of shares that a trader receives q given the amount of currency they have spent x. Indeed, solving (2) for q yields

$$q = \frac{x(n + x + y)}{n + x} \tag{3}$$

Equivalently, after letting $C(q) = x$ denote the total cost of the shares, one sees that

$$C(q) = \frac{\sqrt{(n - q + y)^2 + 4nq} + q - n - y}{2} \tag{4}$$

From this, it is straightforward to derive the marginal and average cost functions. The following lemma characterises their key properties.[9]

---

[9] All proofs are collected in Appendix A.

Lemma 1. *Under the constant product rule, the marginal cost of yes shares $MC(q)$ satisfies (i) $MC(0) = n/(n + y)$, (ii) $MC'(q) > 0$ for all $q \geq 0$, and (iii) $\lim_{q \to \infty} MC(q) = 1$. Similarly, the average cost of yes shares $AC(q)$ satisfies (i) $\lim_{q \to 0^+} AC(q) = n/(n+y)$, (ii) $AC'(q) > 0$ for all $q > 0$, and (iii) $\lim_{q \to \infty} AC(q) = 1$.*

As Lemma 1 emphasises, pricing under the constant product rule is non-linear. For a small purchase of yes shares, both the marginal and average costs are close to $n/(n + y)$ and thus determined by the ratio of yes and no shares held by the AMM (see Figure A1). As the size of the purchase rises, yes shares become more scarce and so both average and marginal costs rise. In the limit as $q \to \infty$, average and marginal costs converge to 1; notice that such a purchase is necessarily unattractive since the expected value of each share cannot exceed 1.

Optimal trading. We now characterise the optimal trading behaviour of the market participants. As the next result shows, the 'marginal price' $n/(n+y)$ plays a key role in pinning down which types of trade participants will make. Specifically, participants with optimistic beliefs ($\pi_i > n/(n + y)$) will bet that the event will take place; whereas participants with pessimistic beliefs ($\pi_i < n/(n + y)$) will bet that event will not take place.

Lemma 2. *Define $p = n/(n + y)$. Then*

- *If $\pi_i > p$, the trader will buy a positive quantity of yes shares.*

- *If $\pi_i = p$, the trader will not hold any shares.*

- *If $\pi_i < p$, the trader will buy a positive quantity of no shares.*

At a technical level, Lemma 2 extends Arrow (1965)'s analysis of optimal betting to the case of non-linear pricing. As pointed out by Arrow (1965), even risk averse traders will choose to make a positive bet if this has positive expected value since they are locally risk neutral around $w_s = 0$. Since pricing is locally linear around $q = 0$, it is unsurprising that this argument extends to our case and that the 'marginal price' at zero $n/(n + y)$ plays a key role in the argument.

Next, we examine how the behaviour of traders responds to an exogenous shock to the market price. More precisely, we suppose that an external participant makes a purchase of yes shares, thereby increasing n, decreasing y and pushing up the marginal price $n/(n + y)$. The next result establishes how traders react qualitatively to the shock.

Lemma 3. *Suppose that the marginal price increases from p to $p + \Delta$. Then*

- *Traders with $\pi_i \geq p + \Delta$ will decrease their holdings of yes shares.*

- *Traders with $\pi_i \in (p, p + \Delta)$ will switch from holding yes shares to holding no shares.*

- *Traders with $\pi_i \leq p$ will increase their holdings of no shares.*

To sketch the proof, consider the case of $\pi_i \geq p + \Delta$. Ignoring the cases of $\pi_i = 1$ and $\pi_i = p + \Delta$ (which require a separate argument), one first shows that optimal purchases must satisfy a first order condition

$$\frac{\pi_i u'(w + q - C(q))}{(1 - \pi_i)u'(w - C(q))} = \frac{C'(q)}{1 - C'(q)} \tag{5}$$

For any fixed $q > 0$, the shock increases marginal costs, thus shifting the right hand side of (5) upwards. Meanwhile, under the assumption of decreasing absolute risk aversion, one can show that the shock shifts the left hand side (i.e. the ratio of marginal utilities) downwards. Each of these forces separately decreases the optimal purchase, which means that traders will unambiguously choose to hold fewer yes shares.[10]

More substantively, Lemma 3 underscores the importance of reversion. It is easy to check that traders who decrease their holdings of yes shares will put downward pressure on the market price. Likewise, the market price will fall if a trader increases their holdings of no shares. Thus, the model predicts that, following an exogenous increase in the price, the behavioural adjustment of an arbitrarily chosen trader will lead the price to revert towards its original value.

Although Lemma 3 captures the logic of reversion, it does not settle the quantitative question of how much prices should revert following a manipulation attempt. In particular, it does not tell us whether reversion should be complete. In addition, Lemma 3 does not describe the speed of reversion and the shape of the price path following the manipulation. To shed light on these more difficult questions, we now turn to simulations.

Simulations. To simulate the market under various conditions, we assume that $u(w_s) = \ln(w_s)$; note that this functional form satisfies our maintained assumptions on u. One advantage of the logarithmic specification is that it allows the optimal trades to be solved explicitly. For example, if $\pi_i > n/(n + y)$, then the optimal spend on yes shares is

$$x^* = \frac{n\left(\sqrt{y[(n + 4(1 - \pi_i)\pi_i w)y + 4(1 - \pi_i)\pi_i w(n + w)]} - 2(1 - \pi_i)w - y\right)}{2(1 - \pi_i)(w + y)} \tag{6}$$

and the optimal purchase q may be computed using (3). Since the simulations involve solving well over 10^7 optimisation problems (see below), avoiding numerical optimisation delivers an

---

[10] As this discussion makes clear, the assumption of decreasing absolute risk aversion is unnecessarily strong. Without it, however, one would need to assume an appropriate bound on the income effect.

important reduction in speed.

In all simulations, the market is initialised and given t periods to reach a stable state. As in our experiments, an external participant then purchases the number of yes shares required to increase the price by 0.05; the market is then given t' periods to adjust. At each time, one trader is randomly selected to re-adjust her holdings. This introduces randomness into the process; thus, we run each simulation 10,000 times and report the average of the realised prices at every time.[11]

In the *baseline case*, we assume that $m = 10$, $w = 100$, and $(n, y) = (1000, 1000)$. We also assume that trader i has a belief $\pi_i = i/m$ for $i \in \{0, 1, ..., m\}$ (i.e. that beliefs are uniform). Finally, we set $t = t' = 100$. The results are shown in Panel A of Figure 1. As can be seen, the average price is stable at 0.5 before the shock; this is unsurprising given that the initial reserves are balanced ($n = y$) and that beliefs are symmetric around 0.5. After the shock increases the price to 0.55, the behavioural responses of the agents induce the price to move back towards its original value of 0.5. However, reversion takes time; and even once prices have stabilised, they have only reverted by around 40%.

It is natural to wonder how these results depend on the assumed market conditions; we thus now report some variations. First, we vary the number of traders, considering $m \in \{10, 20, ..., 60\}$. Second, following Manski (2006), we allow traders to revise their beliefs in light of the market price p. Specifically, we suppose that posterior beliefs are given by $\pi_i' = \lambda p + (1 - \lambda)\pi_i$, where $\lambda \in [0, 1]$ is the 'learning rate'; we consider $\lambda \in \{0, 0.2, ..., 1\}$. Finally, we vary the degree of prior agreement amongst the traders. Specifically, we assume that trader i has a belief $\pi_i = \alpha(i/m) + (1 - \alpha)0.5$ where $\alpha \in [0, 1]$; note that that this nests the baseline case of uniform beliefs ($\alpha = 1$) along with the 'opposite' case of a common prior belief of 0.5 ($\alpha = 0$). We run separate sets of simulations for all $\alpha \in \{0, 0.2, ..., 1\}$.

Figure 1 displays the most 'extreme' version of each variation; see also Table A1 for a more detailed analysis that describes the degree of reversion under all the parameter combinations that we consider. Several results are apparent. First, increasing the number of traders m increases the speed of reversion in the short run as well as increasing the total magnitude of reversion once prices have stabilised. This is intuitive: increasing the number of traders increases the amount of wealth at their disposal, thereby making it easier for them to undo the impact of the manipulative trade. Second, increasing the learning rate λ makes the manipulation more 'sticky'; indeed, in the extreme case of $\lambda = 1$, no reversion is observed whatsoever. Intuitively, increasing λ means that the manipulative trade is 'more persuasive'

---

[11] All simulations can be reproduced using this code.

(i.e. leads to a larger update to trader beliefs) and thus has a larger long-run effect on the market price. Finally, increasing the amount of prior agreement amongst traders makes the market harder to manipulate. In the extreme case where all traders have a common belief of 0.5, moving the price away from 0.5 is especially difficult.

Figure 1: Price adjustment under different market conditions

(a) Baseline

(b) Varying the number of traders

(c) Varying the learning rate

(d) Varying prior agreement
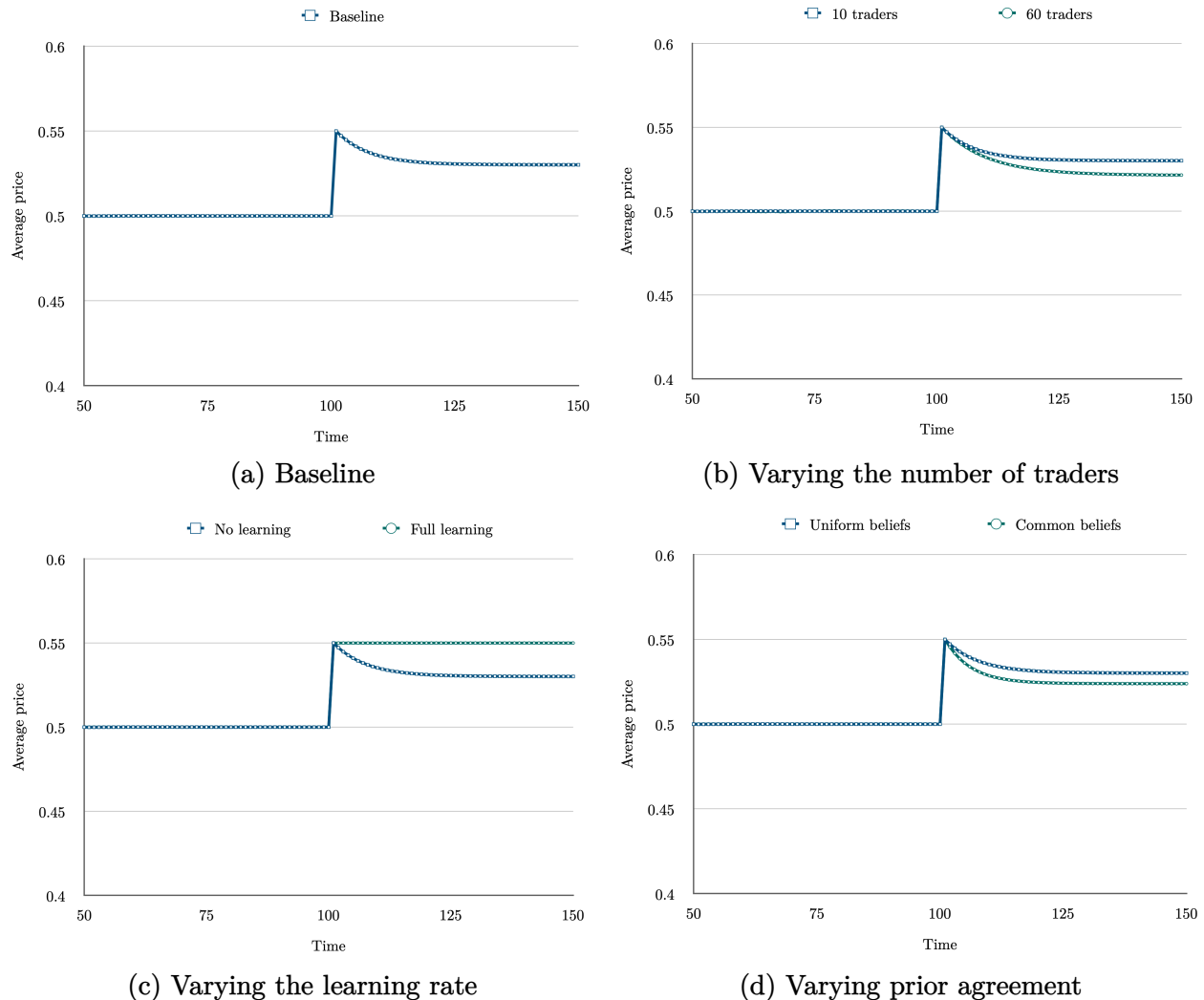
Notes. This figure plots the average price (across 10,000 simulations) from time $t = 50$ to $t = 150$ under different market conditions. Panel (a) displays the baseline simulations; Panel (b) compares $m = 10$ with $m = 60$; Panel (c) compares $\lambda = 0$ and $\lambda = 1$; and Panel (d) compares uniform beliefs with a common belief.

