The Dancer and the Dance: Agents, Beliefs, and Actions in Prediction Markets

A thesis presented

by

Abraham M. Othman

To

Applied Mathematics

in partial fulfillment of the honors requirements

for the degree of

Bachelor of Arts

Harvard College

Cambridge, Massachusetts

April 2, 2007

Labour is blossoming or dancing where
The body is not bruised to pleasure soul.
Nor beauty born out of its own despair,
Nor blear-eyed wisdom out of midnight oil.
O chestnut-tree, great-rooted blossomer,
Are you the leaf, the blossom or the bole?
O body swayed to music, O brightening glance,
How can we know the dancer from the dance?

—W.B. Yeats (1865 - 1939)

Contents

Abstract

Prediction Markets are exchanges that trade on the outcome of future events. With the rise of the internet, these markets have been extended to domains far beyond their traditional sportsbook forebears, including political elections, future product releases, and terrorist attacks. Our work is aimed at giving the theoretical basis behind these markets a solid footing. We carefully construct two fundamentally different models from first principles: a traditional model with a pool of price-taking agents, and a novel agent-based model which simulates transactions. We use these models to analyze the efficiency and design properties of market structure, including intrinsic mispricing. Grounded in practice, our work includes suggestions for traders, market administrators, and economic scientists.

Chapter 1

Introduction

We begin by broadly introducing prediction markets and some of the motivating questions of this work generally, with a particular eye towards the Iowa Electronic Markets and Tradesports. We then discuss some of the previous theoretical models that inspired our analysis, and conclude by providing a guide to the remainder of the work.

1.1 What is a Prediction Market?

Prediction markets are exchanges which trade on the occurrence of future events, and so a typical share on a prediction market will be traded on anything uncertain happening in the near future: the winner of an election, the film that will win the Oscar for Best Picture, or the team that will win the next World Series. Prediction markets represent a hybrid of traditional sportsbooks and asset markets, incorporating the mechanism of the latter with the predictive power (and often dubious legality) of the former.

Because prediction markets are small-scale exchanges, particularly relative to the trillions exchanged on financial markets, they are seen as a more democratic enterprise than traditional asset markets. Every major investment bank has a group devoted to making sure Latin American bond prices are priced efficiently, but prediction market prices on the next winner of the Super Bowl are set by individual traders. Thus, prediction markets are often seen as a force for aggregating the "wisdom of the crowds", and they played a key role in James Surowiecki's 2004 book on the efficiency of collective knowledge [28].

Because they involve real people who often make significant mistakes in trading, (c.f. [9, 10, 5]), prediction markets are the foremost testing ground for the so-called *Hayek hypothesis*: that markets can work efficiently despite a general ignorance on the part of participants in the trading environment [25]. In a certain philosophical sense, then, prediction markets represent a laboratory for theories of the efficiency of free markets as a whole.

It was this sentiment about the power of markets to aggregate information that resulted in the design of the most famous prediction market, which collapsed under the weight of its own notoriety and was never implemented. The Policy Analysis Market was a DARPA-designed program to aggregate the opinions of members in and outside of the intelligence community through the use of a market with futures on geopolitical events, including terrorist attacks, coups, and assassinations. Ethically dubious, the project was scrapped after a congressional inquiry, with Senator Byron Dorgan of North Dakota calling the idea "useless, offensive and unbelievably stupid" [3]. Our work is primarily concerned with two extant prediction markets, the Iowa Electronic Markets (IEM) and Tradesports.

1.1.1 The Iowa Electronic Markets

A project of the University of Iowa Business School, the IEM is the most venerable prediction market, first used to successfully predict the 1988 presidential election. The IEM operates without generating profit for the exchange, and no costs, fees, or commisions are charged. The markets operate under a No-Action Letter from the Securities and Exchange Commission, which allows market participants to have a total of 500 dollars invested in the exchange at a time.

For political issues, the IEM operates two separate exchanges, the *winner-take-all* market, and the *vote-share* market. Imagine we hold a share of a Democratic candidate and a Republican candidate in an election that has concluded with the Democrat winning the election with 52 percent of the vote. Our share in the Democratic candidate in the winner-take-all market pays off at a dollar and pays off at 52 cents in the vote-share market, while our share in the Republican candidate expires worthless in the winner-take-all market and pays off at 48 cents in the vote-share market. A simple interpretation of the prices in the markets are that the winner-take-all market represents the chances a given candidate has of winning the election, while the vote-share market represents the most likely fraction of the vote a candidate will receive. Whether this interpretation is correct, that is, whether prices align with their efficient levels, is a significant motivating question in this work.

Because the IEM is an academic enterprise, it has been the subject of many inquiries. Historically, market prices in the vote-share market are closer to actual election results than national polls [4]. At the same time, individual market participants are not representative of the electorate as a whole - they are younger, whiter, and richer than the average voter and overwhelmingly male [9]. Moreover, by examining historical market prices, individual market participants often make significant trading mistakes; in a study of the 1992 presidential market it was found traders placing market orders left money on the table in nearly 40% of all transactions [21], while in the 1988 market, participants who supported a candidate were far more likely to purchase shares of that candidate [9]. The apparent paradox between individual irrationality and market efficiency is resolved by the *Marginal Trader Hypothesis*, which attributes market efficiency to a small pool of knowledgeable traders who set prices and act without bias. First introduced by Forsythe *et al.* [9], the hypothesis has become a touchstone for works on the IEM (c.f. [4, 21, 5]). Yet the skeptical reader is entirely accurate in recognizing that marginal traders play the role of a *deus ex machina* in explaining the efficiency of the markets: they come from nowhere, self-selected out of a pool of anonymous agents, and serve a crucial role. Another question that motivates our study is whether these marginal traders exist, or if we can explain empirical market results without relying on them.

1.1.2 Tradesports

Tradesports is perhaps the best-known prediction market, with exchanges operating on virtually every political, sporting, economic, or popular-interest issue that could possibly be thought to wager upon. Unlike the IEM, Tradesports operates as a for-profit exchange charging market commissions, and laws against gambling place it in a highly tenuous legal position in the United States, which is why the markets operate out of Ireland and do not advertise directly to Americans.

The IEM and Tradesports differ slightly in terms of market microstructure. The IEM is an example of a *double-share* prediction market, while Tradesports is an example of a *single-share* prediction market. To illustrate the differences in these market formats, consider the example of markets for the 2004 presidential election: The IEM had shares for both Kerry winning the election and Bush winning the election, while Tradesports had shares for only Bush winning. Note that the two market formats are precisely symmetric in the case that all orders in the Kerry market have an analogue at the complementary position within the Bush market, and generally symmetric as long as arbitrage opportunities do not arise between the complementary shares of a double market. Another difference between the two is that positions arise through short sales in a single-share market, while no shorting is necessary in a double-share market. On the whole, however, these differences are minor and are generally abstracted away.

1.2 Previous Work and Our Contributions

1.2.1 The Standard Model

The *standard model* refers to a model of markets as a mechanism for finding the equilibrium price of an asset by equalizing supply and demand. In the standard model, prices are taken as exogenous and agents are price-takers, making investment decisions to maximize utility based on those exogenous prices and individual, private beliefs over future outcomes. Manski [19] provided the first model of prediction markets that suggested inherent structural inefficiency, such that equilibrium prices would not align with their objective probabilities. Starting with a pool of risk-neutral traders in a winner-take-all context, Manski developed two results:

- First, he showed that prices do not necessarily correspond to mean beliefs. Letting F be a distribution function over beliefs, Manski's model suggested that p could be an equilibrium price exactly when:

$$p = 1 - F(p) \tag{1.1}$$

Thus, 30 cents could be an equilibrium price when it corresponds to the 70th percentile of beliefs. Since any arbitrary percentile could correspond to the mean of beliefs, there is no reason to assume prices and means of beliefs are deeply connected.

- Secondly, Manski demonstrated that prices were least accurate around 50 cents - that is to say, the largest possible difference between market prices and actual belief means was found at 50 cents. This implies prediction markets are at their least accurate precisely when we would hope for them to be most accurate - when whether an event will occur or not is most in doubt.

In two papers released independently and almost simultaneously, Wolfers and Zitzewitz [31] and Gjerstad [11] refuted Manski's claims. Both papers showed that Manski's results held only in the case of risk-neutral agents, who would invest their entire endowment if their beliefs were greater than the price of a commodity. Wolfers and Zitzewitz showed that under a logarithmic utility model, equilibrium prices would precisely align with mean beliefs. Gjerstad, going further, examined Manski's results using a number of utility models. Gjerstad demonstrated that, in general, there existed a link between equilibrium prices and belief distribution means, and that Manski's example was an extreme case.

Both papers also examined Manski's second claim regarding the accuracy of various prices. Wolfers and Zitzewitz procured the data from a prediction website in which participants would submit their probabilities for teams winning football games. The website had a complex payoff function for predictions for which participants would expect the greatest reward for correctly identifying objective probabilities. Wolfers and Zitzewitz used these belief data, combined with a variety of homogenous agent utility functions, to simulate the equilibrium prices of a prediction market. Their results indicated that prices around fifty cents were generally quite accurate. Manski's result, they showed, depends on pathological distribution functions in which beliefs are non-zero at only one or two values. Reinforcing this assertion, Gjerstad analyzed beliefs corresponding to a handful of different beta distributions (which are single-peaked and double-bounded) and used these different distributions to analyze various distributional means. It is worth noting, however, that neither of these works were grounded in empirical study of prediction markets themselves - Wolfers and Zitzewitz only *simulated* a prediction market, while Gjerstad arbitrarily assigned distributions. We are particularly concerned with the large concentration of "all-or-nothing" beliefs found in the Wolfers and Zitzewitz data, where predictors would suggest that a team had no chance of winning a particular game. We believe that the presence of such beliefs is not a realistic model for prediction markets, where we imagine seeing market prices at the extremes of the state space may cause agents to reconsider their beliefs. Thus, the rabid Washington Redskins fan who feels his team will certainly beat their rival Dallas Cowboys may reconsider his definition of "certainly" if he observes that market prices for a Redskins victory are at 98 cents.

The first significant advance we make towards modeling prediction markets is in our model of beliefs. Instead of arbitrarily assigning distributions, we ground a theoretical distributional belief model in an empirical study of previous market results. By looking at the historical results of markets, we can determine belief distributions that best fit the behavior observed in the real world.

Furthermore, ours is the first work that provides a theoretical framework that encompasses both winner-take-all as well as vote-share markets. In particular, we assume that individual market participants derive their beliefs from an underlying private distribution of beliefs regarding future outcomes. This allows us to provide the first theoretical model which argues for the efficiency of vote-share markets, such that equilibrium prices align with objective probabilities.

Another advancement we make is to give more flexibility to pricing. Previous modeling papers have all assumed that in equilibrium, market prices for complementary events sum to 1. They justify this claim by a no-arbitrage argument: if prices do not sum to 1 then there exist arbitrage opportunities to either buy or sell no-risk bundles. This simplification, however, is unsupported by many empirical and laboratory studies (c.f. [9, 10, 4, 24]). In fact, the arbitrage conditions that arise from prices not summing to 1 come from the bids on complementary events summing to more than 1, or the asks on complementary events summing to less than 1. Only in the theoretical limit of market thickness, when there are so many orders on the books that the bid-ask spread is negligible, does this "no-arbitrage" condition hold. In our agent-based model for double-share markets (discussed in chapter 8), we provide simulation results that indicate that this arbitrage argument does not hold, that prices can sum to more than 1 even with the presence of an active arbitrageur if agents consistently accept outstanding ask prices rather than outstanding bid prices. We abandon the restriction of prices summing to 1, allowing us to examine for the first time equilibria that result from the sets of "arbitrageable" price vectors so often seen in the real world.

Yet our advancements come at a significant cost. The relaxation to arbitrary price vectors and the development of a model of vote-share markets in which individual decisions do not take closed form mean that the computational issues dealt with in this paper are significantly harder than those seen in previous studies. We make use of heavy computational and mathematical optimizations, employing clever algorithms to speed computation as well as mathematical proofs to ensure rigor, to ensure that we can produce solutions to our problems. With a less nuanced approach, we would be unable to calculate the prices that correspond to equilibria for various belief distributions.

Generating these results is important because our aim is not to create the best theoretical model for prediction markets but to apply our model in the analysis of those markets, particularly with regard to explaining market inefficiency. Perhaps the most widely studied example of internal structural inefficiency in gambling relates to the *longshot bias*. First documented by Ali in 1977 [1], the longshot bias refers to a peculiar phenomenon by which strongly out-of-the-money wagers, i.e. *longshots*, are overpriced relative to their expected payouts, and strongly in-the-money wagers, i.e. *favorites*, are underpriced relative to expected payouts. The effect has been widely observed in many betting markets, including soccer and baseball [6, 32]. An orthodox microeconomic approach is given by Hurley and McDonough [15], who describe it as a product of the interaction between informed and uniformed bettors.

Manski [19] first suggested that the longshot bias might also be intrinsic to prediction markets, a significant observation given the conventional economic reasoning regarding the bias. Note that equation (1.1) suggests the longshot bias by itself, even if mean beliefs align with objective probabilities. Consider an equilibrium price of 10 cents. It can be an equilibrium price only if 90 percent of the belief density lies below it, and thus, almost certainly 10 cents will be above the mean of beliefs. The responding papers from Wolfers and Zitzewitz and Gjerstad indicated that the longshot bias could indeed occur in prediction markets, but their studies presented mixed results as to whether it would occur; for most of their hypothetical belief distributions and for most levels of risk aversion, they determined that the bias would generally be small.

Wolfers and Zitzewitz [30] presented evidence in the form of a historical data analysis, and found that the longshot bias in the IEM winner-take-all market was a significant effect, however, they did not ground their analysis of the phenomenon within an established economic model, relying instead on an ultimately unsatisfying ad-hoc model of behaviors. Thus, despite numerous inquiries, the longshot bias in prediction markets has remained a mysterious effect.

Our work on the longshot bias expands on these previous studies in three important ways: first, we provide additional evidence that the longshot bias is endemic to prediction markets by documenting it in Tradesports contracts from the 2005-06 NBA season; second, we use our work on belief distributions to demonstrate how these observed biases can arise within the standard model of risk-averse price takers; and finally, we examine the implications of our model on the vote-share market, where prices have traditionally not displayed significant one-sided inefficiencies, finding that our model indeed predicts significantly less bias in the vote-share markets than the winner-take-all markets, even for contracts for which prices should line up (i.e. shares in a candidate with a 40% chance of winning compared to shares in a candidate for whom 40% is the most likely vote share).

Additionally, our study is the first to examine equilibrium *quantities* as well as prices, which allows for a deeper exploration of market design issues. Laboratory experiments conducted by the same team of researchers responsible for the IEM have challenged the conventional economic model behind prediction markets by showing that there exists a significant bias towards mispricing of assets to greater than their equilibrium levels. They attribute most of this bias towards agents' psychological preferences for buying rather than selling. Consequently, agents need additional incentive to become sellers, which pushes prices above their expected values.

Such an explanation cannot find support within the standard model, which presupposes a Walrasian tâtonnement process and thus carries the implication that agents should have no preference for either side of a transaction [19]. Because of our methodical analysis with regards to equilibrium prices and quantities, we are able to establish an explanation for this phenomenon that fits in the standard model framework: that economic scientists have generally over-supplied cash to participants in market simulations. We demonstrate that under various plausible belief regimes, this can result in marked overpricing of commodities. This result carries important implications for laboratory studies into prediction markets, and we make suggestions that we hope will improve the effectiveness of future investigations.

1.2.2 The Agent-Based Model

At the same time that we find a possible explanation for the observed *buying bias* - the tendency for participants in a market to buy, rather than sell, shares - within markets, we note that our description of agents as price-taking, and of the market mechanism as a Walrasian auctioneer, is aesthetically unsatisfying in its solution to the observed phenomenon. Moreover, we provide the real-world evidence for the phenomenon by examining outstanding Tradesports orders, which we show skew heavily towards bids rather than asks. This implies the buying bias is a more general phenomenon that extends beyond a simple laboratory experiment giving traders too much cash. The second half of the work explores a model under which phenomena like the buying bias are easy to describe: an *agent-based model*. While the standard model takes prices as exogenous and determines demands of agents based on these prices, the agent-based model uses rules for agent behavior to endogenously determine market prices. Note how the agent-based model allows for a simple embedding of the buying bias: by simply changing the rules of agent behavior to prefer buying to selling, we clearly and effectively simulate the buying bias within the market.

Because of the obvious profits that could arise from the development of new trading strategies, agent-based models have an extensive history in modeling stock markets. The most prominent agent-based market model is the Santa Fe Institute's artificial stock market, which stems from research done in the late 1980s, at precisely the same time that the Iowa Electronic Markets began to operate. By simulating shares, transactions, and proceeds through the behavior of individual programmed market participants, the Santa Fe Institute's artificial stock market explores, simulates, and discovers opportunities in the state space of trading strategies. The market has discovered some remarkable results which correspond to results seen in real financial exchanges, like deviations from rational equilibria in asset prices, heavy-tailed return distributions, and persistent volatility. LeBaron [17] gives an account of these results and the design decisions involved in creating and exploring trading strategies.

We present, to our knowledge, the first agent-based model of a prediction market. Though prediction markets are closely related to traditional financial markets, they differ in subtle but important ways. These differences manifest themselves in our model, which breaks sharply in design from agent-based asset market models.

As an example of the differences in agent-based modeling between traditional and prediction markets, consider one of the seminal results of agent-based modeling, the "boom and busts" result first featured in Arthur *et al.* [2]. Their model attempts to discover phenomenon that arise from agent interactions to price a single asset. Agents operate using a genetic algorithm, where agents have "chromosomes" corresponding to behavior that is either based on fundamentals (the projected objective expectation of a security's value taken from dividends) or on trends (such that agents may be interested in buying or selling based on the immediate history of trades or whether the price is greater or less than a certain ratio with dividends). Successful agents, those that make the most profit, are deemed most fit for the purpose of "breeding" new traders to participate in future trading rounds (or, equivalently, individual traders modify their trading strategies to include elements of successful strategies). Arthur *et al.*'s model strongly influenced the design of many agent-based models for stock exchanges, virtually all of which incorporate the behavioral roles of "fundamentalists" and "trendists" as well as a genetic selection mechanism. An extensive survey of these agent-based models can be found in Chapter 6 of Ralph Grothmann's doctoral dissertation [13].

Arthur *et al.*'s model showed remarkable emergent behavior. Early trading rounds settled into an equilibrium based on agents behaving rationally, but then trend seeking behavior emerged, as agents began to anticipate how others would react to price changes in the rational equilibrium. As more and more agents began incorporating elements of pricing behavior that was not strictly rational, the asset settled into a pricing regime that includes dramatic booms and busts, behavior reminiscent of real markets.

If we attempt to expand this model to examine prediction markets, we are struck by an immediate quandary. Because Arthur *et al.*'s selection mechanism is genetic, the presence of agents that act on trends in long-term equilibrium pools indicates that there is a reward for agents that operate on deviations from expected, "rational" behavior. But prices in prediction markets independent of news are remarkably stable, and the boom-bust behavior is not observed in prediction markets (c.f. [9, 4]). This is because unlike traditional asset markets, prediction markets have an endpoint: a deadline after which contracts expire. As prediction markets evangelist Robin Hanson has expounded, in a philosophical sense, this teleological validation serves as an intrinsic feature which supports the elicitation of "good" information and prevents market manipulation by less objective forces, such as trend traders [14]. In the 2004 IEM presidential winner-take-all market, Koleman Strumpf placed orders of the maximum size, five hundred dollars, several times on random market sides in an effort to manipulate share prices. He found that whatever price manipulation such large orders caused was cleared away within hours of the trade being placed [27]. To put it in a different context, Keynes, who in addition to being a brilliant economist also made a fortune on the stock market, is said to have quipped "Successful investing is anticipating the anticipations of others" and "The market can stay irrational longer than you can stay solvent". Yet these statements simply do not hold in prediction markets like the IEM, where the majority of traders take buy-and-hold positions (c.f. [5]) and at the end of the day those traders who act well on the best information will have the highest returns. At the same time, the concept of having a "true valuation" of a company, which is relatively easy to determine endogenously by an agent with expectations of dividends over the future, is not present within a prediction market. The market most subject and visible to this kind of accurate valuation would certainly be the IEM vote-share markets, where participants are regularly appraised through polls disseminated in the national media of the percentages of likely voters who intend to cast their votes for major candidates. However, Leigh and Wolfers [18] provide data from the most recent Australian election that indicates prediction markets do not lag polls; that is to say, the issuance of a new poll does not provide any information already present in the market. While the model of an infinite time horizon exchange with common knowledge "true valuation" is entirely acceptable for the modeling of an individual asset, we feel that this is not the correct way to think about individual traders in a prediction market. Instead, just as in the standard model, we model agents as having individual beliefs over the probability of an event occurring, and act in the market based on those beliefs.

In this sense, our work draws inspiration from Feigenbaum *et al.* [8]. That paper was framed in a distributed computation context, with each agent possessing an individual piece of binary data that would serve as inputs for a (common knowledge) binary function. By implementing a market structure in which agents would trade based on their expectations, Feigenbaum *et al.* developed sufficient conditions for functions such that the market would converge to the correct return value of the function, as well as bounds for how many rounds of market activity such convergence would require. The analogue to prediction markets is evident in this model - each trader possesses a piece of private information about how an election will turn out, and the "output function" is the result of the election. However, this theoretical model does not provide an adequate basis for agent-based inquiry because of two severe economic flaws in the model: First, the model mandates that all agents have both perfect knowledge and perfect rationality. They are fully aware of the number of participants in the market, of the computational function (which could perhaps be regarded as the mapping from participant beliefs to election outcomes in a prediction market, rather than a computational, context), and are aware that all other agents in the market have the same qualities. Secondly, the agents in the market interact without any regard to incentives - in each round, they place an order at their current expectation of the function's return value. The agents are perfectly rational and back out this information from a bayesian prior on the beliefs of the other agents, updating this prior as they learn the market-clearing price of each successive trading round. Moreover, the agents are forced to trade in every round, in contradiction to the no-trade theorem of Milgrom and Stokey [20], which applies in this hyper-rational context.

Thus, this is a poor model for individual market participants in the real world. Perfect rationality provides a standard that neither humans nor computers can meet generally, while the free surrendering of information without any expectation of profit abstracts away the very motive that drives markets.

Rather than attempting to design an ad-hoc model of human behavior, a technique which is perpetually open to criticism of modeling decisions, we instead adopt *Zero-Intelligence* (ZI) agents. ZI agents were first featured in work by Gode and Sunder [12], who posited that a ZI agent "has no intelligence, does not seek or maximize profits, and does not observe, remember, or learn". Gode and Sunder used their ZI agents to simulate a double-auction market place, where a pool of ZI agents and sellers would place random orders and the market would determine clearing prices and quantities. They found that by imposing what they called a "budget constraint", that is, restricting the space of agent decisions to only those with positive expected payoff, that their markets were virtually as efficient as those featuring profit-motivated humans traders in identical laboratory simulations.

Yet if Feigenbaum *et al.* went too far to the side of rationality with their market model, then it would seem that ZI agents err to the opposite extreme; we cannot truly be suggesting that people do not learn from their past behavior or even recognize their current state. Yet this criticism holds within it an elegant philosophical justification for ZI modeling: because ZI agents are so wildly *irrational*, they provide a means to separate market effects caused by market microstructure decisions from market effects caused by human reasoning or rationality. Put another way, we can attribute differences between market simulations using ZI agents and empirical results to the presence of human rationality. The Marginal Trader Hypothesis is strongly tied to the concept of human rationality, and we are able to provide evidence in support of the theory only because we can compare human performance to a "base case" in which market participants are irrational.

The second half of our work yields several significant advancement towards the modeling of prediction markets:

- We show that in the prediction markets context, an agent-based model is entirely feasible and, where appropriate, confirms our prior theoretical modeling. We discuss the implications of a market comprised only of ZI traders closely exhibiting the same features as real prediction markets, paying particular attention to the presence of the longshot bias within the agent-based model. We discuss our results and their implications for the Hayek hypothesis.

- Agent-based modeling provides the key to investigate the buying bias, which refers to the psychological phenomenon by which people prefer long positions (buying) to short positions (selling). Our work demonstrates the surprising result that the buying bias can actually *increase* the efficiency of the exchange. We contextualize such a result in light of our previous findings and provide an argument as to why the buying bias may be particularly relevant for a Tradesports-style exchange, which we document by showing a surfeit of bid orders as opposed to ask orders in the Tradesports pennant-winner exchanges in the 2007 baseball contracts.

