Automated Market Making: Theory and Practice

Abraham Othman

May 15, 2012
CMU-CS-12-123

School of Computer Science
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Tuomas Sandholm, Chair
Geoffrey Gordon
Albert Kyle
David Pennock
Javier Peña

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Copyright 2012 Abraham Othman

This work has been supported by NSF Grants IIS-0905390, IIS-0964579, CCF-1101668, and IIS-0427858, and by the Google PhD Fellowship in Market Algorithms.

Keywords: Market microstructure, agent design, wagering, risk measures

Abstract

Market makers are unique entities in a market ecosystem. Unlike other participants that have exposure (either speculative or endogenous) to potential future states of the world, market making agents either endeavor to secure a risk-free profit or to facilitate trade that would otherwise not occur. In this thesis we present a principled theoretical framework for market making along with applications of that framework to different contexts. We begin by presenting a synthesis of two concepts—automated market making from the artificial intelligence literature and risk measures from the finance literature—that were developed independently. This synthesis implies that the market making agents we develop in this thesis also correspond to better ways of measuring the riskiness of a portfolio—an important application in quantitative finance. We then present the results of the Gates Hillman Prediction Market (GHPM), a fielded large-scale test of automated market making that successfully predicted the opening date of the new computer science buildings at CMU. Ranging over 365 possible opening days, the market's large event partition required new advances like a novel span-based elicitation interface. The GHPM uncovered some practical flaws of automated market makers; we investigate how to rectify these failures by describing several classes of market makers that are better at facilitating trade in Internet prediction markets. We then shift our focus to notions of profit, and how a market maker can trade to maximize its own account. We explore applying our work to one of the largest and most heavily-traded markets in the world by recasting market making as an algorithmic options trading strategy. Finally, we investigate optimal market makers for fielding wagers when good priors are known, as in sports betting or insurance.

For my parents. Thank you.

Acknowledgments

Tuomas Sandholm, my advisor, has been the strongest influence on my research career. He deserves an incredible amount of credit for allowing a first-year PhD student to pursue a risky course of research, and then for fully supporting that line of research as it began to mature into this thesis. Tuomas is also a talented scientific writer with an incredibly clear grammatical compass. He taught me everything I know about the correct way to express technical ideas to others, and any clarity I've managed in this document is directly a product of his tutelage.

David Pennock has also been a strong influence on my research trajectory. The three weeks in the Summer of 2009 that I spent with his group at Yahoo! Research were extraordinarily fruitful, directly influencing several chapters of this thesis.

Special thanks also goes to the rest of my committee: Geoff Gordon, Javier Peña, and Pete Kyle. This thesis is better as a result of each of your insights and comments.

This thesis is also much better from conversations with my peers in the PhD program. In particular, I would like to thank Andrew Arnold, Mike Benisch, Alex Beutel, Jim Cipar, John Dickerson, Mike Dinitz, Jason Franklin, Sam Ganzfried, Tony Gitter, Andrew Gilpin, Alex Grubb, Severin Hacker, Sid Jain, Elie Krevat, Sarah Loos, Mary McGlohan, Brendan Meeder, Wolfgang Richter, Aaron Roth, Rob Simmons, Jiří Šimša, Mike Tschantz, Kevin Waugh, Gabe Weisz, Erik Zawadzki, and Brian Ziebart.

I have had a wonderful time in the Computer Science PhD program at Carnegie Mellon. The program is unique in two respects: First, in its willingness to take risks by accepting unconventional applicants (including me), and second, in its focus on having those students pursue serious research as quickly as possible. I am indebted in particular to erstwhile department head Peter Lee for being a supportive influence early in my graduate career. The Gates Hillman Prediction Market could not have happened without his assistance and approval.

The support of Intel Corp. and IBM for their machine gifts to the AMEM research group is much appreciated. I would also like to thank Ivan Corwin, who is very good at math and contributed a great deal to my understanding of the higher mathematics that appear occasionally in this document.

Finally, I am so grateful for the support of Dr. Carolyn Roberts, both in this thesis specifically and in my education generally. Auntie Carolyn did the yeoman's work of reading a first draft of this thesis with a detailed eye, finding many awkward phrases and typographical errors.

It is hard for us, without being flippant, to see a scenario within any kind of realm or reason that would see us losing one dollar in any of those transactions.

Joseph Cassano
AIG Financial Products Earnings Call
August 2007

But it is more interesting to consider the possibility that the men could be robots.

Jay McInerney
Bright Lights, Big City

Contents

1 Introduction **1**

2 Related work **7**

3 Theoretical background **15**

3.1 Cost functions and risk measures 15

3.2 Cost function market makers 17

3.2.1 Desiderata 17

3.2.2 Dual space equivalences 19

3.2.3 Coherent risk measures 20

3.2.4 Relaxing desiderata 22

3.2.5 Convex risk measures 23

4 The Gates Hillman Prediction Market **29**

4.1 Market design 30

4.1.1 Incentives and setup 30

4.1.2 The LMSR in the GHPM 31

4.1.3 Span-based elicitation with ternary queries 32

4.2 Problems revealed 35

4.2.1 Spikiness of prices across similar events 35

4.2.2 Liquidity insensitivity 39
4.3 Effective trader strategies 40
4.3.1 Cluster analysis 40
4.3.2 Spike dampening 42
4.3.3 Relative smoothing 42
4.3.4 Information gathering 43
4.4 Analysis of empirical market performance 44
4.4.1 Official communications and the GHPM 44
4.4.2 Self-declared savviness 49
4.4.3 Trade frequencies suggest a power law 51
4.4.4 Trading by a bot 51
4.4.5 Trader-level data is consistent with Marginal Trader Hypothesis 52
4.5 Extensions 59

5 Extensions to convex risk measures **61**
5.1 Market making over infinite spaces 61
5.1.1 Prior work 62
5.1.2 Countable event spaces 63
5.1.3 Uncountable event spaces 65
5.2 Adding profit and liquidity to convex risk measures 70
5.2.1 Desirable properties for a market maker 70
5.2.2 Four oppositional desiderata in the literature 71
5.2.3 How bets are taken 74
5.2.4 Theoretical properties 75
5.2.5 Profit, liquidity, and market depth 77
5.3 Extensions 85

6 **Homogeneous risk measures** **89**
6.1 The OPRS market maker 89
6.1.1 Defining the OPRS 90
6.1.2 Moving forward in obligation space 90
6.1.3 Properties of the OPRS 92
6.2 Dual space results 107
6.2.1 Curvature 108
6.2.2 Divergence from probability simplex 109
6.2.3 The OPRS and its conjugate set 111
6.3 Using the dual space 113
6.3.1 Unit ball market makers 113
6.3.2 Optimal homogeneous risk measures 115
6.4 Extensions 121

7 **Option trading** **125**
7.1 Options 127
7.1.1 What are options? 127
7.1.2 Historical dataset 128
7.2 Agents based on Log-Normal and Normal distributions 128
7.2.1 The BSM model 130
7.2.2 Calculating contract values when the underlying is log-normally distributed 130
7.2.3 Calculating contract values when the underlying is normally distributed . 132
7.3 The LMSR as an option trading agent 133
7.3.1 Compressing the state space 133
7.3.2 Implementation details 134
7.4 Trading agents based on constant exponential utility 135

7.4.1 The undefined prices phenomenon 136
7.4.2 Our Exponential utility trader 141
7.5 Random agent as a control 141
7.6 Experimental setup 141
7.7 Snapshot of a single option chain 143
7.8 Results 148
7.8.1 Quantitative results 149
7.8.2 Qualitative results 153
7.9 Modifying the experiments 160
7.9.1 On the random uniform trading restriction 160
7.9.2 The LMSR with a better discrete prior 162
7.10 Extensions 164

8 Rational market making with probabilistic knowledge 167

8.1 Model 168
8.1.1 Traders 168
8.1.2 Utility and the Bellman equation 169
8.2 Computation of the policy of a Kelly criterion market maker 170
8.2.1 Shape-preserving interpolation 171
8.2.2 Extending the technique 174
8.2.3 Alternative approaches 175
8.3 Experiments 176
8.3.1 Parameterization 176
8.3.2 Optimal risk-neutral policy 177
8.3.3 Optimal log-utility policy 178
8.4 Extensions 182

List of Figures

4.1 A screenshot of the elicitation query for a user-selected span in the GHPM. . . . 34
4.2 A screenshot of the GHPM that shows the spikiness of prices. 36
4.3 A cluster analysis of profitable traders. 41
4.4 Prices over time in the GHPM. 45
4.5 Probability mass on the correct opening day in the GHPM. 46
4.6 Simulated final returns of a trader with inside information of official communications. 48
4.7 The number of trades placed by traders in the GHPM 51
4.8 The distribution of IAR of traders ordered by rank. 55

5.1 Total cost for making a one-ticket bet and marginal prices for an example constant-utility profit-charging market maker. 86
5.2 The profit collected by an example constant-utility profit-charging market maker. 86

6.1 Liquidity-sensitivity in increasing quantities in a two-event market 93
6.2 The cost of a unit bet in a two-event contract when the outcomes have equal quantity 94
6.3 The cost of a unit bet in a two-event contract when the outcomes have unequal quantity 95
6.4 The sum of prices in the OPRS. 100
6.5 Regions of outcome-independent profit for different values of α. 103
6.6 Regions of outcome-independent profit for different starting vectors. 103

6.7 Revenue surplus in the OPRS versus the LMSR. 105
6.8 Convex set in dual space supported by the OPRS. 113
6.9 Convex set supported by the unit ball market maker. 114
6.10 The offset angle t required for a specified maximum sum of prices in the optimal homogeneous risk measure. 120
6.11 The outer shell of the convex set which supports the optimal homogeneous risk measure with maximum sum of prices 1.05. 121
6.12 A comparison of the convex sets of the unit ball market maker and the optimal homogeneous risk measure. 122

7.1 Example of the sensitivity of exponential utility in options trading. 138
7.2 Options trading flowchart. 142
7.3 Performance of each trader over the dataset. 152
7.4 Illustration of the volatility of an underlying during the 2008 financial crisis. . . . 154
7.5 Example of the LMSR trader learning correct distributions. 155
7.6 The K-L divergence of the LMSR from the log-normal distribution taken over four chains. 156
7.7 In this capture of an in-progress run, the heavier tails of the Exponential utility trader relative to the Normal trader are evident. The y-axis is log-scaled. 159

8.1 A sample utility function. 172
8.2 The quilt which matches the function values and partial derivatives. 172
8.3 Evaluating the quilt using Bernstein bases produces a good approximation. 173
8.4 When the market maker's private beliefs align with those of the traders, the optimal ask prices and bid prices do not change significantly over the course of the interaction period. 179
8.5 When the market maker's private beliefs do not align with those of the traders, the optimal policy is highly time and wealth dependent. 180

8.6 The probability of a trader taking each offered bet from a market maker with 25 wealth in each state over the entire trading period. 181

8.7 Simulating the prices and net inventory that result from the interaction of an optimal log utility market maker starting with 25 wealth in both states. 182

LIST OF FIGURES

List of Tables

3.1 Five desiderata for risk measures and their respective expositions in the two literatures. 17

3.2 The equivalences between our desiderata and properties of the convex conjugate in Legendre-Fenchel dual space. 20

4.1 Officially communicated moving dates. 44

4.2 Simulated returns for an inside-information trader in the GHPM. 47

4.3 Counts of self-assessed savviness. 50

4.4 Self-declared savviness versus performance in the GHPM. 50

4.5 Hypothetical prices in a baseball game prediction market. 53

4.6 Gini coefficients of trader performance in the GHPM. 56

4.7 Performance of marginal and non-marginal traders in the 1988 IEM presidential market. 57

4.8 The relative performance of our candidate set of marginal traders in the GHPM. . 57

4.9 The gender composition of the marginal trader pools in the GHPM and IEM. . . 58

5.1 The difference in bid and ask prices in several markets. 71

7.1 Summary of our options dataset. 129

7.2 Daily volatility parameters. 131

7.3 An example option chain at our snapshot. 144

7.4 The prices generated by the Log-normal and Normal traders for our example. . . 145

7.5 The payout vector used in the LMSR for our example. 146
7.6 The option prices for the LMSR trader. 147
7.7 The option prices for the Exponential utility trader. 148
7.8 A summary of comparing each trading agent along a number of dimensions. . . . 149
7.9 Head-to-head comparison of traders' performance in our bootstrap analysis. . . . 150
7.10 Overall head-to-head performance comparison of the trading agents. 150
7.11 Counts of head-to-head performance of traders taking into account statistical significance. 151
7.12 Distribution of the performance of the Exponential utility trader against the Normal and LMSR traders. 153
7.13 Difference in NUPD in our experiments. 161
7.14 Comparison between the performance of random and deterministic trading strategies. 162
7.15 Fraction of options chains on which the LMSR, Log-normal, and Exponential utility traders out-performed the Discrete-prior LMSR. 163
7.16 Comparison of the Discrete-prior LMSR with other trading agents. 164

8.1 Relative errors for shape-preserving interpolation versus linear interpolation on identical rectangles. 174

Chapter 1

Introduction

Automated market makers are algorithmic agents that provide liquidity in electronic markets. In many markets, there may not be enough organic liquidity to support active trade, or the market may encompass enough events that buyers and sellers have trouble matching their orders. With only two events, a bet for one event serves to match against a bet for the other. Consequently, traders submitting bets have no problem finding counterparties: they are just traders betting on the other event. But consider a market with hundreds of events. In this case, a bet for one event serves to match against the *set* of every other event. Two traders with divergent views will not directly be each others' counterparties. Instead, the market will only clear if a set of orders spanning the totality of events—possibly consisting of hundreds of distinct orders—can be matched. So, even clearing a single order could require hundreds of competitive orders placed and waiting. In fact, the problem is even worse than what has been described here. In practice, if traders do not see feedback on their bets, they will likely withdraw from the market entirely, meaning that the hundreds of primed orders required to clear the market will never be present, and the market will fail. Furthermore, if agents are allowed to submit orders on arbitrary combinations of events, the market clearing problem becomes NP-hard (Fortnow et al., 2003). Automated market makers solve these problems by automating a counterparty to step in and price bets for traders. The tradeoff is that this automated agent can, and generally will, run at a loss. This loss can be though of as a subsidy to elicit their information (Hanson, 2003; Pennock and Sami, 2007; Chen and Pennock, 2010).