Taken together, the model outlined in this section generates three key predictions concerning the impact of price manipulation by an external market participant. First, the model predicts that manipulation can systematically affect the market price — both in the short run (due to the need for behavioural adjustment), but also in the long-run (e.g., due to 'learning

effects'). Second, however, the model also predicts that the effect of manipulation should be somewhat 'undone' by future trades; this dynamic is captured qualitatively by Lemma 3 and quantified by the simulation results. Third, the model suggests that the degree of reversion should vary by market type. In particular, markets with more traders, a lower degree of learning, and more prior agreement should be harder to manipulate. We will examine the extent to which these predictions are borne out by the data in the subsequent sections.

3. INSTITUTIONAL BACKGROUND

Manifold markets. We conducted our experiment on the Manifold Markets platform, which was founded in 2021. During our experiment, Manifold was the largest prediction market website in the world as measured by the total number of markets hosted on the platform.[12] In addition, Manifold was one of the largest prediction market platforms as measured by number of users, with around 10,000 'active' users at the start of the experiment in December 2023 (Manifold Markets, 2024).[13] The large number of markets on Manifold was crucial to the design of our large-scale field experiment; such an experiment would not have been possible using more traditional platforms (e.g. the IEM) that have previously been studied in the academic literature (Wolfers and Zitzewitz, 2004).

Trading. Figure A2 provides an example of a particular market on the platform. Traders are shown the full history of market prices and given the opportunity to buy and sell yes and no shares. If they choose to make a trade, they are shown how much this will cost and how much it will move the (marginal) market price. Each market is equipped with a closing date, which usually corresponds to the time at which the relevant question will be resolved. Below the price chart, traders can read a description of the exact resolution criteria as well as any comments that have been made by the traders over the course of the market's lifespan.

Pricing. Trading within each market is conducted via an AMM that implements Maniswap, a generalisation of the constant product rule that was studied in Section 2. Under Maniswap, the AMM adjusts its reserves (y, n) in order to hold the expression $y^p n^{1-p}$ constant. The parameter $p \in (0, 1)$ allows markets to be initialised at prices other than 50% without 'wasting' shares (see Manifold Markets, 2022 for the details). Notice that, in the common case where a market is initialised at 50% (corresponding to $p = 0.5$), the Maniswap algorithm exactly reduces to the constant product rule. Moreover, one can verify that an (appropriately

---

[12] As of August 16 2024, there were 10 markets on PredictIt, 312 markets on Kalshi, 504 markets on Polymarket, and well over 2000 markets on Manifold.

[13] According to Similarweb estimates on website traffic, Manifold was the second most popular prediction market website in April 2024. Specifically, the page view estimates are 1,145,000 for PredictIt, 673,000 for Manifold, 531,000 for Polymarket, and 177,000 for Kalshi.

modified) version of Lemma 1 holds under Maniswap, which allows one to establish analogous versions of Lemmas 2 and 3 in this generalised environment.[14]

Users. According to an informal survey, which should be treated with some caution due to the possibility of self-selection, Manifold users appear to skew heavily male (Manifold Survey, 2024). The majority of users reside in the United States; and these users are substantially more likely to vote for the Democratic than the Republican party. Most users stated that they were either aligned with or interested in the 'rationalist' movement; interest and alignment with the 'effective altruism' movement are also common.

Innovations. In some respects, Manifold is an unusual platform. First, markets are user created and resolved. This allows Manifold to host a large number of markets, which is crucial for the feasibility of our large-scale experiment. Second, Manifold provides rich data on each market's characteristics, which facilitates our heterogeneity analysis. Third, a large portion of trade is conducted by 'bots' (i.e. automated trading algorithms) that attempt to exploit inefficiencies in the markets. Fourth, markets are run on a platform specific currency ('Mana'). While Mana (M) can be purchased using real currency (at a rate of $1 = 100M during the experiment), it is not possible to convert Mana back into dollars.

Incentives. Despite this latter feature, there are several incentives to earn Mana through trading. First, during the experiment, Mana could be converted into donations to a charity of the user's choice at the same $1 = 100M rate; in total, Manifold users had raised $316,000 for charity through this mechanism as of 16 May 2024. This incentive is plausibly especially effective given that many Manifold users are interested in 'effective altruism' and thus may view money and charitable donations as fungible. Second, Manifold ranks users based on their monthly earnings: this may activate social-image incentives. Third, Manifold provides users with personalised information about their trading performance and predictive accuracy through Brier scores and calibration charts; this information may also tap into self-image incentives (Bénabou and Tirole, 2006). Taken together, these incentives help explain why Manifold has over 10,000 active users despite running on a platform specific currency. As one especially successful trader on Manifold put it (EA Forum, 2022),

> In the unusual world in which I find myself, for better or worse, doing well on a prediction markets website is somewhat of a badge of honour . . . I wish I had more noble motivations but, alas, I think that's a good chunk of it . . . Another important motivation for me using Manifold relates to charitable giving.

---

[14] More precisely, all results continue to hold after one has changed the 'marginal price' from $n/(n+y)$ to $np/(np - py + y)$.

The importance of the financial incentives are also underscored by the 'devaluation' of Mana that took place in April 2024. On 23 April, it was announced that Manifold would reduce the rate at which Mana could be converted to charitable donations by a factor of 10. However, users wishing to avoid the devaluation were given a one week 'grace period' in which they could donate their earnings at the old (and substantially better) rate.[15] Figure A3 shows the impact of the announced devaluation on donations to GiveWell's maximum impact fund, which is the charity that received the largest value of donations from Manifold users. As can be seen, the number of donations increased by over a factor of 100 in the week following the announcement; whereas the value of donations increased by over a factor of 200. This suggests that, although Manifold users may be motivated by social and self-image incentives, they are also responsive to the financial incentives offered by the platform.

Figure 2: Calibration
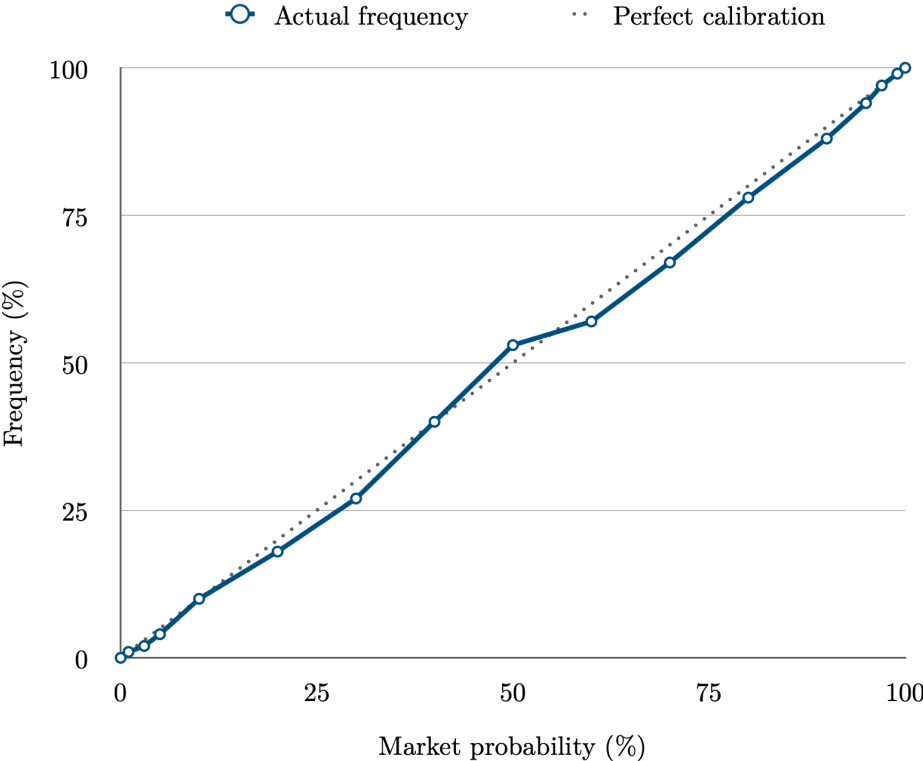
Notes. This figure shows calibration on Manifold as of 20 November 2023, i.e. the relationship between the market price and the probability that a market resolves 'yes'. Source: Manifold Markets (2023).

Predictive performance. Given these incentives, it is not surprising that the predictive performance of Manifold is comparable to that of more traditional platforms. First, as shown

---

[15] This grace period was ultimately extended to 15 May 2024.

by Figure 2, the markets are generally well-calibrated: for example, markets on Manifold with prices close to 40% have historically resolved 'yes' around 40% of the time. Second, in a study of the 2022 US midterm elections, Manifold outperformed the more traditional prediction markets in the sample (Sigma, 2024). Third, Manifold achieves Brier scores that are comparable but slightly worse than Metaculus (EA Forum, 2024). Thus, based on existing evidence, it appears that Manifold markets operate in broadly comparable ways to more traditional prediction markets — a conclusion that is consistent with the findings of previous work on the role of money in prediction markets (Servan-Schreiber et al., 2004).

4. EXPERIMENTAL DESIGN

Intervention. We now outline the structure of our experiment. As discussed earlier, the basic idea was to randomly shock the prices of a large number of prediction markets. More specifically, we either purchased yes shares until the price rose by 5 percentage points (the 'yes' group), purchased no shares until the price fell by 5 percentage points (the 'no' group), or did nothing (the control group). Each given market was equally likely to be assigned to any of these groups. We then observed the prices in every market in order to study whether the effects of our shocks fade over time.