- Apropos of our work on the buying bias, and in terms of practical advice for markets *in situ*, we analyze the effects of commissions and market makers on price formation. We show that the presence of trading fees does not make the exchange less accurate, though we find it may lower trading volume significantly. We also provide evidence and arguments that the role of market makers within prediction markets may be overstated.

- To those concerned about liquidity within an IEM framework, we respond with strong agreement. Our work uncovers the critical role that quantities play in price formation in prediction markets. Moreover, we provide evidence that the IEM may be undersupplied in quantities. If true, we demonstrate that such a result would provide strong evidence in favor of the Marginal Trader Hypothesis.

Outline

Chapter 2 sets up a stylized model for a both the vote-share and winner-take-all market in the IEM. Chapter 3 details the computational innovations required to solve our model numerically. The fruits of our labors are discussed in chapters 4 and 5, which discuss how our model predicts the longshot bias in markets and overpricing in laboratory experiments, respectively.

Our agent-based model is introduced in chapter 6, which models single-share prediction markets. Chapter 7 incorporates stylized commissions and market makers into the model, while chapter 8 extends the model to double-share prediction markets and discusses relevant empirical results. Chapter 9 concludes.

Chapter 2

A Stylized Model

In this chapter we introduce a framework of price-taking traders with independently distributed private valuations. We begin by giving a simple model for a winner-take-all market, and then expanding our definition to include vote-share markets, demonstrating how our vote-share model implicitly induces a winner-take-all model. We conclude by providing empirical reinforcement and simplification to our theoretical model.

2.1 A Simple Winner-Take-All Market

There exist two contracts for two candidates, x and y, with respective exogenous prices p_x and p_y. The contracts are *complementary*, in that either contract x will expire at \$1 and contract y will expire worthless, or contract y will expire at \$1 and contract x will expire worthless. We assume that every trader in the pool is endowed with the same monotonically increasing utility function U and that all traders hold the same amount of funds in the market, w. Trader i holds independent belief $b_i \in [0, 1]$ that the state of the world in which x expires at \$1 will occur.

Traders solve for their demands D by maximizing their expected utility:

$$D(b_i, p_x, p_y) = \arg\max_{x,y \in \mathbb{R}^+} b_i U(w + (1 - p_x)x - p_y y) + (1 - b_i)U(w - p_x x + (1 - p_y)y) \quad (2.1)$$

Subject to the budget constraint that:

$$p_x x + p_y y \leq w \tag{2.2}$$

We assume the trading population consists of M traders with individual beliefs b distributed independently with density function f. Letting D_x and D_y respectively represent demands for x and y, we can define aggregate demands A as:

$$A_x(p_x, p_y) = M \int_{-\infty}^{\infty} D_x(b, p_x, p_y) f(b)\, db \tag{2.3}$$

$$A_y(p_x, p_y) = M \int_{-\infty}^{\infty} D_y(b, p_x, p_y) f(b)\, db \tag{2.4}$$

Shares only enter and exit the IEM in *bundles*, which refer to a single share of both x and y. Note that a bundle is always guaranteed to pay off at a dollar, regardless of whether we are in the winner-take-all or vote-share market. Bundles always cost 1 dollar to purchase from or sell to the exchange, and thus the IEM operates without generating profit. Since the number of outstanding shares of x is always equal to the number of outstanding shares of y, the only feasible set of equilibrium price vectors (p_x, p_y) are those for which $A_x = A_y$.

Note that our traders in the model do not learn or refine their beliefs in the marketplace, that is, the market serves only as a method to solve for the equilibrium posted price for shares of each candidate, and each trader has the utmost certainty that their individual valuation is the correct, objective one. Though simplistic, such a structure is not necessarily a poor approximation; as De Bondt and Thaler [7] asserted, "the most robust finding in the psychology of judgment is that people are overconfident". Furthermore, acting on this overconfidence in individual beliefs is more of an issue in prediction markets as opposed to traditional stock markets. An investor in Microsoft is unlikely to have examined all the public releases of that corporation, but is at least partially aware that the forces shaping price formation have committed resources towards the examination and analysis of such information. In an investment-capped prediction market for the identity of the next president, it is less likely the average investor has qualms about current market prices reflecting actual objective valuations. The lack of institutionalization within prediction markets creates a positive feedback loop for individual investors - the participation of the average trader is elicited by the knowledge that market prices have been set by other average traders, which in turn entices the participation of further overconfident individuals.

2.2 Extending the Model to Vote-Share Markets

By expanding our conception of the way individual beliefs are formed, our model can be extended to vote-share markets. Instead of a single belief value b, let us assume that agent i holds a private belief distribution f_i over the vote share that a candidate will receive.

Subject to the budget constraint in equation (2.2), individual demands in this market are then given by the solution to:

$$D(f_i, p_x, p_y) = \arg \max_{x,y \in \mathbb{R}^+} \int_{-\infty}^{\infty} U(w + zx + (1-z)y - p_x x - p_y y) f_i(z)\, dz \tag{2.5}$$

Letting individual distributions f be parameterized by some θ, and setting the density of private probability functions to be given by $h(\theta)$, then aggregate demands in the vote-share markets are given by:

$$A_x(p_x, p_y) = M \int D_x(f_\theta, p_x, p_y) h(\theta)\, d\theta \tag{2.6}$$

$$A_y(p_x, p_y) = M \int D_y(f_\theta, p_x, p_y) h(\theta)\, d\theta \tag{2.7}$$

As before, (p_x, p_y) is a feasible price level if $A_x = A_y$.

2.2.1 The Embedded Winner-Take-All Market

The IEM defines a candidate to have won an election if he or she receives more than half of the two-party vote share. Given our individual beliefs over prior vote shares f_i, this implies we can solve for individual beliefs in the winner-take-all market as:

$$b_i = \int_{.5}^{\infty} f_i(z)\, dz = 1 - F_i(.5) \tag{2.8}$$

2.3 Experimental Inquiry

Up until now, our treatment has been wholly general, but at this point we shift from a purely theoretical model and start making claims as to specific forms. We begin by specifying the distribution of vote share beliefs and analyzing historical data to produce a more precise model. We conclude by discussing a framework through which we can easily index a range of widely-used utility functions to assign to the agents in our model.

2.3.1 On the Distribution of Vote Share Beliefs

We let $\mu_i \sim N(\mu, \sigma_1^2)$, and $f_i \sim N(\mu_i, \sigma_2^2)$. That is to say, an individual's belief distribution is a normal distribution with mean drawn from a normal distribution with mean μ and variance σ_1^2, and variance σ_2^2.

Claim 1. *The overall distribution of vote-share beliefs is distributed according to* $N(\mu, \sigma_1^2 + \sigma_2^2)$.

Proof: We employ characteristic functions. Note that our method of generating random values is equivalent to summing a draw from two normal distributions: $N(\mu, \sigma_1^2)$ and $N(0, \sigma_2^2)$.

The characteristic function of a random variable X is given by

$$c_X(t) = \mathbb{E}\left(e^{itX}\right)$$

It follows that if $X \sim N(\mu, \sigma^2)$,

$$c_X(t) = \exp\left(it\mu - \frac{\sigma^2 t^2}{2}\right)$$

Now let $A \sim N(\mu, \sigma_1^2)$ and $B \sim N(0, \sigma_2^2)$. Because the characteristic function of the sum of random variables is given by the product of the two characteristic functions, it follows that:

$$\begin{aligned} c_{A+B}(t) &= c_A(t)c_B(t) \\ &= \exp\left(it\mu - \frac{\sigma_1^2 t^2}{2}\right)\exp\left(-\frac{\sigma_2^2 t^2}{2}\right) \\ &= \exp\left(it\mu - \frac{(\sigma_1^2 + \sigma_2^2)t^2}{2}\right) \end{aligned}$$

Thus, $c_{A+B}(t)$ is equal to the characteristic function of a normal random variable with mean μ and variance $\sigma_1^2 + \sigma_2^2$. The Fourier inversion theorem gives us a one-to-one mapping between characteristic functions and distributions, that is to say that if two random variables share the same characteristic function, then they are equivalent. Thus, the overall density of vote share beliefs, f, is $\sim N(\mu, \sigma_1^2 + \sigma_2^2)$. ■

Historical Data and the Model

Let $\sigma_1^2 + \sigma_2^2 = \sigma^2$, such that vote share beliefs are distributed according to $f \sim N(\mu, \sigma^2)$. The challenge is to find an appropriate value for σ based on historical data. Assuming that both the vote-share and winner-take-all markets are efficient, then we should find that the vote-share market should converge to μ, and that the winner-take-all market should converge to:

$$\int_{.5}^{\infty} \mathbf{normpdf}(z, \mu, \sigma)\, dz = 1 - \mathbf{normcdf}(.5, \mu, \sigma)$$

Consider the space of possible distributions indexed by σ by imagining a plot with the vote-share market price on the x axis and the winner-take-all market price on the y. As $\sigma \to 0$, the plot approaches a vertical line at $x = .5$. As $\sigma \to \infty$, the plot approaches a horizontal line at $y = .5$.

Figure 2.1 shows the $\mathcal{L}1$ norm fit for the value of σ, where the horizontal axis shows the vote-share price and the vertical axis shows the winner-take-all price, for prices taken on the eve of election day.

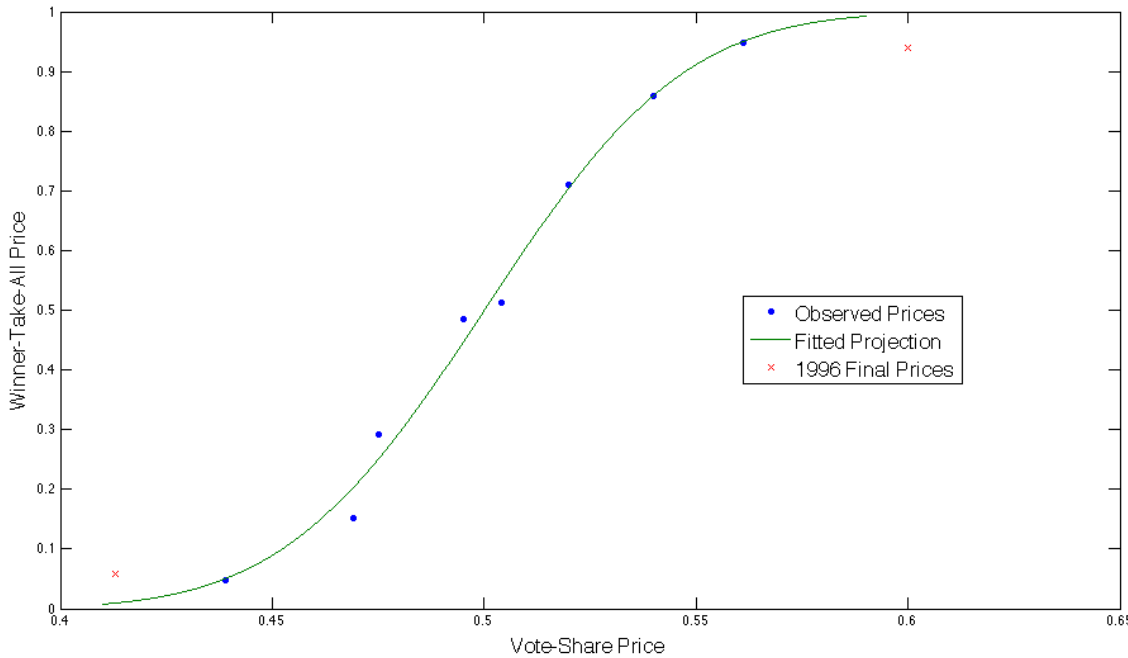

Figure 2.1: Fit of data with $\sigma = .037$

Clearly, the fit appears to be quite good, but the plot deserves further commentary. The red `x`s on the plot represent the eve of the 1996 election. As was discussed in Berg *et al.* [4], the vote-share market in the 1996 election was very accurate until two weeks before the election, when prices began to sharply diverge from the eventual result. The day before the election, the vote-share markets were off by a whopping 4.5 cents. Thus, for our fit, the data from two weeks before the 1996 election was substituted for the 1996 election. That the final 1996 prices on the plot reflects the inaccuracy in the market provides additional support for the vote-share modeling approach. When the markets are accurate, the data fit well; and perhaps more importantly, when the markets are inaccurate, the data fit poorly.

It is worth noting, however, that the fit is not as good if we incorporate data from arbitrary days into the plot. Figure 2.2 shows the same fit using data from the two weeks before an election; the fit is much less precise. This is likely due to the influence of two

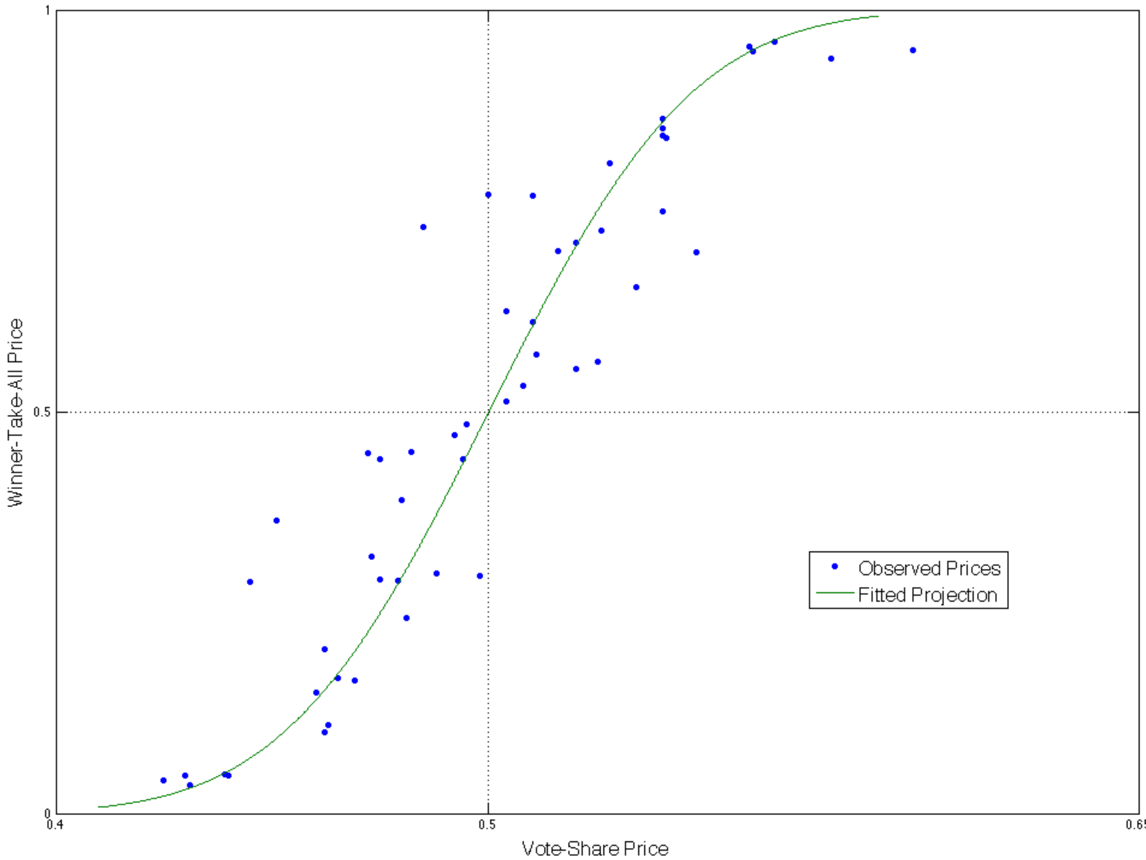

Figure 2.2: $\sigma = .037$ fit using prices from the two weeks before an election.

factors, both of which are minimized by taking data from as close as possible to the election date.

- Though we will not formalize it, the concept of "convergence" relies on information in the market being fully elicited. With more time, more agents are able to participate and more information about the election is revealed. Indeed, the final few days, and the last day in particular, before an election normally see the largest amount of volume in the exchange [4].

- Secondly, we use election-eve results to minimize the effect of the "foot in mouth" hazard, which could cause distributional heavy tails. Months before the election, it might be conceivable that the most-likely vote share that a candidate could receive may not correspond directly to their probability of winning the election - there's always the chance that the candidate could say or do something embarrassing and compromise their election chances. This phenomenon would have the effect of increasing the prices of lower-probability candidates and decreasing the prices of higher-probability candidates from the predicted prices in the model. On figure 2.2, heavy-tails correspond to the data points that appear above the fit line for vote-share prices above .5, and below the fit line for vote-share prices below .5. With time, however, this effect will fade; it is thoroughly unlikely that in the nine hours before polls open any new piece of news or information will be revealed that could have a significant effect on the election.

On Modeling with σ_1 and σ_2

In the computational inquiries that follow, we let $\sigma_1 = \sigma_2$, that is, the normal distribution responsible for selecting a trader's mean belief and the normal distribution over the trader's belief of final vote shares have the same standard deviation. Though any other choice of the two values would be equally arbitrary, note that eliciting such a subtle modeling decision from the beliefs and actions of real people would be virtually impossible. Moreover, there are significant issues involved in skewing the values towards one side or the other: Can we truly think that agents believe that a candidate will receive a certain vote share *exclusively*, or that every agent has identified the same most likely vote share? Given our concerns about the implications of "favoring" one of the two distributions, it is best to go with the most flexible, hardest modeling decision - note that setting $\sigma_1 = \sigma_2$ maximizes $\sigma_1 + \sigma_2$.

This scheme results in $\sigma_1 = \sigma_2 \approx .025$ which seems eminently reasonable; it corresponds to individual traders having 95th percentile beliefs constrained within a 10-percent band of possible vote share outcomes. We note that our winner-take-all belief distributions are more peaked than Gjerstad's distributions in [11] and generally have less density at the edges of the state space than the prediction data used in the market simulations of Wolfers and Zitzewitz [31].

2.3.2 Modeling Utilities

We assume that all traders have the same constant relative risk averse (CRRA) utility function parameterized by $\theta \in (-\infty, \infty)$:

$$U(w) = \begin{cases} \frac{w^{1-\theta}}{1-\theta} & \text{if } \theta \neq 1; \\ \log w & \text{if } \theta = 1. \end{cases}$$

Depending on θ, this functional form allows the capture of a wide variety of utility functions, characterized by the following table:

θ	Type of Agent Utility
$\theta < 0$	Risk-loving
$\theta = 0$	Risk-neutral (linear)
$0 < \theta < 1$	Risk-averse (sub-logarithmic)
$\theta = 1$	Logarithmic
$\theta > 1$	Risk-averse (super-logarithmic)

We will only consider risk-neutral or risk-averse agents ($\theta \geq 0$) for the remainder of this work. As an aside, we can recover Manski's model by fixing $\theta = 0$, such that every agent in the pool is risk-neutral. We can also recover Wolfers and Zitzewitz's [31] result that logarithmic-utility risk-averse bidders yield prices that converge to distributional means by fixing $\theta = 1$.

Chapter 3

Calculating Decisions

As we proceed in the work, we are interested in solving for the equilibrium prices and quantities as they are defined in equations (2.3), (2.4), (2.6), and (2.7), so that we can use these results to drive further inquiry into observed phenomena. Solving for these values is a three-step process: at the top level, we seek to equate aggregate supply and aggregate demand, so we perform a zero-finding using whatever variable we are interested in examining - prices, quantities, belief distribution means, etc. In order to equate aggregate supply and aggregate demand, we must be able to calculate them, which involves calculating the demands for the pool of agents as a whole. Finally, calculating the individual demands for each agent is often non-trivial; recall that individual agents derive their beliefs from an underlying private distribution.

This section details the optimizations, both mathematical and computational, that we performed to generate our results. This section is rather technical and not for the faint of heart; the reader without a predilection for scientific computing may feel comfortable skipping it.

3.1 Reducing the Problem Size

Note that our individual demand equations (2.1) and (2.5) set up two-dimensional search problems, solving over demands for both x and y. Two-dimensional search problems are typically very difficult to solve quickly and accurately, so proving the following two claims allows us to reduce our optimization problem to a one-dimensional search, which makes the problem far more tractable.

Central to many of the proofs involves the concept of a bundle; recall that a bundle consists of a single share of both x and y, which is guaranteed to pay off at a dollar and is always available to buy from or sell to the exchange for a dollar.

Claim 2. *If $p_x + p_y < 1$, then the budget constraint will bind our optimization solution.*

Proof: Assume we have an optimal solution in either market which results in an expenditure of $w - k$, $k > 0$. Then purchasing k bundles guarantees revenue of k from an expenditure of $k(p_x + p_y) < k$. Because U is monotonic, it follows that an agent will achieve higher utility by purchasing the k bundles, and thus only solutions which exhaust the budget constraint can be optimal. ■

Claim 3. *If $p_x + p_y \geq 1$, then there exists a solution to the optimization problem wherein agent demands for one of the contracts is 0.*

Proof: If $p_x + p_y = 1$, it follows that agent are indifferent between holding a bundle consisting of $(\hat{x}, \hat{y})$ shares of x and y respectively, and holding $(\hat{x} + k, \hat{y} + k)$, because the arguments to each utility function are decreased by $k(p_x + p_y) = k$ in expenditure but increases by k. It follows that there exists a *continuum* of solutions to the optimization problem if $p_x + p_y = 1$, and that we may select the minimal solution, such that an agent's holdings of at least one commodity is precisely 0.

Now let $p_x + p_y > 1$ and assume there is a solution (x', y') to the optimization problem, such that wlog $x' > y' > 0$. By the expenditure-revenue argument detailed above, the monotonicity of U shows that $(x' - y', 0)$ has a higher expected utility than our tendered solution. It follows that if $p_x + p_y > 1$, no agent will want to hold a position consisting of non-zero amounts of both x and y. ■

In addition to the previous two claims, the particular form of agent's utility functions allows us to reduce the domain over which we search for optimal solutions.

Claim 4. *Let $p_x + p_y \geq 1$. Then there exists optimal solutions in which agent i has non-zero demand for x if and only if $b_i > p_x$, and non-zero demand for y if and only if $1 - b_i > p_y$.*

Proof: The proof in Gjerstad [11] is given for $p_x + p_y = 1$ and holds also for $p_x + p_y > 1$.
■

Claim 5. *Let $p_x + p_y \geq 1$, and let agent i hold beliefs over the distribution of vote share for candidate x distributed according to $N(\mu_i, \sigma_2^2)$. Then if $p_x > \mu_i$ that agent will not demand any shares of candidate x, and if $p_y > 1 - \mu_i$, that agent will not demand any shares of candidate y.*

Proof: We shall prove the claim for the x case, noting that the y case holds by symmetry because a belief that candidate x will receive vote-share z is identical to a belief that candidate y will receive vote-share $1 - z$.

If $\theta = 1$, then our agent has logarithmic utility, and our expected utility from holding a position of x given prior wealth w is:

$$EU = \frac{1}{\sqrt{2\pi}\sigma_2} \int_{-\infty}^{\infty} \log(w + (z - p_x)x) e^{\frac{(z-\mu_i)^2}{2\sigma^2}} \, dz$$

Otherwise, if $\theta \neq 1$ our expected utility is:

$$EU = \frac{1}{\sqrt{2\pi}\sigma_2} \int_{-\infty}^{\infty} \frac{(w + (z - p_x)x)^{1-\theta}}{1-\theta} e^{\frac{(z-\mu_i)^2}{2\sigma_2^2}} \, dz \tag{3.1}$$

We shall prove the claim by showing that $\frac{\partial EU}{\partial x} < 0$ at $x = 0$ and that the second-order condition, $\frac{\partial^2 EU}{\partial x^2} < 0$ for all $x \in [0, w/p_x]$.

Our first partial is (regardless of the particular value of θ):

$$\frac{1}{\sqrt{2\pi}\sigma_2} \int_{-\infty}^{\infty} \frac{z - p_x}{(w + (z - p_x)x)^{\theta}} e^{\frac{(z-\mu_i)^2}{2\sigma_2^2}} \, dz \tag{3.2}$$

Evaluating at $x = 0$ and letting $m = z - p_x$ we find that equation 3.2 collapses to:

$$\frac{1}{\sqrt{2\pi}\sigma_2} \frac{1}{w^{\theta}} \int_{-\infty}^{\infty} m e^{\frac{(m-(\mu_i-p_x))^2}{2\sigma_2^2}} \, dm$$

We note that the expression inside the integral is simply the principal moment of a normal distribution with mean $\mu_i - p_x$ and so this reduces further to:

$$\frac{\mu_i - p_x}{w^\theta}$$

which is negative if $p_x > \mu$.