Markets mediated by automated agents have successfully predicted the openings of buildings (Othman and Sandholm, 2010a), provided detailed point spreads in sports matches (Goel et al., 2008), anticipated the ratings of course instructors (Chakraborty et al., 2011), etc. Automated market makers are also used by a number of companies (e.g., Inkling Markets) that offer private corporate prediction markets to aggregate internal information. For instance, a company could run an internal market to estimate when a new product line will ship, or whether a new initiative will increase profitability. Automated market makers are generally a necessity for this type of setting, because these corporate markets are populated by non-experts and run over an arbitrary event space.

Internet prediction markets are just one application of automated market making. The market makers we describe in this thesis are appropriate for use with any assets that trade off a *binary payoff* structure, in which the future can be partitioned into a number of states, exactly one of which will be realized. For instance, companies like WeatherBill (weather insurance) and Bet365 (sports betting) are beginning to use proprietary automated market makers to offer instantaneous price quotes across thousands or millions of highly customizable assets. These kinds of binary payout structures are also becoming more prominent within traditional finance. The Chicago Board Options Exchange (CBOE) now offers binary options on the S&P and Volatility indices. While currently lightly traded relative to standard options, the integration of these contracts into the largest options exchange in the U.S. augurs well for their future. Credit default swaps (CDS), which resemble insurance on bonds, have this kind of binary payout structure as well, in which the underlying bond either experiences a default event or does not. The total size of the CDS market was recently estimated at about 28 trillion dollars, making it one of the largest markets in the world (Williams, 2009).

Automated market makers are related to the discipline of mechanism design, but the setting is not fully analogous. Even though a market as a whole is a multi-agent system, automated market making centers on the design of a single agent. This reduced focus means that certain concerns are no longer relevant. Most notably, *why* counterparties choose to trade with the market maker is bracketed out of the design of market making agents. However, some analogues to conditions in traditional mechanism design do affect the design of market makers. For instance, we want the trades offered by our market maker to not violate individual rationality, and we would like to incentivize traders to move directly to their desired allocations, without taking on a roundabout path of intermediate allocations.

Intriguingly, many of the structures and results of the artificial intelligence (AI) literature were developed independently by the academic finance community. In that literature, automated market makers are known as *risk measures*, and rather than used constructively to create agents, they are described as regulatory controls to determine whether a portfolio is acceptable or not. While the AI literature grew out of the need to provide automated pricing in a huge variety of markets, the finance literature grew out of experiences with the failure of naïve techniques to fully comprehend risk. For instance, one of the simplest risk control techniques is *VaR* (Value at Risk). VaR assesses a portfolio by its expected performance measured at its 10th (or 1st, or 5th) percentile. This unsophisticated technique was one of the failures of risk management blamed for facilitating the recent financial crisis (Nocera, 2009). By advancing the state of the art in automated market making, this thesis also creates sophisticated tools for measuring and assessing risk.

The automated market makers of the AI literature generally function in public contexts, and so it is easy to see and track their adoption. In contrast, risk controls are normally firewalled at proprietary desks at large institutions like banks. Consequently, it is difficult to gauge the current level of acceptance and importance of these techniques within applied finance from the outside.

Our personal communications with academics working inside banks suggest that these techniques currently play some role, but not a crucial one, in generating trading decisions (Carr, 2010; Madan, 2010).

In the traditional automated market maker setting, the future state of the world is divided into a finite event partition. Interacting traders then make bets with the market maker that pay out various amounts depending on the realized future state of the world. Three examples will help to elucidate the flexibility of the setting:

- A baseball match played between the Red Sox and Yankees. Here, there are two events—"Red Sox Win" and "Yankees Win"—and a trader might request a ten dollar bet if the Red Sox win.

- A political nominating process, where the events are a range of named candidates and then an otherwise-encompassing "Field" candidate, which pays off if none of the named candidates win the nomination. In this setting, a trader might make a ten dollar bet that none of the named candidates will win the nomination.

  (Observe that in this setting, the "Field wins" event can split off into multiple named candidates, as long as traders that have made bets on this event also get shares in the split off candidate. For instance, consider a trader that holds a ten dollar payout if the Field candidate wins. If candidate X is split from the Field, that trader should also get a ten dollar payout on candidate X.)

- Insurance on a bond, where the events are if a bond experiences a default event in year one, two, three, four, or five, or does not experience a default event. A bet in this setting could be a trader requesting a payout equal to the bond's face value if the bond experiences a default event in the next five years. This bet would be roughly equivalent to a credit default swap on the underlying bond. (The event space could also be extended to cover the possibility that the bond defaults but is not completely recoverable; for instance, the event space could include events of the form "The bond defaults in year three with 57% recovery".)

In their simplest form, automated market makers work by summing the bets the market maker has made with traders (the market maker's *portfolio* or *inventory*), and mapping that vector of payouts in possible future states in the world to a single value. The market maker then prices a bet by charging the trader the difference between evaluating their current portfolio and evaluating their portfolio if the trader were to take the offered bet. Consequently, in this simple incarnation, the design of an automated market maker is just given by the behavior of a single function, a *cost function*, that maps from vectors of payouts to a single value. Immediately, there are several reasonable desiderata we would want the cost function to have. For instance, we would not want the corresponding automated market maker to be able to lose an arbitrary amount of money. We formally describe the traditional setting from both the artificial intelligence and finance literature perspectives in Chapter 3. We then extend the standard model in several ways in Chapter 5, including relaxing the finite event partition into infinite event spaces.

As we have discussed, one of the advantages of automated market makers is their ability to mediate markets with a large event space. In Chapter 4, we present the *Gates Hillman Prediction Market (GHPM)*, which at the time featured the largest event partition ever to appear in a prediction market: 365 events, generating a complete distribution over the probability the new computer science buildings at CMU would open on each day of a year. Our study is split between two parts. First, we describe the advances required to facilitate a market running over such a large event space. Second, we perform an in-depth analysis of how the market worked and how traders behaved, leveraging both the large corpus of identity-linked trades generated by the market as well as interviews with participants.

Since we ran the GHPM, prediction markets over much larger event spaces have been created. Most notably, Predictalot, a project from Yahoo! Research that we had the pleasure of consulting on, was a prediction market that ran over all 2^{63} outcomes of the annual Men's NCAA basketball tournament. We view the GHPM as a bridge between markets running over a small number of events, mediated by humans, and the later development of exponentially larger combinatorial markets, mediated with automated market makers.

In addition to enabling larger event spaces, automated market makers can also be used as an algorithmic alternative to the human (or human-constructed) market makers that currently populate existing markets. While effective on average, the worst-case performance of these agents is questionable (e.g., in the recent "Flash Crash" (U.S. Commodity and Futures Trading Commission and U.S. Securities & Exchange Commission, 2010)). Human-controlled market makers often withdraw from uncertain or volatile markets, yielding catastrophic consequences (MacKenzie, 2006; Taleb, 2007). In contrast, the market makers we construct in this thesis have well-defined performance characteristics, including bounds on loss independent of the behavior of counterparties, and need not panic in the face of uncertainty. Providing an algorithmic alternative to more fickle human-mediated market makers should be considered a long-term goal of this line of research.

One obstacle that has held this line back is that algorithmic market making agents generally do a poor job of matching many attributes of human market makers, particularly those related to profitability. To put it bluntly, the AI agents of the literature are virtually assured of losing money. As we have mentioned, automated market making agents have been employed where the goal is information elicitation, and the market maker's losses can be rationalized as subsidies to induce elicitation. However, real markets seldom run at a loss. The agents of the literature have not been successfully employed in any real-money markets where profit and loss is an important consideration.

In addition to poor average-case profitability, the most popular market making agents are burdened by a fixed market depth that is set *a priori*. In practice, this means that there may not be enough money invested in unpopular markets to reach their correct marginal prices, and that even small bets in popular markets may result in enormous changes in marginal prices.

The market makers we introduce in Chapters 5 and 6 provide solutions to these concerns about profitability and liquidity. In Chapter 5 we introduce an extension to a class of market makers from the literature that can expand market depth and produce a profit. In Chapter 6 we introduce a new class of market makers that trade according to the *relative* amounts wagered on the events in question. Consequently, with these market makers it takes a small amount of money to move marginal prices in lightly traded markets and a large amount of money to move marginal prices in heavily traded markets.

The differences between the market makers presented in the two chapters is subtle. The market makers in Chapter 5 have more descriptive power (e.g., they are able to stop increasing market depth or stop taking an additional profit cut on top of quoted prices), but the market makers in Chapter 6 are simpler to configure in existing markets. In fact, we present a market maker in Chapter 6 that has a closed-form price response, which makes it simple to implement in practice even by unsophisticated market administrators.

In the traditional model of automated market making, the market maker offers prices on the bets presented to market makers. But this same pricing logic can be used to decide whether the bet a market maker could take is profitable. For instance, imagine that a market maker would price the addition of a bet $\mathbf{x}$ to their portfolio at a price of p. If an agent offers to sell them the bet $\mathbf{x}$ at a price less than p, the market maker could take that bet and book a (subjective) profit in the amount of the difference. In this way, we can view market makers as trading agents that take or reject offered orders by comparing them to computed, subjective, fair prices.

We apply this logic to the widely traded options markets in Chapter 7. In that chapter we simulate the performance of automated traders on a large database of recent options data. Two traders we consider in that chapter are from the finance literature and price options contracts based on some distribution over the future expiration price. Another agent prices contracts based only on the inventory it holds from past trades. The key insight of this chapter is that these notions—having good priors, and learning from past actions—are not oppositional. We combine the two ideas to create a trading agent that acts based on both factors. We show that this synthesized trading agent has higher performance over our dataset measured along a number of axes, and is the Condorcet winner among our trading agents over the dataset, beating each other trader we consider in head-to-head matchups.

The significance of this chapter is not that the synthesized trading agent is the best possible options trading agent. Rather, it is the more general point that an agent's performance is enhanced by paying attention to its past actions. In contrast to this finding, models of pricing in the finance literature are generally based around the notion of *autarky* that puts prices as philosophically prior to the agents that trade on them. Put another way, in an autarky model the prices of options exist platonically, even without a market to trade on them. Since these prices exist without agents trading on them, the holdings of those agents are immaterial. But our results in Chapter 7 show that incorporating past actions into good forecasts of the future can increase performance without adding any new information into the trading process—just by paying attention to the information that autarky models ignore.

In the automated market making models we have discussed so far, our view of the setting is self-consciously uninformed. We are cognizant that traders could be better informed than the market maker, and that an arbitrary future event could be the one that is realized. Consequently, considerations like worst-case loss play an important role in the design of our agents. However, if we have more advanced information over our trading counterparties and over the probabilities of the realization of each outcome, our pricing policy can become much more sophisticated and aggressive in its pursuit of profits.

Chapter 8 explores this setting by contrasting the optimal policy of a logarithmic-utility agent and a linear-utility agent. In that chapter, we consider a two-event setting where the market maker has accurate priors over the probabilities that each event is realized and how traders respond to the bets a market maker offers. We show that while a linear-utility agent follows a simple policy that does not depend on time or the market maker's past history, the logarithmic-utility agent follows a sophisticated policy that depends on both. Surprisingly, we show that it can be optimal for a logarithmic-utility agent to offer bets that are myopically irrational (e.g., selling a contract for less than it is worth according to the market maker's beliefs) for the entire trading period. In contrast, a risk-neutral agent never offers such a bet. Computing the optimal policy of the logarithmic-utility agent is challenging and requires extending advanced techniques from the computational economics literature.