As explained in the pre-registration (Rasooly and Rozzi, 2023), we focus on the difference in mean prices between the 'Yes' and 'No' groups since this is our best powered comparison. Two points should be noted. First, since the treatment is randomly assigned, the difference in average prices between these groups has a causal interpretation. Second, if prices in the markets can be approximated by a random walk, then one would expect the variance of the prices (and thus our standard errors) to increase over time. For this reason, we focus on effects within 7 days of our intervention; although we also conduct analyses with longer time horizons.

Exclusion criteria. We only bet on binary markets, i.e. markets that must resolve as 'Yes' or 'No' (or N/A). We excluded various types of market from our sample. First, we excluded markets that do not resolve based on an external event by the end of 2025 (e.g. self-referential markets). While self-referential markets (e.g. "will this market resolve at above 50%?") along with very long-run markets (e.g. resolving in 2030) may be relatively easy to manipulate, they are also less representative of the markets that operate on other platforms. Second, since very small markets are less likely to form the basis for influential probability estimates, we excluded any market that had fewer than 10 traders at the time of our bet. Third, to obtain a large sample with our experimental budget, we excluded any market which would have

cost over 200M to move in either direction by 5 percentage points.[16] Fourth, since markets can be volatile at the start of their life-cycle (which increases standard errors), we excluded markets that started within the last 7 days; to satisfy our data collection requirements, we also excluded markets that ended within 30 days of our bet. Fifth, to reduce the scope for arbitrage and informational effects (discussed in detail later), we excluded any market that was closely related to another market that was already in our sample (e.g. "Will Trump win?" vs "Will Trump lose?").

Data. We collected hourly price data, starting 24 hours before the bet and continuing for 30 days after the bet. Since we collected data on 817 markets, this means that we have $817 \times 24 \times 31 \approx 600,000$ price observations in total. We also collected rich data on each market's characteristics. First, we recorded the number of traders within each market at the time of our bet; our model suggests that this variable should influence the market's manipulability. Second, we recorded whether each market's question was also present on another prediction platform (Metaculus) that is frequently discussed on Manifold and sometimes used as the source for Manifold questions. If a question is on Metaculus, then one would expect that 'price learning' should be lower and thus the market should be harder to manipulate (see Section 2). Third, we collected three measures of the 'activity of a market' — the total volume of trade, the volume of the trade within the last 24 hours, and the number of comments left by traders. Since our model identifies a 'time period' with a trade, it suggests that more active markets should effectively run on a faster 'clock speed' and thus revert more quickly. For completeness, we also recorded some less interesting variables that were also available on the platform; see Rasooly and Rozzi (2023) for a full description.

Power. Based on the calculations in Rasooly and Rozzi (2023), we needed $n = 849$ markets in order to have 90% power to detect a difference of 3 percentage points between the average prices in the 'yes' and 'no' groups after 1 week.[17] Since the observed reversion was ultimately substantially lower than we had anticipated, this sample size turned out to be unnecessarily large for this purpose. However, the size of our sample remains useful for the heterogeneity and longer-run analyses.

---

[16] In practice, the effect of this restriction is to reduce the number of markets in our sample with initial prices close to 0% to 100%. To see why this is true, note that, under the constant product rule, it is infinitely expensive to increase the price from 95% to 100% (or to decrease the price from 5% to 0%).

[17] When analysing the data, we realised that some of the 849 markets on which we had bet inadvertently violated at least one of our exclusion criteria. After discarding these markets, we ended up with $n = 817$ markets in our sample. We should emphasise that our main results are entirely unaffected by the inclusion (or lack of inclusion) of these markets.

Timelines. We pre-registered our experiment (with an analysis plan) in December 2023.[18] We made the bets over a 5 month period from December 2023 until April 2024. We concluded our main data collection in May 2024 (30 days after the final bet).

5. Experimental results

5.1 Sample

We now describe the results of the experiment, beginning with an overview of the markets in our sample. Table 1 reports some descriptive statistics. As can be seen, around 9% of the market questions could also be found on the Metaculus platform. On average, users had left 3.5 comments at the time of our bet; however, this varies substantially across markets with a minimum of 0 and maximum of 83. At the time of our bets, 27 traders had participated in each market on average. If one considers the markets that had resolved as of 10 July 2024, one sees that this number had risen to 47 traders on average by the end of the markets' lifespan. Meanwhile, the total number of trades had increased on average to 136, with a minimum of 0 and maximum of 1,400.

Table 1: Descriptive statistics

Variable	Mean	Std. dev.	Minimum	Maximum
24 hour volume	24.7	119.9	0	1,701
Total volume	1,879	4,279	89	56,485
Metaculus	0.093	0.291	0	1
Comments	3.54	7.26	0	83
Traders	27.0	25.0	10	300
Final traders	47.1	44.9	10	321
Final trades	136.0	189.7	13	1,400

Notes. This table shows descriptive statistics for the markets in our sample. The first 5 variables show statistics for the full sample ($n = 817$). The last two variables show statistics for the 127 markets that had resolved as of 10 July 2024.

Table A2 provides an overview of the main topics addressed by the markets in our sample; see also Fig A4 for a visual illustration. As can be seen, markets on politics (especially the

---

[18] In general, we conformed closely to the analyses that we had pre-registered. However, we ultimately deviated from the plan in a few (generally minor) ways. All deviations from the pre-registered plan can be viewed in our pre-registration; see the document 'Analysis plan (updated after the experiment)'.

US 2024 Presidential election) and political conflict (especially the conflicts in Gaza and Ukraine) are common. In addition, many markets concern artificial intelligence and the speed of its development. Our sample also includes a variety of markets on sport, popular culture (e.g. the outcomes of Oscar nominations), and various macroeconomic indicators (e.g. inflation and interest rates).

5.2 MAIN RESULTS

7 day effects. We now examine the extent to which our manipulations had a persistent effect over time. To obtain some initial graphical evidence on this question, Figure 3 plots the average price in each of the three treatment groups, starting 24 hours before the trade and continuing for 7 days after the trade. As can be seen, the average prices in each group are stable in the 24 hours before the trade. By design, our intervention immediately increases the average price in the 'yes' group by 5 percentage points and immediately decreases the average price in the 'no' group 5 percentage points, thus creating a gap of 10 percentage points in average prices between the groups.[19] As time goes by, the gap between the 'yes' group and 'no' group becomes smaller, implying that some reversion is taking place. However, even after 7 days have elapsed, a substantial gap of approximately 7.3 percentage points can be observed between groups.

To examine this issue more formally, we estimate models of the form

$$p_{t,i} = \beta_0 + \beta_1 \mathbb{1}_{Y,i} + \beta_2 \mathbb{1}_{C,i} + \beta_3 p_{-1,i} + u_i \tag{7}$$

where $p_{t,i}$ is the price in market i at time t, $\mathbb{1}_{Y,i}$ is a dummy variable that equals 1 if market i is in 'Yes' group, $\mathbb{1}_{C,i}$ is a dummy variable that equals 1 if market i is in the control group, and $p_{-1,i}$ is the price in the market in the hour before our bet. Thus, the 'No' group is the omitted category. As explained in our pre-registration, we include the baseline control $p_{-1,i}$ to increase statistical power. In subsequent robustness checks, we also control for the full suite of baseline variables that are available in our dataset.

Figure 4 presents the estimated $\hat{\beta}_1$ coefficients obtained from estimating this regression for all $t \in \{0, 1, ..., 167\}$. Note that this estimate is simply the difference in average prices between the 'yes' and 'no' groups, correcting for any baseline imbalance in the $p_{-1,i}$ variable. Again, one can see that the difference starts at around 10 percentage points (immediately after the intervention) and declines over the course of the 1 week period. However, there is a clear

---

[19] Since our data is hourly, there is some scope for reversion even in the hour in which the trade is made: for this reason, the initially observed impacts of the yes and no shocks are slightly below 5 percentage points.

Figure 3: Average prices over time
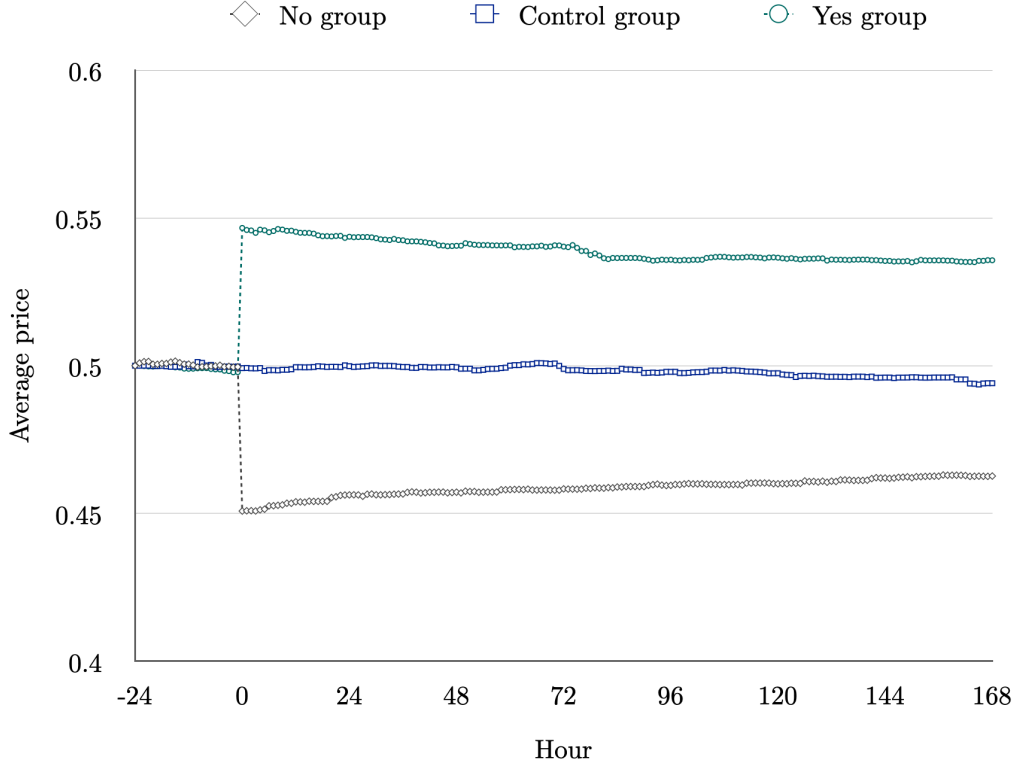
Notes. This figure shows average prices over time in the three treatment groups. All series are rebased (by adding a constant) so that they start at 0.5.

difference between the groups of around 7.5 percentage points even after 1 week. As one can see from the 95% confidence intervals (also depicted in the figure), this difference is statistically distinguishable from zero ($p < 0.01$).

Figure A5 presents analogous results for the $\hat{\beta}_2$ coefficients, i.e. the comparison of the 'no' group with the control. As can be seen, the results are broadly similar: some reversion is observed, but a sizeable (and statistically significant) difference between the groups can be observed even after 7 days. Note that both of these findings — some reversion, along with a persistent effect of our manipulations — are consistent with the model that is outlined in Section 2.