Now consider the second partial. We must show that

$$\frac{1}{\sqrt{2\pi}\sigma_2}\int_{-\infty}^{\infty} -\frac{m^2}{(w+mx)^{\theta+1}} e^{\frac{(m-(\mu-p_x))^2}{2\sigma_2^2}}\, dm < 0 \tag{3.3}$$

holds at all $x \in [0, w/p_x]$. Because $-m^2 e^{\frac{(m-(\mu-p_x))^2}{2\sigma_2^2}} < 0$ and $\frac{1}{\sqrt{2\pi}\sigma_2} > 0$, it is sufficient that, $\forall z$ and x:

$$(w+mx)^{1+\theta} \geq 0$$

for which it is sufficient that:

$$w + (z - p_x)x \geq 0$$

which follows if $z \geq 0$, which makes sense given the physical intuition behind the concept of vote share, namely, that it is a value between 0 and 1. Thus, as long as we are careful to ensure that only an exceedingly small portion of any private belief has a candidate violating these physical constraints, the claim holds. ■

The reader will be sure to note that we are always mindful of this constraint, taking care to ensure that we do not make claims to results based on negative or greater-than-one vote shares.

The proofs in this section greatly reduce the amount of work we will need to do to compute and aggregate optimal solutions to the demand problem. We now need to focus only on specific types of demands we know to be optimal, and we only need to look for agents which have those demands on a particular part of our problem domain.

3.2 Individual Demands - Vote-Share Market

We will focus on individual demands in the vote-share market in only the case where $p_x + p_y \geq 1$. The proof of claim 3 gives us that we only need to consider an agent holding shares of either x or y. However, we still need to find a way to efficiently calculate the integral in the first-order condition (3.2) derived in claim 5.

Because U is a continuous, smooth function, its integral will be well-behaved and so Gauss-Hermite quadrature provides a way to compute the demand integral effectively. Gauss-Hermite quadrature consists of a set of nodes $x = \{x_1, \ldots, x_n\}$ and weights $\omega = \{\omega_1, \ldots, \omega_n\}$, available from tables like Stroud and Secrest [26], such that:

$$\int_{-\infty}^{\infty} f(x)e^{-x^2}\,dx \approx \sum_{i=1}^{n} \omega_i f(x_i)$$

What is remarkable about the scheme is that the integral can be computed over an infinite domain with a small number of nodes; the error is on the order of $|f^{(2n)}|$, which is very small for functions that are not pathological. Practically, Judd [16] shows an example of Gauss-Hermite quadrature on an otherwise-difficult integral yielding essentially machine precision at just 13 interpolation nodes.

Let trader i have vote-share beliefs $f_i \sim N(\mu_i, \sigma_2^2)$. First assume that $p_x + p_y \geq 1$ and that wlog trader i is interested in holding only shares of x, that is, that $\mu_i > p_x$. The first partial κ is:

$$\kappa(x, p_x, \mu_i) = \frac{1}{\sqrt{2\pi}\sigma_2} \int_{-\infty}^{\infty} \frac{z - p_x}{(w + (z - p_x)x)^{\theta}} e^{\frac{-(z-\mu_i)^2}{2\sigma_2^2}}\,dz$$

And the second partial κ' is:

$$\kappa'(x, p_x, \mu_i) = \frac{1}{\sqrt{2\pi}\sigma_2} \int_{-\infty}^{\infty} -\frac{(z - p_x)^2}{(w + (z - p_x)x)^{\theta+1}} e^{\frac{-(z-\mu_i)^2}{2\sigma_2^2}}\,dz \qquad (3.4)$$

Now consider how we can apply Gauss-Hermite quadrature to these equations. By substitution and the change of variables $\zeta = \frac{z-\mu_i}{\sqrt{2}\sigma_2}$, we note that:

$$\begin{aligned}
\kappa(x, p_x, \mu_i) &= \frac{1}{\sqrt{2\pi}\sigma_2} \int_{-\infty}^{\infty} \frac{z - p_x}{(w + (z - p_x)x)^\theta} e^{\frac{-(z-\mu_i)^2}{2\sigma_2^2}} \, dz \\
&= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \frac{\sqrt{2}\sigma_2\zeta + \mu_i - p_x}{(w + \sqrt{2}\sigma_2\zeta + \mu_i - p_x)^\theta} e^{-\zeta^2} \, d\zeta \\
&\approx \frac{1}{\sqrt{\pi}} \sum_{i=1}^{n} \omega_i \frac{\sqrt{2}\sigma_2 x_i + \mu_i - p_x}{(w + \sqrt{2}\sigma_2 x_i + \mu_i - p_x)^\theta}
\end{aligned}$$

And we can follow an identical substitution in the second-order equation to obtain the approximation:

$$\kappa'(x, p_x, \mu_i) \approx -\frac{\theta}{\sqrt{\pi}} \sum_{i=1}^{n} \omega_i \frac{(\sqrt{2}\sigma_2 x_i + \mu_i - p_x)^2}{(w + \sqrt{2}\sigma_2 x_i + \mu_i - p_x)x)^{\theta+1}}$$

Now we have a method to quickly evaluate the first- and second-order conditions. Through empirical analysis, we determined that the second partial (3.4) is roughly constant for $x \in [0, w/p_x]$. This implies that Newton's Method will be a particularly effective way of computing an optimal solution to the demand problem, because Newton's Method uses the derivative at a guess to develop an improvement to that guess; if the derivative (in this case κ') does not change substantially the algorithm will converge quickly. We employ the following algorithm:

```
Input: Prices p_x and p_y, Individual distribution parameter μ_i.
Output: Individual demands x and y.
x ← 0
y ← 0
if p_x < μ_i then
    if κ(w/p_x, p_x, μ_i) > 0 then x ← w/p_x
    else
        while |κ(x, p_x, μ_i)| > ε do
            x ← x − κ(x, p_x, μ_i)/κ'(x, p_x, μ_i)
        end
    end
end
else if p_y < 1 − μ_i then
    if κ(w/p_y, p_y, 1 − μ_i) > 0 then y ← w/p_y
    else
        while |κ(y, p_y, 1 − μ_i)| > ε do
            y ← y − κ(y, p_y, 1 − μ_i)/κ'(y, p_y, 1 − μ_i)
        end
    end
end
return x, y
```

Algorithm 1: Newton's Method Individual Demand Solver for $p_x + p_y \geq 1$

An Aside - Verification

There are always risks associated with the fact that computers cannot represent infinite-precision numbers, and so numerical precision errors always a threat to the computational economist. We note that algorithm 1 does not solve for the zero of the first-order condition, but merely returns a value that solves for optimal allocations within ϵ. In practice and throughout this work, we set ϵ to be small - 10^{-10} - in all of our functions that require zeroing, including the eventual calculation of excess demand. Moreover, whenever polynomial interpolation is required, we use many more nodes than is necessary. For instance, our Gauss-Hermite quadrature scheme was generating accurate values to 10^{-10} with 16 interpolation nodes, but for production code we use 50 nodes.

As Judd [16] points out, however, the absolute precision of an error term pales in importance to the *relative* precision of the error. An error on the order of .01 might be trivial, or it might be significant, depending on the context of the problem. Since we are interested in results only to the fourth decimal place, we are confident that our results are robust to the desired precision.

3.3 Individual Demands - Winner-Take-All Market

Unlike in the vote-share market, there exists a closed form expression for demand in the winner-take-all market. Letting $p_x + p_y = 1$, Gjerstad [11] gives the *difference* in demands for x and y as:

$$D_x - D_y = \frac{w\left((b - bp_x)^{\frac{1}{\theta}} - (p_x - bp_x)^{\frac{1}{\theta}}\right)}{(1 - p_x)(p_x - bp_x)^{\frac{1}{\theta}} + p_x(b - bp_x)^{\frac{1}{\theta}}} \tag{3.5}$$

As we are interested in the minimal solution to the demand problem, such that at least one of D_x or D_y is 0, we can replace this expression as follows:

$$D_x(b, p_x, p_y) = \begin{cases} \frac{w\left((b-bp_x)^{\frac{1}{\theta}} - (p_x-bp_x)^{\frac{1}{\theta}}\right)}{(1-p_x)(p_x-bp_x)^{\frac{1}{\theta}} + p_x(b-bp_x)^{\frac{1}{\theta}}} & \text{if } b > p_x; \\ 0 & \text{otherwise.} \end{cases} \tag{3.6}$$

$$D_y(b, p_x, p_y) = \begin{cases} \frac{w\left(((1-b)-(1-b)p_y)^{\frac{1}{\theta}} - (p_y-(1-b)p_y)^{\frac{1}{\theta}}\right)}{(1-p_y)(p_y-(1-b)p_y)^{\frac{1}{\theta}} + p_y((1-b)-(1-b)p_y)^{\frac{1}{\theta}}} & \text{if } 1-b > p_y; \\ 0 & \text{otherwise.} \end{cases} \tag{3.7}$$

Now imagine that we are at some optimal position $(D_x, 0)$ when $p_x + p_y = 1$. Since $D_y = 0$ and p_y does not appear in our solution for D_x, increasing p_y has no effect on our solution to the demand problem. A symmetric argument holds from starting at an optimal position $(0, D_y)$, and consequently we can use the closed-form expressions in equations (3.6) and (3.7) to solve our demand equations as long as $p_x + p_y \geq 1$. Note that if $p_x + p_y > 1$, there will exist some belief $\hat{b}$ satisfying $\hat{b} < p_x$ and $1 - \hat{b} < p_y$, such that $D_x(\hat{b}, p_x, p_y) = D_y(\hat{b}, p_x, p_y) = 0$.

Now consider the case where $p_x + p_y \leq 1$. Recall from the proof of claim 2 that our budget constraint binds us in this case. Therefore, we substitute $p_y y = w - p_x x$ into our optimization equation to produce:

$$D_x(b, p_x, p_y) = \arg\max_{x \in \mathbb{R}^+} bU(x) + (1 - b)U\left(\frac{w - p_x x}{p_y}\right)$$

For utilities corresponding to $\theta > 0$, this equation has an interior maximum that solves the first-order condition:

$$\frac{\partial D_x}{\partial x} = bU'(x) + (1-b)U'\left(\frac{w - p_x x}{p_y}\right)\left(-\frac{p_x}{p_y}\right) = 0$$

Recognizing that $U'(x) = x^{-\theta}$, this equation has the solution:

$$x^* = \frac{w}{\left(\frac{1-b}{b} p_y^{\theta-1} p_x\right)^{\frac{1}{\theta}} + p_x}$$

Which implies the following solutions to the demand problem for $p_x + p_y \leq 1$:

$$D_x(b, p_x, p_y) = \frac{w}{\left(\frac{1-b}{b} p_y^{\theta-1} p_x\right)^{\frac{1}{\theta}} + p_x} \tag{3.8}$$

$$D_y(b, p_x, p_y) = \frac{w - p_x D_x(b, p_x, p_y)}{p_y} \tag{3.9}$$

Note that we have two different expressions for demand at $p_x + p_y = 1$. Equations (3.6) and (3.7) represent the *minimal* solution to the demand problem, while equations (3.8) and (3.9) represent the *maximal* solution to the problem. At $p_x + p_y = 1$ both solutions will return the same payoff; recall that agents are indifferent between holding k shares of both x and y and holding k dollars.

Thus, we have closed-form solutions for any meaningful vector of prices in the winner-take-all market.

3.4 Aggregating Demands - Vote-Share Market

As it turns out, individual demands in the vote-share market, even for very large values of θ, resemble step functions.

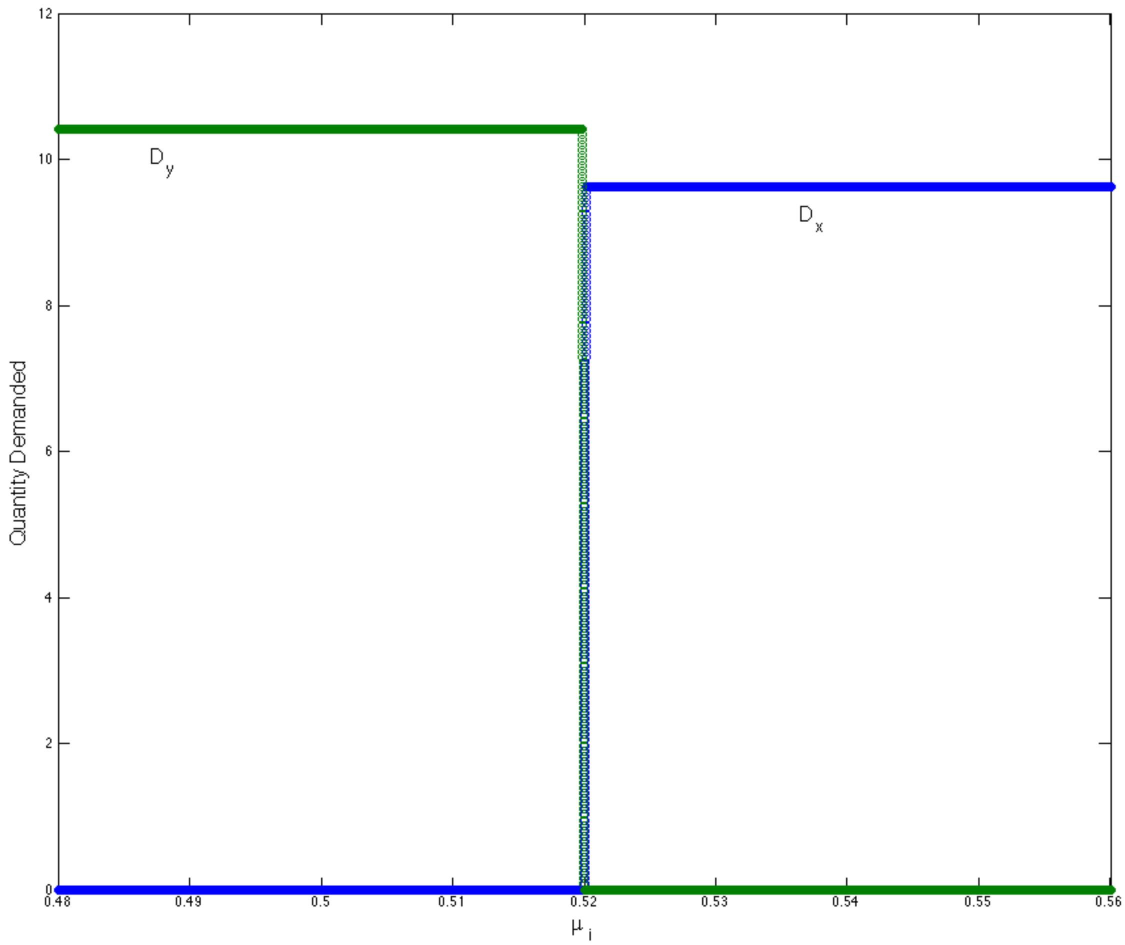

Figure 3.1: Example of Individual Demands with $p_x = .52$, $p_y = .48$, and $\theta = .1$

Figure 3.1 shows an example of individual demands in the vote-share market. Note that if the mean of agent's beliefs is even fractionally over the price, they will invest their entire endowment into shares of that candidate. This suggests the approach used in algorithm 2, which works by backing away in small steps from the prices until we find the value at which demand hits the budget constraint.

Input: Prices p_x and p_y, Distribution parameters σ_1 and σ_2.
Output: Aggregate demands A_x and A_y.
$A_x \leftarrow 0$
$A_y \leftarrow 0$
$y \leftarrow 1 - p_y$
while $D_y(N(y, \sigma_2), p_x, p_y) < w$ **do**
| $A_y \leftarrow A_y + \epsilon D_y(N(y, \sigma_2), p_x, p_y)\mathbf{normpdf}(y, \mu, \sigma_1)$
| $y \leftarrow y - \epsilon$
end
$A_y \leftarrow A_y + w\mathbf{normcdf}(y, \mu, \sigma_1)$
$x \leftarrow p_x$
while $D_x(N(x, \sigma_2), p_x, p_y) < w$ **do**
| $A_x \leftarrow A_x + \epsilon D_x(N(x, \sigma_2), p_x, p_y)\mathbf{normpdf}(x, \mu, \sigma_1)$
| $x \leftarrow x + \epsilon$
end
$A_x \leftarrow A_x + w(1 - \mathbf{normcdf}(x, \mu, \sigma_1))$
return A_x, A_y

Algorithm 2: Vote-Share Aggregate Demand Calculator for $p_x + p_y \geq 1$

3.5 Aggregating Demands - Winner-Take-All Market

On first glance, we might be tempted to try Gauss-Hermite quadrature to aggregate demands in the winner-take-all market. Such a scheme would involve generating a large number of nodes from the distribution of means (the normal distribution with mean μ and variance σ_1^2), then using those nodes as the means of agents, calculating the demands of those agents, and then multiplying by the appropriate weights and summing.

Unfortunately, as $\theta \to 0$, the solutions to the demand equations begin to resemble step functions. This is easily observed by noting that a risk-neutral ($\theta = 0$) agent will spend their entire endowment if their private belief is greater than the market price, while spending nothing if their b_i is even marginally less than the market price. We might call the aggregating code many times with different pricing parameters, and as such Gauss-Hermite quadrature will work appropriately as long as small changes in price do not result in dramatic shifts in demands. Step functions, however, do not meet this requirement, as small changes in prices can lead to a dramatic change in demands if an agent with belief mean at a quadrature node switches from spending their endowment on shares of x instead of y.

We instead replace quadrature over the whole real line with quadrature over $[0, 1]$. Letting $g(b)$ represent the distribution of beliefs, we thus seek to calculate:

$$A_x(p_x, p_y) = M \int_0^1 D_x(b, p_x, p_y) g(b)\, db \tag{3.10}$$

$$A_y(p_x, p_y) = M \int_0^1 D_y(b, p_x, p_y) g(b)\, db \tag{3.11}$$

This method becomes viable only because we can find a easily calculable, closed-form expression for g, the distribution of beliefs in the winner-take-all market.

3.5.1 Distribution of Beliefs in the Winner-Take-All Market

To simplify our notation, we define a helpful function. Let $\phi(x)$ return the solution μ that satisfies $x = 1 - \mathbf{normcdf}(.5, \mu, \sigma_2)$, that is, the mean of the individual normal distribution that corresponds with an individual belief that the candidate has x probability of winning the election.

Note that:

$$\phi(x) = .5 - \mathbf{norminv}(1 - x, 0, \sigma_2)$$

Where $\mathbf{norminv}(y, \mu, \sigma)$ returns the argument x that satisfies $\mathbf{normcdf}(x, \mu, \sigma) = y$.

In order to find the cumulative density function $G(x)$ of the new distribution, that is, the probability that any individual feels the candidate will win with probability x or less, we should find the mean of the distribution that corresponds to a belief that the candidate will win with probability x, and then find the probability that any value not greater than that mean is selected by the initial draw. That is,

$$G(x) = \textbf{normcdf}(\phi(x), \mu, \sigma_1)$$

Now consider the density function of the new distribution, $g(x) = G'(x)$. By the chain rule:

$$g(x) = \textbf{normpdf}(\phi(x), \mu, \sigma_1)\phi'(x)$$

To develop this further we must find the derivative of $\phi(x)$. First, we solve for the inverse function of $\phi(x)$:

$$\begin{aligned}
x &= .5 - \textbf{norminv}(1 - y, 0, \sigma_2) \\
\textbf{norminv}(1 - y, 0, \sigma_2) &= .5 - x \\
-\textbf{norminv}(y, 0, \sigma_2) &= \\
y &= \textbf{normcdf}(x - .5, 0, \sigma_2) \\
&= \textbf{normcdf}(x, .5, \sigma_2)
\end{aligned}$$

Therefore $\phi^{-1}(x) = \textbf{normcdf}(x, .5, \sigma_2)$. Recalling that:

$$\left[f^{-1}\right]'(x) = \frac{1}{f'(f^{-1}(x))}$$

It follows that:

$$\phi'(x) = \frac{1}{\textbf{normpdf}(\phi(x), .5, \sigma_2)}$$

Thus:

$$g(x) = \frac{\textbf{normpdf}(\phi(x), \mu, \sigma_1)}{\textbf{normpdf}(\phi(x), .5, \sigma_2)} \tag{3.12}$$

Equation (3.12) gives us a simple, computationally inexpensive way to calculate the distribution of beliefs in the winner-take-all market.

3.5.2 Quadrature over Finite Domains

We consider two different quadrature methods for integrating over finite domains, that is, sets of nodes $x = \{x_1, \ldots, x_n\}$ and weights $\omega = \{\omega_1, \ldots, \omega_n\}$ such that:

$$\int_0^1 f(x)\,dx \approx \sum_{i=1}^{n} \omega_i f(x_i)$$

One option for integrating over finite domains is to use Gaussian Quadrature. Over finite domains, Gaussian quadrature means doing Gauss-Legendre quadrature, which uses the zeroes of orthonormal polynomials over the space of real continuous functions endowed with the inner product $\langle f, g\rangle = \int_0^1 f(x)g(x)\,dx$ to guarantee an error on the order of $\Theta(|f^{(2n)}|)$.

Instead, we use Clenshaw-Curtis quadrature to calculate the integral. Based around the zeroes of Chebyshev polynomials, Clenshaw-Curtis quadrature only provides a guarantee of error on the order of $\Theta(|f^{(n)}|)$. However, Clenshaw-Curtis has two significant advantages that more than make up for any lack of accuracy:

- The interpolation nodes and weights for Clenshaw-Curtis quadrature can be quickly computed by a discrete Fourier transform in time $O(n \log n)$, as compared to $O(n^2)$ for the nodes and weights of Gauss-Legendre quadrature. Because MATLAB already features a heavily optimized implementation of the discrete Fourier transform, Waldvogel [29] showed the speedup gained by employing this method is astounding for large numbers of nodes, an improvement of nearly three orders of magnitude at $n = 1024$.

- Because the interpolation nodes of Clenshaw-Curtis quadrature include the endpoints, we can easily apply a method of *adaptive* quadrature, so that the proposed integral value between any two points can be itself the result of a recursive, lower-level Clenshaw-Curtis quadrature. We can thus obtain an arbitrarily high degree of precision by recursively calculating between interpolation nodes. Such an algorithm is given in Press *et al.* [23].

By using a fast adaptive quadrature method, Clenshaw-Curtis quadrature gives us a quick method for computing integrals while at the same time giving us confidence that our result will be numerically robust enough to handle even the most unwieldy function behavior.

Putting it all together

As was discussed in the introduction to this section, solving for missing variables in equilibrium is a three-step process: zeroing aggregate demands, computing aggregate demands, and computing individual demands. We have provided enough computational optimization in the bottom two levels to simply hand the top level off to MATLAB's general `fzero` solver, with the goal of zeroing excess aggregate demand and searching over the relevant variable.

Chapter 4

The Longshot Bias

Recall that the longshot bias refers to a market inefficiency by which likely events are underpriced relative to their objective probabilities and unlikely events are overpriced relative to their objective probabilities.

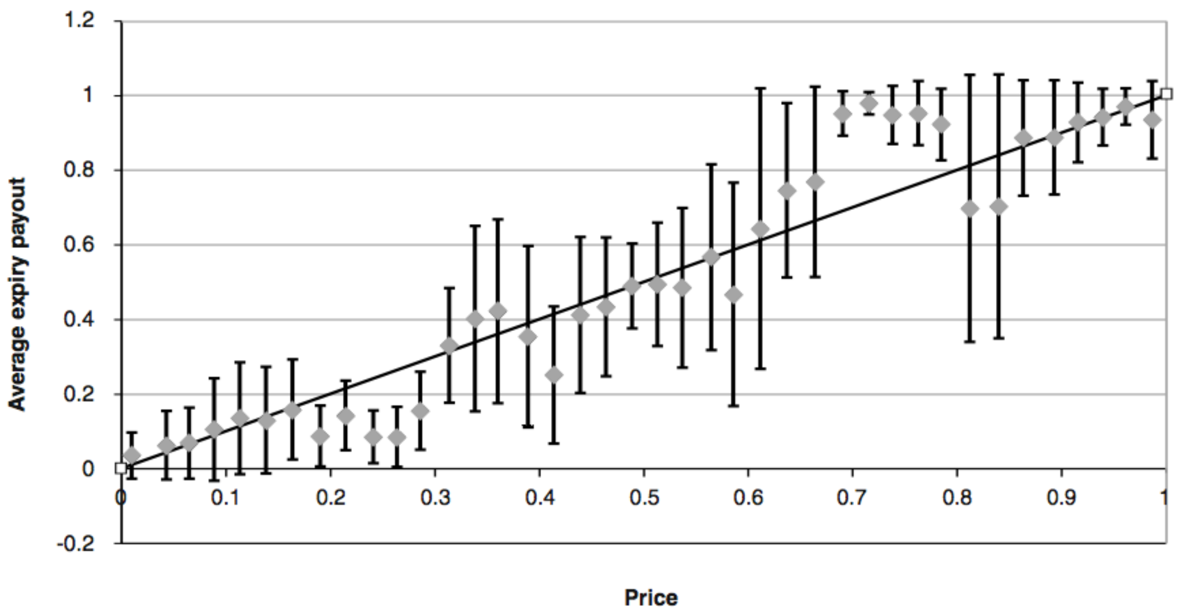

Figure 4.1: Market prices versus expected payouts in the IEM. Taken from Wolfers and Zitzewitz [30].