Aside from the contrast over an uninformed versus well-informed setting, there are several other tensions in the thesis. One is over the kind of utility function used. We leverage properties of logarithmic utility in Chapters 3, 5, and 7. In contrast, we show in Chapter 7 that logarithmic utility is not appropriate for trading options because it produces undefined prices for simple contracts. Another contrast is between path-independent and path-dependent market makers. In the standard setting of Chapters 3, 4, and 7, the market makers we consider are path independent. Chapter 5 presents a path dependent market maker, and in Chapter 6 we introduce market makers that could function either as path independent or path dependent. Path independent market makers allow an agent to buy and then immediately sell back a bet without cost. This is an unrealistic condition for trading in a market, because such an atomic buy-and-sell operation generally costs the trader a small amount through a bid/ask spread. However, path independence makes more sense in the context of risk measurement, because a firm is exposed identically to risk no matter how that risk was obtained.

Chapter 2

Related work

In this chapter we contextualize the thesis by providing links, comparisons, and contrasts to the literature.

General overviews

Probably the most extensive survey of the prediction market literature, and the computer science literature's relation to it, is given by Pennock and Sami (2007). Chen and Pennock (2010) is a high-level survey, focusing in particular on more recent results. Surveys of prediction markets specifically are given by Tziralis and Tatsiopoulos (2007); Wolfers and Zitzewitz (2004) and Berg and Rietz (2003), which contain extensive literature reviews of the successes and failures of prediction markets in practice. From the finance literature, Carmona (2009) is a compilation of recent research on theoretical aspects of indifference pricing. O'Hara (1995) provides a review of the theoretical market making literature and provides cogent illustrations of the foundational models in that literature.

Three popular science books cover topics related to the thesis. Surowiecki (2004) discusses the power of prediction markets for forecasting and decision making. Abramowicz (2008) advocates an extrapolation of prediction markets, suggesting their use as a new, more accountable way to make government policy. Finally, Poundstone (2006) is an accessibly-written introduction to the Kelly criterion, a key concept in the last chapter of the thesis, and its use in practice by the mathematician and hedge fund manager Ed Thorpe.

Empirical studies

Chapter 4 describes our work designing and analyzing the Gates Hillman Prediction Market. We were motivated in that study by the success of the Iowa Electronic Markets (IEM) in practice. The

IEM are the longest-running electronic prediction markets. It was originally designed for use in the 1988 presidential election, and has participated in every presidential election, and many other political events, since then. Berg et al. (2001) is a survey of the results of the IEM produced by its administrators.

Of particular significance to the study of trader behavior in Chapter 4 are the studies of trader behavior in the IEM (Forsythe et al., 1992, 1999; Oliven and Rietz, 2004). Those surveys promulgate and promote the *Marginal Trader Hypothesis (MTH)*, the idea that the IEM is so successful because a small group of rational, well-informed traders (the *marginal traders*) essentially arbitrage the much larger pool of poorly informed traders. Of course, the IEM is not always successful in its predictions, like in the 1996 markets when prices for the Clinton contract became unmoored from accurate predictions in the final weeks of the election (Berg et al., 2001). Othman and Sandholm (2010b) propose an alternative mechanism, based around agent ordering, in an effort to explain the IEM's success and failure. Because there are so few large-scale laboratory prediction market studies with identity-linked trades, and no publicly available data on those studies, studying the MTH was a major goal of the GHPM.

We also make the claim in Chapter 4 that the Gates Hillman Prediction Market was, at the time, the largest event partition ever elicited in a prediction market. Ledyard et al. (2009) and Healy et al. (2010) feature combinatorial markets with 256 events (the GHPM had 365) but the markets are not comparable in their scope, and intent. The markets described in both of those works featured fewer than ten traders participating in a laboratory environment, while the GHPM featured hundreds of traders and was publicly observable.

Chapter 7 studies options trading agents. In that chapter we simulate trading agents placing orders in options markets every 15 minutes for complete options chains. Other empirical studies have also looked at options in the context of large datasets moving through time. Schwert (1990) examines the 1987 stock market crash using a dataset of close to a hundred years of stock market data as well as options prices around the crash itself, using daily prices. Dumas et al. (1998) study the predictive power and performance of an options pricing methodology over several years of weekly pricing data. This makes it very close in form to our own study, as the profit or loss of a trading agent is directly tied to its predictive ability.

One of the central components of Chapter 7 is the study of risk-averse trading agents. This focus on risk aversion is closely related to the studies of Jackwerth and Rubinstein (1996) and Jackwerth (2000), which discuss how to back out the overall level of risk aversion embedded in the market from options prices. Specifically, those works model underlyings as having some risk-neutral distribution, which is then subjectively distorted by every trader having the same level of risk aversion. Put another way, risk aversion divorces a trader's beliefs from their actionable beliefs. This is closely related to our own design of risk-averse traders, but rather than studying the effect of all market participants having the same level of risk aversion, we instead examine the effect of a single trader's risk aversion on the way the trader prices options (which then determines the contracts that the trader finds profitable).

Cost-function-based market maker design

Most of the agents we explore in the thesis are extensions and derivations of the same automated market making methodology, *cost function* market makers. These market makers work by having a scalar field C that maps payout vectors (the amount the market maker owes to traders in each possible future state) to a real number; when a trader makes a bet that changes the market maker's payout vector from $\mathbf{x}$ to $\mathbf{y}$ that trader is charged the difference $C(\mathbf{y}) - C(\mathbf{x})$.

The original cost function market maker is the *Logarithmic Market Scoring Rule (LMSR)*, developed in Hanson (2003, 2007). The market maker was originally developed as an extension to *proper scoring rules*, which are ways of rewarding agents for making accurate forecasts (Winkler, 1969; Savage, 1971; Winkler, 1994; Gneiting and Raftery, 2007). Lambert et al. (2008b) provides a formal mathematical treatment of the links between scoring rules and cost functions.

The work of Chen and Pennock (2007) is foundational for this thesis. That work developed *constant-utility cost functions*, market makers that price contracts to maintain constant utility. We utilize constant-utility cost functions in several ways in this thesis:

- In Chapter 5, we relax the constraint that the event partition is finite, allowing arbitrary separable measure spaces instead. A major concern in this setting is that the market maker has unbounded loss. For instance, extending the LMSR over a continuous space produces a market maker with unbounded loss (Gao et al., 2009). However, by using constant-utility cost functions equipped with a special class of utility functions, *barrier* utility functions that go to negative infinity as their arguments get close to zero, we are able to produce market makers that retain bounded loss.

- We leverage barrier utility functions in constant-utility cost functions another way in Chapter 5, when we produce a market maker that expands market depth and takes a profit cut on offered prices. Here, the concern is that when we expand market depth, we could expand it too fast, producing unbounded losses. By using barrier utility functions, we are able to expand market depth in a controlled and precise way.

- We use the fact that the LMSR is equivalent to a constant exponential utility cost function in Chapter 7 in order to create a continuous version of the LMSR that operates with a specific prior belief distribution

An alternative to constant-utility cost functions as a cost function framework are *Sequential Convex Pari-mutual Mechanisms (SCPMs)*, which are intimately related to *optimized certainty equivalents* (Ben-Tal and Teboulle, 2007). SCPM market makers include the market makers in Peters et al. (2007), Agarwal et al. (2008), and their more general expansion in Agrawal et al. (2009). Here, the term *pari-mutual* is a misnomer; the term implies a mechanism in which the market maker does not risk loss, but in general, SCPMs do run at a risk of loss. One of the advances in Agrawal et al.

(2009) is the integration of limit orders within an automated market maker. In contrast, the market makers we develop here generally work only with market orders. An earlier precursor to the SCPM idea is developed in Bossaerts et al. (2002), who present what they call a "combined value" mechanism. Their mechanism is a basic version of the SCPM where the center periodically clears pooled limit orders and the market maker does not take on risk.

One of the most intriguing recent developments in automated market making is the link between cost functions and experts algorithms from learning theory, specifically between cost functions and online follow-the-regularized-leader experts algorithms. This link first appeared in a supporting role in Chen et al. (2008), and was significantly expanded in later work by those authors (Chen and Vaughan, 2010; Abernethy et al., 2011). Formally, cost functions whose marginal prices form a probability distribution are equivalent to the the class of no-regret experts algorithms. In this context, the loss of a market maker is equivalent to the regret of the experts algorithm, and the marginal price on each event is equal to the weight placed on each expert in the online learning algorithm. These online learning algorithms are conventionally expressed not as cost functions (or, in the machine learning literature, *potential functions*), but rather in dual space (Shalev-Shwartz and Singer, 2007). The dual-space formulation is a powerful way of interpreting and constructing automated market makers that we will leverage in Chapters 3 and 6.

However, cost functions and online learners are not fully analogous, and it is only a particular class of cost functions that are analogous to online expert algorithms (specifically, only convex risk measures, a class we explore in detail in Chapters 3 and 5). More general concepts from the online learning literature do not have analogues in cost functions. For instance, consider that in the standard online learning setting, it is impossible for an experts algorithm to have zero regret, but it is easy to construct an automated market maker that has zero worst-case loss—simply charge a trader more than they could win in every state of the world for each bet. This argument implies that there is no way to fit market makers whose marginal prices do not form a probability distribution (like those in Chapters 5 and 6) into the standard experts algorithm framework.

Alternative mechanisms

There are several other approaches to market making in the recent literature that we do not extend or further consider in the thesis. However, it is valuable to contrast these mechanisms with the automated market makers we do develop.

Lambert et al. (2008a) consider pari-mutual mechanisms in which participants report a probability distribution and a wagered amount, and are paid according to the amount they wager and the accuracy of their forecast. This mechanism is truly pari-mutual, because the amount it pays out is equal to the amount that is paid in by all agents. The contrast between this work and our own is that the market making agents we consider here price bets that have fixed payouts at the time the agents make them.

The Dynamic Pari-mutual Mechanism introduced in Pennock (2004) is a hybrid mechanism, part-way between the fixed-odds bets the market makers in this thesis offer and a fully variable-payout pari-mutual setting. Traders participating in this mechanism receive more "shares" when they invest in currently-unpopular outcomes, but these shares do not correspond to a fixed payout. Furthermore, the market maker can run at a loss because it must seed the initial amount of liquidity on each event.

Bayesian market makers offer an alternative way to price bets that are not related to cost functions. These market makers use Bayes rule to compute optimal bid and ask prices with traders over time, using trader interactions to learn appropriate values. One of the fundamental contrasts between these market makers and the market makers in this thesis is that Bayesian market makers essentially function by heuristics, without any worst-case guarantees on performance or behavior. Das (2008) and Das and Magdon-Ismail (2009) explore how Bayesian market makers can learn correct values from interactions with traders. Brahma et al. (2010) and Chakraborty et al. (2011) examine the relative performance of Bayesian market makers and the LMSR, finding that Bayesian market makers are generally much more profitable.

Risk measures and indifference pricing

The parallels between automated market makers and a relatively new branch of the finance literature, risk measures, are manifold. Essentially, both fields aim to price assets that take on different values in different future states of the world. Whereas the AI literature motivates these applications in the creation of trading agents, the finance literature motivates these applications by suggesting that these techniques are to be used to value untradeable assets. In these incomplete markets, risk can be priced but not fully hedged. We formally contrast the relationship between cost functions and risk measures in Chapter 3.

Despite this similarity in outward appearance, the different motivations of the two disciplines leads to very different presentations in the literature. The finance literature stresses what this thesis (and the prediction market literature broadly) refers to as the dual, price-space form of a cost function. This dual form features an explicit optimization, while cost functions themselves do not. Combined with natural differences in terminology and notation, this difference can make even close readings of the risk measure literature look unrelated to the prediction market literature.

However, in the process of examining the dual problem (to the finance literature's orientation, which recall is itself dual to the prediction market literature's orientation, and is therefore equivalent to the primal perspective of the prediction market literature), Elliott and van der Hoek (2009) explicitly develop several constant-utility cost functions. For instance, those authors solve for what the AI literature would call the LMSR as the solution to a constant-exponential-utility cost function.

In our reading, the closest that the financial literature on risk measures came to creating the automated market makers of the prediction market literature was in Carr et al. (2001). The last section of that work discusses how trading agents operating from risk measures would endogenously produce a bid/ask spread. Barrieu and Karoui (2009) also (somewhat skeptically) discuss the link between the two concepts, writing (after defining ρ as some risk measure) that

> [W]hen paying the amount $-\rho(X)$, the new exposure $X-(-\rho(X))$ does not carry any risk with positive measure, i.e., the agent is somehow indifferent using this criterion between doing nothing and having this "hedged" exposure.