Before proceeding, we use the control group to investigate whether 'yes' and 'no' bets have symmetric effects (as one might expect). To do this, we notice that the hypothesis of symmetry is equivalent to the hypothesis that $\beta_1 = 2\beta_2$, which can be tested as a linear restriction on the statistical model. Using an F-test, we are unable to reject this restriction at conventional significance levels ($p = 0.124$), which suggests that the observed effects are reasonably symmetric.

18

Figure 4: Comparing the 'yes' and 'no' groups ($\hat{\beta}_1$)
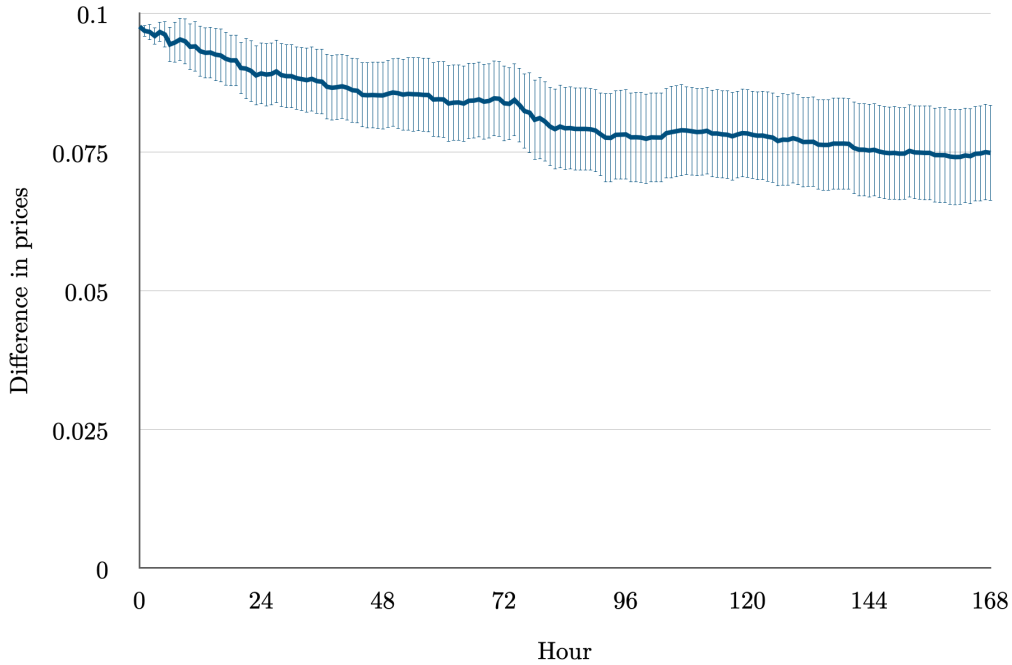
Notes. This figure plots the $\hat{\beta}_1$ coefficients obtained from estimating (7) for all $t \in \{0, 1, ..., 167\}$. The 95% confidence intervals are computed using robust standard errors.

30 day effects. Although our pre-registration focused on 7 day effects for reasons of statistical power, one can also conduct similar analyses for the full 30 day horizon by estimating the regression for all $t \in \{0, 1, ..., 719\}$. As shown by Figure A6, one observes some additional reversion over the longer time horizon. However, even after 30 days, the estimated difference between the 'yes' and 'no' groups has only fallen to 6.9%. Remarkably, even though standard errors have unsurprisingly grown due to the accumulation of shocks of time, the effect remains highly significant ($p < 0.01$). It is also worth noticing that the speed of reversion significantly slows over time: we observe 25% reversion in the first week, but just an additional 6% reversion over the next three weeks. This reduction in reversion speed is exactly the pattern that is predicted in the model of Section 2 (see Figure 1).

60 day effects. Although we had originally planned to only collect data for 30 days after the shocks (as outlined in our analysis plan), the strength of the results after 30 days suggested that very long-run effects might be observable. As a result, we decided to additionally record the price in every market 60 days after the bets. Note that, although our dataset contains hourly data over the course of the first 30 days, we only recorded a single 'price snapshot' to measure 60 day effects.

Table 2 displays the results. As one would expect, the standard error of $\hat{\beta}_1$ has risen further over time (increasing from .004 after 1 week to .009 after 30 days and .014 after 60 days). Despite this, however, significant effects (with $p < 0.01$) are clearly visible in the data even 60 days after the bets have been made. As can be seen from Column (1), the coefficient $\hat{\beta}_1$ has fallen to .056; that is, prices have now reverted by approximately 44%. However, this includes markets that had already resolved within 60 days; and one would not expect to be able to manipulate the '60 day price' of these markets (which should generally end up close to 0% or 100% regardless of our intervention). Indeed, no significant effects are observed on the '60 day price' of already resolved markets; see Column (2). Once these markets are excluded from the sample, the coefficient on $\hat{\beta}_1$ becomes .059, which suggests that only 41% of reversion has taken place; see Column (3) for the details. Thus, the effects of our manipulations can be clearly seen in the data even 60 days after they have taken place, especially if one restricts attention to markets that had not resolved by that time.

Table 2: 60 day effects

Variable	(1) 60 day effects	(2) 60 day effects	(3) 60 day effects
Yes	0.0559***	0.0855	0.0586***
	[0.0135]	[0.183]	[0.0114]
Control	0.0122	-0.155	0.0258**
	[0.0135]	[0.138]	[0.0103]
$p_{-1,i}$	1.020***	0.860***	1.035***
	[0.0258]	[0.277]	[0.0209]
Constant	-0.0474***	0.0835	-0.0596***
	[0.0141]	[0.135]	[0.0110]
n	817	46	771
R^2	0.617	0.156	0.723

Notes. This table shows the results of estimating regression (7) after 60 days. Column (1) reports results for the full sample; Column (2) for markets that had already resolved; and Column (3) for markets that had not already resolved. Robust standard errors in parentheses (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

Robustness. The previous sections demonstrate that, on average, market prices somewhat revert following manipulative trades. However, even 60 days after the trades have taken place, their impacts can still be clearly seen in the data. Since this result contrasts with the general message from the previous literature, it is important to now assess its robustness. To do this, we focus on the comparison of the 'yes' and 'no' groups within 30 days of our

bets (in line with our pre-registered analysis plan).

First, we re-estimate our main specification while controlling for *all* baseline variables that are available. The results are shown in Columns (2) and (4) of Table 3; as a comparison, Columns (1) and (3) show the results from the baseline specification (which just controls for $p_{-1,i}$). As can be seen, adding the baseline controls makes very little difference to the estimates: for instance, the estimated level of reversion changes from 25% to 24% after 1 week. In addition, the standard errors of $\hat{\beta}_1$ remain very stable once the controls are added, suggesting that controlling for variables above and beyond the previous price $p_{-1,i}$ adds little statistical power.

Table 3: Results with all baseline controls

	(1)	(2)	(3)	(4)
Variable	1 week effects	1 week effects	30 day effects	30 day effects
Yes	0.0748***	0.0759***	0.0686***	0.0697***
	[0.00446]	[0.00431]	[0.00908]	[0.00912]
Control	0.0313***	0.0305***	0.0239***	0.0235***
	[0.00484]	[0.00490]	[0.00853]	[0.00854]
$p_{-1,i}$	1.000***	1.003***	1.028***	1.030***
	[0.00781]	[0.00781]	[0.0161]	[0.0155]
Constant	-0.0371***	-0.0297***	-0.0525***	-0.0644***
	[0.00547]	[0.0112]	[0.00910]	[0.0198]
n	817	817	817	817
R^2	0.933	0.937	0.798	0.803

Notes. This table shows the estimated $\hat{\beta}_1$ coefficients obtained for $t = 167$ (Columns 1 and 2) and $t = 719$ (Columns 3 and 4). Columns (1) and (3) present the results under the main specification; whereas Columns (2) and (4) control for all baseline variables. Robust standard errors in parentheses (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

Next, we examine whether our results might be biased by spillovers; i.e. by the possibility that betting on one market influences the price of another market in our sample due to either arbitrage or informational effects. As our analysis in Appendix C makes clear, it is extremely unlikely that spillovers are responsible for our results. First, since we deliberately avoided selecting 'highly correlated' markets, it is unclear *a priori* whether our interventions generated any spillovers. Second, even if they did generate spillovers, these should cancel out due to the randomness of the treatment assignment (see Appendix C for elaboration). Third, and perhaps most importantly, if spillovers are driving the results, then one would expect that dropping the more closely related markets would change the estimates. However,

dropping such markets turns out not to have any appreciable effect on our estimates.

5.3 HETEROGENEITY

While the previous section reports the results for the full sample, the model we develop in Section 2 suggests that the degree of reversion that is observed should systematically depend on the market type. First, the model predicts that markets with less scope for learning should be harder to manipulate. To operationalise this notion, we consider whether each market question was 'duplicated' on Metaculus since this provides an external probability estimate and thus should lessen the need for traders to base their personal probability estimates on Manifold's market price.[20] Second, we consider each market's level of 'activity' as measured by its total trading volume, trading volume in the last 24 hours, and total number of user comments. As discussed earlier, more active markets should revert more quickly since they effectively run on a faster clock speed. Third, we consider each market's number of traders; as discussed in Section 2, markets with more traders should also be harder to manipulate.

As outlined in our pre-registration (Rasooly and Rozzi, 2023), we approach this question through median splits. That is, for every continuous moderating variable v_i, we construct a variable x_i that equals 1 if and only if a market's v_i value is at least the median of v_i.[21] When v_i is binary, we simply set $x_i = v_i$. We then estimate the regression

$$p_{167,i} = \beta_0 + \beta_1 \mathbb{1}_{Y,i} + \beta_2 \mathbb{1}_{C,i} + \beta_3 p_{-1,i} + \beta_4 \mathbb{1}_{Y,i} x_i + \beta_5 \mathbb{1}_{C,i} x_i + \beta_6 p_{-1,i} x_i + u_i \qquad (8)$$

where $\mathbb{1}_{Y,i}$ and $\mathbb{1}_{C,i}$ are defined as before. Notice that the coefficient β_4 is precisely the difference in treatment effects between markets with $x_i = 0$ (below median) and markets with $x_i = 1$ (above median). By obtaining the standard error of $\hat{\beta}_4$, one can then assess if any differences that are observed are statistically significant.

Table 4 reports the 1 week treatment effects conditional on $x_i = 0$ and $x_i = 1$ (Columns 1 and 2); it also displays the differences between treatment effects $\hat{\beta}_4$ along with the p-values corresponding to the test that these differences are zero. Three results are apparent. First, markets that are also present on Metaculus appear to be harder to manipulate: after 1 week, they exhibit 47% reversion in contrast to the 23% reversion that is observed for the non-Metaculus markets. While this result is in line with theoretical expectations, it should be noted that, statistically, the difference between the two groups is only significant at the

---

[20] Whether a market is duplicated on Metaculus is also a reasonable proxy for whether the market can also be found on other prediction market platforms, which in turn should also blunt price learning effects.

[21] In the case where the majority of entries are zero, we define the median as the smallest non-zero number and split the sample at this point.