Figure 4.1 shows the longshot bias in the IEM market. To make this plot, Wolfers and

Zitzewitz took the data from every pricing day in every IEM winner-take-all election. They then partitioned the space into "buckets" of size 2.5 cents, so that shares of approximately the same price are grouped together. The grey data points represent the average payout shares in each bucket, while the black bars represent 95th percentile confidence intervals.

Note how shares priced around 25 cents pay off only 10 percent of the time, and shares around 75 cents pay off with near certainty. But why does the longshot bias occur? It is not conceivable that the belief distributions of separate shares which happen to be priced similarly would all show the same systematic bias; the cause must instead be intrinsic to the market.

The explanation of this phenomenon in Wolfers and Zitzewitz is not satisfactory. In the work, the authors provide the example of a market maker's spread on a share that has an objective value of three cents - a spread of a cent, such that the bid is two cents and the ask is four cents, clearly cuts away at a great deal of possible belief density on the buying side.

Figure 4.2 provides a visual guide their argument by illustrating density function for a random variable with a mean of three cents. The pool of buyers in the market is represented by the density to the left of 3 cents, while the pool of sellers is to the right. The dark grey area represents prospective buyers who will not trade, while the light grey area represents prospective sellers that will not trade with a spread of two cents. Since the dark grey area is larger, the buying side is more affected by the market maker's spread than the selling side. Thus, market makers might place the bid/ask midpoint higher than the objective probability to encourage a balance of buyers and sellers.

However, this explanation is not suitable for two potent reasons:

- First, the values in question here are not particularly small - twenty-five cents is far enough away from 0 to avoid problems that might plague analysis at the borders of the state space.

- The explanation relies, at least in part, on the role of the market maker. The IEM, and several Tradesports markets, have no traders elected to serve as market makers,

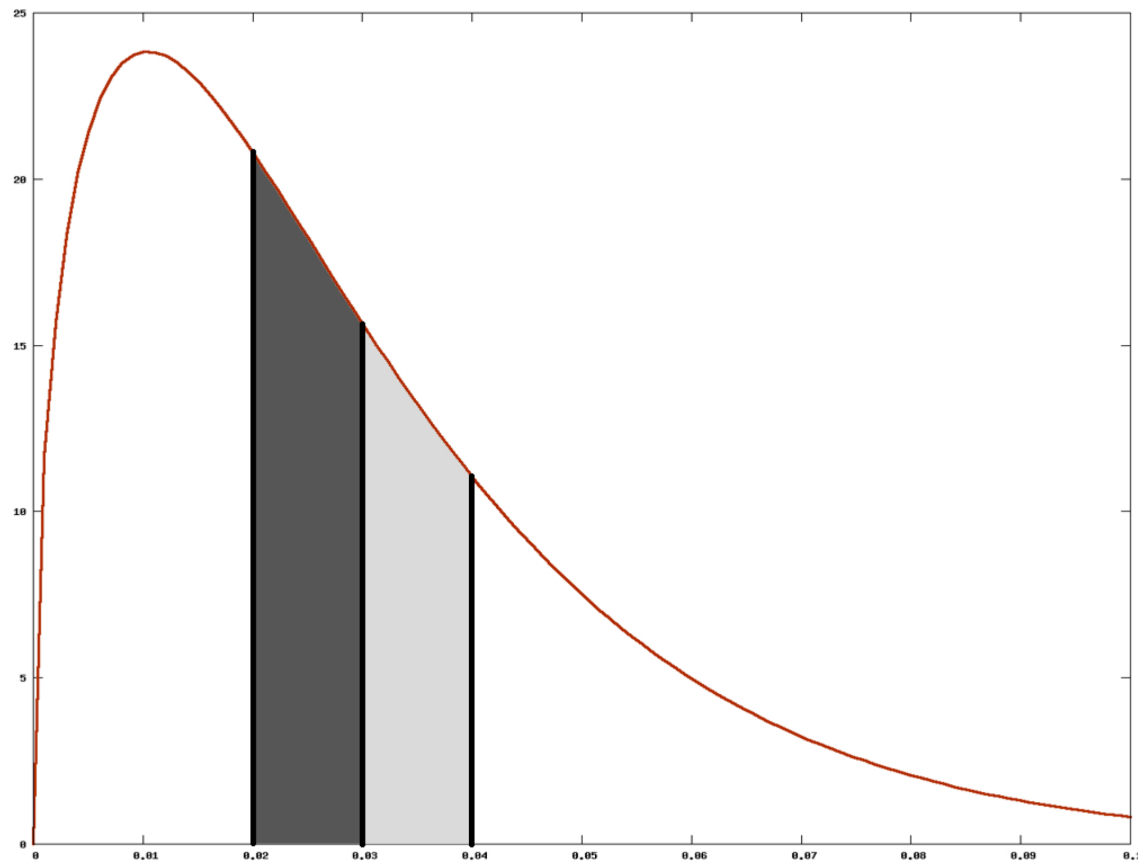

Figure 4.2: A market-maker's spread cuts away from proportionally greater density on the buying side.

no nexus of power ultimately responsible for price setting or ensuring volume. Indeed, such a concept is antithetical to the democratic mores of the IEM.

In this section, we aim to find a better explanation for the longshot bias in observed markets. We begin by documenting the longshot bias in a market completely separate from the IEM - the futures market run by Tradesports in basketball games. We then offer an explanation grounded in a lack of risk-aversion that yields the observed market inefficiencies, and discuss the practical effects of such reasoning, including justifying the presence of weak risk aversion within the markets. We conclude by examining what weak risk aversion entails for our vote-share market model relative to historical results.

4.1 Tradesports NBA Contracts

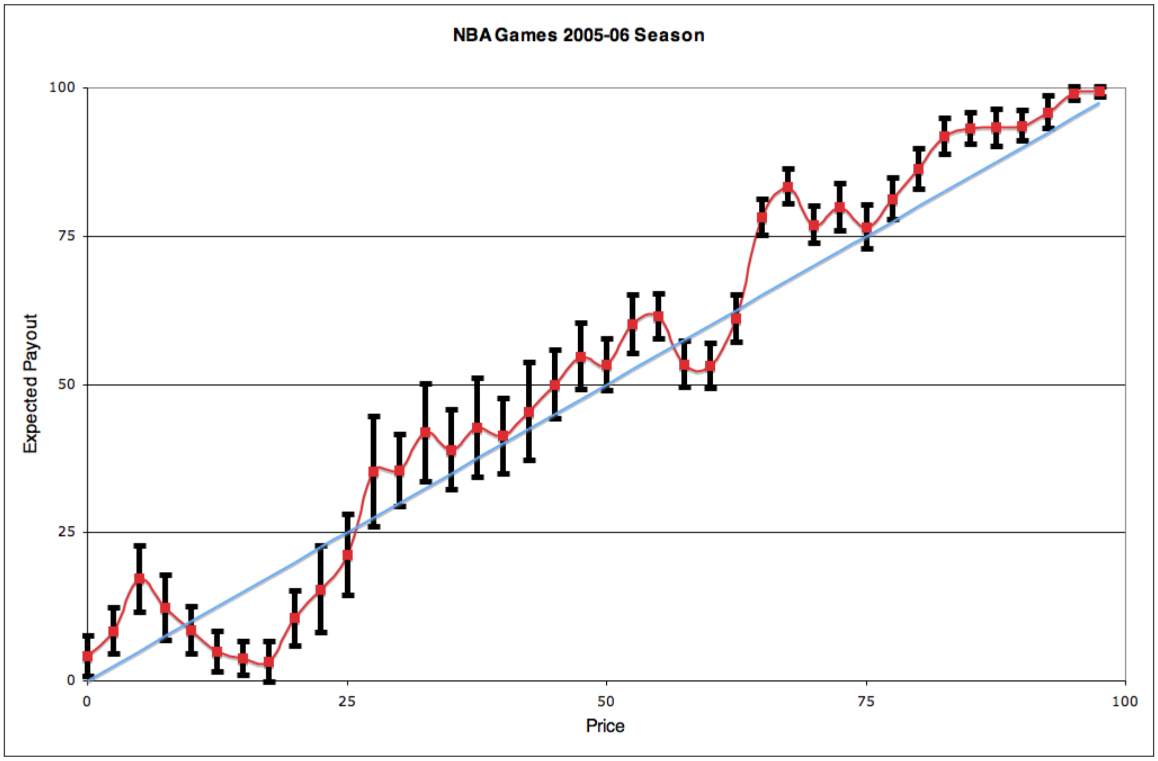

Figure 4.3: Prices and Expected Payouts for Tradesports contracts in the 2005-06 NBA Season.

Figure 4.3 shows aggregated Tradesports contracts taken from the 2005-06 NBA season, plotting the price of a particular trade against its expected payout. Following the methodology used by Wolfers and Zitzewitz [30] to produce figure 4.1, we partition trades into 2-cent "buckets", such that each trade belongs to only one bucket. The thick black bars represent 95th percentile confidence intervals, noting that buckets can involve sampling a single contract many times, and that nearby buckets often have many of the same contracts within them.

Here, the NBA markets show much of the same bias that was found in the IEM winner-take-all markets; note how contracts around 20 cents are significantly overvalued, and contracts around 80 cents are significantly undervalued. Given that this data was taken from a completely separate market, with different trading rules, for a completely different set of events, finding precisely the same bias provides strong evidence that the longshot bias is structurally endemic to prediction markets.

4.2 The Winner-Take-All Market

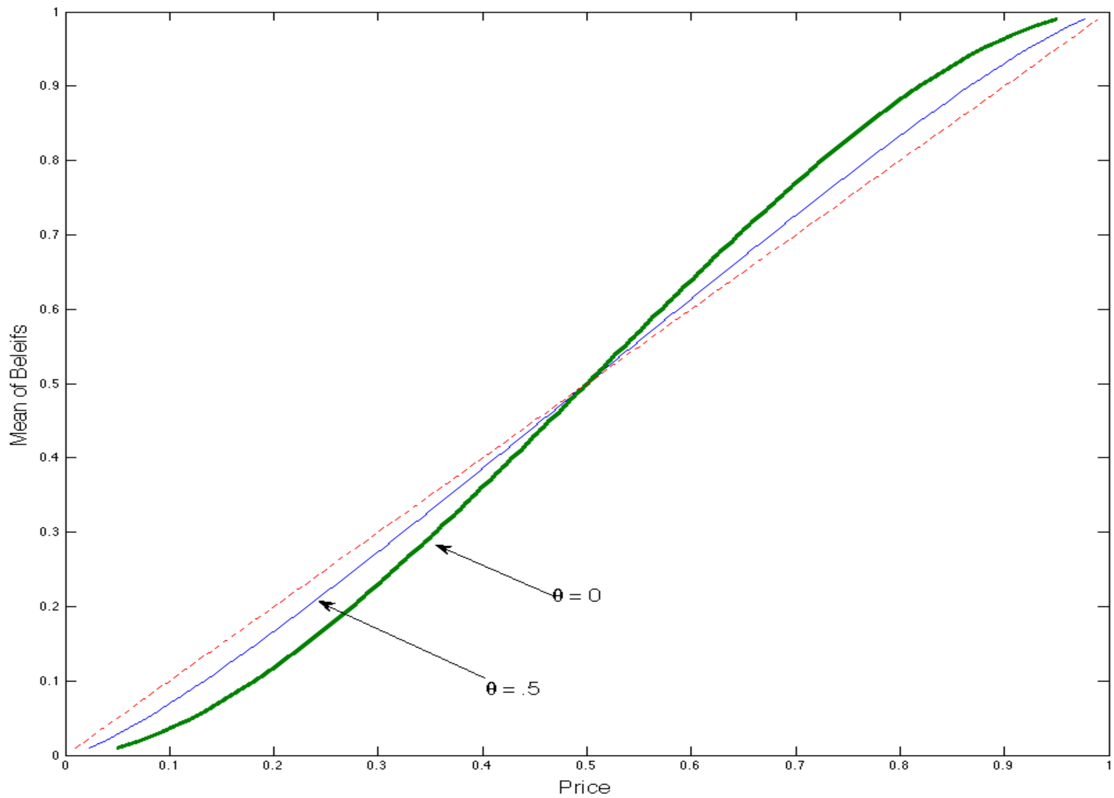

Figure 4.4: Calculated winner-take-all equilibrium prices when $\theta = 0$ and .5.

Returning to the model we developed in chapter 2, figure 4.4 shows the equilibrium solutions for risk-aversion parameter $\theta = .5$ and $\theta = 0$. Recall that $\theta = 0$ corresponds to risk-neutrality, while $\theta < 1$ corresponds to weak (sub-logarithmic) risk aversion. To calculate the plot value at point z, we solve for the p_x that equalizes aggregate demands A_x and A_y subject to the constraint that $p_x + p_y = 1$, such that beliefs are distributed with mean z.

On the plot, the dashed line represents perfect accuracy, while the thick line represents prices when $\theta = 0$ and the thinner line prices when $\theta = .5$. Note how, just as in figure 4.1 and figure 4.3, the greatest deviations from objective probabilities arise between 20 and 30 cents and between 70 and 80 cents.

We are thus able to draw the link between the observed longshot bias in prediction markets and general weak risk aversion. Unlike the studies of Wolfers and Zitzewitz [31] and Gjerstad [11], our distributional analysis allows us to develop a model based on empirical study of prediction markets in which participant risk aversion yields a significant longshot bias. Wolfers and Zitzewitz's figure 4.1 and our figure 4.3 demonstrates that the longshot bias is a real effect within prediction markets, and we are able to provide an explanation for that bias within the standard model. We feel that this is a substantial contribution to the literature.

Note also how prices for all levels of risk aversion show the most accuracy around 50 cents. This result should be viewed in the same vein as the work of Wolfers and Zitzewitz and Gjerstad in rejecting Manski's assertion that prediction markets are least accurate around prices of 50 cents [19]. That result was based around pathological distributions of beliefs and should be regarded as a thought experiment rather than an empirical reality.

4.2.1 On Partisanship and Risk Aversion

If weak risk aversion bordering on risk neutrality is the answer to why the IEM displays a longshot bias, we must consider the causes and practical effects of such behavior. We tender two distinct but not exclusive explanations for the phenomenon. The explanations can be thought of as reflecting the duality of the term *prediction markets* - one ties in with traditional economics literature related to risk and uncertainty, while the other is more psychological, related to the act of prediction itself.

The total amount of money a prospective trader can have in the Iowa Electronic Markets as a whole is 500 dollars. As a result, traders' investments in the market are capped at a relatively small level compared to their total wealth. A range of examples documenting risk-neutral or slightly risk-averse behaviors are summarized in Palacios-Huerta *et al.* [22], and in general, the authors point out that for small gambles, the overwhelming evidence seems to point towards the average individual having a relative risk-aversion parameter θ of less than one, with many studies settling on $\theta \approx .5$, which corresponds to sub-logarithmic risk aversion. The relatively small investment levels of prediction markets should certainly be given serious consideration as the cause of the general lack of risk-aversion among participants.

But prediction markets are also futures markets, in the sense that they have payouts contingent on future events. Futures markets are used by traders to hedge their risk against these future events. One can see an easy analogue here with political markets. Consider an alternative energy consultant whose personal beliefs and economic interests side with the Democratic candidate winning a presidential election. This trader could hedge against the sharp variance in their post-election states of the world by investing in the Republican candidate. However, Wolfers and Zitzewitz [30] note that the small stakes of current prediction markets preclude any meaningful economic hedging. If prediction markets continue to grow, however, the hedging of individual traders against their economic risk will certainly have an intriguing impact on the efficiency of future markets and is a promising avenue for future theoretical study.

Note that hedging runs counter to our perception about how people would make investment decisions in prediction markets - our fictional alternative energy consultant would place orders based not on their beliefs about the election but rather on their specific circumstances. However, we are more interested in exploring the opposite effect, a phenomenon that I will refer to as *partisanship*, which can help explain low levels of risk aversion. To illustrate the practical effects of the risk-aversion parameter, figure 4.5 displays various levels of individual demand at different belief levels when $p_x = .3$ and $p_x + p_y \geq 1$.

As is evident, weaker levels of risk aversion correspond to investing more at lower belief levels. The idea behind partisanship is that agents have state-dependent utilities based on the outcome of the event being wagered upon, and that these state-dependent utilities may cause them to invest more than if utility was based only on wealth. Thus, it is possible that a trader may be weakly risk-averse (say, $\theta = .5$), but behave in such a way that they are much closer to risk-neutrality ($\theta = 0$).

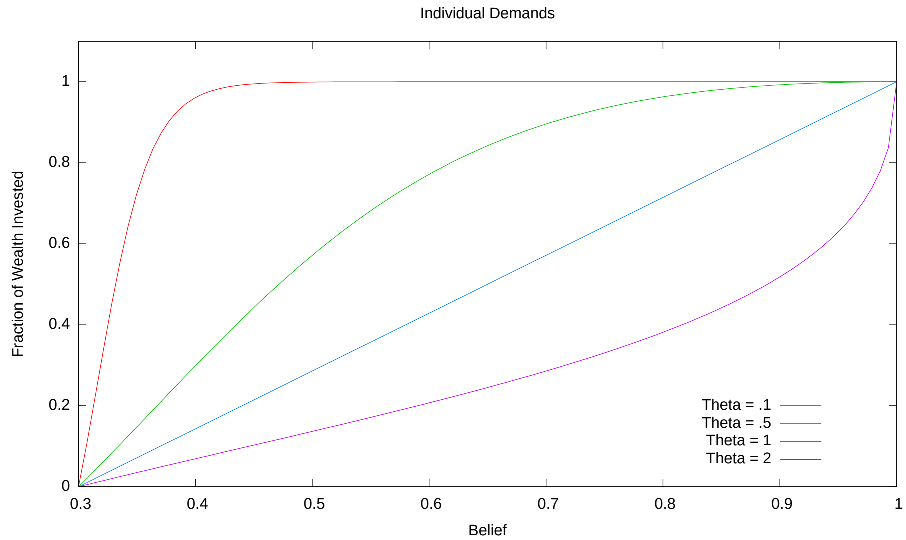

Figure 4.5: Individual demands relative to beliefs for different levels of risk aversion.

To formalize this, imagine that we are dealing with a trader biased towards candidate x with private belief b_i, such that that the trader has utility function $U(w, I_x)$, where I_x represents an indicator variable for the victory of candidate x. By letting $p_x + p_y \geq 1$ and $b_i > p_x$, we assert that our biased trader is interested in holding only shares of x. The agent then solves:

$$D_x = \arg \max_{0 \leq x \leq w/p_x} b_i U(w + (1 - p_x)x, 1) + (1 - b_i)U(w - p_x x, 0) \tag{4.1}$$

We abuse notation by letting $U'(.,.)$ represent the partial derivative of U with respect to its first argument. Our first order condition is then the x^* that solves:

$$b_i(1 - p_x)U'(w + (1 - p_x)x^*, 1) = (1 - b_i)p_x U'(w - p_x x^*, 0) \tag{4.2}$$

Now let $\hat{x}$ represent the solution to the demand equation for an unbiased agent with identical private beliefs. In order for $x^* > \hat{x}$, that is, for the biased agent to demand more than the unbiased agent, the following condition is necessary and sufficient:

$$\frac{U'(w+(1-p_x)\hat{x},1)}{U'(w+(1-p_x)\hat{x})} > \frac{U'(w-p_x\hat{x},0)}{U'(w-p_x\hat{x})} \tag{4.3}$$

What is the intuition behind equation (4.3)? The term on the left represents the we change the way we view our endowment in the state of the world in which our chosen candidate wins, while the term on the right represents the change in the way we view our endowment in the state of the world in which our chosen candidate loses.

To further elucidate this point, consider an illustrative biased trader. That trader feels positively reinforced by the state of the world in which their candidate wins, basking in the victory of their chosen party as well as their own predictive power, while feeling so badly in the state of the world that their candidate loses that even the loss of additional funds in a prediction market matters little. Thus, for this agent, $U'(w+(1-p_x)\hat{x},1)$ is much greater relative to $U'(w+(1-p_x)\hat{x})$ than $U'(w-p_x\hat{x},0)$ is to $U'(w-p_x\hat{x})$. Equation (4.3) therefore holds, and the trader will invest more at lower belief levels than their unbiased counterparts, essentially simulating the effects of having a lower level of risk aversion.

But recall that we started our inquiry with the assumption that $b_i > p_x$, that is, that the agent biased towards candidate x has a higher belief that candidate x will win the election. If biased agents do not believe their favored candidate will win, then we cannot argue for the partisanship effect. There is significant evidence, however, suggesting that agents biased in this way do exist. In particular, a survey of the 1988 IEM presidential market found that the market could be divided into three groups based on trading behavior - agents biased towards Dukakis, agents biased towards Bush, and agents biased towards neither [9]. Forsythe *et al.* [10] cite a survey in which for every presidential election since Stevenson/Eisenhower in 1952 voters were asked whether they thought the candidate they supported would win the election. For no election did fewer than 71% of voters think their candidate would win, and the average figure was 79%. Thus, people who support a candidate have an optimistic view of that candidate's chances of victory.

Though the examples we have used and referenced are political in nature, we do not feel that the partisanship phenomenon need be strictly confined to that arena. One could imagine agents placing orders on sports games displaying preferences for certain teams; not necessarily in that they would actively be fans of a particular team, but perhaps that they would display preferences for the play style, coaching, or possible playoff implications for their own favorite team that would cause them to be biased in the direction of one team or another.

4.3 The Vote-Share Market

Unlike in the winner-take-all market, the IEM vote-share market does not appear to hold a longshot bias.

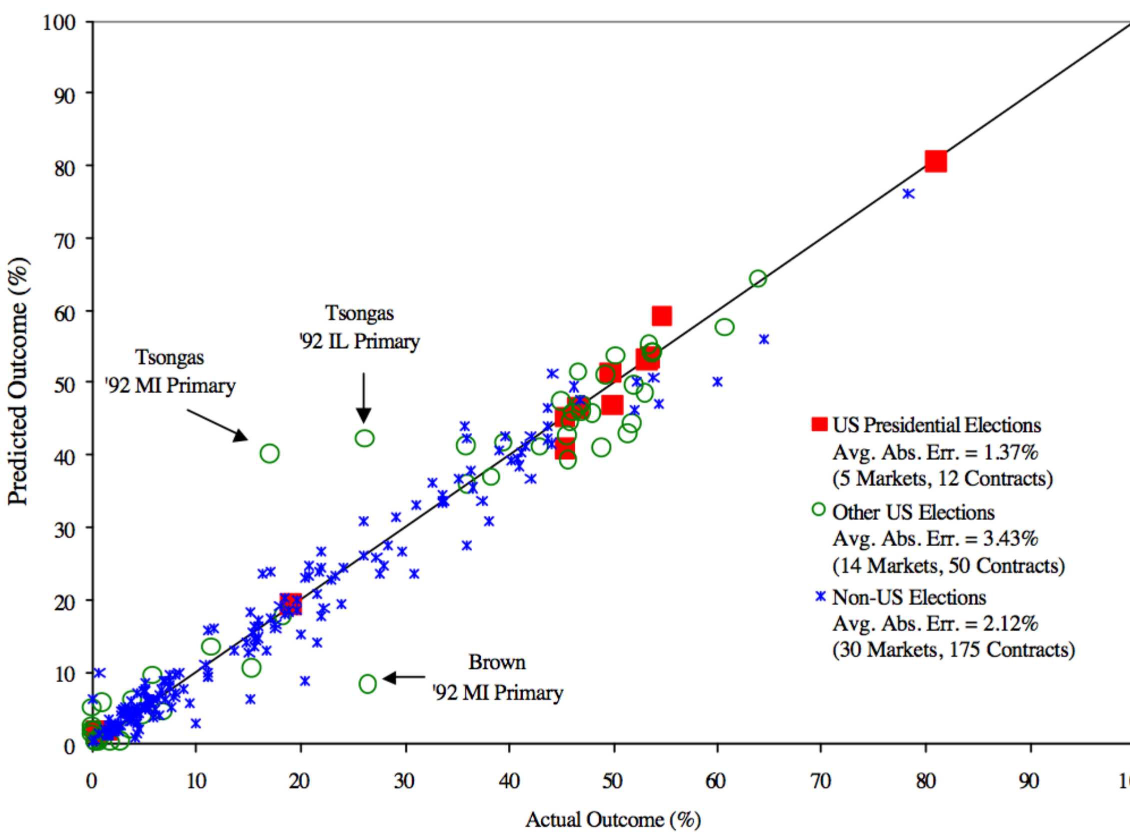

Figure 4.6: IEM Vote-Share Predictions and Results. Taken from [4].