The foundational work on risk measures is Artzner et al. (1999), who introduce *coherent risk measures*, with an eye on applications to regulatory policing of risk. As we will discuss in Chapter 3, the requirements of a coherent risk measure are very stringent and can be relaxed meaningfully. One particular relaxation led to the development of *convex risk measures* in Carr et al. (2001), Föllmer and Schied (2002a), and Föllmer and Schied (2002b) a class that includes the LMSR. Observe that these were developed independently from, and several years before, their corresponding construction as agents in the prediction market literature.

As we have mentioned, the LMSR is the most commonly used cost function in the AI literature. Interestingly, the risk measure analogue to the LMSR, the *entropic risk measure* formed by examining constant exponential utility, is the most frequently used risk measure in the finance literature (Barrieu and Karoui, 2009). The relationship here appears to be coincidental; the LMSR is popular in practice because of its closed-form expression and its simple bound on worst-case loss, while the entropic risk measure is popular in the finance literature because exponential utility corresponds to agents which operate in a wealth-independent manner. Examples of studies using the entropic risk measure include Delbaen et al. (2002), who provide a rigorous mathematical derivation of the duality between exponential utility and its entropic dual, Musiela and Zariphopoulou (2004), who suggest using the entropic risk measure to price assets, and Mania and Schweizer (2005), who examine using the entropic risk measure in a dynamic setting.

In this thesis, we consider only a single time period. Much of the more recent finance literature concerns *dynamic* risk measures, with explicit dependence on time, and on changing values and preferences over time, leading to studies that involve heavy use of stochastic calculus. Surveys of this literature are given by Henderson and Hobson (2009) and Barrieu and Karoui (2009).

Market microstructure theory

O'Hara (1995) divides the finance literature on market microstructure into the older study of *inventory-based market makers*, and the more recent study of *information-based market makers*. Inventory-based models center around the idea of a market maker clearing or balancing orders in an attempt to match the opposing views of traders. Information-based models center around the idea that the trades an agent makes with a market maker contain information about the underlying being traded. We contend that the automated market markets we develop in this thesis fall somewhere between the two models and have characteristics of both.

Stoll (1978) presents a foundational inventory-based model that has several parallels to this thesis. In the Stoll model, just as in much of this thesis, the market maker has some static true beliefs over the future state of the world, and has an exposure cost to taking on inventory. The Stoll model, however, involves making markets simultaneously on several stocks with known (to the market maker) rates of return. Furthermore, the exposure cost in the model of Stoll (1978) is exogenously defined, while this thesis involves the literal use of a cost function which satisfies desirable properties.

Two information-based models are especially relevant to this thesis. Glosten and Milgrom (1985) defines the basic framework for a market maker interacting sequentially with an anonymous pool of traders. In their model, the market maker sets prices so that they are exactly equal to the expectation of the market maker's posterior belief provided they are accepted. One important difference between the model of Glosten and Milgrom (1985) and this thesis is that the Glosten and Milgrom model assumes that there is competition between market makers, and so the policy of a single market maker is given by solving for behavior with zero expected profit. In contrast, Kyle (1985) considers an interaction between a monopolistic market maker and a mix of noise traders and informed traders. Our setting in Chapter 8 is similar, in that we consider utility-maximizing pricing by a monopolistic market maker. Another more general link between our thesis and the work of Kyle (1985) is that both of the models are perhaps best described as *quasi game theoretic*. We have already mentioned that the market makers in this thesis are sensitive to some game theoretic concerns, e.g., they would like to disincentivize myopic traders from taking on roundabout intermediate allocations. Furthermore, with the exception of the more informed setting in Chapter 8, the market makers in this thesis do not directly map to any simple utility function or maximization problem with explicit economic intuition. Similarly, Kyle writes that his "model is not quite a game theoretic one because the market makers do not explicitly maximize any particular objective".

We consider the market makers in this thesis to derive from both inventory-based as well as information-based models. On the one hand, a cost-function-based market maker is literally a mapping between inventories and prices. This would suggest that the market makers in this thesis are primarily inventory-based. On the other hand, as we have discussed, there are close links between cost-function-based market makers and online learning algorithms. We therefore consider the market makers in this thesis to be derived from both inventory-based as well as information-based precepts. Perhaps it is most accurate to say that the automated market makers in this thesis *learn* the correct values for the underlying event space *through* their inventories.

In Chapter 8, we solve for the pricing policy of a rational monopolistic market maker endowed with some utility function as it faces a series of traders. This setting can be considered a discrete version of the continuous-time model originally considered by Ho and Stoll (1981). In order to arrive at an analytical solution, that model used many (possibly unrealistic) simplifications, such as symmetric linear supply and demand, and also only considers the final time period of the market maker's optimization. The most challenging part of that chapter involves computing the policy of a risk-averse market maker. The notion of an explicitly risk-averse, rather than risk-neutral, market maker was introduced in Rock (1996).

Numerical dynamic programming

Chapter 8 features the computation of optimal policy of a rational market maker with probabilistic knowledge. In order to solve for the optimal policies of a Kelly criterion market maker through time, we use backwards induction to approximate the value function, working from the explicit, closed-form termination state.The backwards induction process is far from straightforward, because it carries the concern that any approximation errors are amplified—that a small error in the first approximation becomes a larger error in the next approximation, eventually causing the entire value function to be massively inaccurate.

Gordon (1995) provides a formal and experimental treatment of the problem of the expansion of error caused by approximation. That work emphasizes that any approximation needs to not expand the maximum error between the actual function and its approximation in each iteration, and that traditional approximation techniques, like a regression that minimizes sum-of-squared-distance, fail in this regard. Guestrin et al. (2001) explores using these *non-expansive approximations* to solve factored MDPs, while Stachurski (2008) experiments with non-expansive approximations for dynamic programming problems with continuous state, similar to the setting we consider in Chapter 8.

The technique we use to approximate the value function is *Constantini shape-preserving interpolation*, originally developed by Constantini and Fontanella (1990). The most prominent use of this technique in the literature is in Wang and Judd (2000), who study portfolio allocation over time between a risky stock and a risk-free bond. Similar to our approach in Chapter 8, those authors use Constantini shape-preserving interpolation to approximate the value function in each time step of a dynamic program.

Chapter 3

Theoretical background

This chapter introduces the mathematical framework we will use in the rest of the thesis. We begin by formally treating the differences and similarities between *cost functions* from the artificial intelligence literature and *risk measures* from the finance literature. Then we explore a set of desiderata from both literatures and examine the simple forms that these desiderata take in Legendre-Fenchel dual space. Finally, we explore combinations of these desiderata, producing an impossibility result that motivates Chapters 5 and 6.

3.1 Cost functions and risk measures

We consider a general setting in which the future state of the world is exhaustively partitioned into n events, $\{\omega_1, \ldots, \omega_n\}$, so that exactly one of the ω_i will occur. This model applies to a wide variety of settings, including financial models on stock prices and interest rates, sports betting, and traditional prediction markets. For instance, the events could be which of two sports teams will win their next match, or which candidate from an exhaustive set of candidates (i.e., including a "Field" candidate) will win their party's presidential nomination. In Chapter 7 the events we consider are the possible expiration prices of an underlying stock. In Chapter 5 we relax the finite nature of the partition and explore an extension to market making over an infinite number of events.

In our notation, $\mathbf{x}$ is a vector and x is a scalar, $\mathbf{1}$ is the n-dimensional vector of all ones, $\mathbf{0}$ is the n-dimensional vector of all zeros, and $\nabla_i f$ represents the i-th element of the gradient of a function f. (More traditionally, this would be denoted as $\frac{\partial}{\partial x_i}$.) Occasionally, and with obvious and clear distinction, we will abuse notation and instead let ∇_i be a *subgradient* operator that takes an arbitrary function and returns its set of subgradients along the i-th coordinate. The non-negative orthant is given by $\mathbb{R}^n_+ \equiv \{\mathbf{x} \mid \min_i x_i \geq 0\}$.

At its most abstract, this thesis involves looking at a distribution of payouts over these events and evaluating whether or not that payout is acceptable to take on (a binary, ordinal, qualitative question) and evaluating the value of that payout (a continuous, cardinal, quantitative question). Both the artificial intelligence and finance literature independently built their own frameworks for discussing these questions. The finance literature has focused on the former question through a device called *risk measures*. The artificial intelligence literature has focused on the latter question through *cost functions*.

We suggest that the reason the two concepts have been developed independently is a fundamental philosophical difference motivated by their respective applications. Unlike cost functions in the AI literature, which are directly oriented towards creating agents in markets, risk measures in the financial literature are concerned with assessing the riskiness of a portfolio. Cost functions are explicitly a cardinal measure—the values they produce matter a great deal to the resulting market maker, in that they determine the prices charged to participating traders. In contrast, risk measures are often used as binary measures: a consistent way to determine if a risk is acceptable or not, or which risk profile from a set of options is most acceptable. This binary orientation is emphasized by Artzner et al. (1999) in their foundational paper on risk measures:

> It has been pointed out to us that describing risk "by a single number" involves a great loss of information. However, the actual decision about taking a risk or allowing one to take it is fundamentally binary, of the "yes or no" type...this is the actual origin of risk measurement.

The notation traditionally used in risk measures and cost functions differs slightly but in an ultimately inconsequential way.

- Risk measures are conventionally defined as operating on a random variable X, rather than a vector. Over discrete event spaces the conversion between the two is straightforward because we can arbitrarily label the random variable's underlying set of events as $\omega_1, \ldots, \omega_n$ and assign each vector index x_i to the value of the random variable if the event corresponding to that index is realized (i.e., $x_i \equiv X(\omega_i)$).
- Risk measures have traditionally (although not exclusively, cf. Barrieu and Karoui (2009)) considered positive arguments as gains, while cost functions have considered positive arguments as losses (amounts that need to be paid out). Therefore, a cost function $C(\mathbf{x})$ would be equivalent to a risk measure $-C(-\mathbf{x})$.

In this thesis, we will use the cost function notation of the prediction market literature, and we will use the terms *cost function* and *risk measure* interchangeably, generally using cost function to denote a market maker using the function as a tool for automated pricing and risk measure to denote classes of such functions.

3.2 Cost function market makers

Let U be a convex subset of $\mathbb{R}^n$. Our work concerns functions $C : U \mapsto \mathbb{R}$ which map vector payouts over the events to scalar values. We sometimes refer to the market maker's current vector of payouts as its *state*.

Traders make bets with the market maker by changing the market maker's state. To move the market maker from state $\mathbf{x}$ to state $\mathbf{x}'$, traders pay $C(\mathbf{x}') - C(\mathbf{x})$. For instance, if the state is $x_1 = 5$ and $x_2 = 3$, then the market maker needs to pay out five dollars if ω_1 is realized and pay out 3 dollars if ω_2 is realized. If a new trader wants a bet that pays out one dollar if event ω_1 occurs, then they change the market maker's state to be $\{6, 3\}$, and pay $C(\{6, 3\}) - C(\{5, 3\})$. Arbitrary bets by traders can be represented by changing the market maker's state. The market maker starts from some initial state $\mathbf{x^0}$, generally taken to be $\mathbf{0}$ (no payouts on any event).

The *prices* or *marginal prices* of a cost function are given by the gradient of the cost function: $p_i(\mathbf{x}) = \nabla_i C(\mathbf{x})$. These values are the instantaneous cost of a bet on each event.

3.2.1 Desiderata

In this section we introduce five desiderata for cost functions: Convexity, Monotonicty, Translation invariance, Positive homogeneity, and Bounded loss. Interestingly, each of these properties has been acknowledged as desirable in both the prediction market and finance literatures (with the exception of bounded loss in the finance literature). Table 3.1 shows the appearance of these desiderata in the two literatures.

Desideratum	Finance literature	Prediction market literature
Convexity	Föllmer and Schied (2002a)	Agrawal et al. (2009)
Monotonicity	Artzner et al. (1999)	Hanson (2003); Othman et al. (2010)
Translation invariance	Artzner et al. (1999)	Hanson (2003); Agrawal et al. (2009)
Positive homogeneity	Artzner et al. (1999)	Othman et al. (2010)
Bounded loss	(does not appear)	Hanson (2003)

Table 3.1: Five desiderata for risk measures and their respective expositions in the two literatures.

We proceed to formally define all five properties and briefly describe why they should be considered valuable.

Desideratum 1 (Monotonicity). For all $\mathbf{x}$ and $\mathbf{y}$ such that $x_i \leq y_i$, $C(\mathbf{x}) \leq C(\mathbf{y})$.

Monotonicity prevents simple arbitrages like a trader buying a zero-cost contract that never results in losses but sometimes results in gains. Monotonicity also ensures that myopic traders (provided they have sufficient capital) are incentivized to directly trade with the market maker until the market maker offers no bets they find agreeable, because the marginal price of a bet never decreases.