10% level. Second, markets with higher levels of 'activity' (e.g., higher trading volume) are also harder to manipulate. For example, markets with above median total volume exhibit 31% reversion over the course of a week; whereas markets with below median total volume exhibit just 19% reversion. Again, these results are in line with theoretical expectations but not always significant: one observes highly significant results when examining 24 hour volume, but somewhat less significant results when examining total volume and the total number of comments. Third, as one would expect, markets with a larger number of traders are also harder to manipulate: reversion for markets with an above median number of traders is 33%, whereas reversion for markets with a below median number of traders is 17%.[22]

Table 4: Heterogeneity in treatment effects

Variable	Above median	Below median	Difference	p-value
Metaculus	0.053***	0.077***	0.024*	0.096
24 hour volume	0.049***	0.081***	0.032**	0.041
Total volume	0.069***	0.081***	0.012	0.191
Comments	0.069***	0.084***	0.015*	0.071
Total traders	0.067***	0.083***	0.017*	0.056

Notes. This table shows how results depend on market type. The first two columns reports 1 week effects for sub-samples of the data that are above and below median values of the moderating variable. The next two columns report the difference between these effects along with the p-value corresponding to the test that this difference is zero. The asterisks correspond to the hypothesis that the relevant coefficient equals zero (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

While Table 4 reports how the estimates change following median splits, one can also split the data in a more granular way to obtain a more detailed understanding of the heterogeneities at play. Tables A3 and A4 conduct this exercise for the two moderating variables with well-defined percentile values: the number of traders and the total volume of trade. Specifically, we split the sample according to whether each of these moderating variables exceeds their 0th percentile (the full sample), their 25th percentile, their 50th percentile (the median split), and their 75th percentile values.[23] In line with the predictions of our model (see Table A1), this produces an apparently monotone pattern: as the number of traders or total volume in the sample rises, the 1 week treatment effect falls. For example, while reversion is 25% after a week for the full sample, reversion rises to 37% for the sample of markets that are

---

[22] While all three results are consistent with our model, they come with the usual caveat concerning causality: since market characteristics are not randomly assigned, one cannot be sure that any observed heterogeneities reflect the causal impact of these characteristics.

[23] Note that this exercise cannot be conducted for all variables since they need not have well-defined percentiles: for example, the Metaculus variable is binary.

in the top 25% in terms of number of traders; note that such markets have 56 traders on average and 349 final trades. However, even once one restricts attention to subgroups with a very high total volume or number of traders (which involves a substantial reduction of the sample), one can still observe clearly significant treatment effects ($p < 0.01$).

In summary, the evidence presented here suggests that the manipulability of markets varies across the sample in just the way that our model predicts. In all sub-samples that we consider, we observe the same qualitative pattern: some reversion over time, but also some scope for manipulation. However, the degree of reversion varies in the expected ways. In particular, markets with an external source of probability estimates (namely, Metaculus), markets with higher 'activity' (measured by market volume and the number of comments) and markets with a larger number of traders appear to be harder to persistently manipulate.

6. Conclusion

In their review of the existing evidence, Wolfers and Zitzewitz (2004) state that manipulation attempts do not have 'much of a discernible effect on prices, except during a short transition phase'. Our large-scale field experiment challenges this conclusion: we can detect the effects of our manipulations even 60 days after they were made. However, as predicted by our model, we also find substantial reversion ($\sim$25% after a week) and important heterogeneities in the expected directions.

Our findings somewhat confirm the concerns raised by prediction markets' critics. However, they do *not* imply that prediction markets are unhelpful or that they should be regulated further: even if manipulable, their prices can still be somewhat informative (Hanson, 2004). Moreover, although non-causal, our heterogeneity results may suggest that making prediction markets more 'active' (by encouraging higher volume, more traders, etc.) can make them more robust to manipulation attempts.

Our experiment also opens the door to a lot of future work. First, although we study 'manipulation through trade', it may also be interesting to study 'manipulation through buzz': by leaving appropriate comments, it may also be possible to systematically shift the market price.[24] Second, it would be interesting to study optimal manipulation. While very small trades cannot have a large impact on the market price, very large trades also appear unlikely to have persistent effects since they are unlikely to appear 'credible' — for this reason, one might expect a 'U-shaped' relationship between the long-run effect on prices and size of the initial shock. Although our design does not address this question since it holds

---

[24] We thank Koleman Strumpf for this suggestion.

the size of the initial shock fixed, it would be straightforward (if logistically demanding) to adapt our design so that it considers a range of initial shock sizes.

REFERENCES

Allen, F. and Gale, D. (1992). Stock-price manipulation. *The Review of Financial Studies*, 5(3):503–529.

Arrow, K. J. (1965). The theory of risk aversion. In *Aspects of the Theory of Risk Bearing*, chapter 2. Yrjo Jahnsonin Saatio, Helsinki.

Aumann, R. J. (1964). Markets with a continuum of traders. *Econometrica*, 32:39–50.

Bénabou, R. and Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review*, 96(5):1652–1678.

Berg, J., Forsythe, R., Nelson, F., and Rietz, T. (2008). Results from a dozen years of election futures markets research. *Handbook of experimental economics results*, 1:742–751.

Brown, A. and Yang, F. (2017). Anchoring and manipulation in speculative markets: A field experiment. *Available at SSRN 2898142*.

Buckley, P. and O'Brien, F. (2017). The effect of malicious manipulations on prediction market accuracy. *Information Systems Frontiers*, 19:611–623.

Camerer, C. F. (1998). Can asset markets be manipulated? A field experiment with racetrack betting. *Journal of Political Economy*, 106(3):457–482.

Choo, L., Kaplan, T. R., and Zultan, R. (2022). Manipulation and (mis) trust in prediction markets. *Management Science*, 68(9):6716–6732.

Deck, C., Lin, S., and Porter, D. (2013). Affecting policy by manipulating prediction markets: Experimental evidence. *Journal of Economic Behavior & Organization*, 85:48–62.

EA Forum (2022). Manifold Markets interview with Joel Becker.

EA Forum (2024). Predictive Performance on Metaculus vs. Manifold Markets.

Fama, E. F. (1970). Efficient capital markets. *Journal of Finance*, 25(2):383–417.

Figlewski, S. (1979). Subjective information and market efficiency in a betting market. *Journal of Political Economy*, 87(1):75–88.

Filippin, A. and Mantovani, M. (2023). Risk aversion and information aggregation in binary-asset markets. *Quantitative Economics*, 14(2):753–798.

FT (2023). If prediction markets can tell the future, why is the US so afraid of them? *Financial Times Ltd.*

Gjerstad, S. (2005). Risk aversion, beliefs, and prediction market equilibrium. *Economic Science Laboratory, University of Arizona.*

Hansen, J., Schmidt, C., and Strobel, M. (2004). Manipulation in political stock markets–preconditions and evidence. *Applied Economics Letters*, 11(7):459–463.

Hanson, R. (2004). Foul play in information markets. *George Mason University.*

Hanson, R. and Oprea, R. (2009). A manipulator can aid prediction market accuracy. *Economica*, 76(302):304–314.

Hanson, R., Oprea, R., and Porter, D. (2006). Information aggregation and manipulation in an experimental market. *Journal of Economic Behavior & Organization*, 60(4):449–459.

Jian, L. and Sami, R. (2012). Aggregation and manipulation in prediction markets: Effects of trading mechanism and information distribution. *Management Science*, 58(1):123–140.

Kumar, P. and Seppi, D. J. (1992). Futures manipulation with "cash settlement". *The Journal of Finance*, 47(4):1485–1502.

Manifold Markets (2022). Maniswap.

Manifold Markets (2023). Track Record and Accuracy.

Manifold Markets (2024). Active Users.

Manifold Survey (2024). Survey Results Pt. 1.

Manski, C. F. (2006). Interpreting the predictions of prediction markets. *Economics Letters*, 91(3):425–429.

Mantovani, M. and Filippin, A. (2024). When do prediction markets return average beliefs? Experimental evidence. *University of Milan Bicocca Department of Economics, Management and Statistics Working Paper.*

NYT (2023). The Wager That Betting Can Change the World. *The New York Times.*

Oprea, R., Porter, D., Hibbert, C., Hanson, R., and Tila, D. (2008). Can manipulators mislead prediction market observers? Working Paper 08-01, Chapman University, Economic Science Institute.

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

Pennock, D. M., Lawrence, S., Giles, C. L., Nielsen, F. A., et al. (2001). The real power of artificial markets. *Science*, 291(5506):987–988.

Rasooly, I. and Rozzi, R. (2023). Experimental evidence on price manipulability (pre-registration). *AEA RCT Registry*.

Rhode, P. W. and Strumpf, K. S. (2004). Historical presidential betting markets. *Journal of Economic Perspectives*, 18(2):127–142.

Rhode, P. W. and Strumpf, K. S. (2006). Manipulating political stock markets: A field experiment and a century of observational data. *University of Arizona, mimeo*.

Roll, R. (1984). Orange juice and weather. *The American Economic Review*, 74(5):861–880.

Rothschild, D. M. and Sethi, R. (2016). Trading strategies and market microstructure: Evidence from a prediction market. *The Journal of Prediction Markets*, 10(1):1–29.

Servan-Schreiber, E., Wolfers, J., Pennock, D. M., and Galebach, B. (2004). Prediction markets: Does money matter? *Electronic markets*, 14(3):243–251.

Sigma, F. (2024). Comparing election forecast accuracy.

Stiglitz, J. E. (2003). Terrorism: there's no futures in it. *Los Angeles Times*.

Veiga, H. and Vorsatz, M. (2009). Price manipulation in an experimental asset market. *European Economic Review*, 53(3):327–342.

Veiga, H. and Vorsatz, M. (2010). Information aggregation in experimental asset markets in the presence of a manipulator. *Experimental Economics*, 13(4):379–398.

Vox (2024). Why prediction markets are bad at predicting who'll be president. *Vox*.

Wolfers, J. and Leigh, A. (2002). Three tools for forecasting federal elections: Lessons from 2001. *Australian Journal of Political Science*, 37(2):223–240.

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

Wolfers, J. and Zitzewitz, E. (2006). Interpreting prediction market prices as probabilities. *National Bureau of Economic Research*.

Proof of Lemma 1. We begin by considering the marginal cost function. As shown in the main text, the total cost function is given by

$$C(q) = \frac{\sqrt{(n-q+y)^2 + 4nq} + q - n - y}{2} \tag{9}$$

This implies that marginal costs are given by

$$MC(q) = \frac{1}{2}\left(\frac{n+q-y}{\sqrt{(n-q+y)^2 + 4nq}} + 1\right) \tag{10}$$

The three properties follow from this formula:

(i) To show that $MC(0) = n/(n+y)$, one simply substitutes $q = 0$ into (10) and simplifies the resulting expression.

(ii) To show that $MC'(q) > 0$ for all $q \geq 0$, one computes the derivative

$$MC'(q) = \frac{2ny}{\left((n-q+y)^2 + 4nq\right)^{3/2}} \tag{11}$$

which is positive under the restrictions that $n > 0$, $y > 0$ and $q \geq 0$.