Figure 4.6 shows the collected results of a dozen years of IEM vote share markets, as is evident, we see no correlation between share prices and bias. We restrict our analysis, however, to vote-share markets with outcomes between 35 and 65 percent. Recall that our model for vote-share is based around the unbounded normal distribution rather than a double-bounded distribution that would allow for evaluation at the edges of the state space (it is the embedded winner-take-all market induced by the vote-share framework that is double bounded and does not suffer from this problem). The values of .35 and .65 were chosen because they are far from the boundaries of successes and failure at which our model delivers results based on murky assumptions (e.g. negative vote share). Still, the restricted domain gives us cover for the most important and widely traded vote-share markets - those for presidential elections. The most lopsided election in modern history was Richard Nixon's 1972 victory in which he captured 61.8 percent of the 2-party vote share[1] and for a candidate to win a presidential election by more than 30 percent is virtually inconceivable.

Figure 4.7 shows the calculated equilibrium prices for the vote share market between .35 and .65. To generate the equilibrium price at actual outcome z, we fix the mean of the distribution of vote share beliefs to be z and solve for the p_x that equalizes aggregate demands subject to the constraint that $p_x + p_y = 1$. The observant reader will note that our axes have "flipped" relative to the graphs in the previous subsection - this was done to facilitate comparisons with figure 4.6.

The black line indicates perfect accuracy, the red line represents equilibrium solutions with risk aversion parameter $\theta = .5$, while the red circles represent solutions for risk-averse agents, with $\theta = 0$. As is evident, the solutions corresponding to the different levels of risk aversion are virtually identical, differing by no more than 2 thousandths of a cent. Furthermore, even at the boundaries of the plot, at 35 and 65 cents, the solutions differ from the actual outcome by less than a cent.

We might ask how these results compare to the winner-take-all market on the restricted domain; one might argue that a lack of significant bias is a result of operating only on this smaller set of values.

[1]Readers with an interest in politics will recall that in 1964 Lyndon Johnson captured the largest fraction of the popular vote in the modern era, but third parties played a larger role in the 1972 election and so Nixon's share of the *2-party* vote share was actually marginally greater.

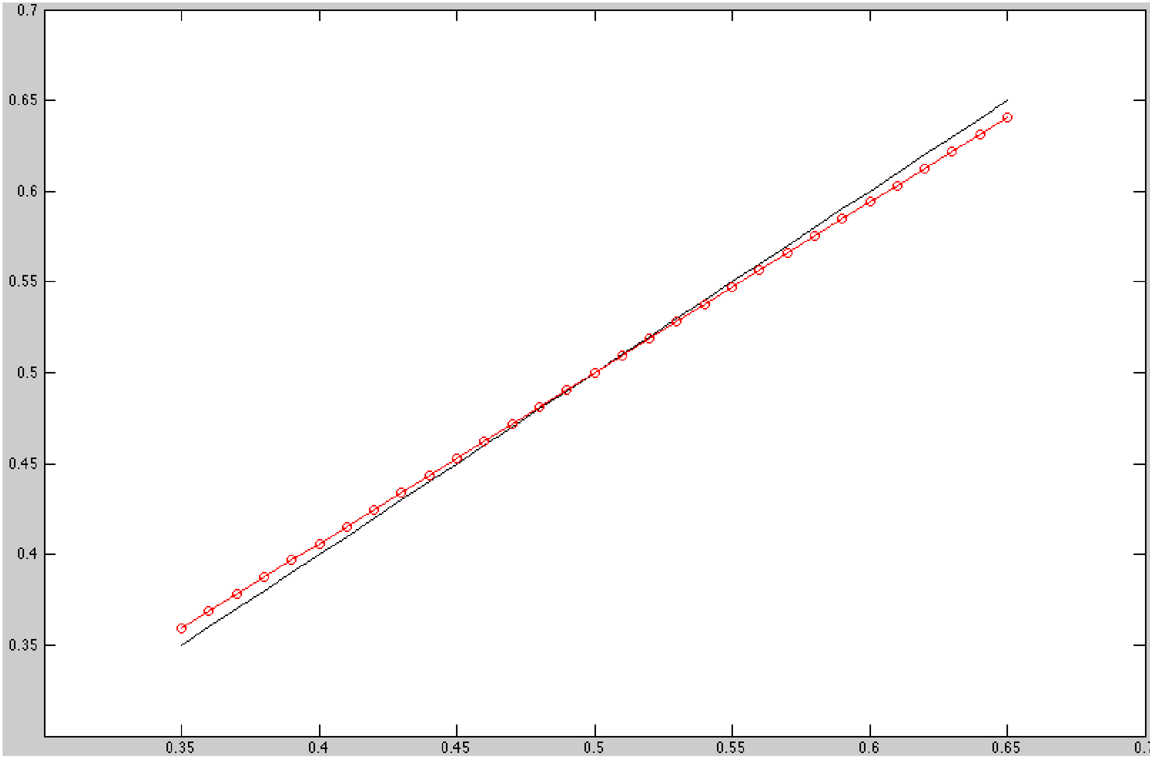

Figure 4.7: Calculated vote-share equilibrium prices for different levels of risk aversion.

Figure 4.8 refutes such a claim. It displays the equilibrium prices for the vote-share and winner-take-all markets with risk neutral agents relative to perfect accuracy, where perfect accuracy in the winner-take-all market refers to a candidate having a certain percentage chance of winning an election, while perfect accuracy in the vote-share market refers to a candidate begin most likely to capture a certain percentage of the vote share.

Here, the light blue line is the vote-share equilibrium price, while the orange line is the winner-take-all equilibrium price. The winner-take-all market has deviation from accuracy of nearly 4.5 cents at the edges of this reduced domain, and it fits the observed data much worse than the vote-share market. Thus, it is evident that the lack of deviation from equilibrium solutions is a property of our vote-share market model, rather than an artifact of the reduced domain.

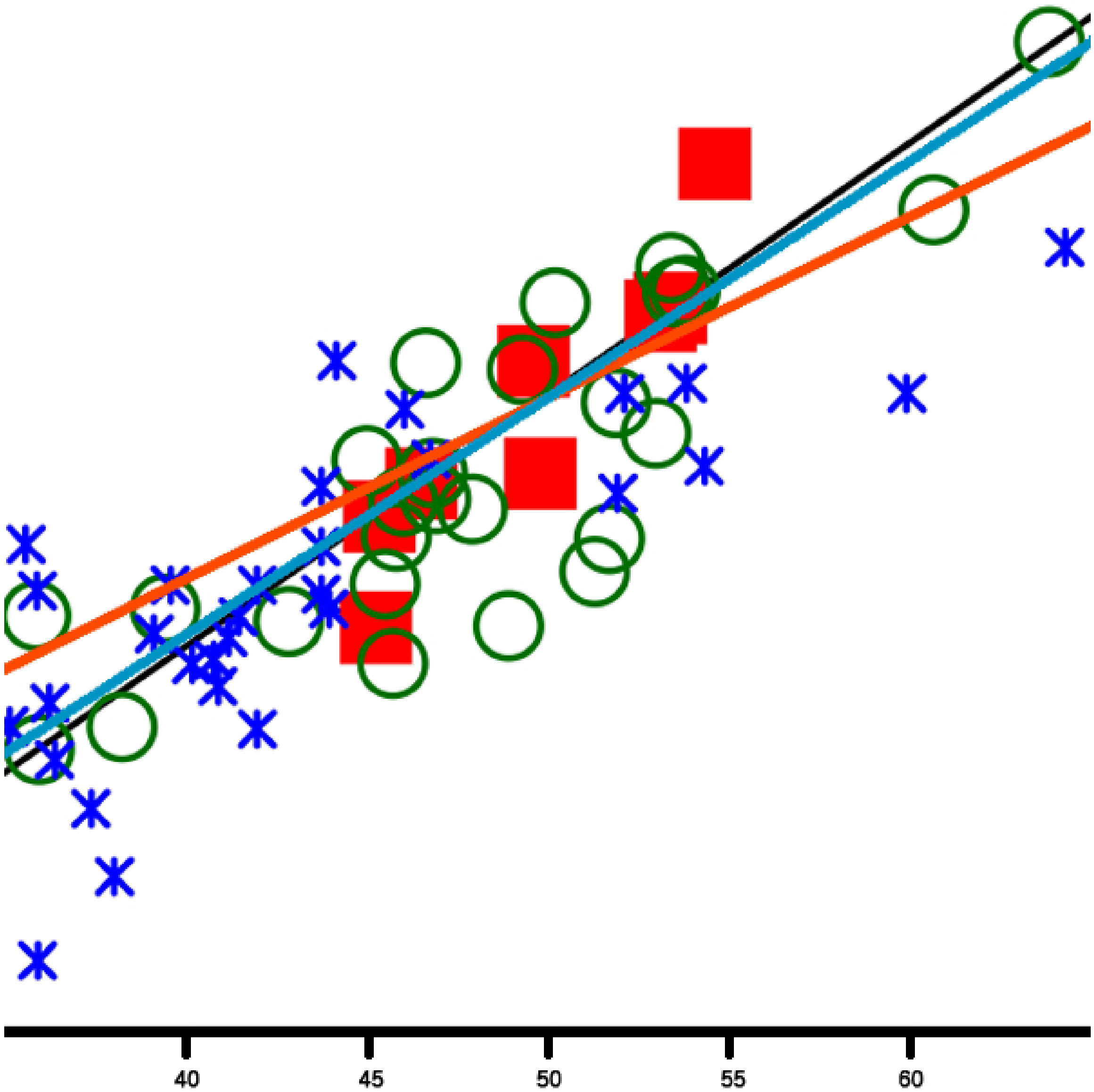

Figure 4.8: Comparison of winner-take-all (orange) and vote-share (light blue) equilibrium prices for outcomes between .35 and .65.

Chapter 5

Equilibrium Quantities and Overpricing

In two probing and extensive laboratory studies, Forsythe *et al.* [10] and Rietz [24] consider a winner-take-all market under which a "green" candidate wins 70 percent of the time and a "blue" candidate wins 30 percent of the time. Traders are given fixed, but unbalanced, amounts of the assets - five traders started with six shares of green and two shares of blue, and five traders started with two shares of green and six shares of blue, for a total of 40 overall shares. Traders have an initial wealth of 40 dollars each. Both studies found that final prices consistently exceeded arbitrage levels. For instance, Rietz found that prices exceeded expected values significantly in 83% of trading sessions, and that the shares summed, on average, to 1.145 to over 1.2, depending on small changes in market rules (like allowing short sales). The results persisted even when experiment participants were experienced in the market, and had heard an explanation of how to exploit arbitrage opportunities.

These results pose serious problems to the theoretical basis of our model. The probabilities of .7 and .3 were commonly known, and only risk-loving agents would make an investment at a price greater than their belief. In order to break through this conundrum, we assume that agents misjudge the true probabilities - that is, that they act in such a way consonant with having a noisy belief over the true outcome. This can be due to psychological effects such as "hunches", a misunderstanding of applied probabilities, or an endowment effect, as traders began the simulation with unbalanced holdings of the assets. We assume, however, that agents correctly understand the complementary nature of the contracts, such that an "applied" belief that green will win of p implies an"applied" belief that blue will win of $1 - p$. We thus preserve the property that the beliefs of each individual agent sum to precisely 1.

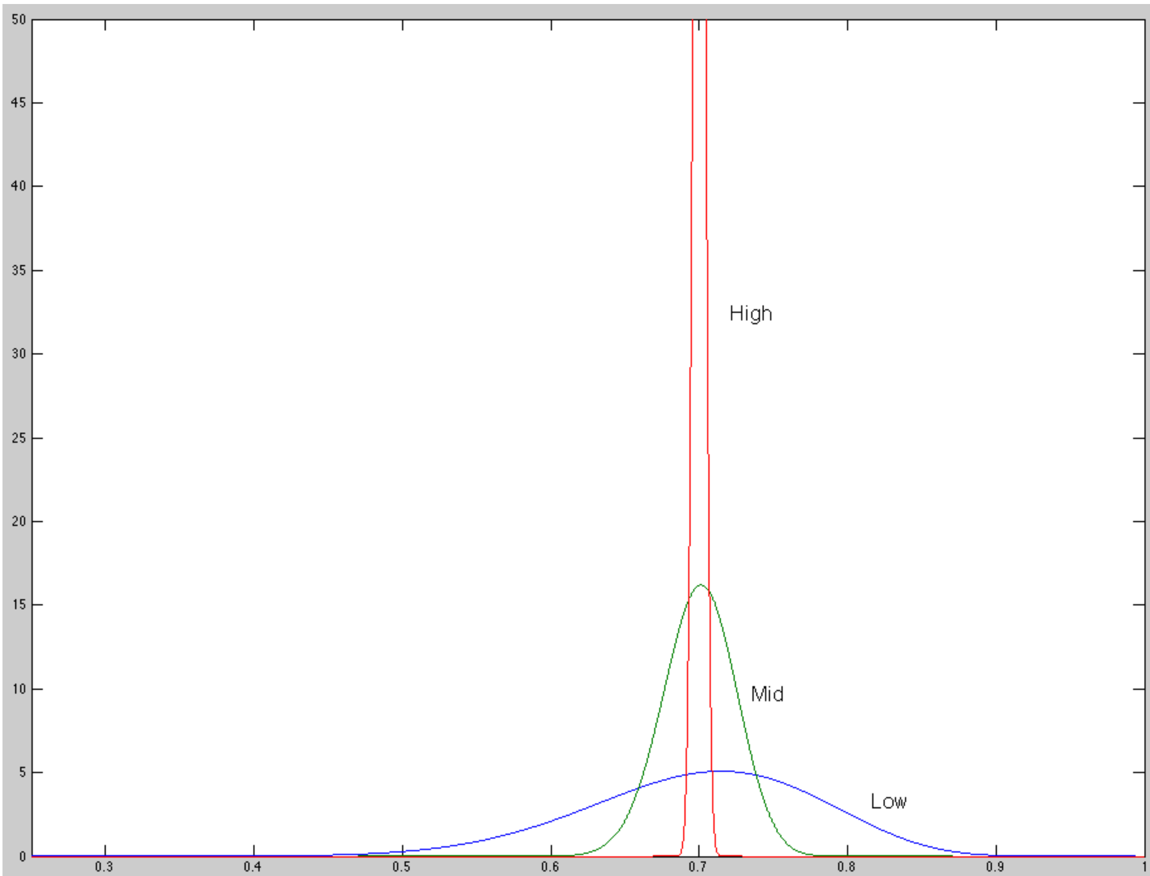

Figure 5.1: Belief densities for different levels of accuracy.

We examine equilibrium pricing under the three different belief regimes for the green candidate depicted in figure 6.4. The "high" accuracy pdf shows non-zero density only in a band about two cents from true accuracy, while "low" accuracy pdf is more diffuse. The mean of each distribution is .7. Note that we do not claim the actual participants in the study had beliefs corresponding to one of these densities, instead, we only present a set of distributions to use as a basis for further computational study.

One advantage in the very meticulous approach to the equilibrium concept in the model of chapter 2 is that it allows us to investigate equilibrium quantities, something that has not been examined in previous literature.

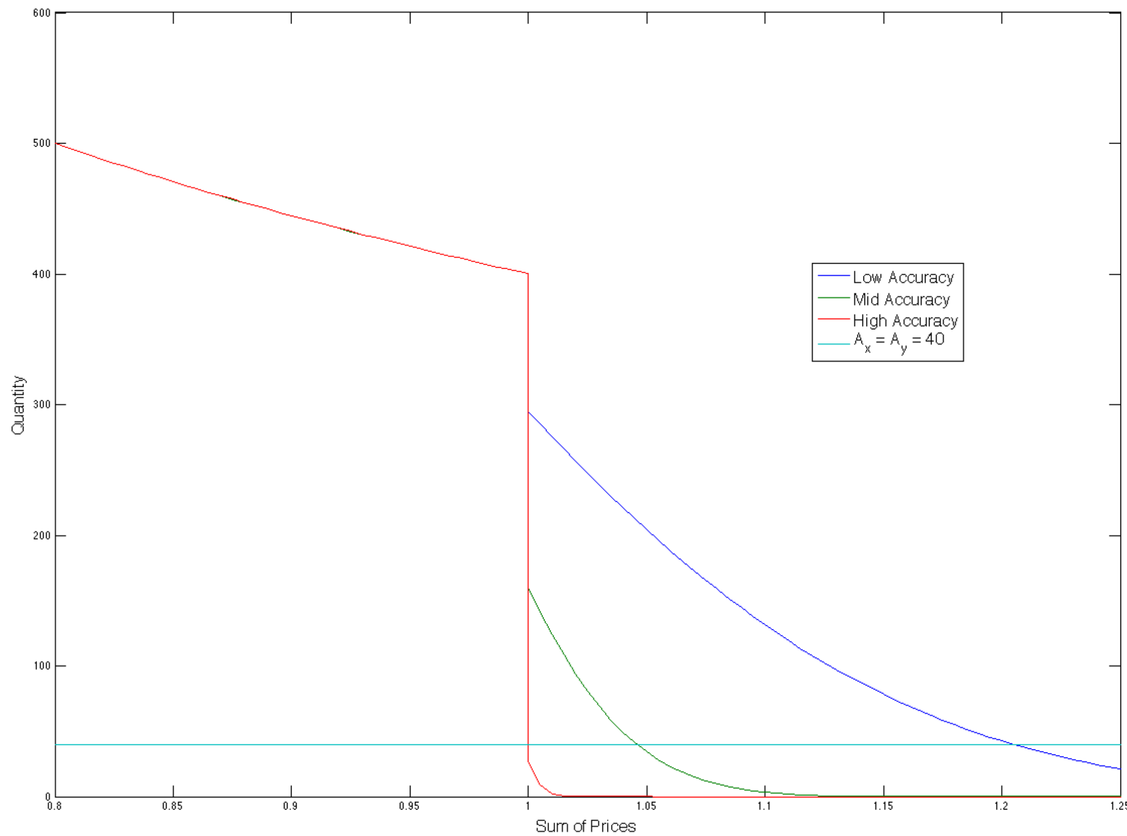

Figure 5.2: Undersupplied quantities and diffuse beliefs lead to overpricing in the Forsythe *et al.* and Rietz experiments.

Figure 5.2 shows equilibrium quantities as a function of the sum of prices - to generate the plot at point z the equilibrium price for the green candidate p_x was found subject to the constraint that the price for the blue candidate would be $z - p_x$. The horizontal line at $A_x = A_y = 40$ corresponds to the (fixed) quantities available in Forsythe *et al.* [10] and Rietz [24]. As is evident from the plot, overpricing at $A_x = A_y = 40$ occurs for a range of belief levels, including around 5 cents over with "middle" beliefs and more than 20 cents over at the "low" accuracy level. We set $\theta = .1$ (very weak risk aversion) for the plot, though the results are qualitatively similar for all values of θ, with the discontinuity at 1 increasing with θ. This discontinuity corresponds to the continuum of solutions that exist when $p_x + p_y = 1$; recall that at that price agents are indifferent between holding k dollars and holding k bundles of shares. Finally, note that even if the discontinuity at 1 did not exist, the derivative of the demand curves with respect to the sum of prices would not be continuous. Claim 6 in the following section shows that when the prices sum to less than 1, all agents maximize their holdings, and thus the equilibrium quantities correspond to $\frac{wM}{p_x+p_y}$. As the price increases past one, we not only see this price effect, but in addition there are also agents that demand no holdings whatsoever, and as such the slope of the demand curves past 1 is steeper than the slope for prices that sum to less than 1.

We feel that figure 5.2 represents a significant contribution to the literature. Not only does it explain why overpricing in laboratory experiments occurs, but it also elucidates some of our intuition behind equilibrium prices. Arbitrage arguments in the limit of market thickness are traditionally made to necessitate that prices sum to 1 in a double-auction prediction market. However, the computational approach undertaken in this work shows that for many quantities, prices should sum to 1 *even in the absence of arbitrageurs.* The discontinuity in the demand curve at $p_x + p_y = 1$ provides an alternative explanation for why market prices should align to this "equilibrium" price level.

5.1 Implications for Experimental Design

Perhaps wanting to encourage flexibility for experimental market participants, economic alchemists generally provide their test subjects with a level of wealth that forces prices above efficient levels. To formalize this, let M market participants each hold w dollars at the start of the economic simulation.

Claim 6. *Regardless of the particular experiment or associated probabilities, for any $s < 1$, if individual demands are continuous and monotonically decreasing with respect to their indexed input prices, and* $A_x(0, s) \geq A_y(0, s)$ and $A_y(s, 0) \geq A_x(s, 0)$ *then there exists a unique price vector $(\hat{p}_x, \hat{p}_y)$ such that $\hat{p}_x + \hat{p}_y = s$ and:*

$$A_x(\hat{p}_x, \hat{p}_y) = A_y(\hat{p}_x, \hat{p}_y) = wM/s$$

Proof: The function $A_x(p_x, s - p_x) - A_y(p_x, s - p_x)$ is positive at $p_x = 0$ and negative at $p_x = s$, and thus because A_x is a monotonically decreasing function and A_y is a monotonically increasing function in p_x, there is precisely one value of $\hat{p}_x$ (and $\hat{p}_y = s - \hat{p}_x$) such that $A_x = A_y$. Since $s < 1$ agents maximize their holdings by utility monotonicity, and as such each agent will spend w/s regardless of the particular value of p_x and p_y adopted. It follows that at $(\hat{p}_x, \hat{p}_y)$, $A_x = A_y = wM/s$. ■

We can thus assert that in equilibrium:

$$\lim_{p_x+p_y \to 1^-} A_x = \lim_{p_x+p_y \to 1^-} A_y = wM$$

We note that Forsythe *et. al* [10] and Rietz [24] fix their quantities at one-tenth of wM. It is our suggestion that this level is too low to force prices into summing to 1 for many belief distributions. Consequently, we suggest fixing quantities to be one-half of wM. We imagine that it will be easier to accomplish this goal by reducing w rather than increasing available quantities, for two reasons:

- To avoid "flooding" the market in the eyes of market participants, it might be easier for them to think about having less money than having fewer shares. We imagine it is much simpler for participants have 8 dollars, rather than 20, than initial allocations of 40 shares, rather than 8.

- With less cash on hand, simulations will run faster and perhaps participants will be more judicious with their spending choices. With more quantities, we can imagine exactly the opposite effect - longer, more costly simulations with less value being placed on each individual purchase or sale.

Chapter 6

Single-Share Prediction Markets

This chapter begins our agent-based inquiry into prediction markets. As opposed to the standard model in which prices are exogenous and actions (demands) are endogenous, the agent-based model determines prices endogenously from behavioral rules. Additionally, agent-based modeling can be used to simulate a market over time, as opposed to only the static analysis provided by the standard model.

In this chapter we adapt Gode and Sunder's Zero-Intelligence agents to a single-share (Tradesports-style) prediction market. Somewhat counter-intuitively, we find that our agent-based model exhibits the longshot bias, which we explicate both numerically and qualitatively. We discuss the philosophical implications of this finding and compare it to our previous results in the standard model. We conclude with a discussion of the buying bias, the tendency for agents to place bid rather than ask orders, providing empirical evidence for the phenomenon and implementing it within our model.