Desideratum 2 (Convexity)**.** For all $\mathbf{x}$ and $\mathbf{y}$ and $\lambda \in [0, 1]$

$$C(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda C(\mathbf{x}) + (1-\lambda)C(\mathbf{y}).$$

Convexity can be thought of as a condition that encourages diversification. The cost of the blend of two payout vectors is not greater than the sum of the cost of each individually. Consequently, the market maker is incentivized to diversify away its risk. The acknowledgment of diversification as desirable goes back to the very beginning of the mathematical finance literature (Markowitz, 1952).

A risk measure that is not convex can produce bizarre degeneracies. For instance, Artzner et al. (1999) point out that a non-convex risk measure can produce cases of intra-firm arbitrage. If a firm uses a non-convex risk measure, it can be advantageous for two traders to report their specific portfolios $\mathbf{x}$ and $\mathbf{y}$ separately, rather than to report their desk's common exposure as $\mathbf{x}+\mathbf{y}$. Since the firm itself exists as (and is exposed to) the sum total of its risks, this division should be discouraged.

Desideratum 3 (Bounded loss)**.** $\sup_{\mathbf{x}} [\max_i (x_i) - C(\mathbf{x})] < \infty$.

A market maker using a cost function with bounded loss can only lose a finite amount to interacting traders, regardless of the traders' actions and the realized outcome.

A useful notion that quantifies bounded loss is the cost function's *worst-case loss*.

Definition 1. The *worst-case loss* of a cost function C which begins from initial payout vector $\mathbf{x^0}$ is

$$\sup_{\mathbf{x}} \left(\max_i x_i - C(\mathbf{x}) + C(\mathbf{x^0}) \right).$$

If a cost function has bounded loss, it has finite worst-case loss.

Desideratum 4 (Translation invariance)**.** For all $\mathbf{x}$ and scalar α,

$$C(\mathbf{x} + \alpha\mathbf{1}) = C(\mathbf{x}) + \alpha.$$

Translation invariance ensures that adding a dollar to the payout of every state of the world will cost a dollar.

Desideratum 5 (Positive homogeneity)**.** For all $\mathbf{x}$ and scalar $\gamma > 0$, $C(\gamma\mathbf{x}) = \gamma C(\mathbf{x})$.

Positive homogeneity ensures a scale-invariant, currency-independent price response. From a risk measurement perspective, positive homogeneity ensures that doubling a risk doubles its cost.

3.2.2 Dual space equivalences

The desiderata are global properties that need to hold over the entire space the cost function is defined over. It is often difficult to verify that a given cost function satisfies these desiderata directly, and inversely, it is difficult to construct new cost functions that satisfy specific desiderata. Remarkably, each of these desiderata have simple representations in Legendre-Fenchel dual space.

Definition 2. The *Legendre-Fenchel dual* (aka *convex conjugate*) of a convex cost function C is a convex function $f : \mathbb{S} \mapsto \mathbb{R}$ over a convex set $\mathbb{S} \subset \mathbb{R}^n$ such that

$$C(\mathbf{x}) = \max_{\mathbf{y} \in \mathbb{S}} [\mathbf{x} \cdot \mathbf{y} - f(\mathbf{y})]$$

We say that the cost function is "conjugate to" the pair $\mathbb{S}$ and f. Convex conjugates exist uniquely for convex cost functions defined over $\mathbb{R}^n$ (Rockafellar, 1970; Boyd and Vandenberghe, 2004).

We will refer to the convex optimization in dual space as the "optimization" or "optimization problem", and the maximizing $\mathbf{y}$ as the "maximizing argument".

One way of interpreting the dual is that it represents the "price space" of the market maker, as opposed to a cost function which is defined over a "quantity space" (Abernethy et al., 2011). The only prices a market maker can assume are those $\mathbf{y} \in \mathbb{S}$, while the function f serves as a measure of market sensitivity and a way to limit how quickly prices are adjusted in response to bets. As we have discussed, in the prediction market literature "prices" denote the partial derivatives of the cost function (Pennock and Sami, 2007; Othman et al., 2010). When it is unique, the maximizing argument of the convex conjugate is the gradient of the cost function, and when it is not unique, then the maximizing arguments represent the subgradients of the cost function. Consequently, the unique maximizing argument provides the market maker's marginal prices over the events. A fuller discussion of the relation between convex conjugates and derivatives is available in convex analysis texts (Rockafellar, 1970; Boyd and Vandenberghe, 2004).

With these interpretations in mind, we proceed to show the power of the dual space: we can represent desiderata simply and easily by the respective properties of their convex conjugates. The relations between convex and monotonic cost functions, convex and positive homogeneous cost functions, and their respective duals are a consequence of well-known results in the convex analysis literature (Rockafellar, 1966, 1970).

Proposition 1. *A risk measure is convex and monotonic if and only if the set* $\mathbb{S}$ *is exclusively within the non-negative orthant.*

Proposition 2. *A risk measure is convex and positive homogeneous if and only if its convex conjugate has compact* $\mathbb{S}$ *and has* $f(\mathbf{y}) = 0$ *for every* $\mathbf{y} \in \mathbb{S}$.

Desideratum	Conjugate Equivalence
Convexity (C)	Automatic
Monotonicity (M)	C,M $\Leftrightarrow$ Only defined in non-negative orthant
Translation invariance (TI)	C,M,TI $\Leftrightarrow$ Only defined on probability simplex
Positive homogeneity (PH)	C,PH $\Leftrightarrow$ $f(\mathbf{y}) = 0$
Bounded loss (BL)	C,BL $\Leftrightarrow$ Defined over whole probability simplex

Table 3.2: The equivalences between our desiderata and properties of the convex conjugate in Legendre-Fenchel dual space.

The following results can be derived from convex analysis and the work of Abernethy et al. (2011).

Proposition 3. *A risk measure is convex, monotonic, and translation invariant if and only if the set* $\mathbb{S}$ *lies exclusively on the probability simplex.*

Proposition 4. *A risk measure is convex and has bounded loss if and only if the set* $\mathbb{S}$ *includes the probability simplex.*

We will use these dual spaces properties constructively in Chapter 6, where we build new cost functions directly from their convex conjugates.

Cost functions can be classified by the desiderata they satisfy. We now proceed to describe two particular classes of risk measures that have appeared in the literature, *coherent risk measures* and *convex risk measures*. A third class of risk measure, *homogeneous risk measures*, are introduced in Chapter 6.

3.2.3 Coherent risk measures

A cost function that satisfies all of the desiderata except bounded loss is called a *coherent risk measure*. Coherent risk measures were first introduced by Artzner et al. (1999).

Definition 3. A *coherent risk measure* is a cost function that satisfies monotonicity, convexity, translation invariance, and positive homogeneity.

An impossibility result

Probably the most prominent concern after introducing our desiderata is the existence of properties of the set of functions which satisfy all five of them. In this section, we prove an important impossibility result: that only cost function which satisfies all five of our desiderata is max.

Proposition 5. *The only coherent risk measure with bounded loss is* $C(\mathbf{x}) = \max_i x_i$.

Proof. Suppose there exists a coherent risk measure C and a vector $\mathbf{x}$ with $\max_i x_i = \bar{x}$ but $C(\mathbf{x}) \neq \bar{x}$. Since C is convex, it is continuous. Therefore $C(\mathbf{0}) = 0$, because the function is positive homogeneous and for every $\mathbf{z}$

$$\lim_{\gamma \downarrow 0} C(\gamma \mathbf{z}) = 0$$

So by translation invariance:

$$C(\bar{x}\mathbf{1}) = \bar{x},$$

and so by monotonicity,

$$C(\mathbf{x}) \leq \bar{x}.$$

However, if

$$C(\mathbf{x}) < \bar{x}$$

then the loss is unbounded, because

$$\lim_{k\to\infty} k\bar{x} - C(k\mathbf{x}) = \lim_{k\to\infty} k\bar{x} - kC(\mathbf{x}) = \lim_{k\to\infty} k\left(\bar{x} - C(\mathbf{x})\right) = \infty.$$

So since the loss must be bounded,

$$C(\mathbf{x}) = \bar{x}$$

which is a contradiction. ■

The max market maker corresponds to an order-matching, risk-averse cost function that either charges agents nothing for their transactions, or exactly as much as they could be expected to gain in the best case. For instance, a trader wishing to move the max market maker from state $\{5, 3\}$ to state $\{7, 3\}$ would be charged 2 dollars, exactly as much as they would win if the first event happened—which means taking the bet is a dominated action. On the other hand, a trader wishing to move the market maker from state $\{5, 3\}$ to state $\{5, 5\}$ pays nothing! These two small examples suggest that max is a poor risk measure in practice, and therefore Proposition 5 should be viewed as an impossibility result.

Remark. *This result, in that only a single function with impractical properties satisfies a set of desiderata, is similar in flavor to the canonical possibility/impossibility result of Arrow (1963).*

3.2.4 Relaxing desiderata

Now that we have hit an impossibility result by accumulating desiderata, it is straightforward to begin examining what happens when we relax these desiderata.

We will explore two such relaxations. The first is to relax positive homogeneity. This produces a class of risk measures known as *convex risk measures*. Convex risk measures were first introduced in Carr et al. (2001).

Definition 4. A *convex risk measure* is a cost function which satisfies monotonicity, convexity, and translation invariance.

The second relaxation is to relax translation invariance. This produces a class of risk measures which we dub *homogeneous risk measures*.

Definition 5. A *homogeneous risk measure* is a cost function which satisfies monotonicity, convexity, and positive homogeneity.

Both of these relaxations have intuitive interpretations in dual space.

Dual space interpretation

Consider the convex conjugate to max. Combining all of our dual-space equivalences, we have that the conjugate of max is defined exclusively on the whole probability simplex, where it is identically 0. In dual price space, the maximizing argument to the max cost function can always be represented as a point on one of the axes.

Recall that the conjugacy operation is

$$C(\mathbf{x}) = \max_{\mathbf{y} \in \mathbb{S}} \mathbf{x} \cdot \mathbf{y} - f(\mathbf{y})$$

and for the max cost function, $\Pi = \mathbb{S}$ and f is the zero map.

In dual space, the problem with using max in practice is that its maximizing argument does not smoothly move with respect to the payout vector. It either stays at its current axis, or abruptly jumps to another axis.

When we relax positive homogeneity to produce a convex risk measure, the result is that the f function no longer needs to be the zero map. f can be a convex regularizing function which smooths out the price response. Since the set of valid price vectors is still restricted to the probability simplex, the prices of a convex risk measure will form a probability distribution.

When we relax translation invariance to produce a homogeneous risk measure, the result is that $\mathbb{S}$ no longer needs to be restricted to the probability simplex. As we show in Chapter 6, the cost function will have a smooth price response if the set $\mathbb{S}$ is curved. (The probability simplex is planar and is not curved.) Put another way, the shape of the space itself serves as an implicit regularizer over the prices. Furthermore, by relaxing the set of valid pricing vectors away from the probability simplex, the prices of the resulting homogeneous risk measure will not generally form a probability distribution.

Convex risk measures feature prominently in both the artificial intelligence and finance literatures. We proceed to provide a formal introduction to several of the more prominent risk measures from the literature. We reserve a fuller discussion of homogeneous risk measures to Chapter 6.

3.2.5 Convex risk measures

Convex risk measures are prominent in both the artificial intelligence and finance literatures. They were originally introduced in the finance literature by Carr et al. (2001). They feature very prominently in the prediction market literature; virtually every cost function market maker is a convex risk measure (Hanson, 2003; Ben-Tal and Teboulle, 2007; Hanson, 2007; Chen and Pennock, 2007; Peters et al., 2007; Agrawal et al., 2009; Abernethy et al., 2011).

In this section, we formally discuss two convex risk measures that will be important for the remainder of the thesis, the LMSR, and constant-utility cost functions.

The LMSR

The most popular cost function used in Internet prediction markets is Hanson's logarithmic market scoring rule (LMSR) (Hanson, 2003, 2007). There are at least three main reasons why the LMSR is so widely used: (1) it was the first automated market maker for prediction markets, (2) it has a simple analytical form, and (3) it has bounded loss.

The LMSR is defined as

$$C(\mathbf{x}) = b \log \left(\sum_i \exp(x_i/b) \right)$$

for fixed $b > 0$. b is called the *liquidity parameter*, because it controls the magnitude of the price response of the market maker to bets. For instance, consider an agent wishing to move the market maker from state $\{5, 3\}$ to state $\{6, 3\}$ (i.e., by making a bet that pays out one dollar if the first event occurs). If the LMSR is used with $b = 10$, $C(\{6, 3\}) - C(\{5, 3\}) \approx .56$, and so the market maker would quote a price of 56 cents to the agent for their bet. If $b = 1$, the same bet would cost 92 cents. With $b = 1$, the LMSR is equivalent to the *entropic risk measure* of the finance literature (Föllmer and Schied, 2002b).

The LMSR has a finite bound in loss which increases in the liquidity parameter b. In particular, the LMSR has a worst-case loss of $b \log n$, achieved by setting the market maker's initial payout vector to a scalar multiple of 1.