(iii) To show that $\lim_{q\to\infty} MC(q) = 1$, one notes that

$$
\begin{aligned}
\lim_{q\to\infty} MC(q) &= \frac{1}{2}\lim_{q\to\infty}\frac{n+q-y}{\sqrt{(n-q+y)^2 + 4nq}} + \frac{1}{2}\\
&= \frac{1}{2}\lim_{q\to\infty}\sqrt{\frac{(n+q-y)^2}{(n-q+y)^2 + 4nq}} + \frac{1}{2}\\
&= \frac{1}{2}\sqrt{\frac{1}{1}} + \frac{1}{2}\\
&= 1
\end{aligned}
\tag{12}
$$

where the penultimate equality follows after expanding out the terms and dividing all terms by the leading power q^2. This establishes the final property of the marginal cost function.

We now show that these properties transfer to the average cost function:

(i) To show that $\lim_{q \to 0^+} AC(q) = n/(n+y)$, note that

$$\lim_{q \to 0^+} AC(q) = \lim_{q \to 0^+} \frac{C(q)}{q} = \lim_{q \to 0^+} \frac{C(0+q) - C(0)}{q} = MC(0) = \frac{n}{n+y} \tag{13}$$

where the second equality uses the fact that $C(0) = 0$.

(ii) To show that $AC'(q) > 0$ for all $q > 0$, we explicitly compute the derivative

$$\frac{\partial AC}{\partial q} = \frac{1}{2q^2} \left(-\sqrt{(n-q+y)^2 + 4nq} + \frac{q(n+q-y)}{\sqrt{(n-q+y)^2 + 4nq}} + n + y \right) \tag{14}$$

Multiplying by $2q^2\sqrt{(n-q+y)^2 + 4nq} > 0$, this has the same sign as

$$-(n-q+y)^2 - 4nq + q(n+q-y) + (n+y)\sqrt{(n-q+y)^2 + 4nq} \tag{15}$$

We want to show that this expression is positive, or

$$(n+y)\sqrt{(n-q+y)^2 + 4nq} > -[q(n+q-y) - (n-q+y)^2 - 4nq] \tag{16}$$

After squaring both sides and simplifying, the condition becomes $nq^2y > 0$, which holds since $n, q, y > 0$. Thus, $AC'(q) > 0$ for all $q > 0$ as claimed.

(iii) To show that $\lim_{q \to \infty} AC(q) = 1$, notice that, as $q \to \infty$, both $C(q) \to \infty$ and $q \to \infty$. Since both are differentiable, we can evaluate the limit of $AC(q)$ using L'Hôpital's rule:

$$\lim_{q \to \infty} AC(q) = \lim_{q \to \infty} \frac{C(q)}{q} = \lim_{q \to \infty} \frac{C'(q)}{1} = \lim_{q \to \infty} MC(q) = 1 \tag{17}$$

This establishes the final property of the average cost function. $\qquad \square$

Proof of Lemma 2. We consider the cases separately:

Case 1: $\pi_i > p$. First, suppose that the agent purchases a positive quantity $q_n > 0$ of no

shares. Then the expected value of their holdings W would be:

$$
\begin{aligned}
\mathbb{E}[W] &= \pi_i(w - q_n AC(q_n)) + (1 - \pi_i)(w - q_n AC(q_n) + q_n) \\
&< \pi_i\left(w - q_n\frac{y}{n+y}\right) + (1 - \pi_i)\left(w - q_n\frac{y}{n+y} + q_n\right) \\
&= w + q_n\left(\frac{n}{n+y} - \pi_i\right) \\
&< w
\end{aligned}
\tag{18}
$$

Here, the first inequality holds since $AC(q_n) > y/(n+y)$ for any $q_n > 0$ (by the no shares analogue of Lemma 1); the next equality holds via algebraic manipulation; and the final inequality holds since $\pi_i > p \equiv n/(n+y)$. Thus, holding no shares reduces the expected value of the agent's holdings. Since it also exposes the agent to risk, and the agent is risk averse, it therefore cannot be optimal.

Next, suppose that the agent purchases a quantity $q_y \geq 0$ of yes shares (possibly, $q_y = 0$). Their expected utility is then

$$
\mathbb{E}[u(W)] = \pi_i u(w + q_y - C(q_y)) + (1 - \pi_i)u(w - C(q_y))
\tag{19}
$$

Using the chain rule, this has derivative

$$
\frac{\partial \mathbb{E}[u(W)]}{\partial q_y} = \pi_i u'(w + q_y - C(q_y))(1 - C'(q_y)) - (1 - \pi_i)u'(w - C(q_y))C'(q_y)
\tag{20}
$$

Recalling Lemma 1, we see that

$$
\frac{\partial \mathbb{E}[u(W)]}{\partial q_y}\bigg|_{q_y=0} = \pi_i u'(w)\left(1 - \frac{n}{n+y}\right) - (1 - \pi_i)u'(w)\left(\frac{n}{n+y}\right)
\tag{21}
$$

$$
= u'(w)\left(\pi_i - \frac{n}{n+y}\right)
\tag{22}
$$

Since $u'(w) > 0$ and $\pi_i > n/(n+y)$, this expression is positive and so $q_y = 0$ cannot be optimal: there must exist some $\epsilon > 0$ such that $q_y = \epsilon$ generates higher expected utility.

The previous arguments show that neither $q_n > 0$ nor $q_y = 0$ are optimal choices. However, since the trader's choice set is compact (recall the budget constraint) and their objective function is continuous (recall that u is continuous and that the cost function is polynomial), an optimal choice must exist. From this, it follows that some $q_y > 0$ must be optimal.

Case 2 $\pi_i = p$. Using the same reasoning as before, we see that, if the agent purchases no

shares, then the expected value of their holdings will be

$$\mathbb{E}[W] < w + q_n \left(\frac{n}{n+y} - \pi_i \right) = w \tag{23}$$

Thus, buying no shares reduces the expected value of their holdings. Since it exposes them to risk, and they are risk averse, this cannot be optimal. Analogously, if they purchase yes shares, the expected value of their holdings becomes

$$\mathbb{E}[W] < w + q_y \left(\pi_i - \frac{n}{n+y} \right) = w \tag{24}$$

Thus, purchasing yes shares also reduces the expected value of their holdings and thus cannot be optimal. However, an optimal choice does exist (see above). From this, it follows that the optimal choice is $q_n = q_y = 0$.

Case 3 $\pi_i < p$. This case can be handled in a similar way to the case of $\pi_i > p$. That is, one first shows that purchasing yes shares cannot be optimal since it reduces the expected value of the agent's holdings. By examining the derivative of expected utility when $q_n = 0$, one then shows that $q_n = 0$ cannot be optimal either. $\qquad \square$

Proof of Lemma 3. First, we use Lemma 2 to pin down the types of share that each type of trader must buy. Notice that:

- If $\pi_i \geq p + \Delta$, then $\pi_i \geq p$ (since $\Delta > 0$). Thus, Lemma 2 implies that such types must hold a weakly positive quantity of yes shares both before and after the price change.

- If $\pi_i \in (p, p + \Delta)$, Lemma 2 implies that such types would buy yes shares before the price change but buy no shares after the price change.

- If $\pi_i \leq p$, then $\pi_i \leq p + \Delta$. Thus, Lemma 2 implies that such types must hold a weakly positive quantity of no shares both before and after the price change.

Next, we need to show that types with $\pi_i \geq p + \Delta$ will reduce their holdings of yes shares q. If $\pi_i = p + \Delta$, the statement is trivial: by Lemma 2, such types will choose $q > 0$ before the price change but $q = 0$ after the price change. Next, suppose that $\pi_i = 1$. For such types, expected utility reduces to $u(w + q - C(q))$, which can be transformed to $w + q - C(q)$. Differentiating with respect to q, this becomes $1 - C'(q) > 0$ since $C'(q) < 1$ for any finite q (by Lemma 1). Thus, such types will purchase the maximum feasible number of yes shares, i.e. $\bar{q}$ such that $C(\bar{q}) = w$. Since C is strictly increasing in q and so invertible, $\bar{q} = C^{-1}(w)$. Moreover, one can verify (see below) that increasing the price — which is equivalent to

increasing n and decreasing y while holding the product ny constant — increases $C(q)$ for any $q \geq 0$, thus decreasing $C^{-1}(q)$. Hence, increasing the price decreases $\bar{q} = C^{-1}(w)$; that is, it induces the trader to reduce their holdings of yes shares (as claimed).

Having shown that the statement holds when $\pi_i = p + \Delta$ or $\pi_i = 1$, we now consider traders whose beliefs satisfy $\pi_i \in (p + \Delta, 1)$. As shown by Lemma 2, $q = 0$ cannot be optimal for such types. Meanwhile, since $\lim_{w_s \to 0^+} u'(w_s) = \infty$, it cannot be optimal to choose $q = \bar{q}$, thereby obtaining zero wealth in one of the states. Hence, $q \in (0, \bar{q})$. Since expected utility is differentiable, this means that any optimal q must satisfy the first order condition

$$\pi_i u'(w + q - C(q))(1 - C'(q)) - (1 - \pi_i)u'(w - C(q))C'(q) = 0 \tag{25}$$

Moreover,

$$
\begin{aligned}
\frac{\partial^2 \mathbb{E}[u(W)]}{\partial q^2} = &\; \pi_i \left[u''(w + q - C(q))(1 - C'(q))^2 - C''(q)u'(w + q - C(q)) \right] \\
&- (1 - \pi_i) \left[-C'(q)^2 u''(w - C(q)) + u'(w - C(q))C''(q) \right]
\end{aligned}
\tag{26}
$$

which is negative under the restrictions that $u' > 0$, $u'' < 0$, $C' > 0$ and $C'' > 0$. Thus, the first order condition (25) has a unique solution.

Rewriting (25), we obtain

$$\frac{\pi_i u'(w + q - C(q))}{(1 - \pi_i)u'(w - C(q))} = \frac{C'(q)}{1 - C'(q)} \tag{27}$$

Notice that, as the price rises — which is equivalent to increasing n and decreasing y while holding the product $k = ny$ constant — both $C(q)$ and $C'(q)$ may change. To make this dependence explicit, we can write the first order condition as $f(q, y, n) = g(q, y, n)$, where f and g are respectively defined as the left hand side and right hand side of (27). Since y is a function of n (specifically, $y = k/n$), this can be simplified to $f(q, n) = g(q, n)$. To see how the optimal q varies with the price (i.e. n), one can apply the implicit function theorem

$$\frac{dq}{dn} = \frac{\frac{\partial g(q,n)}{\partial n} - \frac{\partial f(q,n)}{\partial n}}{\frac{\partial f(q,n)}{\partial q} - \frac{\partial g(q,n)}{\partial q}} \tag{28}$$

To compute the sign of (28), we now compute the signs of every term:

⬦ Observe that

$$\frac{\partial g(q,n)}{\partial n} = \frac{\partial g(q,n)}{\partial C'(q)} \frac{\partial C'(q)}{\partial n} = \frac{1}{(1 - C'(q))^2} \frac{\partial C'(q)}{\partial n} \tag{29}$$

Moreover, after substituting $y = k/n$ into the marginal cost function, one finds that

$$\frac{\partial C'(q)}{\partial n} = \frac{2kn\left(k + n^2\right)}{\left(k^2 + 2kn(n-q) + n^2(n+q)^2\right)^{3/2}} > 0 \tag{30}$$

From (29) and (30), it follows that $\frac{\partial g(q,n)}{\partial n} > 0$.