6.1 Designing Zero-Intelligence Agents

Gode and Sunder's "budget constrained" ZI agents followed the design maxim of "when in doubt, decide randomly and uniformly" in all matters except one: agents were never allowed to perform an action with a negative expected payout (this being the "budget constraint" of the ZI agents).

Following their example, we constructed the following trading rules: When called to market, an agent chooses to buy with probability .5 or sell with probability .5, essentially flipping a coin to determine their market role. The agent then either places a standing bid order at a value uniform on $[0, b]$ for one share, or a standing ask order uniform on $[b, 1]$ for one share. If the order crosses the queues, an exchange takes place at the earlier price. Unlike in Gode and Sunder's work, a successful trade does not empty the order queues, which is behavior more consonant with the functioning of actual prediction markets.

Agents are called to market in random order, but they have no memory of their prior actions, nor do they attempt to discover anything that might be revealed in past or present market prices.

6.2 Initial Results and Reactions

10,000 simulations were run with a pool of 250 agents having valuations independently drawn (for each trial) from the "low accuracy" distribution of chapter 5. Recall that that distribution was fairly diffuse around its mean of .7.

Figure 6.1 shows the results of the simulation. The thick red line represents the mean price at each market round (market rounds are demarcated by a successful trade), while the thin red lines represent 95th percentile confidence intervals. This plot has several salient features.

- First, note that prices very quickly settle. The data for the 50th trade is virtually identical to the data for the 500th trade. This rapid convergence is remarkable given two factors: the spread out distribution of beliefs, and that our ZI agents do not learn or back out values from trades that have occurred or gain any information from current order books.

- However, even though prices very quickly settle into what appears to be a long-term equilibrium, there is still significant volatility within the market, as evinced by the relatively large variance in prices. This is likely the result of the relatively small pool of agents participating, combined with trading rules that mandate uniform placement

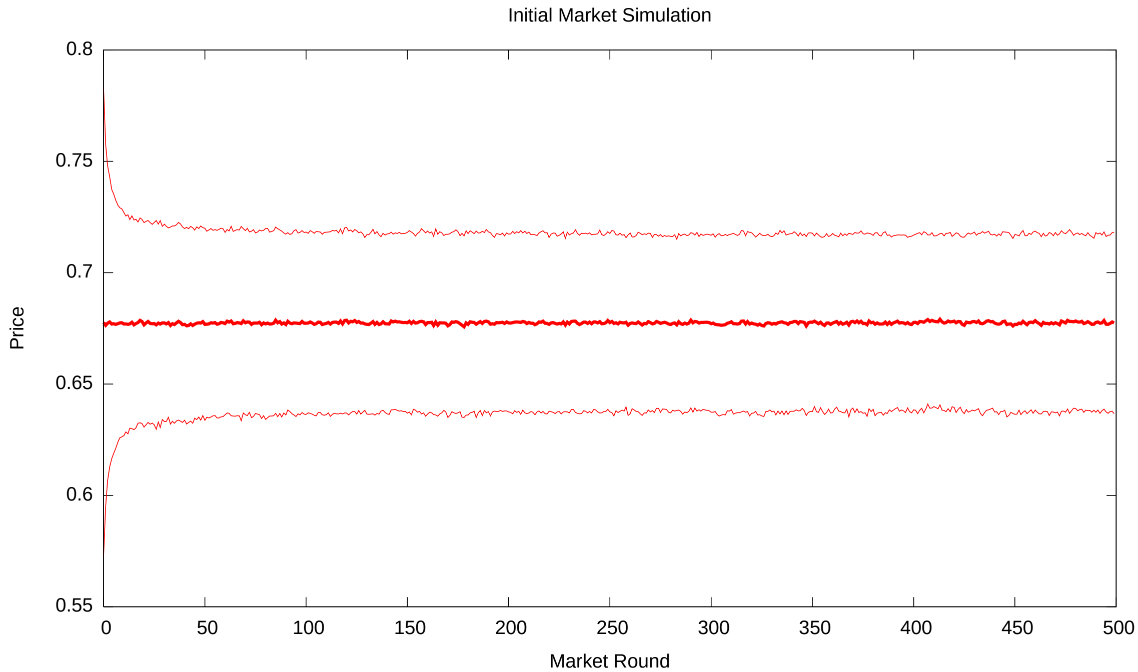

Figure 6.1: Market Prices over Time with ZI agents.

of orders rather than orders closer to beliefs.

- Most prominently, however, the prices that emerge are consistently lower than the mean of beliefs at .7 (and, for that matter, also lower than the median and mode of beliefs, which are at .7038 and .7139 respectively). This systemic mispricing deserves further attention.

6.3 Explicating the Longshot Bias

We provide two different routes for explaining why the agent-based model displays the longshot bias. One explanation is primarily numerical, while the second is more intuitive. Two sides of the same coin, both explanations are the result of discarding the assumption in the traditional model that agents are price takers, instead finding them to be more active market participants.

6.3.1 A Numerical Inquiry

In long-run equilibrium, such that order books are sufficiently saturated, we should converge to the price p^* such that in expectation over all the agents in the trading pool:

$$\mathbb{P}\left(\text{Placing a bid order at price} \geq p^*\right) = \mathbb{P}\left(\text{Placing an ask order at price} \leq p^*\right)$$

Designate the density function over beliefs of agents in the market to be f. Given our particular rules that we ascribe to our ZI agents, that is, uniform pricing and selection probabilities, by continuous exhaustive partitioning over the space of beliefs this equation becomes:

$$\int_{p^*}^{1} \left(\frac{b - p^*}{b}\right) f(b)\, db = \int_{0}^{p^*} \left(\frac{p^* - b}{1 - b}\right) f(b)\, db \tag{6.1}$$

Solving (6.1) numerically with our simulated distribution of beliefs, we find that $p^* \approx .6799$. The average price of trades 400 through 500 in the simulation was .6776. We can attribute the small difference between these values to our assumption of full saturation of the order books according to expectations. With a smaller pool of agents and a smaller number of orders, we note that there is more price space below p^* than above p^*, which leads us to believe that the second-highest bid price should be further away from the last trade than the second-highest ask price. In addition, because p^* is smaller than the mean, median, and mode of the distribution, we may run into possible discretization issues related to the pool of trader values, such that agents with beliefs greater than p^* are over-represented. As a result, there is a small downward pressure on prices, which is evidently a second-order effect.

6.3.2 An Intuitive Explanation

Another way to think of the buying bias is to consider the behavior of an individual agent placing an order in the market.

Figure 6.2 shows the fraction of trades that resulted from agents placing orders a variable number of cents from their valuation. It was generated by noting the orders which resulted

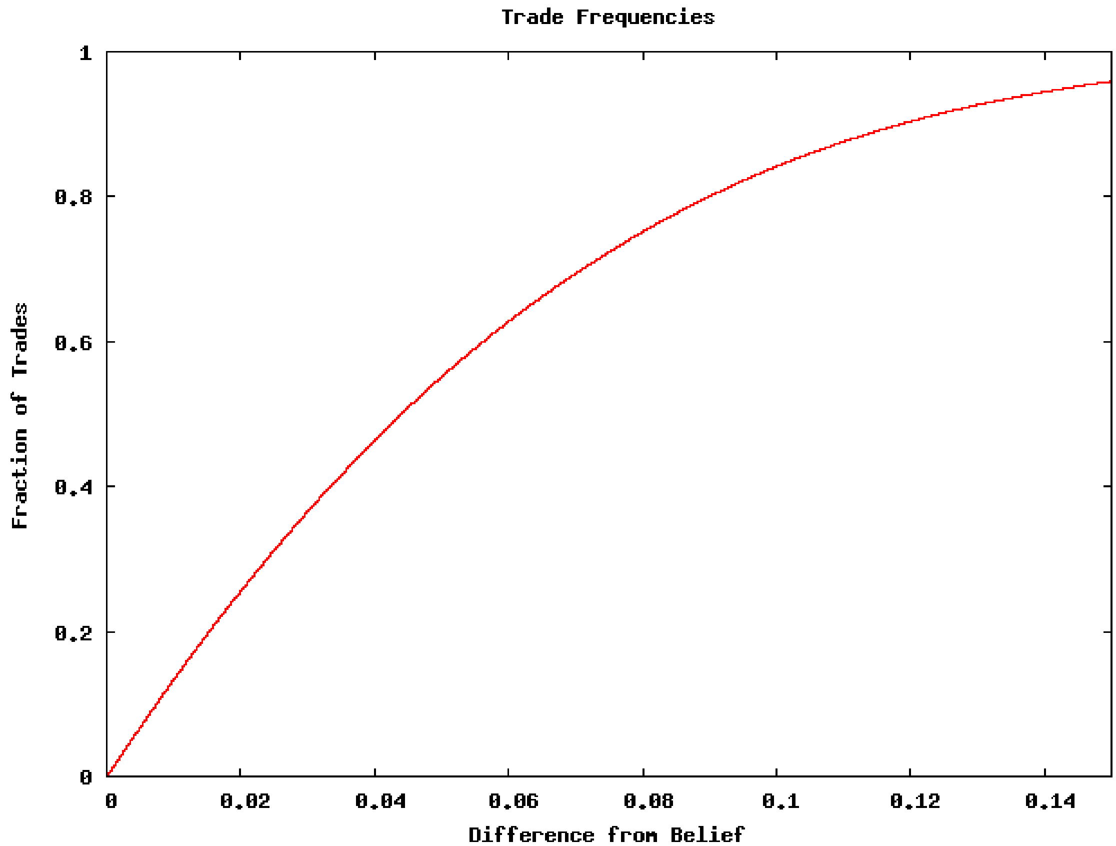

Figure 6.2: Distribution of trades in the Market Simulation.

in trades and calculating the difference in order prices from the beliefs of the agents that placed them for 1000 simulations.

As is evident from the plot, orders that were placed close to agents' beliefs had a disproportionate effect on market prices. But because the mean agent selects uniformly on $[0, .7]$ to place a bid order and uniformly on $[.7, 1]$ to place an ask order, more ask orders than bid orders will be "closer" to an agents' valuation. As a result, prices are pushed downwards. Note that the same intuition holds if the mean of beliefs is less than one-half; in this case, agents will be more likely to place bid orders closer to their true valuations, creating an upward pressure on prices.

A Surprising Parallel

Observant readers will note that the concave shape of figure 6.2 recalls the earlier figure 4.5. Is there a connection between the two figures?

The answer, perhaps unexpectedly, is that there is. The work of Gjerstad [11] showed that if individual demands in the standard model of winner-take-all markets are concave, the longshot bias occurs, if they are linear, then prices converge to distributional means, and if they are convex, the reverse longshot bias is present.

Figure 6.2 indicates that the orders which are "meaningful" are those closest to individual valuations. Since, as has been discussed, the valuation space is necessarily smaller on one end than the other (if the distributional mean is not one-half), this creates pressure away from the truncated side, leading to the longshot bias. However, if trades placed at any distance from individual beliefs were equally important in setting market prices, figure 6.2 would be a straight line and it would not matter that one side of the price space was larger than the other, and prices would converge to distributional means. Finally, if figure 6.2 were convex, then orders far away from values would matter more than orders closer to values, which would create pressure on prices from the larger side of the price space (because an agent is more likely to place an order far away from trading prices selecting uniformly on $[0, .7]$ as opposed to $[.7, 1]$). As a result, equilibrium prices would exhibit the reverse longshot bias.

Thus, plots generated from two different models, using two different methods, with two different rationales produce the same geometrical intuition!

6.3.3 Comparing the Biases

Figure 6.3 provides a comparison of the longshot bias in the standard model and the agent-based model, where in the standard model we set our pool of traders to be risk-neutral ($\theta = 0$). Values in the plot were calculated by fixing a distribution, and then using equations (1.1) and (6.1) to solve for equilibrium prices. With risk-neutral agents, the standard model shows a slightly greater longshot bias than the agent-based model.

What is the significance of the longshot bias being found in the ZI agent-model? As

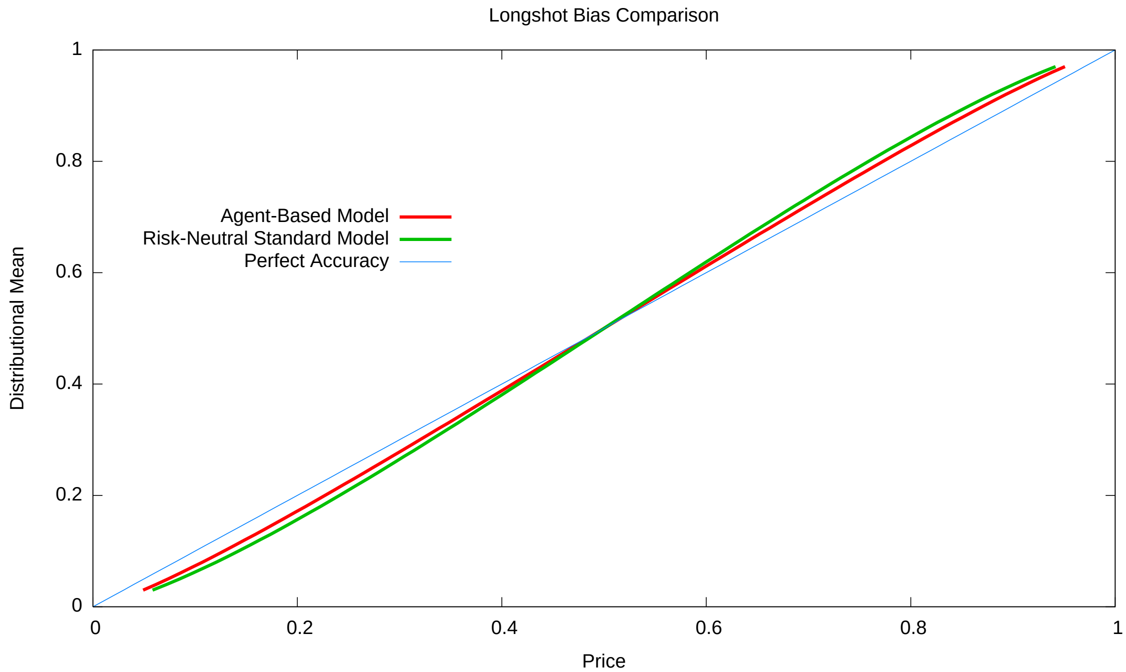

Figure 6.3: The standard model shows a slightly greater longshot bias than the agent-based model.

we have discussed, the ZI model represents equilibria for a market functioning without the presence of human rationality. It allows us to separate the constructive forces of market microstructure from the constructive forces of human behavior. The presence of the longshot bias within the ZI model suggests that the bias is intrinsic to market structure given a framework of belief-based investing. To consider why the longshot bias may occur intrinsically, consider a trader arriving to market with a certain amount they want to invest in a sports team which is currently priced at 5 cents to win an upcoming game. For the same amount of money, that trader can buy 19 times more of that team than they can sell. The net effect of many such traders would be to drive prices up. This is a stylized example, but it holds the intuition as to why both the agent-based model and standard model show the longshot bias: because both models have traders investing on beliefs, and the expression of those beliefs can be more or less pronounced dependent on something unrelated to beliefs, market prices.

We note, however, that perhaps the longshot bias of the agent-based model is not as pronounced as it should be to align with empirical results. This carries with it the surprising implication that the presence of human reasoning actually *decreases* the efficiency of the exchange. We refer the reader to our explanation of partisanship in section 4.2.1 as a possible cause of this result. If people change the way in which they view wealth (or the way in which they view their own act of correct prediction) if their chosen candidate wins, then they may invest more in a prediction market than they would in a context-free wager.

Will the longshot bias ever be defeated? We feel so. In the next five years, as studies depicting the bias in real-world markets gain wider visibility, traders essentially acting as arbitrageurs will act to drive prices to efficient levels. So far, only one study published late last year [30] has demonstrated empirically the longshot bias in prediction markets. Our work has demonstrated again the bias in a different prediction market, while also providing a theoretical justification for why it could be occurring. We imagine that as the longshot bias gains wider visibility, it will be eliminated by profit-seeking traders. As it stands now, it is a startling proposition to many traders and followers of prediction markets that certain contracts could be intrinsically mispriced in certain ways. After all, in a market, wouldn't those biases be inherently wiped out? Yes, but only if the information about those biases being present is in wide-enough dissemination.

Now imagine that we are market administrators for a new internal prediction market for a large-size company; our trading pool does not have enough contracts or savvy enough traders to counter the longshot bias. In the next section, we discuss how we can structure contracts so that this trader inexperience can in effect cancel out the longshot bias.

6.4 Agent-Based Modeling and the Buying Bias

As was discussed in chapter 1, agent-based modeling holds the key to modeling and analyzing the buying bias, the tendency for agents to place more bid orders than ask orders for a commodity. Yet the buying bias has a very limited theoretical history: the only significant work to make use of it is Rietz's study [24] where it was used to explain why prices for complementary outcomes consistently sum to more than 1. Of course, in chapter 5 we analyzed an alternative explanation to this overpricing that did not rely on the buying bias. Thus, our study of the buying bias is as foundational as it is inquisitive.

Our work on the buying bias is in three parts: We provide empirical evidence from Tradesports exchanges that indicate traders are more likely to place bids rather than asks. We implement the buying bias within our agent-based model, both mathematically and in software, and analyze its effect on prices. Finally, we conclude with an intriguing study of how the buying bias can increase market efficiency for likely events.

6.4.1 Does the Buying Bias Exist?

To examine whether or not the buying bias actually occurs in real prediction markets, we examined the "division winner" markets on Tradesports for the 2007 Major League Baseball season. Each of the thirty MLB teams is assigned to a division with four other teams, and these are the contracts for each team to finish the season ahead of the other teams in their division. The most competitive outstanding market prices for teams ranged from over 55 cents for the New York Yankees to win the American League East division, to under a cent and a half for the Baltimore Orioles to win that same division.

As of 7pm on March 27, 2007 there were 11,043 outstanding orders across all 30 markets. 8,192 of these (74.1%) were bid orders, and bid quantities outnumbered ask quantities for 23 of the 30 teams (76.7%).

Why might the buying bias exist on Tradesports? Consider a trader entering the market who wishes to express a belief that a share is overpriced. In a single-share prediction market, that trader must short sell to express this position, because new shares enter the market only by sellers creating them with ask orders. Particularly for less-skilled traders, short selling is not a "natural" position with regards to asset markets; when the average individual investor "takes a position" on Microsoft they are, with near certainty, taking a *long* position, that is, buying the asset. We posit therefore that it is easier, and in a sense more comfortable, for traders to place a bid order rather than an ask order. Such a hypothesis deserves further laboratory and empirical study.

6.4.2 Implementing the Buying Bias

We incorporate the buying bias by making our ZI agents k times more likely to buy rather than to sell. Thus, when deciding whether to buy or sell within the agent activation code, if a value generated uniformly on $[0, 1]$ is less than $\frac{k}{k+1}$ our agent places a bid order, if not then they will place an ask order.

Our mathematical intuition of equation (6.1) changes slightly in this new framework. Since in expectation agents will place bid orders k times more frequently than ask orders, the equation becomes:

$$k \int_{p^*}^{1} \left(\frac{b - p^*}{b} \right) f(b)\, db = \int_{0}^{p^*} \left(\frac{p^* - b}{1 - b} \right) f(b)\, db \tag{6.2}$$

Solving (6.2) for $k = 2$, we find that $p^* \approx .7004$. To confirm our mathematical intuition, we run a simulation incorporating the buying bias, once again with 10,000 trials and 250 agents twice as likely to buy as to sell.

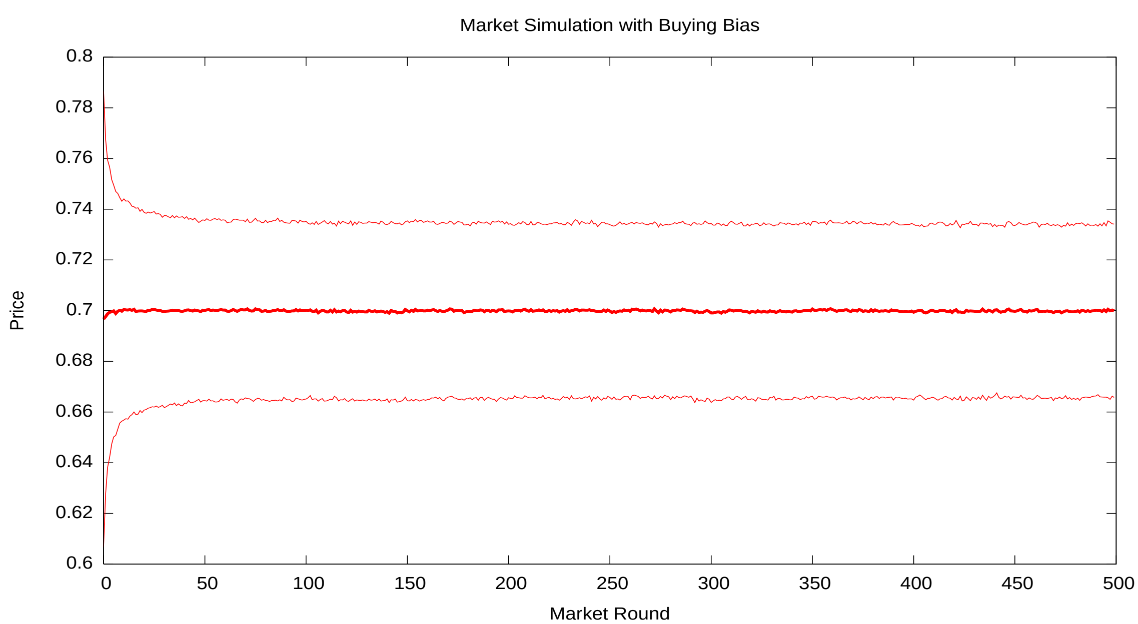

Figure 6.4: Prices with $k = 2$ buying bias.

Figure 6.4 shows the results of our simulation; the average price for the last hundred trades was .6998. Note that this plot exhibits similar qualities to figure 6.1, namely rapid convergence towards a relatively volatile pricing regime.

6.4.3 An Intriguing Observation

We note that the longshot bias pushes prices below efficient levels for events with distributional means above .5, but that the buying bias pushes prices upwards. This is the foundation of an idea with significant policy implications: that for likely events, the two biases can essentially serve to counteract one another.

Figure 6.5: The buying bias counters the longshot bias for likely events.

In order to evaluate this hypothesis, we must find a way of shifting our distributional mean so that our distributions remain qualitatively similar. In this case, we use the distribution defined by (3.12), but fixing it so that $\sigma_2^2 = .95$ and $\sigma_1^2 = .05$, as is the case that defines the "low accuracy" distribution we have been using. We note in passing that as σ_2 increases relative to σ_1 the distribution narrows, while as σ_1 increases relative to σ_2 we get most of the density located around 0 or 1.

Figure 6.5 shows the result of our study with a bias of $k = 2$ (agents twice as likely to buy as to sell). The buying bias effectively counters the longshot bias for likely events; between objective probabilities .6 and .95 the largest amount of error is 1.3 cents. To confirm our theoretical intuition we simulated markets with the various distributional means and found the results there identical to our proposed values.

We return now to our internal prediction market discussed at the close of the last section. We suggest to market administrators that they structure their single-share contracts so that the most likely outcome is the one being traded. Thus, if two sports teams were playing, the contract would be for the favorite to win, or if we have an internal question on whether, say, an upgrade for an important product will be finished by June, we could perform a small-scale internal survey and structure the contract so that it trades on the more-popular result. It is worth noting that Tradesports contracts for events are already structured this way, so that in sports predictions the most favored team is always the one being traded upon.

Chapter 7

Commissions and Market Makers

Our simple single-share ZI agent model does not include two of Tradesports' most prominent features, the commissions that drive revenues, and the market makers who are responsible for setting most of the prices on the exchange. In this chapter, we extend our model by adding commissions and market makers. We analyze the results of our simulations and discover some surprising results - among them that commissions do not decrease efficiency or volatility, and that market makers do not increase profit for the exchange.

7.1 Adding Commissions

Tradesports commissions are notoriously complicated and have changed for the more convoluted in the past year. Price taking behavior is charged either no fee, or a three or five cent fee per share, depending on the share price and the order time, while price making behavior (i.e. an order that does not cross books and fire a trade) costs nothing, but positions that end up in-the-money are assessed a fee of ten cents per share at expiry. Prior to 2007, Tradesports operated by charging either a two or four cent commission, with an expiry fee of four cents per contract being assessed for positions regardless of their expiration value.