The prices of the LMSR also have a very simple analytical form:

$$p_i(\mathbf{x}) = \frac{\exp(x_i/b)}{\sum_j \exp(x_j/b)} \tag{3.1}$$

Since the LMSR is a convex risk measure, the prices in the LMSR sum to one.

Constant-utility cost functions

In this section we describe *constant-utility cost functions*, an existing framework for building cost functions that was introduced by Chen and Pennock (2007). We apply and extend this framework in both Chapter 5 and Chapter 7.

Definition 6. Let $U \subset \mathbb{R}$ be an open interval on the real line that includes all positive values. A *utility function* is a strictly increasing, concave function $u : U \mapsto \mathbb{R}$.

A constant-utility cost function works by charging traders the amount that keeps the market maker's utility at a constant level. Put another way, the market maker prices each bet so that he is indifferent between a trader declining and accepting it.

Definition 7. Let $x^0 \in \textbf{dom}\ u$ and π_i be the market maker's (subjective) probability that ω_i will occur. A *constant-utility cost function* $C : \mathbb{R}^n \mapsto \mathbb{R}$ is defined implicitly as the solution to

$$\sum_i \pi_i u(C(\mathbf{x}) - x_i) = u(x^0)$$

Since the cost function is given implicitly by a root-finding problem, rather than explicitly as a function of the vector **x**, it is not immediately clear that costs exist, are unique, and are easily computable. However, because the utility function is strictly increasing, the cost function exists for any input vector and is unique. Furthermore, because the utility function is increasing, we can compute b bits of the value of the cost function in b steps of a binary search over possible values.

Constant utility cost functions start from the initial payout vector **0**. Observe that $C(\mathbf{0}) = x^0$, because

$$\sum_i p_i u(x^0 - 0) = u(x^0)$$

Where u is differentiable the prices of a constant-utility cost function have a closed form (Jackwerth, 2000; Chen and Pennock, 2007):

$$p_i(\mathbf{x}) = \frac{\pi_i u'(C(\mathbf{x}) - x_i)}{\sum_j \pi_j u'(C(\mathbf{x}) - x_j)}. \tag{3.2}$$

Observe that here the prices form a probability distribution because they sum to one and are non-negative (the utility function is strictly increasing and so its derivative is always positive).

For completeness, we proceed to prove formally that constant-utility cost functions are indeed convex risk measures.

Proposition 6. *Constant-utility cost functions are convex risk measures.*

Proof. Recall that in order to be a convex risk measure a cost function must satisfy monotonicity, convexity, and translation invariance.

Monotonicity holds because the utility function is increasing. Consider $\mathbf{y} \geq \mathbf{x}$. Then because the utility function is increasing:

$$\sum_i \pi_i u(C(\mathbf{x}) - x_i) \geq \sum_i \pi_i u(C(\mathbf{x}) - y_i)$$

and so

$$C(\mathbf{y}) \geq C(\mathbf{x})$$

so the cost function is monotonic.

Translation invariance is straightforward. To prove translation invariance, observe that

$$\sum_i \pi_i u(C(\mathbf{x} - x_i)) = \sum_i \pi_i u((C(\mathbf{x}) + \alpha) - (x_i + \alpha))$$

and therefore

$$C(\mathbf{x} + \alpha\mathbf{1}) = C(\mathbf{x}) + \alpha$$

Finally, the cost function is convex, which can be seen by noting that it corresponds to the iso-utility graph of a concave utility function. Proofs of the convexity of such functions are well-known and can be found in standard microeconomics texts (Varian, 1992). ■

In some parts of the thesis we will be particularly concerned with a special class of utility functions called *barrier* utility functions. These are functions which have a barrier to going into negative wealths.

Definition 8. A *barrier* utility function u is a utility function that has $\lim_{x \downarrow 0} u(x) = -\infty$.

Examples of barrier utility functions include $u(x) = \log(x)$ and $u(x) = -1/x$, both of which are defined over the open interval $(0, \infty)$.

Remark. *There is nothing special about the lower bound of zero; all our results for barrier utility functions still hold qualitatively as long as* some *lower bound exists. However, restricting the class to only those functions with a barrier of zero is without loss of generality, because utility functions with a different barrier (say, a function u with a barrier x) could be translated into a barrier function u' with a barrier of 0 by the translation $u'(z) = u(z + x)$.*

One advantage of using constant-utility cost functions with barrier utility functions is that their worst-case loss can be bounded simply.

Proposition 7. *Let C be a constant-utility cost function that employs a barrier utility function. If $\pi_i > 0$ for every i, the worst-case loss of the cost function is bounded by x^0.*

Proof. Suppose x^0 were not an upper bound. Then there exists some $\mathbf{x}$ such that

$$\max_i x_i - C(\mathbf{x}) - C(\mathbf{0}) > x^0$$

and because $C(\mathbf{0}) = x^0$ by definition, this implies

$$\max_i x_i > C(\mathbf{x})$$

but by definition

$$\sum_i \pi_i u(C(\mathbf{x}) - x_i) = u(x^0)$$

but then for the maximal i, this equation requires the evaluation of $u(z)$ for $z < 0$. Since we are using a barrier utility function, this value is undefined. Consequently, we have that x^0 is an upper bound.

■

Despite their different forms, the LMSR and constant-utility cost functions have much in common. In fact, it is possible to fully express the LMSR as a constant-utility cost function.

Priors and constant utility in the LMSR

Equations 3.1 and 3.2 suggest that the LMSR, starting from initial payout vector **o**, is equivalent to a constant-utility cost function with $\pi_i = 1/n$ and $u(x) = -\exp(-x/b)$ (that is, exponential utility). This link is for a prior-free, maximum-entropy LMSR. There are two equivalent ways to incorporate a prior $\{\pi_1, \ldots, \pi_n\}$ into the LMSR.

The first way is to go back to the definition of marginal prices within constant-utility cost functions:

$$\nabla_i C(\mathbf{x}) = \frac{\pi_i u'(C(\mathbf{x}) - x_i)}{\sum_j \pi_j u'(C(\mathbf{x}) - x_j)}$$

This equation gives an easy way to incorporate the prior into prices. Since in the LMSR, $u(x) = -\exp(-x/b)$, these prices correspond to the following cost function

$$C(\mathbf{x}) = b \log \left(\sum_j \pi_j \exp(x_j/b) \right)$$

The second way is based on the relation of the LMSR to the logarithmic proper scoring rule. The connection between automated market makers and scoring rules is deep, and is the focus of much prior literature (Hanson, 2007; Pennock and Sami, 2007; Lambert et al., 2008b; Chen and Pennock, 2010).

Here, we initialize an initial payout vector $\mathbf{q^0}$ as the payouts implied by the logarithmic scoring rule:

$$q_i^0 = b \log(\pi_i)$$

Then the cost function proceeds as usual, but with the addition of the $\mathbf{q^0}$ vector of payouts, so that

$$C(\mathbf{x}) = b \log \left(\sum_j \exp((x_j + q_j^0)/b) \right)$$

These two constructions are in fact equivalent because

$$\pi_i \exp(x_i/b) = \exp(\log(\pi_i)) \exp(x_i/b) = \exp((x_i + b \log(\pi_i))/b) = \exp((x_i + q_i^0)/b)$$

Now that we have introduced the LMSR formally, we proceed to discuss a practical application of the cost function.

CHAPTER 3. THEORETICAL BACKGROUND

Chapter 4

The Gates Hillman Prediction Market

Prediction markets are powerful tools for aggregating information. Most prediction markets in use today, however, only generate a single data point. For simple binary events, like the probability of a sports team winning its next match, this is entirely satisfactory. However for more complex events this can be inappropriate. Consider a prediction market to estimate the expected number of US casualties in Afghanistan over the next year. Conceivably, market participants could be split between a very low estimate and a very high estimate. The resulting consensus of a middle value could be an accurate estimate of the expectation, but would be misleading to design policy around.

Recent theoretical work has suggested that eliciting interesting distribution properties (like the element that has maximum probability) is as difficult as eliciting an entire distribution (Lambert et al., 2008b). In this paper, we discuss the design of a market, the *Gates Hillman Prediction Market (GHPM)*, that generated a complete distribution over a fine-grained partition of possibilities, while retaining the interactivity and simplicity of a traditional market.

The GHPM market was designed to elicit the opening day of the new computer science buildings at Carnegie Mellon University, from a universe of 365 potential days. At any snapshot in time, the market forecast a probability distribution over the building opening on each of these days. At a high level, we were motivated to build the GHPM to answer three kinds of questions. Two of them are specific to the GHPM: (1) What extensions are necessary to traditional theory in order to run a market over a large number of possible events? and (2) What breaks in practice when a market runs over hundreds of events, hundreds of traders, and hundreds of days? The third question is common to any market case study: How did the market work?

Fundamental to our design is an automated market maker (Hanson, 2003). It has three primary benefits. First, the market maker provides a rich form of liquidity: it guarantees that participants can make any self-selected trade at any time. Second, it allows instant feedback to traders, rather than delayed, uncertain, potential feedback. A trader can always get actionable prices both on any potential trade she is considering and on the current values of the bets she currently holds. Third, the automated market maker obviates the need to match combinations of user-generated buy and sell orders—a problem that can be combinatorially complex (Fortnow et al., 2003; Chen et al., 2008)—making a large event space computationally feasible.

Equally important to the success of the GHPM was the user interface of the trading platform. Interfaces with the same expressive power in theory can perform quite differently in practice, particularly given well-documented shortcomings in human reasoning. In particular, a large event space implies that the average probability of an event is small, and people have great difficulty discriminating between small probabilities (e.g., Ali (1977)). To solve this problem, the GHPM used a span-based interface with ternary elicitation queries, which we discuss in Section 4.1.3.

As the first test of automated market making in a large prediction market, the GHPM allowed us to discover two flaws in current automated market makers, which will help focus future design of market makers. Section 4.2 discusses the two flaws, *spikiness* and *liquidity insensitivity*, in detail and explores their theoretical roots.

Traditional laboratory experiments are generally small due to practical constraints like subject payments, training effort, and the viable duration of an experiment. For example, Healy et al. (2010) study behavior and prices in laboratory prediction markets in detail, but their experiment only had three traders. The Gates Hillman Prediction Market involved hundreds of traders making thousands of trades, and so provides an unusually rich data set, particularly when combined with interviews with traders about the strategies they employed. Sections 4.3 and 4.4 use the data generated by the market to examine its performance and characteristics in depth.

4.1 Market design

The GHPM used a raffle-ticket currency tied to real-world prizes, an automated market maker, and a novel span-based ternary elicitation interface. In the following sections we discuss each of these in turn.

4.1.1 Incentives and setup

Due to legal concerns, the GHPM used raffle tickets as currency rather than real money. Thanks to generous grants from Yahoo! and other sources, we secured the equivalent of about $2,500 in prizes to distribute. At the close of the market, prize selection slots were allocated randomly, in proportion to the number of tickets each user amassed. Traders then selected their prizes in descending order until all the prizes were exhausted. This gives (risk-neutral) participants the same incentives as if real money were used—unlike the approach where the best prize is given to the top trader, the second-best prize to the second-best trader, etc.

The GHPM was publicly accessible on the web at `whenwillwemove.com`, but trading accounts were only available to holders of Carnegie Mellon e-mail addresses. For fairness, we did not allow people with direct control over the building process (e.g., members of the building committee) to participate. Upon signup, each user received 20 tickets, and each week, if that user placed at least one new trade, she would receive an additional bonus of two tickets. In a market with real money, we would expect that traders more interested or knowledgeable would stake more of their personal funds in the market. However, in a fake-money setting, we do not have this option. For instance, a mechanism that asked users if they were "very interested" in the market, and promised to give them extra tickets if they answered affirmatively would obviously not be incentive compatible. The two ticket weekly bonus was intended to give more interested traders more influence in the market and to encourage traders to be more involved in the market over time.

One of the most challenging parts of running a prediction market over real events is defining contracts so that it is clear which bets pay out. For example, *InTrade*, a major commercial prediction market, ran into controversy over a market it administered involving whether North Korea would test missiles by a certain date. When North Korea putatively tested missiles unsuccessfully, but the event was not officially confirmed, the market was reduced to a squabble over definitions. We set out to study when the Computer Science Department would move to its new home in the Gates and Hillman Centers (GHC), but *move* is a vague term. Does it indicate boxes being moved? Some people occupying new offices? The last person occupying a new office? The parking garage being open? From discussions with Prof. Guy Blelloch, the head of the building committee, we settled on using "the earliest date on which at least 50% of the occupiable space of the GHC receives a temporary occupancy permit". Temporary occupancy permits are publicly issued and verifiable, must be granted before the building is occupied, and are normally issued immediately preceding occupancy (as was the case in the GHC).

The market was active from September 4th, 2008 to August 7th, 2009. On this latter date, the GHC received its first occupancy permit, which covered slightly over 50% of the space in the building. The price of a contract for August 7th, 2009 converged to 1 about five hours before the public announcement that the building had received its permit.