$\diamond$ Next, observe that

$$\frac{\partial f(q,n)}{\partial n} = \frac{\partial f(q,n)}{\partial C(q)}\frac{\partial C(q)}{\partial n} \tag{31}$$

Since $C(q) = \int_0^q C'(q)dq$ and $\frac{\partial C'(q)}{\partial n} > 0$, $\frac{\partial C(q)}{\partial n} > 0$. Moreover,

$$\begin{aligned}\frac{\partial f(q,n)}{\partial C(q)} &= \frac{\pi_i}{(1-\pi_i)}\frac{\partial}{\partial C(q)}\frac{u'(w+q-C(q))}{u'(w-C(q))}\\ &= \frac{\pi_i}{(1-\pi_i)}\frac{u''(w-C(q))u'(w+q-C(q)) - u''(w+q-C(q))u'(w-C(q))}{u'(w-C(q))^2}\end{aligned} \tag{32}$$

which has the same sign as

$$u''(w-C(q))u'(w+q-C(q)) - u''(w+q-C(q))u'(w-C(q)) \tag{33}$$

We claim that this expression is negative, or

$$-\frac{u''(w-C(q))}{u'(w-C(q))} > -\frac{u''(w+q-C(q))}{u'(w+q-C(q))} \tag{34}$$

which holds under the assumption of decreasing absolute risk aversion. Hence, $\frac{\partial f(q,n)}{\partial C(q)} < 0$. From this and (31), it follows that $\frac{\partial f(q,n)}{\partial n} < 0$.

$\diamond$ Next, observe that

$$\frac{\partial f(q,n)}{\partial q} = \frac{\pi_i}{1-\pi_i}\frac{\partial}{\partial q}\frac{u'(w+q-C(q))}{u'(w-C(q))} \tag{35}$$

Since $\pi_i/(1-\pi_i) > 0$, this has the same sign as

$$\frac{u''(w+q-C(q))u'(w-C(q))(1-C'(q)) + u''(w-C(q))u'(w+q-C(q))C'(q)}{u'(w-C(q))^2} \tag{36}$$

which is negative given that $u' > 0$, $u'' < 0$ and $C'(q) \in (0,1)$. Hence, $\frac{\partial f(q,n)}{\partial q} < 0$.

$\diamond$ Finally, observe that

$$\frac{\partial g(q,n)}{\partial q} = \frac{\partial g(q,n)}{\partial C(q)}C'(q) = \frac{C'(q)}{(1-C'(q))^2} > 0 \tag{37}$$
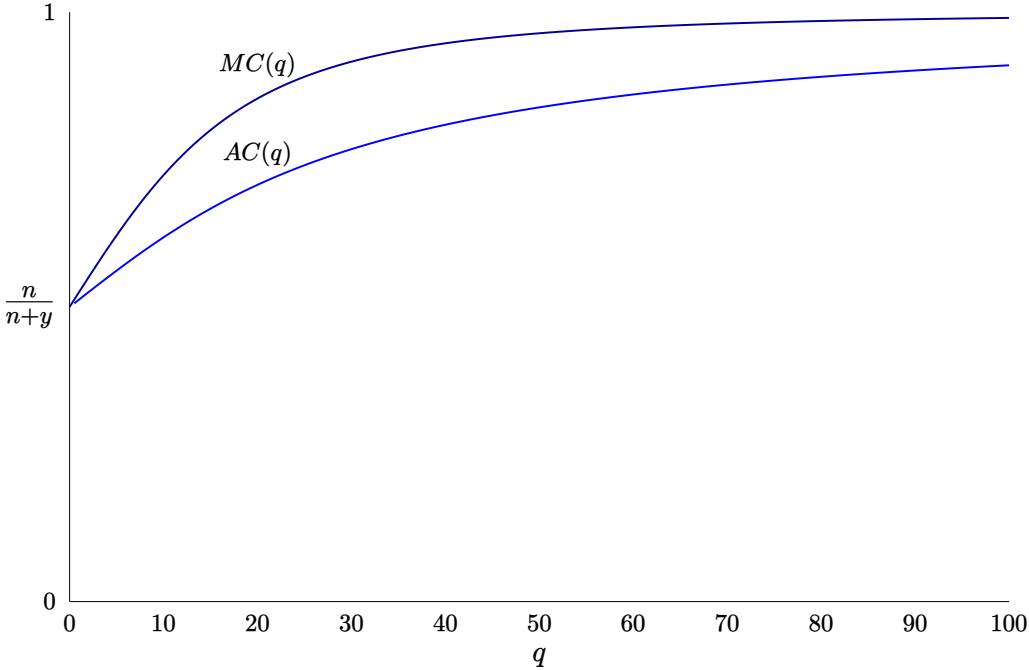
In summary, $\frac{\partial g(q,n)}{\partial n} > 0$, $\frac{\partial f(q,n)}{\partial n} < 0$, $\frac{\partial f(q,n)}{\partial q} < 0$ and $\frac{\partial g(q,n)}{\partial q} > 0$. It follows that

$$\frac{dq}{dn} = \frac{\frac{\partial g(q,n)}{\partial n} - \frac{\partial f(q,n)}{\partial n}}{\frac{\partial f(q,n)}{\partial q} - \frac{\partial g(q,n)}{\partial q}} < 0 \tag{38}$$

Thus, any trader with $\pi_i \geq p + \Delta$ will decrease their holdings of yes shares. By an exactly symmetric argument (which we omit for brevity), one can also show that any trader with $\pi_i \leq p$ will increase their holdings of no shares. $\qquad\square$

Figure A1: Costs under the constant product rule $(n = y = 10)$

Notes. This figure displays marginal and average costs under the constant product rule with initial reserves $(y, n) = (10, 10)$.

Table A1: Reversion coefficients under different market conditions

λ	0	0.2	0.4	0.6	0.8	1
SR reversion	14.5%	13.3%	11.2%	8.3%	4.5%	0.0%
LR reversion	39.8%	37.3%	33.0%	26.2%	15.9%	0.0%
α	0	0.2	0.4	0.6	0.8	1
SR reversion	22.7%	22.4%	21.5%	19.9%	17.6%	14.5%
LR reversion	52.3%	51.8%	50.6%	48.4%	44.9%	39.8%
m	10	20	30	40	50	60
SR reversion	14.5%	16.0%	16.7%	16.9%	18.5%	18.6%
LR reversion	39.8%	57.1%	67.1%	74.4%	79.3%	82.0%

Notes. This table shows the short-run (SR) and long-run (LR) reversion coefficients under different market conditions. SR reversion is defined using $(\bar{p}_3 - \bar{p}_0)/0.05$ and LR reversion using $(\bar{p}_{100} - \bar{p}_0)/0.05$, where $\bar{p}_t$ is the average price across the 10,000 simulations at time t and $t = 0$ is the time of the manipulation. The coefficients λ, α and m respectively denote the learning rate, the belief parameter and the number of traders.

Figure A2: An example of a market on Manifold.

Notes. This figure shows an example of a market on the Manifold platform.

Figure A3: The impact of the Mana devaluation on GiveWell donations.
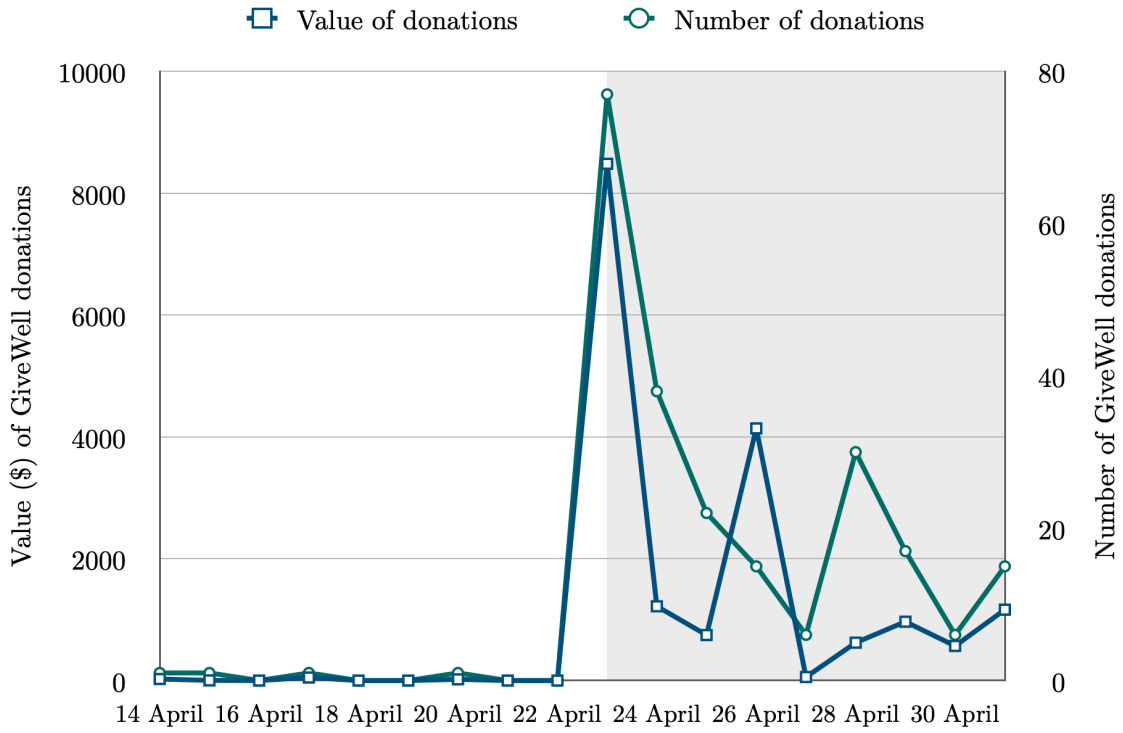
Notes. This figure shows the daily number and value (in USD) of donations to GiveWell's Maximum Impact Fund by Manifold users from 14 April 2024 to 1 May 2024. Dates after and including 23 April (the shaded area) represent dates after the announcement of the devaluation.