Implementing Tradesports complex commission regime would wreak havoc upon our ZI framework. In particular, in order to avoid violating our maxim of "budget constraint", of never placing a negative expectation trade, agents would need to solve a complex stochastic equation that was very reliant on time, current market prices, and their own beliefs. Maybe price making behavior will produce positive expected payout, as opposed to price taking behavior, or perhaps the other way round, depending on the individual circumstances. Modelling Tradesports' arcane rules with bounded rational agents is certainly a promising route for future investigation, but for now we resign ourselves to a ZI regime.

In order to make commissions simple but still expressive, we say that a k cent fee is levied on both parties when a trade occurs, regardless of which party serves as a price taker or price maker. We also charge no expiry fee. As a result, agents will simply place their bid orders uniform on $[0, b-k]$ and their ask orders uniform on $[b+k, 1]$ to ensure that they never place an order with negative expected payout. We feel that this scheme manages to capture a great deal of the essence of trading commissions, while at the same time easily retaining our implicit conception of how simple our ZI agents should be.

7.1.1 Results

Table 7.1: The effect of commissions on transactions

Cent Commission	Transactions per Round
0	.0446
2	.0323
5	.0176

Table 7.1 shows the average number of transactions placed per round before 500 transactions are completed in the market. As is evident, adding commissions increases the time between orders significantly. Because agents are unable to place orders within k cents of their values, we lose a significant fraction of trades that would normally occur; figure 6.2 indicates that about 20% of trades occur within two cents of an agents' belief, and about half of trades occur within five cents of an agents' belief.

Figure 7.1 shows the results of running the market with commissions and the standard low accuracy distribution. Note that there is no buying bias, so that agents are just as likely to buy as to sell. As is evident, accuracy increases and volatility decreases with commission

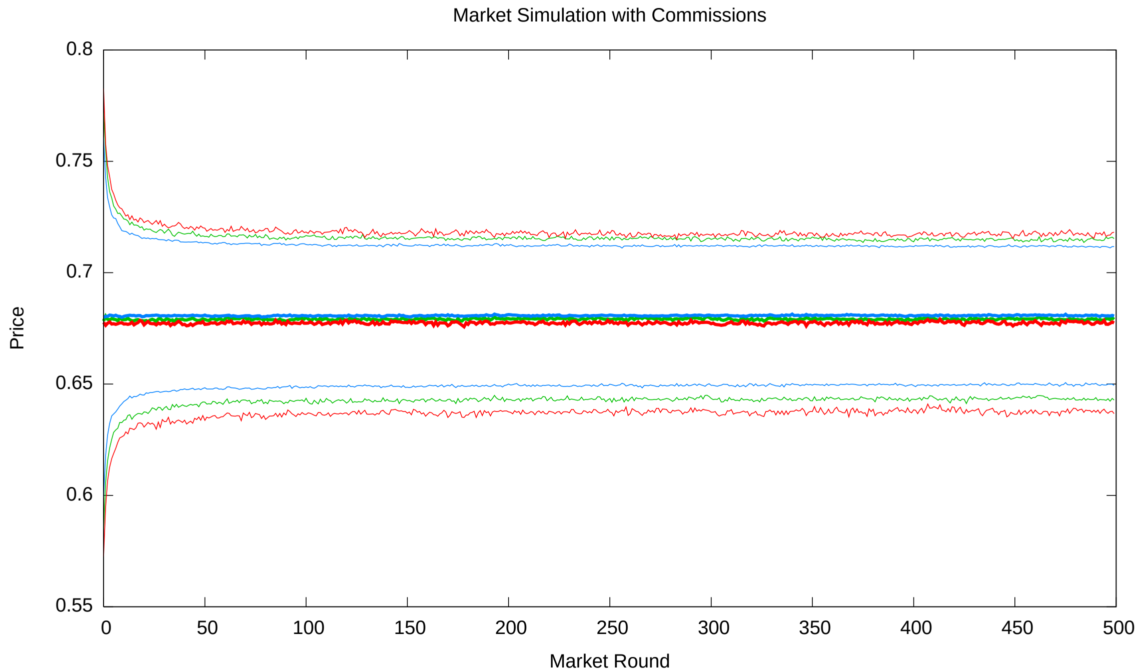

Figure 7.1: Prices with Commissions: Blue Five Cents; Green Two Cents; Red No Fees.

price. We attribute this to the increased number of trades in the order books for the higher-commission order books. More trades means a smaller gap between the best and second-best orders, so that there is less of a chance of large price scale movements. In addition, there might be less volatility because the pool of traders that can place relevant orders at the equilibrium price shrinks dramatically with commissions, because a large amount of the density of agent beliefs lies within a few cents of the equilibrium price.

Note that we can mathematically model long-term equilibrium prices with commissions, based off of equation (6.1). Since agents now do not trade within a k-cent neighborhood of their private belief, our long-term equilibrium price is the p^* that satisfies:

$$\int_{p^*+k}^{1} \left(\frac{b - p^* - k}{b - k} \right) f(b)\, db = \int_{0}^{p^*-k} \left(\frac{p^* - b - k}{1 - b - k} \right) f(b)\, db \tag{7.1}$$

Let $p^*(k)$ represent the numerical solution to equation (7.1) given a k-cent commission, while $\hat{p}(k)$ represent the average of the average price of the last hundred trades in our simulations. Our results are summarized in table 7.2:

Table 7.2: Numerical Solutions to Equation (7.1) versus Simulation Results.

k	$p^*(k)$	$\hat{p}(k)$
0	.6799	.6776
2	.6809	.6791
5	.6812	.6808

We suggest that the decreasing gap between the numerical and simulated solutions to the equilibrium price problem may be due to the higher variance in price associated with lower-commission simulations. However, note that this modest increase in accuracy and decrease in volatility comes at a significant cost in terms of trading volume - table 7.1 indicates that five cent commissions induce a more than 60% decrease in trading volume.

7.2 Incorporating Market Makers

One peculiar feature of Tradesports that is not incorporated with the IEM are market makers. These are special agents that enter in a contract with the exchange to provide liquidity. Market makers are responsible for ensuring that there are always bid and ask orders in queues are are responsible for setting the initial trading prices. In return for their services, market makers trade for free on the exchange.

7.2.1 Design and Implementation

How do we create an agent-based model for market makers? We consider three different behavioral rules, ranging from most rational to least rational:

Bounded Rational A profit-maximizing market-maker has either an omniscient knowledge of the distribution of private beliefs or uses market prices and a model of ZI behavior to back out such beliefs. Based on this distribution and the current order books, the market maker sets prices optimally in order to maximize their profits over the course of the simulation.

Omniscient The omniscient market-maker has knowledge of the efficient outcome of the markets (that is to say, the objective probability that an event will occur) and sets a bid/ask spread around this value.

Reinforcing After each trade, the reinforcing market maker establishes a bid/ask spread around the mean value of the highest outstanding bid and lowest outstanding ask orders.

We decided to model reinforcing agent behavior for our simulation, because we felt that it was the most realistic model for the way in which market makers behave within actual markets. The analysis of the longshot bias within the Tradesports NBA contracts suggests that there is no force pushing the markets towards objective probabilities, which would rule out omniscient market makers. Furthermore, we felt that creating bounded rational market makers gives those agents too much credit for their role, that essentially market makers make money not by optimizing over the space of beliefs and the time of the market but instead through the bid/ask spread. The design and implementation of more complex, profit-seeking bounded-rational market makers, bound by contract to deliver liquidity, is a promising future direction for research.

After each trade, our reinforcing market maker with a bid/ask spread of s cents operates on the following rules:

1. Removing all outstanding bids and asks placed by the market maker.

2. Letting the highest bid be at price b and the highest ask be at price a, the market maker calculates the mean price $m = (a + b)/2$.

3. If $s/2 < a - m$, then the market maker places a bid order at $m - s/2$ and an ask order at $m + s/2$. Thus, a market maker will not place an order in the market if the bid/ask spread is less than s.

Market makers generally place orders for hundreds of shares, while our market makers place orders for only one share at a time. But because we set our market makers to operate after every trade, out market makers will always have a presence in the markets unless the bid/ask spread is smaller than s.

7.2.2 Results

Market Prices

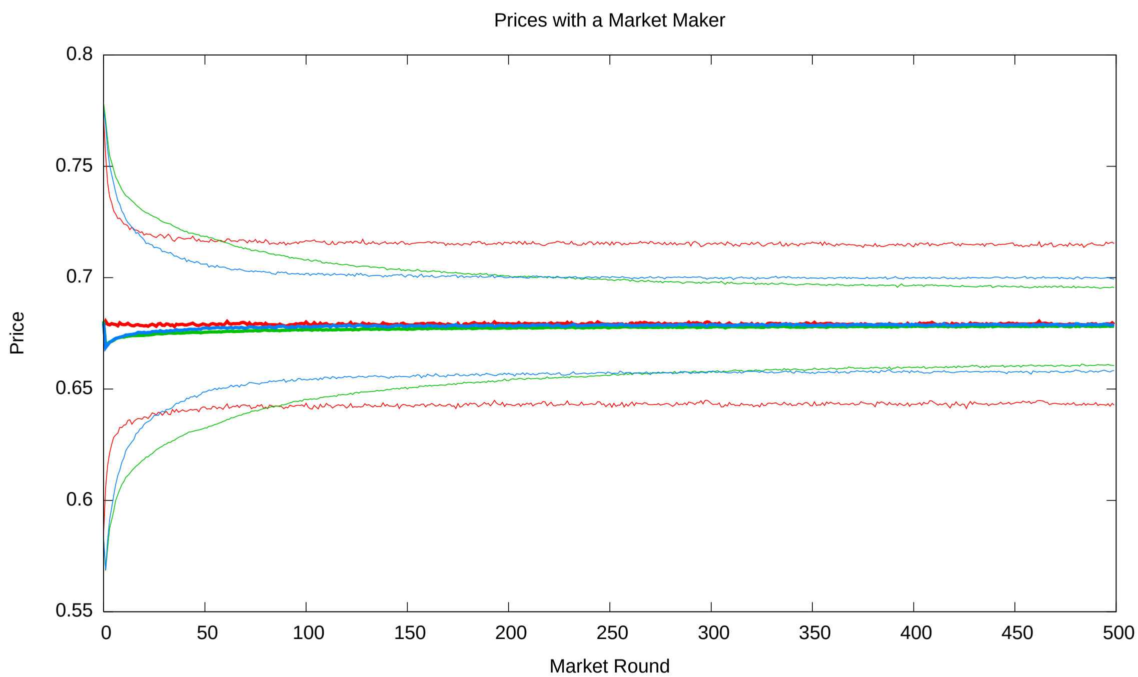

Figure 7.2: Prices induced by Market Makers: Red: No market maker; Green: bid/ask spread of 1 cent; Blue: bid/ask spread of 2 cents.

Figure 7.2 shows the average prices from 10,000 simulations with market makers. The presence of a market maker induces a fair amount of initial volatility, in the long term market makers reduce volatility significantly, with both effects increasing in the tightness of the bid/ask spread, s.

The systematic under-pricing of shares with a market maker is a function of our design rule - market makers take the mean price of the best bid and ask orders, and because the mean of the distribution of beliefs is greater than one-half, there is more space below the mean than above. Consequently, it is more likely that the lowest ask order is closer to the long-term equilibrium price than the highest bid order, and so the prices set by the market maker will therefore be spread around a lower mean than is appropriate. As is evident from the plot, however, the difference in mean prices with and without market makers falls with time.

Market Proceeds

Despite the fact that market makers trade without paying any commissions, it is possible that their presence actually increases revenue for the exchange by tightening the bid/ask spread, inducing more trades.

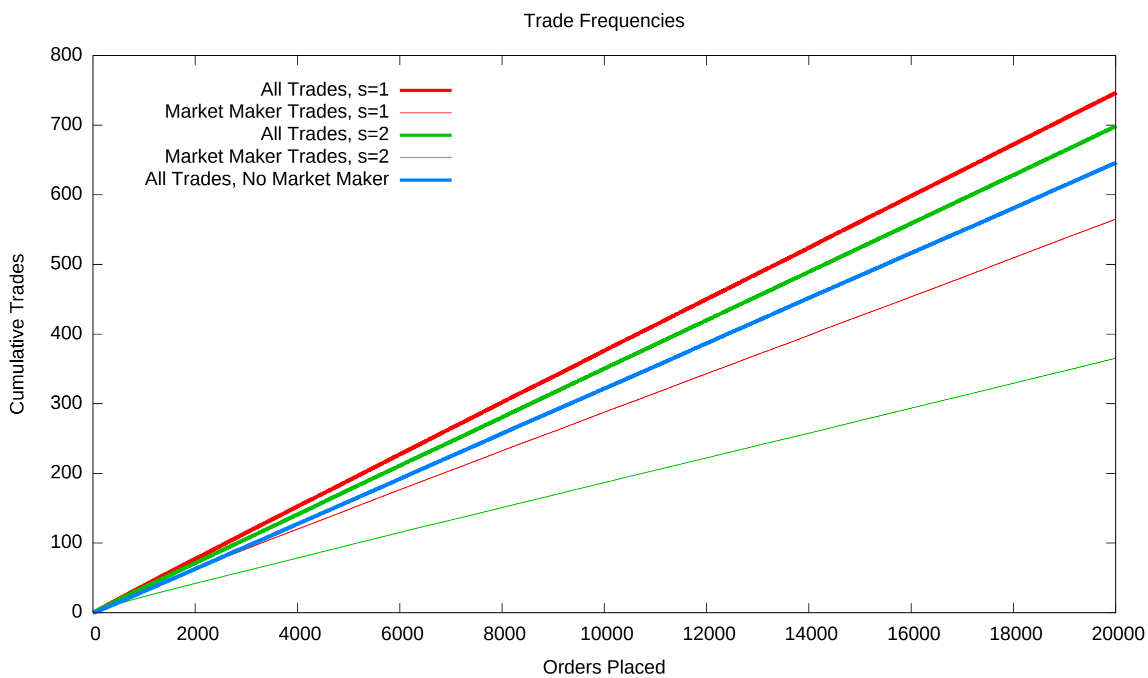

Figure 7.3: Trade Frequencies with Market Makers.

Table 7.3: Average exchange profits after 20,000 rounds.

Market	Dollars
$s = 1$	18.55
$s = 2$	20.62
No Market Maker	25.84

Figure 7.2 and table 7.3 indicate that, while market makers do increase the amount of trade significantly, the increase in trade is not enough to offset the revenue lost by having half of a transaction not pay a commission.

Figure 7.3 was generated by taking the average number of cumulative trades over time (which is equivalent to the number of orders placed) over 10,000 unique market simulations. Note how trading frequency is roughly constant in time, which is why the red, green, and blue lines are linear over their domain. The chief difference is the slope of these lines, the steeper red and green lines indicate that the trading frequency when the market maker is present is higher than without, and that the trading frequency increases with the market maker setting a smaller bid/ask spread. In fact, there are roughly 15% more trades with a market maker setting a one cent spread than without a market maker.

But when market makers are present, the fraction of the trades they perform is much higher in the marketplace. When $s = 1$, market makers perform more than three-quarters of all trades, while when $s = 2$ market makers are involved in roughly half of all trades. Table 7.3 shows the profits accrued by the exchange through trading commissions. Profits are highest in the absence of a market maker and increase with the bid/ask spread set by the market maker because the market makers is involved in fewer trades.

7.3 A Qualitative Assessment

Our results show that in a ZI framework, market makers decrease the volatility of the exchange and increase the trading frequency, with both effects decreasing in intensity with the size of the market maker's spread. Under a simple model of commissions, we have also demonstrated that exchange profits actually decrease in the presence of a market maker, despite the increased number of trades. The most pertinent question, however, is how applicable our analysis is to real-world behavior. What conclusions can we draw about market makers in the real world from our analysis?

Some of our results, indeed, are the product of the specific design decisions we have made and are not applicable for the real world. Because our commission regime differs so much from the real-world commission model used in Tradesports, our claims to the detrimental profitability of market makers within the exchange need not apply in the real world. Additionally, the initial volatility of the exchange in early trading rounds is a direct function of how our market makers set prices, so we offer no suggestion that the presence of market makers makes real-world Tradesports markets more volatile.

That said, we argue that the role played by market makers within a prediction market is fundamentally different from the role played by market makers within a traditional exchange. In a traditional exchange, market makers function to provide liquidity in markets to induce investment. Our suggestion is that a prediction market will generally not need such inducement. We attribute this to overconfidence and wish fulfillment - in general, people are willing to privilege their own opinions over the opinions of others. As a thought experiment, imagine a new trader logging in to the Iowa Electronic Markets for the first time. She sees that the last trade on his favorite candidate was at only 30 cents. Is she more likely to revise her own opinions on her favorite candidate and reconsider her strongly partisan expectation of the future, or decide that the market is significantly underpricing her candidate?

Moreover, we suggest that liquidity concerns are much less prominent in prediction markets than in traditional exchanges. If a trader takes a position in a traditional share, or even more so for a future, they will want to convert that investment into cash in the future, either to secure gains or losses or in the case of futures, to avoid having to take ownership of 20 tons of frozen pork bellies. Without a market maker, investors may be skittish about future liquidity and may thus avoid taking a position entirely. In prediction markets, however, we can be assured of a finite time horizon at which shares will expire, and thus we avoid the liquidity concerns, particularly for the majority of agents interested in buying and holding a position in a candidate (c.f. [5]).

The ZI agent framework is not nuanced enough to accommodate all of Tradesports' complex pricing rules, and as a consequence our profitability result should not be viewed rigorously. We suggest, however, that market makers in prediction markets may be a vestigial holdover from traditional exchanges, and that they may not ultimately play a large role in price formation or efficiency.

Chapter 8

Double-Share Prediction Markets

Recall that double-share prediction markets like the IEM feature contracts for complementary events, with separate contracts for Kerry and Bush to win an election, as opposed to a single contract trading on Bush. In this chapter we implement ZI agents for double-share prediction markets and run simulations to examine market prices. We find evidence that the 1992 IEM presidential market may have been under-supplied in quantities and use our model to analyze the effects of quantities on market prices. We conclude with a discussion of the implication of our findings on the Marginal Trader hypothesis.

8.1 Designing Zero-Intelligence Agents

Our model of ZI agents in double-share prediction markets follows the model of our single-share prediction market ZI agents with a few important changes. First, and most obviously, we need to expand our agents to interact with two shares rather than one. We incorporate this change by simply having agents flip a coin to decide in which share to participate. We retain our requirement that agents feel that complementary contracts have complementary values, so that a belief of b that candidate x will win implies a belief of $1-b$ that candidate y will win.

But the differences between single-share and double-share prediction markets go beyond the trivial. Shares enter a single-share market through a transaction - without any transactions, there are no shares in the marketplace. In contrast, shares enter a double-share prediction market through purchase of risk-free market bundles. The amount of cash an agent has on hand thus becomes of great import, because it relates directly to these quantities.

Our ZI agents have initial holdings of s market bundles and c cash. They are required to be able to cover all their outstanding orders in the market, so each time they place an ask order their holdings of the appropriate share is decremented by one and each time they place a bid order for price p, c is decremented by p. If our ZI agents randomly select to place an ask order in a share they do not hold, or if they attempt to place a bid order for price $p > c$, their attempt at ordering fizzles, no order is placed, and they exit the marketplace without trying to place another order.

We should consider the very important question with regards to whether our agents' cash and share holdings compromise the guiding principles behind ZI agent design. Are not holdings of cash and shares a form of state? Indeed they are. However, our ZI agents do not use their holdings in any way when making an ordering decision, which implies efforts to do infeasible things like place an ask order without holding any shares. Because our ZI agents do not use their private holdings to develop any information about the past, present, or future, we can regard cash and share reserves in the same way as we view market prices - information a more savvy agent could use to develop better trading strategies.

8.2 Thinking about Endowments

It becomes an open question as to how we should endow our ZI agents. On the one hand, we are tempted to give them large quantities of both cash and shares, in order to make the model as close as possible to the unboundedness of our single-share prediction markets. On the other hand, it is important to note that these markets exist in the real world, where issues like quantities could, and as we have and will argue, do play a crucial role in determining prices. The question then arises: In a normal IEM presidential election, what do individual investment levels look like?

8.2.1 Historical Evidence from the IEM

The IEM has a standing policy to not release information at the user-level. Thus, in order to examine the markets for user-level information, such as investment amounts and holdings, we examine the papers that have been written by members of the IEM board. Two studies in particular are helpful, which analyzed in-depth the 1988 presidential market [9] and the 1992 presidential market [21]. We will look at each of these in turn.

The 1988 Election

The 1988 election was open only to members of the University of Iowa community. Forsythe *et al.* [9] record a total of 192 traders holding 1,462 shares and investing a total of $4,967. The average trader thus invested about 26 dollars in the market. Since market shares cost $2.50 rather than the traditional $1, the figures imply that the average trader invested 19 of these dollars to hold about 7.6 market shares. The issue of market share pricing complicates how we think about wealth and endowments, but we can assume that agents desired to hold k dollars *worth* of shares in a set of candidates rather than hold k *shares* of those candidates. Thus, the average agent committed about three-quarters of their wealth in the form of investments in the market.

The 1992 Election

The data from the 1992 election is much less helpful. Oliven and Rietz [21] list the average daily total volume as $142.08. Because the market ran from July 17th to November 3rd, 80 days, this implies that the total volume was $11,360. The total amount of funds allocated by the 500 market participants is given as more than $83,000. As an upper bound on quantities, we calculate the figure for market shares as if each dollar of daily volume resulted from the sale of a new market bundle. This implies a total 11,360 market shares within the system. We can thus assert that as an upper bound, the average agent invested about 166 dollars, holding about 23 market shares and about 143 dollars in cash. This represents agents holding about 13.7% of their wealth in the form of investments in the market.

8.2.2 Complications

Unfortunately, because of the IEM's restrictions on releasing detailed trading data, we have no way of knowing if our estimates of how much of the wealth of the "average agent" was invested in the marketplace. In particular, it is entirely possible that many market participants simply deposited money in the system and never logged in to trade - their money would then be reflected in the count of total funds invested but they would not be true market participants. On the other hand, it is also very likely that all of the daily volume in the 1992 election was not directed towards trading new shares that arose by unique market bundle purchases, but rather shares that would exchange hands several times. This would have the effect of reducing the percent of total wealth invested in shares in the 1992 election.

8.3 Results

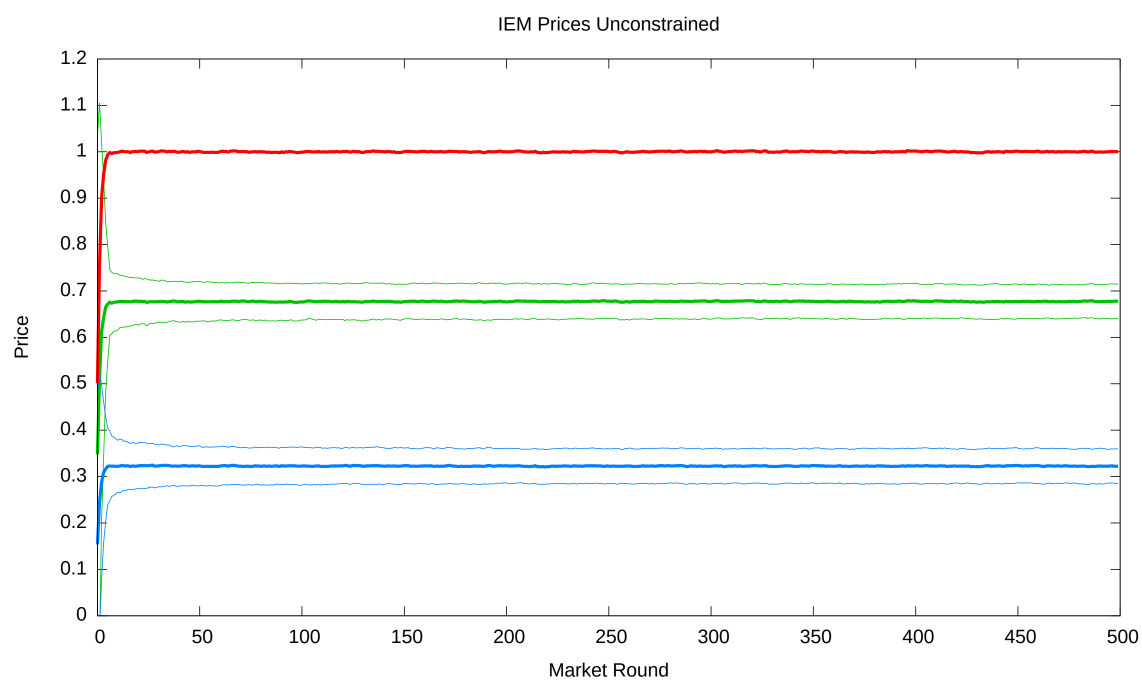

Figure 8.1: Prices for a double-share market with unconstrained allocations.