In total, 210 people registered to trade and 169 people placed at least one trade. A total of 39,842 bets were placed with the market maker, with about two-thirds of the trades in the market being placed by a single trading bot (further discussed in Section 4.4.4). Following the conclusion of the market, we conducted recorded interviews with traders we deemed interesting about their strategies and their impressions of the GHPM. Excerpts of some of these conversations appear in Section 4.3.

4.1.2 The LMSR in the GHPM

We began by partitioning the event space into $n = 365$ events, one for each day from April 2, 2009 to March 30, 2010 with the addition of "April 1, 2009 and everything before" and "March 31, 2010 and everything after", to completely cover the space of opening days. At the time, the

GHPM was by far the largest market (by event partition size) ever conducted. The largest prior prediction markets fielded in practice were, to our knowledge, markets over candidates for political nominations, where as many as 20 candidates could have contracts (of course, only a handful of candidates in these markets are actively traded). Previous laboratory studies have involved limited trials with as many as 256 events (Ledyard et al., 2009), but those studies involved very few traders and are not commensurable with the GHPM, which was designed as a publicly visible mechanism. Since the GHPM concluded, markets with much larger event spaces have been fielded. *Predictalot*, a product of Yahoo! Research, was a public prediction market that fielded bets on the 2010 and 2011 NCAA men's basketball tournament, a setting with 2^{63} events. As the first large-scale test of automated market making, the GHPM is the link between smaller, human-mediated markets and the later development of exponentially larger combinatorial markets, mediated with automated market makers.

Recall that the cost function for the LMSR is

$$C(\mathbf{q}) = b \log \left(\sum_i \exp(q_i/b) \right)$$

where $b > 0$ is a constant fixed *a priori* by the market administrator. As Pennock and Sami (2007) discuss, the b parameter can be thought of as a measure of market liquidity, where higher values represent markets less affected by small bets. In the GHPM we fixed $b = 32$, and since the LMSR has worst-case loss of $b \log n$ (Hanson, 2003, 2007), at most about 190 surplus tickets would be won from the market maker by participating traders. (This is indeed the amount actually transferred from the market maker to the participants because probability mass converged to the correct day before the market ended.) The *ad hoc* nature of selecting the liquidity parameter is an intrinsic feature of using the LMSR, as it comes with little guidance for market administrators (Othman et al., 2010). The parameter we selected turned out to be too small, which led to some problems in the market (see Section 4.2.2).

Recall that prices are defined by the gradient of the cost function, so that

$$p_i(\mathbf{q}) = \frac{\exp(q_i/b)}{\sum_j \exp(q_j/b)}$$

is the price of the i-th event. We call these p_i *pricing rules*. Recall that the LMSR is a differentiable convex risk measure, and so the prices can also be directly thought of as event probabilities, because they define a probability distribution over the event space: they sum to unity, are non-negative, and exist for any set of events.

4.1.3 Span-based elicitation with ternary queries

In this section, we present the novel elicitation mechanism used in the GHPM. A similar interface was developed independently and contemporaneously by Yahoo! Research for Yoopick, an application for wagering on point spreads in sporting events that runs on the social network Facebook (Goel et al., 2008).

The major problem in implementing fine-grained markets in practice is one of elicitation: they are too fine for people to make reliable point-wise estimates. Consider the GHPM, which is divided into 365 separate contracts, each representing a day of a year. Under a traditional interaction model, traders would act over individual contracts, specifying their actionable beliefs over each day. But with 365 separate contracts, the average estimate of each event is less than .3%. People have great difficulty reliably distinguishing between such small probabilities (Ali, 1977), and problems estimating low-probability events have been observed in prediction markets (Wolfers and Zitzewitz, 2006).

We solve this problem by simple span-based elicitation, which makes estimation of probabilities easy for users. In our system, the user can select a related set of events and gauge the probability for the entire set. Spans are a natural way of thinking about large sets of discrete events: people group months into years, minutes into hours, and group numbers by thousands, millions, or billions. The key here is that spans use the concept of distance between events that is intrinsic to the setting.

For example, let the market be at payout vector $\mathbf{q^0} = \{q_1^0, \ldots, q_n^0\}$. A user's interaction begins with the selection of an interval from indices s to t. This partitions the indices into (at most) three segments of the contract space: $[1, s)$, $[s, t]$, and $(t, n]$. The user then specifies an amount r to risk. Our market maker proceeds to offer the following alternatives to the user:

- The "for" bet. The agent bets *for* the event to occur within the contracts $[s, t]$. The user's payoff if he is correct, π_f, satisfies

$$C\left(q_1^0, \ldots, q_{s-1}^0, q_s^0 + \pi_f, \ldots q_t^0 + \pi_f, q_{t+1}^0, \ldots, q_n^0\right) = C(\mathbf{q^0}) + r$$

- The "against" bet. The agent bets *against* the event occurring within the contracts $[s, t]$. The user's payoff if he is correct, π_b, satisfies

$$C\left(q_1^0 + \pi_b, \ldots, q_{s-1}^0 + \pi_b, q_s^0, \ldots, q_t^0, q_{t+1}^0 + \pi_b, \ldots, q_n^0 + \pi_b\right) = C(\mathbf{q^0}) + r$$

As long as the cost function is strictly increasing, as is the case for the LMSR, these values uniquely exist. However, solving for π_f and π_b is not generally possible in closed form. These equations can be solved numerically using, for example, Newton's method. Depending on the specific cost function and numerical solution method, there might be issues with solution instability that should be addressed; for instance, the GHPM used Newton's method with a bounded step size at each iteration to discourage divergence.

Given a selected set of events, the simplest way to represent a bet for that set is to have each event in the set pay out an identical amount if the event is realized, as we do in the two equations above. This simplicity means we can significantly condense the language we use when eliciting a

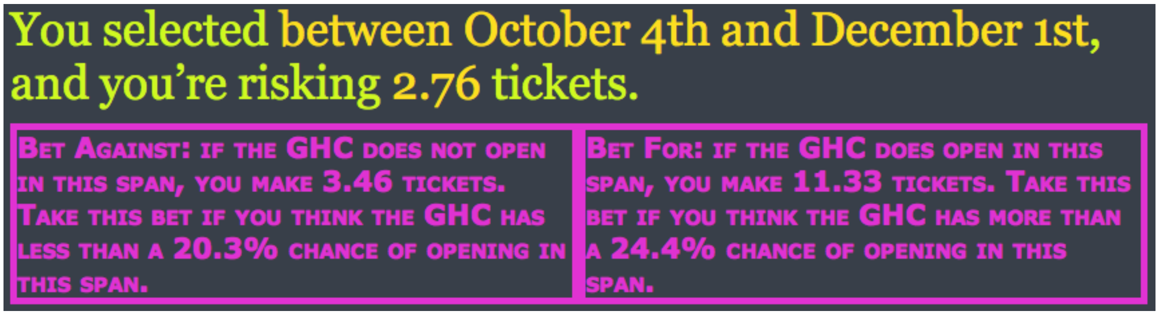


Figure 4.1: A screenshot of the elicitation query for a user-selected span in the GHPM. The query is ternary because it partitions the user's probability assessment into three parts. *The GHC* is the Gates and Hillman Centers, the new computer science buildings at Carnegie Mellon. Because of legal concerns, the market used raffle tickets rather than money.

wager from a trader. Instead of asking the trader whether he would accept an n-dimensional payout vector, we need only present a single value to him. A screenshot of the elicitation process in the GHPM appears as Figure 4.1.

Yahoo! Research's Yoopick does not have "against" bets, but the GHPM does. From discussions with traders in the GHPM, against bets were used frequently to bet against specific (single) contracts they feel are overvalued. Several successful traders had a portfolio consisting solely of bets against a large number of single contracts. The success of these traders was likely a combination of the misjudging of small probabilities by other traders as well as the spiky price phenomenon discussed in the next section.

There are several relevant pieces of information the market administrator could provide the users for each potential bet:

- The agent's direct payout if he is correct, π_f (or π_b). This is the amount a trader wins if he bets on a span including the event that occurs and he holds the contract through to expiry. Both Yoopick and the GHPM display this information.
- The *averaged* payout probability on the span, r/π_f or $1 - r/\pi_b$. This value is the actual odds at which a bet is being made. The GHPM displays this as a *ternary* (three-way) query, where traders can select whether their probability estimate lies in one of three partitions, as in Figure 4.1. (Yoopick does not display this information.) Faced with the ternary query, the user selects whether his probability for the span is less than $1 - r/\pi_b$, greater than r/π_f, or in-between. (If the prices are increasing in quantities, as they are in the LMSR, then $r/\pi_f \geq 1 - r/\pi_b$, with equality in the limit as $r \to 0$.) If a trader's belief lies in the middle partition, presumably they could reduce their bet size or find another span on which to wager.
- The *marginal* payout probability, which is the sum of the prices on the relevant span after

the π_f or π_b of additional quantity. Since agents who are acting straightforwardly will not want to move marginal prices beyond their private valuation, marginal prices could be more informative to decision making. Neither the GHPM nor Yoopick displays this information. Early trials of the GHPM included marginal prices in the interaction interface, but testers found the information confusing when combined with the averaged payout probabilities and so we removed the marginal payout probabilities from later versions of the interface. Even though they were not explicit in the interface, sophisticated traders could still produce marginal prices either by explicitly knowing the pricing rule or by making small tweaks in the number of tickets risked and observing how prices changed. We believe that for a market exclusively populated by mathematically adept traders, explicit marginal prices would be a helpful tool.

Finally, even though it simplifies interactions, the span-based elicitation scheme is arbitrarily expressive. If the users are sophisticated enough to make discriminating judgments over small probabilities, to the point that they can express their actionable beliefs over every contract, then they can still express this sophistication using spans—e.g., by trading spans that contain only one element (one day in the case of the GHPM).

4.2 Problems revealed

There are two key findings from our study. The first is a large and interesting data set of trades, which we analyze in Sections 4.3 and 4.4. The second is that we discovered two real-world flaws in the automated market-making concept. These were the *spikiness* inherent in prices and the *liquidity insensitivity* that made prices in the later stages of the market change too much. We proceed to discuss each of these flaws.

4.2.1 Spikiness of prices across similar events

A phenomenon that quickly arose in the GHPM was how spiky the prices were across events at any snapshot in time. There was extraordinary local volatility between days that one expects should have approximately the same probability. This volatility is far more than could be expected from a rational standpoint—e.g., weekends and holidays could be expected to have much lower probability than weekdays—and it persisted even in the presence of profit-driven traders whose inefficiency-exploiting actions made for less pronounced, but still evident, spikes. (These traders' strategies are discussed in detail in Section 4.3.2.) Figure 4.2 is a screenshot of the live GHPM where spiky prices are evident. Spikiness here refers to the exaggerated sawtooth pattern of prices in a fixed snapshot in time, not to how prices moved or changed over time. It is a distinct phenomenon from both real markets with sharply rising or falling prices (e.g., electricity spot markets) or from the "mirages" observed in laboratory studies of the LMSR (Healy et al., 2010).

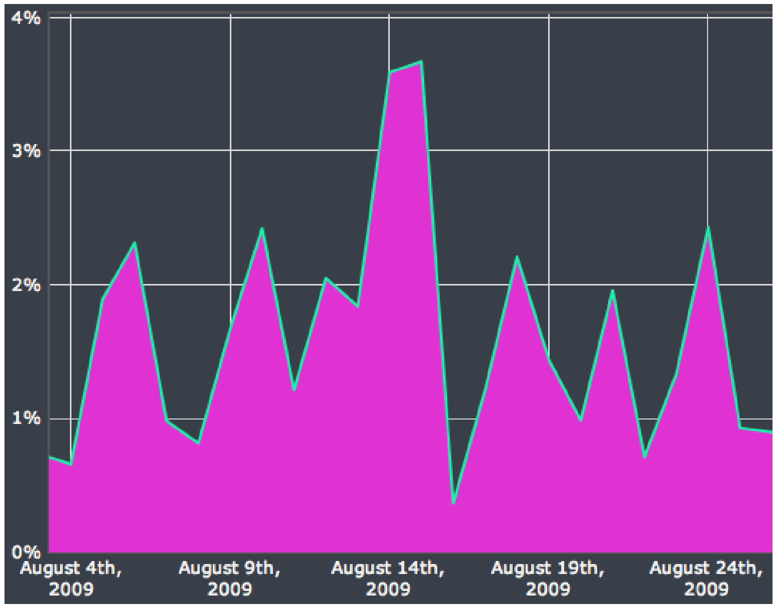


Figure 4.2: A screenshot of the GHPM that shows the spikiness of prices. The x-axis ranges over a set of potential opening days. The y-axis displays prices (as percentages; e.g., 1% means a price of 0.01).