Table A2: Topics

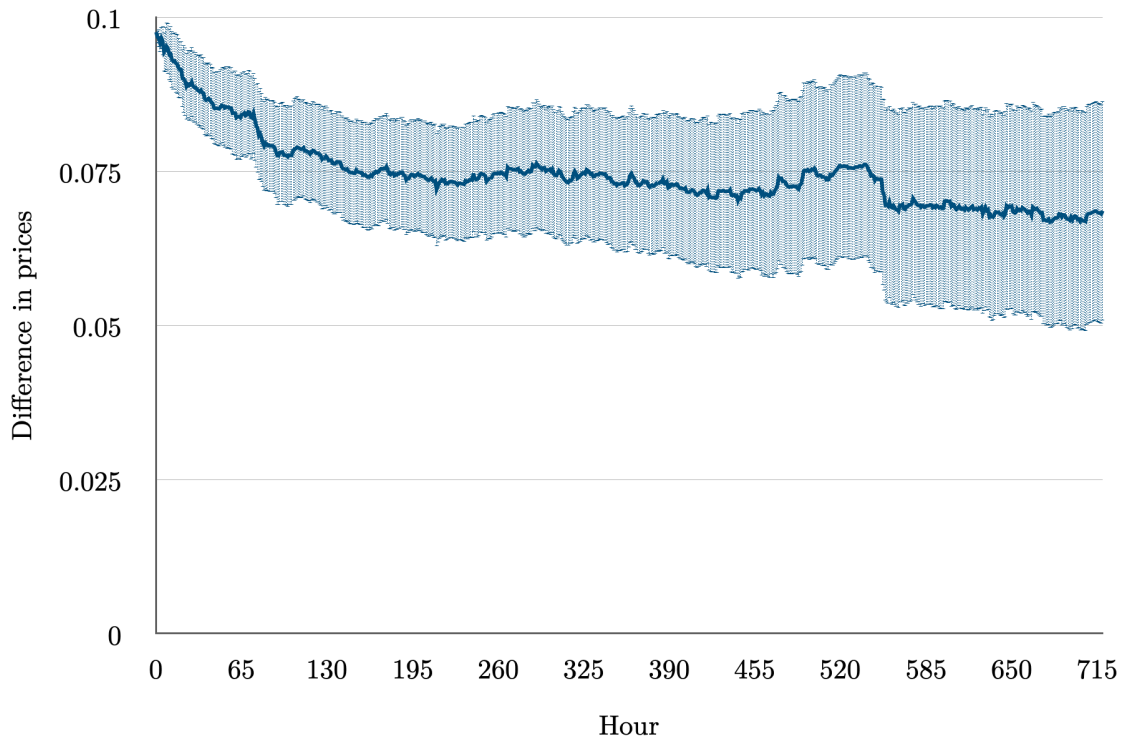| Topic	Frequency	Percentage
Artificial intelligence	103	12.61
US politics	50	6.12
Macroeconomics	31	3.79
Israel and the Palestinian territories	30	3.67
YouTube and its streamers	27	3.3
Video games	24	2.94
Manifold Markets	22	2.69
The Russo-Ukrainian War	21	2.57
Football	19	2.33
American football	17	2.08
Legal rulings	16	1.96
Television	14	1.71
Basketball	12	1.47
X (formerly Twitter)	11	1.35
Taylor Swift	10	1.22
UK politics	10	1.22
Cryptocurrency	9	1.1
Elon Musk	9	1.1
The 2024 Paris Olympics	9	1.1
The Oscars	9	1.1
Russian politics	9	1.1
Apple Inc.	8	0.98
Science	8	0.98
Nuclear energy and weapons	7	0.86
Satellites and space	7	0.86
Amazon.com, Inc.	6	0.73
Canadian politics	6	0.73
Effective altruism	6	0.73
Tesla, Inc.	6	0.73
Other topics	203	36.87
Total	719	100

Notes. This table shows the topics of the questions in the sample; the topic of each question is manually assigned and can be viewed using the replication package. Topics with 5 or fewer questions are omitted from the table.

Figure A4: Markets in the sample

Notes. This figure shows a 'word cloud' formed from the market questions in our sample.

Figure A5: Comparing the 'no' and 'control' groups $(\hat{\beta}_2)$

Notes. This figure plots the $\hat{\beta}_2$ coefficients obtained from estimating (7) for $t \in \{0, 1, ..., 167\}$ along with (robust) 95% confidence intervals.

Figure A6: Comparing the 'yes' and 'no' groups ($\hat{\beta}_1$) after 30 days

Notes. This figure plots the $\hat{\beta}_1$ coefficients obtained from estimating (7) for all $t \in \{0, 1, ..., 719\}$ along with (robust) 95% confidence intervals.

Table A3: Heterogeneity by number of traders

Variable	(1) Full sample	(2) 25% split	(3) 50% split	(4) 75% split
Yes	0.0748***	0.0693***	0.0666***	0.0625***
	[0.00446]	[0.00565]	[0.00602]	[0.00999]
Control	0.0313***	0.0268***	0.0264***	0.0259**
	[0.00484]	[0.00594]	[0.00673]	[0.0108]
$p_{-1,i}$	1.000***	0.997***	1.009***	1.003***
	[0.00781]	[0.0103]	[0.0157]	[0.0283]
Constant	-0.0371***	-0.0325***	-0.0406***	-0.0378**
	[0.00547]	[0.00719]	[0.00824]	[0.0149]
n	817	640	419	207
R^2	0.933	0.915	0.908	0.838

Notes. This table shows the results of estimating regression 7 on: the full sample (Column 1), the sample with an above 25th percentile number of traders (Column 2), the sample with an above median number of traders (Column 3), and the sample with an above 75th percentile number of traders (Column 4). Robust standard errors in parentheses (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

Table A4: Heterogeneity by total volume

	(1)	(2)	(3)	(4)
Variable	Full sample	25% split	50% split	75% split
Yes	0.0748***	0.0706***	0.0692***	0.0539***
	[0.00446]	[0.00485]	[0.00623]	[0.0106]
Control	0.0313***	0.0284***	0.0374***	0.0195*
	[0.00484]	[0.00575]	[0.00661]	[0.0112]
$p_{-1,i}$	1.000***	0.999***	1.012***	0.998***
	[0.00781]	[0.0103]	[0.0124]	[0.0245]
Constant	-0.0371***	-0.0359***	-0.0479***	-0.0315**
	[0.00547]	[0.00604]	[0.00699]	[0.0130]
n	817	613	410	205
R^2	0.933	0.926	0.918	0.858

Notes. This table shows the results of estimating regression 7 on: the full sample (Column 1), the sample with an above 25th percentile total volume (Column 2), the sample with an above median total volume (Column 3), and the sample with an above 75th percentile total volume (Column 4). Robust standard errors in parentheses (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

In this section, we consider whether our results might be affected *spillovers* between the markets. In theory, there are two ways in which this might occur:

- *Informational spillovers* arise when betting on one market provides information about another market, thereby influencing its price.

- *Arbitrage spillovers* arise when betting on one market creates an arbitrage opportunity, thereby inducing trade on a closely related market.

Since we avoided betting on closely related markets, it is not clear that either type of spillover would have arisen during our experiment. For completeness, however, we now consider the possibility (and relevance) of such spillovers in some detail.

Theoretical observations. In the main analysis, we estimate the model

$$p_{t,i} = \beta_0 + \beta_1 \mathbb{1}_{Y,i} + \beta_2 \mathbb{1}_{C,i} + \beta_3 p_{-1,i} + u_i \tag{39}$$

where $\mathbb{1}_{Y,i}$ and $\mathbb{1}_{C,i}$ are the treatment dummies, $p_{-1,i}$ is the price in market i just before the bet, and u_i captures the effect on subsequent trades on the price at time t. To make the analysis more tractable, we remove the control group from consideration and study the simpler model

$$p_{t,i} = \beta_0 + \beta_1 \mathbb{1}_{Y,i} + \beta_2 p_{-1,i} + u_i \tag{40}$$

This model does not allow for the possibility of spillovers. To allow for this possibility, we suppose that a subset of the markets S generates spillovers in another subset of the markets S'. For simplicity, we assume that S and S' are disjoint and let $\sigma_{i,j}$ denote the effect of market j on market i. (Thus, $\sigma_{i,j} = 0$ if $i \notin S'$ or $j \notin S$.) We then generalise (40) to

$$p_{t,i} = \beta_0 + \beta_1 \mathbb{1}_{Y,i} + \beta_2 p_{-1,i} + \sum_{j \in S} \sigma_{i,j} \mathbb{1}_{Y,j} - \sum_{j \in S} \sigma_{i,j}(1 - \mathbb{1}_{Y,j}) + u_i \tag{41}$$

To understand (41), suppose that market i and market j are positively correlated. In that case, betting *yes* on j should increase the price of j and thereby increase the price of i: this explains the $\sigma_{i,j} \mathbb{1}_{Y,j}$ term. For the same reason, however, betting *no* on j should decrease the price of j and thereby decrease the price of i: this explains the $\sigma_{i,j}(1 - \mathbb{1}_{Y,j})$ term. Given the random assignment of markets into either the yes group Y (which increases $p_{t,i}$) or the no group N (which decreases $p_{t,i}$), there is thus no reason to expect betting on j to influence

the price of i on average. More formally, one sees that

$$\mathbb{E}[p_{t,i}] = \beta_0 + \beta_1 \mathbb{1}_{Y,i} + \beta_2 p_{-1,i} + \sum_{j \in S} \sigma_{i,j} \mathbb{P}(j \in Y) - \sum_{j \in S} \sigma_{i,j}(1 - \mathbb{P}(j \in Y)) \qquad (42)$$

and so the last two 'spillover' terms cancel (given that $\mathbb{P}(j \in Y) = 0.5$, which in turn holds since the control group is omitted from the analysis). In addition, it is worth noting that, while spillovers could be positive (in the case of positively correlated markets), they could just as well be negative — which provides an additional reason for scepticism about whether they plausibly bias the results.

Empirical results. The previous observations suggest that, even if spillovers did arise (which seems unlikely), they should not systematically bias the results. To investigate this issue empirically, we now consider the impact of dropping the 'most related' markets from our sample. If spillovers are responsible for our results, then dropping these markets should substantially change our estimates. Conversely, if spillovers are absent (as we suspect) or irrelevant (as our theoretical analysis suggests), dropping these markets should not have a large impact on our estimates.

To do this, we considered sets of markets on which we had bet within a one-week period. Notice that, if one considers 7 day effects as we do here, it is not possible for spillovers to bias the results for markets whose betting times differ by more than 7 days. We gave every such set of questions to GPT-4.0 along with the prompt:

> Some of these questions may be on the same topic. These questions have the property that, if you know the answer to one of them, you will learn a lot about the answer to another of them. If you find any such questions, can you flag them?

In a handful of cases, GPT-4.0 flagged questions that were extremely unlikely to be connected via spillovers (e.g. 'Will Jimmy Carter Die on a weekday in 2024?' *vs* 'Will Manifold require phone verification for all new users at any time in 2024?'). In such cases, we ignored GPT's suggestion. In all other cases, however, we dropped all but one of the flagged set of markets from the dataset, thereby generating a 'reduced dataset' that can be viewed in the replication package.

Table A5 displays the results of re-estimating our main analysis on the reduced dataset; to further reduce to the possibility of spillovers, we exclude the control group. As can be seen, dropping the (possibly) related markets reduces the sample from 556 to 482 markets. While this is not an especially large reduction, it should be remembered that we deliberately selected our markets to be unrelated. As can also be seen from comparing the two columns,

the results are essentially identical after dropping the 'most related' markets. This provides further evidence that spillovers are extremely unlikely to be generating our results.

Table A5: 7 day effects on the reduced sample

Variable	Main results	Reduced sample
Yes	0.075***	0.075***
	[0.004]	[0.005]
$p_{-1,i}$	1.003***	1.003***
	[0.008]	[0.009]
Constant	-0.038***	-0.037***
	[0.006]	[0.006]
R^2	0.937	0.942
n	556	482

Notes. This table shows the results of estimating regression (40) on the full sample (Column 1) and the reduced sample (Column 2), excluding the control group from all analyses. Robust standard errors in parentheses (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).