For figures 8.1, 8.2, and 8.3, the thick green lines represent the average price for shares that have beliefs distributed according to the "low accuracy" distribution used previously, while the thick blue lines represent average prices for the complementary shares, and thick red lines represent the sum of the prices of these complementary shares. Thin lines represent 95th percentile confidence intervals. To clarify, a "market round" is considered to be triggered whenever a share trades in *either* commodity. In a sense then, one can regard each round as being half as long in these simulations as in a single-share simulation.

Figure 8.1 shows a market simulation run with 500 agents and unconstrained initial allocations. These can be considered "unconstrained" because agents' initial endowments are set high enough to ensure that no agent will ever be unable to place a bid or ask order, and thus the simulation should run, like the single-share simulation, without any regard for endowments. Much of the initial variance in trading is due to our definition of a "market round" - in the first market round only one share will have traded and the price of the other is set to zero by default.

The plot perfectly shows the theoretical symmetry between single- and double-share prediction markets; note how prices sum to 1 and that shares prices and variances are identical to those featured in the single-share markets.

Figure 8.2 displays share prices for agents acting with a buying bias such that they aim to purchase twice as frequently as they would sell. Prices remain relatively stable at summing to about 1.04, and the prices and confidence intervals for the "green" shares are identical to those for the single-share market simulation, while the "blue" shares reflect the longshot bias of shares with distributional means under one-half.

Figure 8.3 displays share prices induced by initial quantities set up to replicate the conditions of our estimate of the 1992 IEM. In this case, agents each have an initial endowment of 13 or 14 market bundles, and 87 or 86 dollars cash, respectively. The lower plot shows the average number of market bundles (half of the sum of the shares in each candidate) held by each agent. The red crosses are provided as a reference line at 1. Around round 300, when the average agent holds about 4 market shares, prices begin to rise steadily; by the 500th round the sum of share prices rises to more than 1.06.

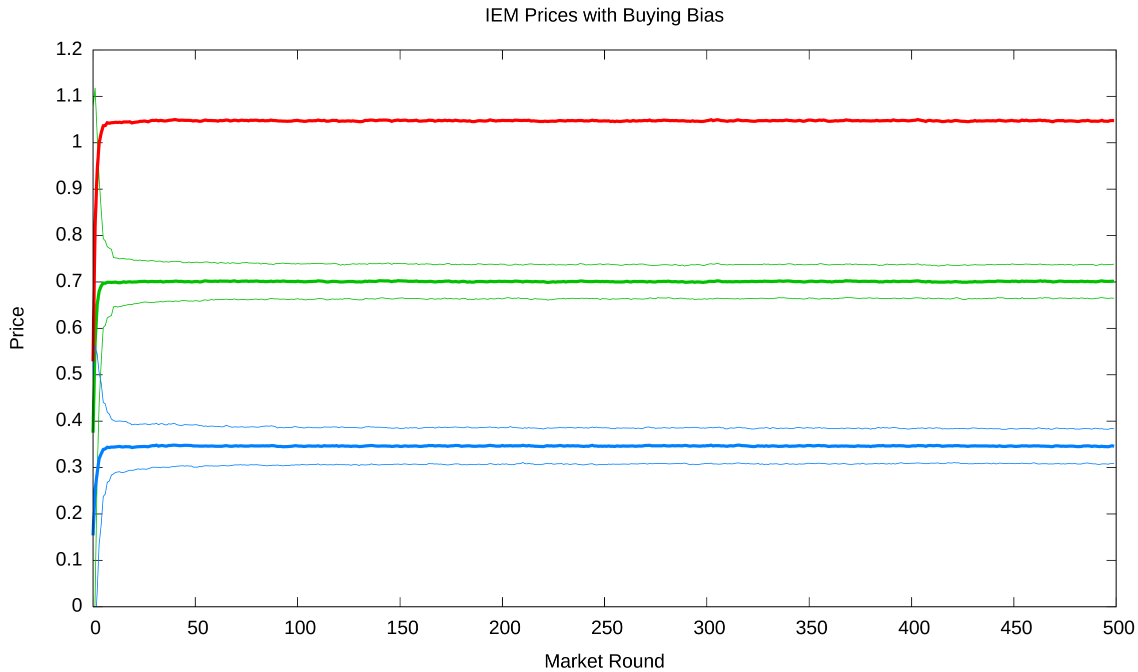

Figure 8.2: Prices for a double-share market with buying bias and unconstrained allocations.

8.4 Return of the Marginal Trader

The most intriguing of our above results is figure 8.3. Under the ZI framework, we quickly see that the average holding of agents drops dramatically. Once agents can no longer place ask orders because they have exhausted their quantities, prices rise as higher-priced ask orders are met.

What is most provoking about the simulation is that the initial endowments that induced the price-rising levels are roughly comparable to our estimation of the ratio of shares to cash holdings of the agents participating in the 1992 election. In fact, our estimation of this ratio was in some sense an upper bound on the quantity holdings of agents in the 1992 market because we assumed that every traded share came from a new market order. Thus, it is entirely conceivable that the actual quantity levels in the 1992 election were lower, and perhaps substantially lower, than our estimates.

But the 1992 market did not display the inexorable price raising that we are linking

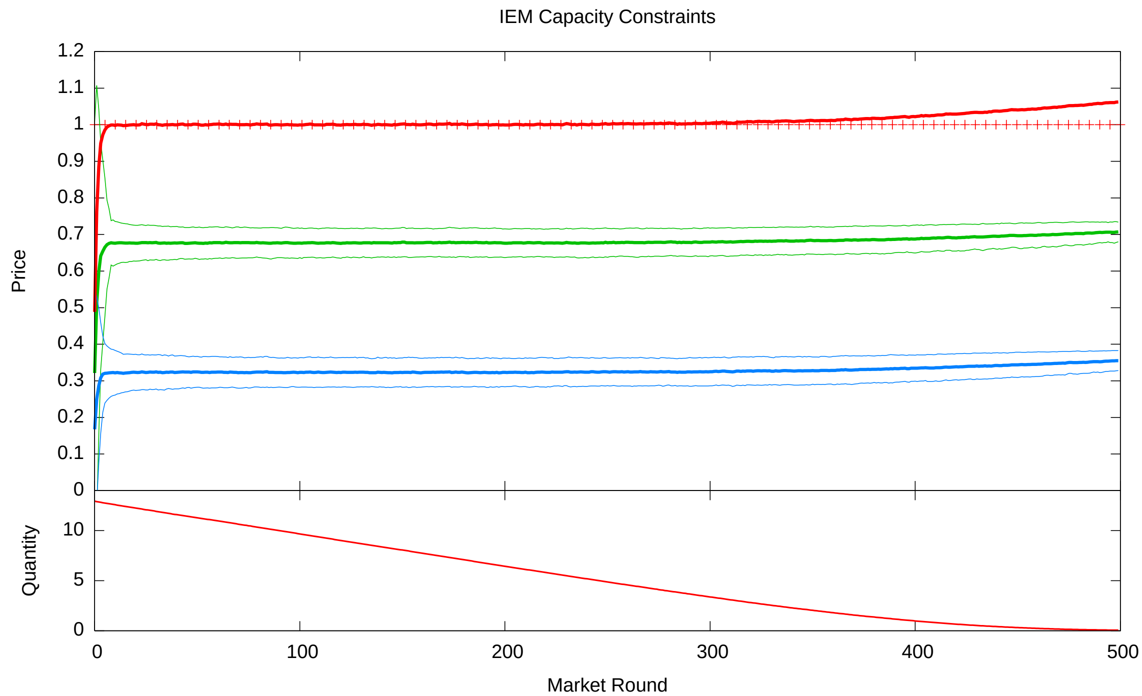

Figure 8.3: Prices for the IEM with constrained allocations.

to quantity constraints. Figure 8.4 displays the average prices for each day from the 1992 winner-take-all market for the last two months before the election. Blue represents shares of Clinton (the eventual winner), brown Bush, and green Perot. The red line indicates the sum of the three prices, while the red crosses provide a visual cue at 1.

Clearly, the prices for the three assets stay sum to nearly 1 for the virtually the entirety of the final two months of trading. This is despite the rather massive price changes in the assets. We thus arrive at our principal hypothesis of the chapter: that if capacity constraints did indeed exist in the 1992 markets, it provides a strong argument for the existence of the apocryphal "marginal traders", the small pool of savvy traders to whom it is attributed responsibility for setting prices and the markets' accuracy [9, 4, 21].

One could suggest the counter-argument, however, that the marginal traders are merely the amalgamated effects of various arbitrageurs, no better informed than the rest of the market traders. Under this argument, prices are ensured to sum to 1 despite capacity constraints because of the actions of arbitrageurs that visit the market and exploit the

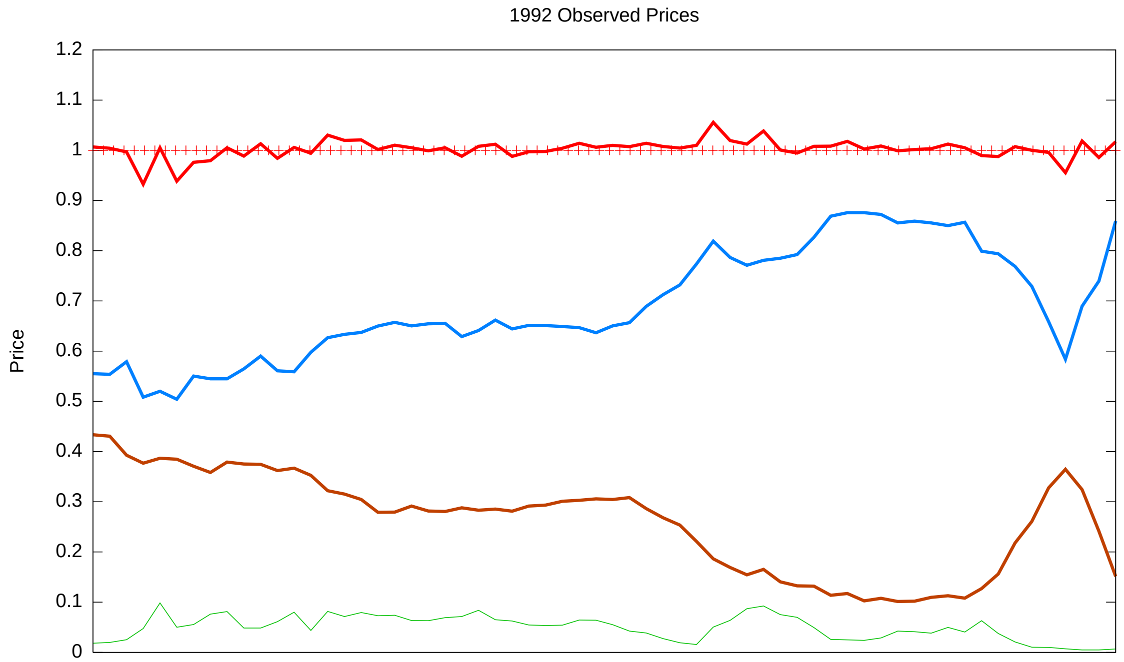

Figure 8.4: Average prices from the last two months of the 1992 election.

arbitrage opportunities that cause prices to sum to more than 1. Unlike the stronger Marginal Trader Hypothesis, such a theory would not depend on having agents be better informed or more rational, simply capable of acting upon the arbitrage opportunities which arise when the highest complementary bid orders sum to more than 1, or when the lowest complementary ask orders sum to less than 1.

Such a hypothesis, however, does not turn out to provide a compelling explanation of the observed behavior. That is to say, the presence of an active arbitrageur that exploits every arbitrage opportunity within the market does *not* lower trading prices to their efficient levels. At first glance, such a result seems paradoxical - how could an active arbitrageur not drive prices to 1? The answer comes by noting that, when capacities are constrained within the ZI framework, more than 90% of trades come from buyers accepting ask prices. Though an arbitrageur can ensure that the highest bid prices sum to less than 1, ask prices summing to more than 1 do not indicate an arbitrage opportunity. Because so many of these prices are accepted, the trading prices of shares will continue to be at prices that sum to more than 1, even though arbitrage opportunities no longer exist within the market.

We have not proven that capacity constraints existed in the 1992 market, indeed doing so would require careful examination of proprietary user-level data. However, we have provided strong evidence from published research quantities were undersupplied in the 1992 winner-take-all market. Yet we have also demonstrated that the market was quite efficient. If quantities were indeed undersupplied, the failure of our ZI traders, even with the presence of an active arbitrageur, to keep prices trading prices at 1 indicates the presence of a human level of rationality beyond rote, rule-based behavior. Who kept the market efficient? It is a glass slipper inside of which the marginal trader's foot slips effortlessly.

Chapter 9

Conclusion

In this work, we developed two distinct models for prediction markets. The first was a mathematical model for prediction markets based on previous work by Manski [19], Gjerstad [11] and Wolfers and Zitzewitz [31]. Our model differed from theirs in three key areas: we allowed much more flexibility in possible equilibrium prices, we based our distributional assumptions on empirical data, and our model also accommodated both vote-share as well as winner-take-all prediction markets. These changes made the task of analyzing equilibria significantly harder from a computational perspective, and we used algorithmic optimizations to solve our computing problems quickly, while we used mathematical proofs to ensure rigor.

We used our results to analyze two phenomena that had arisen and were as-yet unexplained in real-world markets: first, the longshot bias, by which likely events are underpriced and unlikely events are overpriced, and second, the overpricing observed in laboratory experiments. We provided additional evidence that the longshot bias occurs in practice by documenting it in contracts from the Tradesports 2005-06 NBA season. We concluded that weak risk aversion was responsible for causing the longshot bias seen in real markets, and also proposed a new theory based on partisanship, biased individual traders, that could be an entangled observation which drives traders towards risk neutrality. Finally, we demonstrated that the overpricing phenomenon observed in laboratory experiments could be a result of under-supplying quantities relative to assets, and we provided advice to economic scientists that we feel will eliminate this overpricing in future studies.

The second model we considered was an agent-based model, to our knowledge the first agent-based model for prediction markets. By adapting the Zero-Intelligence agents of Gode and Sunder [12] to a prediction markets context, we were effectively able to resolve the effects seen in prediction markets from human rationality, giving us a base case to contrast against empirical results. We found that the same longshot bias, albeit to a smaller extent, is present in our ZI model, implying that not only is the bias intrinsic to prediction markets, but also that human behavior amplifies this bias. In addition, agent-based modeling allowed us to examine a new phenomenon, the buying bias, where traders are more likely to place buy order rather than to sell. We provide the first empirical evidence that the buying bias exists within prediction markets, and analyze its effect on equilibrium prices, finding that for likely events the buying bias can serve to counteract the longshot bias and actually increase efficiency.

In addition, we developed a stylized model of market commissions and market makers. We found that commissions significantly decreased trading volumes and market makers lowered market volatility. Despite this finding, we argued that market makers may be less important in prediction markets functioning well than in traditional markets, particularly futures markets.

We also analyzed double-share prediction markets like those found in the IEM. Our work demonstrated the theoretical similarity between single-share and double-share prediction markets in all cases except in quantity-constrained double-share markets, where we demonstrated that prices would rise above efficient levels. We provided evidence that the 1992 IEM presidential markets were quantity constrained, but we also show that those markets functioned efficiently. We show that if the 1992 markets were indeed quantity constrained that simple arbitrageur behavior would not be enough to bring the prices to efficient levels, implying that the presence of a greater level of human rationality and decision-making would be required to explain the efficiency of the 1992 markets. We conclude that the presence of quantity constrains in the 1992 market provides a strong argument that bolsters the marginal trader hypothesis.

Future Work

Our work has opened far more doors than it has closed.

One issue we have not dealt with in our markets is the concept of time, particularly in terms of information shocks - in the political context these could be political events or polls, while in a sports context it could be a team going on a scoring run or hitting a shot to send a game into overtime. Forsythe *et al.* [9] demonstrated that traders they identified as marginal in the 1988 election were capable of correctly discriminating, in general, between important and unimportant news events over the course of the election. We imagine future studies which seek to formalize the concept of how news is absorbed into the markets, perhaps in an agent-based model in which we see the emergence of overreaction or underreaction to news.

We have also demonstrated two new ideas relating to prediction markets that deserve further attention from behavioral economists. The first is the partisanship phenomenon. How do people react to the act of a successful or unsuccessful prediction, and how does this influence their investment decision? Is there a social norms aspect to the act of prediction? How do people react to gambles that involve dramatic changes in the way they view the world, and are these reactions different from gambles that exist in isolation?

The second behavioral phenomenon is the buying bias. We provided the first empirical evidence that traders prefer to buy rather than to sell in prediction markets, but the issue merits further study. Do double-share as well as single-share markets also show the buying bias? Laboratory study may also be a useful tool here.

In the same spirit, we would like to see more study and publication on the longshot bias in prediction markets. By documenting the bias in as many markets as possible, and making that work as visible as possible, savvy traders will be able to act on these potential opportunities. Examining the longshot bias in a variety of contexts also serves as a test of the partisanship phenomenon - partisanship would predict that the longshot bias is most prevalent for events in which people are passionate about the outcome (like presidential politics) and least prevalent for events in which outcomes do not alter their state of the world (like whether the Dow Jones average will be move up less than 25 points in day).

We stand at a curious vista in the study of prediction markets - the markets have matured enough to have enough volume and data to begin to draw conclusions based on theory, but they still lack the kind of detailed study that would drive prices towards efficiency. Ironically, the greatest contribution of this work may be to invalidate its conclusions, as knowledge of the longshot bias drives prediction market prices to efficiency.

Finally, in the context of our agent-based model, we view our work as the first step towards developing more nuanced models of prediction markets. We imagine the development of more human-like strategies, and using these strategies to develop optimal trading strategies and market-maker strategies. When combined with the Tradesports API, it will be possible to create completely automated profitable trading agents.

Bibliography

[1] Ali, M. Probability and Utility Estimates for Racetrack Bettors. *The Journal of Political Economy 85*, 4 (1977), 803–815.

[2] Arthur, W., Holland, J., LeBaron, B., Palmer, R., and Tayler, P. Asset Pricing Under Endogenous Expectations in an Artificial Stock Market in The Economy as an Evolving Complex System II. Arthur WB, Durlauf S. and Lane D. Editors. *SFI Studies in the Sciences of Complexity 27* (1997).

[3] BBC News. Pentagon axes online terror bets, July 2003.

[4] Berg, J., Forsythe, R., Nelson, F., and Rietz, T. Results from a Dozen Years of Election Futures Markets Research. *Handbook of Experimental Economics Results* (2001).

[5] Berg, J., and Rietz, T. The Iowa Electronic Markets: Stylized Facts and Open Issues. *Information Markets: A New Way of Making Decisions, edited by Robert W. Hahn and Paul C. Tetlock. AEI-Brookings Joint Center for Regulatory Studies* (2006).

[6] Cain, M., Law, D., and Peel, D. The Favourite-Longshot Bias and Market Efficiency in UK Football betting. *Scottish Journal of Political Economy 47*, 1 (2000), 25–36.

[7] De Bondt, W., and Thaler, R. Financial decision-making in markets and firms: A behavioral perspective, in Finance. *Handbooks in Operations Research and Management Science(RA Jarrow, V. Maksimovic and WT Ziemba, Eds.) 9* (1995).

[8] Feigenbaum, J., Fortnow, L., Pennock, D., and Sami, R. Computation in a distributed information market. *Proceedings of the 4th ACM conference on Electronic commerce* (2003), 156–165.

[9] Forsythe, R., Nelson, F., Neumann, G., and Wright, J. Anatomy of an Experimental Political Stock Market. *The American Economic Review 82*, 5 (1992), 1142–1161.

[10] Forsythe, R., Rietz, T., and Ross, T. Wishes, expectations and actions: a survey on price formation in election stock markets. *Journal of Economic Behavior and Organization 39*, 1 (1999), 83–110.

[11] Gjerstad, S. Risk Aversion, Beliefs, and Prediction Market Equilibrium. *Economic Science Laboratory, University of Arizona* (2005).

[12] Gode, D., and Sunder, S. Allocative Efficiency of Markets with Zero-Intelligence Traders: Market as a Partial Substitute for Individual Rationality. *Journal of Political Economy 101*, 1 (1993), 119.

[13] Grothmann, R. Multi-agent market modeling based on neural networks, 2002.

[14] Hanson, R., and Oprea, R. Manipulators Increase Information Market Accuracy. Tech. rep., George Mason University, 2004.

[15] Hurley, W., and McDonough, L. A Note on the Hayek Hypothesis and the Favorite-Longshot Bias in Parimutuel Betting. *The American Economic Review 85*, 4 (1995), 949–955.

[16] Judd, K. *Numerical Methods in Economics*. MIT Press, 1998.

[17] LeBaron, B. Building the Santa Fe Artificial Stock Market. *Physica A* (2002).

[18] Leigh, A., and Wolfers, J. Competing Approaches to Forecasting Elections: Economic Models, Opinion Polling and Prediction Markets. *The Economic Record 82*, 258 (2006), 325–340.

[19] Manski, C. Interpreting the Predictions of Prediction Markets, 2004.

[20] Milgrom, P., and Stokey, N. *Information, Trade, and Common Knowledge*. Northwestern University, Graduate School of Management, 1979.

[21] Oliven, K., and Rietz, T. Suckers Are Born but Markets Are Made: Individual Rationality, Arbitrage, and Market Efficiency on an Electronic Futures Market. *Management Science 50*, 3 (2004), 336–351.

[22] Palacios-Huerta, I., Serrano, R., and Volij, O. Rejecting Small Gambles Under Expected Utility. *Economics Letters 91*, 2 (2006), 250–259.

[23] Press, W., Teukolsky, S., Vetterling, W., and Flannery, B. *Numerical recipes in C. The art of scientific computing*, 2nd ed. Cambridge University Press, 1992.

[24] Rietz, T. Behavioral Mis-pricing and Arbitrage in Experimental Asset Markets. Tech. rep., University of Iowa, Iowa City, 2005.

[25] Smith, V. Markets as Economizers of Information: Experimental Examination of the "Hayek Hypothesis". *Economic Inquiry 20*, 2 (1982), 165–79.

[26] Stroud, A., and Secrest, D. *Gaussian quadrature formulas*. Prentice-Hall Englewood Cliffs, NJ, 1966.

[27] Strumpf, K. Manipulating the Iowa Political Stock Market. Tech. rep., University of North Carolina, 2004.

[28] Surowiecki, J. *The wisdom of crowds: why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations*. Doubleday, 2004.

[29] Waldvogel, J. Fast Construction of the Fejér and Clenshaw–Curtis Quadrature Rules. *BIT Numerical Mathematics 46*, 1 (2006), 195–202.

[30] Wolfers, J., and Zitzewitz, E. *Five Open Questions about Prediction Markets*. National Bureau of Economic Research Cambridge, Mass., USA, 2006.

[31] Wolfers, J., and Zitzewitz, E. Interpreting Prediction Market Prices as Probabilities, 2006.

[32] Woodland, L., and Woodland, B. Market Efficiency and the Favorite-Longshot Bias: The Baseball Betting Market. *The Journal of Finance 49*, 1 (1994), 269–279.

Acknowledgments

Getting this chimaera off the ground has taken the committed efforts of many, many people and I am deeply thankful for their support. David Parkes has listened to my weekly blathering for years. Ulrich Doraszelski taught me how to truly *use* a computer. Eric Budish, Al Roth, Peter Coles, and Paul Bamberg were there to answer my silly questions. John Delaney of Tradesports provided me with the NBA contract data, while Greg Jensen of Bridgewater Associates suggested how to use it.

My roommates, Chris, James, Savi, and Steve, are the smartest, most vibrant guys I know. I'm not sure how they've put up with me for so long.

My parents taught me tenacity and patience. To them I can only say: I'm sorry it's not Biology.