Spikiness in theory

Why did spikiness occur? Is there an automated market maker with a different pricing rule that would have resulted in a market where prices were not spiky? In this section, we show that a large class of market makers will tend to induce spiky prices. Specifically, our result concerns pricing rules which satisfy three conditions: differentiability, non-negativity, and a technical condition we call *pairwise unboundedness*.

Definition 9. When $n \geq 3$, a pricing rule is *pairwise unbounded* if, when enough is bet against any pair of events, the prices on both of those events goes to zero. Formally, let the set I consist of any two events i and j, and let $\mathbf{I}$ be the indicator vector for the set I, so that $\mathbf{I}_i = \mathbf{I}_j = 1$ and $\mathbf{I}_{\neq i \vee j} = 0$ otherwise. Then for any $\mathbf{x}$

$$\lim_{k \to \infty} p_i(\mathbf{x} - k\mathbf{I}) = \lim_{k \to \infty} p_j(\mathbf{x} - k\mathbf{I}) = 0$$

Of the three properties, non-negativity and pairwise unboundedness are the most natural. A negative price would make the marginal prices not form a probability distribution, and would imply a trader could arbitrage the market maker by buying the event with a negative price. The pairwise unbounded condition also is also natural, because a market maker that is not pairwise unbounded will keep non-vanishing prices on pairs of events no matter how much is bet against them by traders.

To our knowledge, every market maker in the literature satisfies both non-negativity and pairwise unboundedness. In contrast, there do exist market makers that do not have a differentiable price response (such as max). However, a smooth price response is desirable for interacting traders (Othman and Sandholm, 2011b), and the bulk of market makers from the literature do have differentiable prices. It is easy to see that the LMSR satisfies all three conditions.

What does it mean for a market maker to produce spiky prices? From examining transactions in the GHPM, we found that spiky prices arose most often from tiny differences in the prices of nearby days being amplified by a bet span that included both of them. These tiny initial price differences seem to arise endogenously from traders selecting slightly different intervals to bet on, although there is no guarantee that they will be present. With this in mind, consider a bet placed for the interval $I = \{i \text{ and } j\}$, when $p_i > p_j$. Recall that the derivative of a differentiable function $g : \mathbb{R}^n \mapsto \mathbb{R}$ along a vector **x** (the *vector derivative*) is given by the scalar $\nabla_{\mathbf{x}} = \nabla g \cdot \mathbf{x}$. The vector derivative $\nabla_{\mathbf{I}}$ of the price functions p_i and p_j is

$$\nabla_{\mathbf{I}} p_i = \nabla_i p_i + \nabla_j p_i$$

and

$$\nabla_{\mathbf{I}} p_j = \nabla_j p_j + \nabla_i p_j$$

The market maker has spike-inducing behavior if betting on the bundle that consists of both i and j amplifies the difference between the prices on those events, that is, if

$$\nabla_{\mathbf{I}} p_i > \nabla_{\mathbf{I}} p_j$$

By the symmetry of second derivatives, $\nabla_j p_i = \nabla_i p_j$. Therefore we have spike-inducing behavior for this bet if

$$\nabla_i p_i > \nabla_j p_j$$

This leads us to the following definition.

Definition 10. In a setting where $n \geq 3$, a market maker *induces spikes* if

$$\nabla_i p_i(\mathbf{x}) > \nabla_j p_j(\mathbf{x})$$

for some **x** and some i and j such that

$$p_i(\mathbf{x}) > p_j(\mathbf{x})$$

Proposition 8. *Every differentiable, non-negative, and pairwise unbounded pricing rule induces spikes.*

Proof. Suppose there is a pricing rule satisfying the three conditions that does not induce spikes. Then, for every $\mathbf{x}, i, j$ tuple we have

$$\nabla_i p_i(\mathbf{x}) \leq \nabla_j p_j(\mathbf{x})$$

Select some $\mathbf{x}$, some indices i and j, and some $\epsilon > 0$ such that

$$p_i(\mathbf{x}) \geq p_j(\mathbf{x}) + \epsilon$$

Because the market maker has differentiable, pairwise unbounded prices, such an $\mathbf{x}, i, j$ tuple must exist.

Now consider the vector $\mathbf{I}$ which will be the indicator vector for indices i and j. Because the market maker does not induce spikes we have that, for $k \geq 0$

$$p_i(\mathbf{x} - k\mathbf{I}) \geq p_j(\mathbf{x} - k\mathbf{I}) + \epsilon$$

But because the market maker has pairwise unbounded prices,

$$\lim_{k \to \infty} p_i(\mathbf{x} - k\mathbf{I}) = 0$$

Consequently there exists some $K > 0$ such that

$$p_i(\mathbf{x} - K\mathbf{I}) \leq \epsilon/2$$

But then at this K,

$$p_j(\mathbf{x} - K\mathbf{I}) \leq -\epsilon/2$$

so the market maker has prices that are negative, but this is a contradiction, because we assumed the pricing rule was non-negative. ■

The impact of spikiness

Traders were aware of spikiness and this knowledge affected their behavior. In Section 4.3, we discuss and analyze interviews with traders which suggest that spikiness played a large role in determining the way agents behaved in the GHPM.

Spiky prices are a problem because they create a disconnect between the user and the elicitation process. Users feel that the spiky prices they observe after interacting with the market maker do not reflect their actual beliefs. This is because users agree only to a specified bet rather than to an explicit specification of prices after their interaction. Moreover, because the difference between spiky prices and (putatively) efficient prices is so small, traders have little incentive to tie up their capital in making small bets to correct spikiness; there is almost certainly another interval where their actionable beliefs diverge more from posted prices. Our interview with Brian, a PhD student in the Machine Learning Department and the market's best-performing trader, was informative. He described a sophisticated strategy where he would check the future prospects of his current holdings against what he viewed as a risk-free rate of return—for instance, by betting against the building opening on a weekend. If the risk-free rate of return was higher, he would sell his in-the-money holdings and buy into the risk-free asset. So once a spike is small enough, damping it out can be less lucrative than other opportunities. (This argument also implies that spikiness could be diminished by supplying traders with more capital.)

Finally, to bet against a spike, a trader accepts an equal payout on every other day. But increasing the quantity on events by the same amount is what caused spiky prices in the first place. Put another way, betting against a spiky price will have the tendency to create spiky prices elsewhere in the probability distribution. Consequently, when using a differentiable market maker, while it might be possible for savvy traders to diminish spikiness, it seems unlikely that it could ever be fully eliminated.

4.2.2 Liquidity insensitivity

Recall that the LMSR is translation invariant and differentiable. Consequently

$$p_i(\mathbf{q}) = p_i(\mathbf{q} + \alpha\mathbf{1})$$

for scalar α. A practical interpretation of this result is that the market maker is *liquidity insensitive*, so that quoted responses do not respond to the level of activity seen in the market. This implies that prices change exactly the same amount for a one dollar bet placed at the start of the market (say, at $\mathbf{q} = (0, 0, \ldots, 0)$) as after the market maker has matched millions of dollars ($\mathbf{q} = (1000000, \ldots, 1000000)$).

This is not the way we think of markets in the real world, operated by humans, as working. As markets grow larger with more frequent trading, they become deeper so that small bets have vanishingly small impact on prices. When using the LMSR, because the liquidity parameter b is a constant, the market's reaction to bets at increasing levels of volume is constant, too.

Impact on trader behavior

Liquidity insensitivity had an impact on traders in the GHPM, but unlike spikiness, which was publicly visible and a source of frequent consternation in our interviews, it appears that only the most active and savvy traders were aware of liquidity insensitivity. Brian, the market's best trader, said this about the way he approached the market in its final weeks:

> *One thing I noticed was that at the end, these small bets would still make big jumps in the prices. So I would try to keep the amount that I bet really small...to try and minimize what would happen to the prices.*

So, at least the savviest traders were aware of the disconnect between the automated market maker and the way a traditional market would function.

Relation to spikiness

Although spikiness and liquidity insensitivity appear quite different, they are actually related. A market maker that is sensitive to liquidity would be able to temper spikiness, because in more liquid (deeper) markets, the market maker could move prices less per each dollar invested. Since spikes are the product of discrepancies in the amount that prices move, if prices move less, spikiness will be diminished.

4.3 Effective trader strategies

In this section, we discuss the strategies employed by profitable traders. We found these strategies can be grouped into three categories, *spike dampening*, *relative smoothing*, and *information gathering*. We begin by performing a cluster analysis of the set of profitable traders to identify the qualitative attributes of successful strategies.

4.3.1 Cluster analysis

There were 49 traders who ended the market with more than 20 tickets and who made at least five trades (the median number over the whole set of traders). Consider grouping those traders along two criteria: First, the fraction of bets they made which were negative (i.e., *against* the span they selected, as in Figure 4.1), and the number of total trades they made. Figure 4.3 shows the results of clustering these successful traders into two groups.

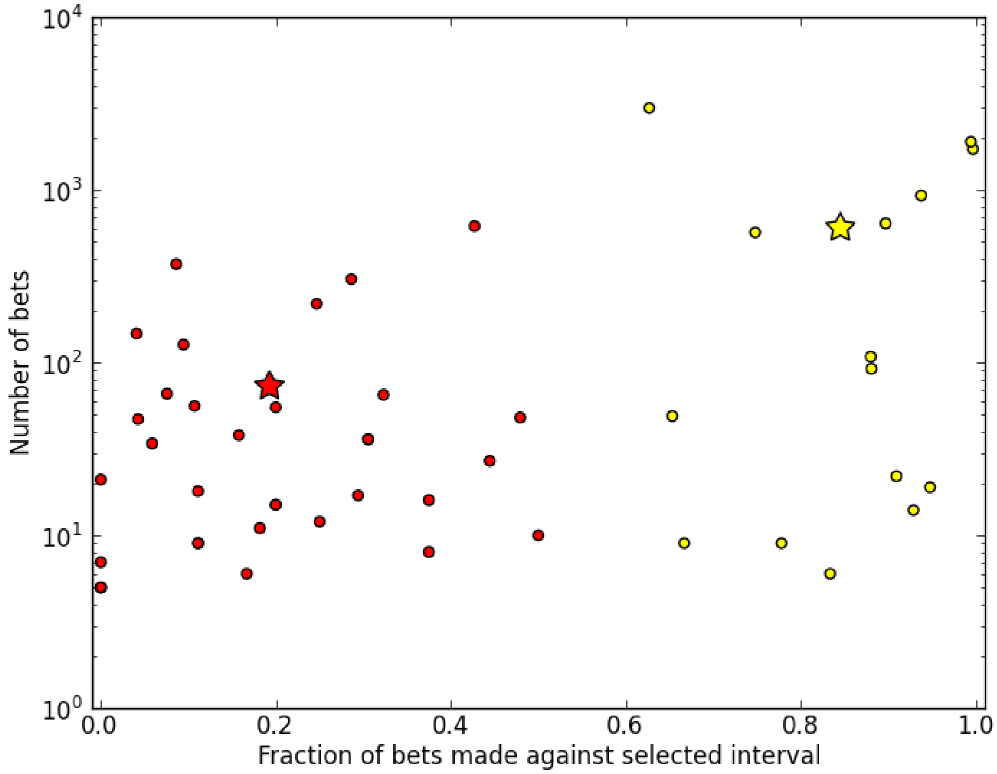


Figure 4.3: A cluster analysis of profitable traders. The y axis is log-scaled.

The large stars shows the centroid of each group, suggesting two groups of traders: one that made a large number of negative bets, and another that made roughly an order of magnitude fewer bets, most of them positive.

The negative and positive bets themselves were also quite different. The mean negative bet was over 16.3 days and the median negative bet was on just 1 day. The mean positive bet was over 20.3 days and the median positive bet was on 4 days. (Here, the means are much larger than the medians because the samples were skewed by large outliers.) So the negative bets were over much smaller intervals than positive bets.

This analysis suggests that successful traders followed one of two distinct strategies. One group of traders made large numbers of negative bets over very small intervals; and the other made smaller numbers of positive bets on larger intervals. We call the former strategy *spike dampening* and the latter strategy *relative smoothing*. We proceed to discuss these strategies in the context of interviews with key traders we conducted immediately after the market concluded.

4.3.2 Spike dampening

Recall that in Section 4.2.1 we discussed how and why the observed market prices were spiky, with unjustifiably large variance in the prices of putatively similar individual days. Several successful traders based their strategies entirely around betting against spikes. Rob, a PhD candidate in the Computer Science Department, ended with about 256 tickets and finished in fourth place overall. In our interview with him, he described his strategy as follows:

> *I knew that the market was presumably figuring out the probabilities of events, and early on, those predictions were* very *uneven. I supposed some people were setting all their money down on a single day or small set of days, and that this was causing the probability graph to be very "spiky." I bet against the spikes.*
>
> *Presuming (and I was correct) that as new people entered the market, the spikes would change radically and I'd cash out on the old spikes (making money) and bet against the new spikes.*

Of course, on the other side of Rob's actions were traders like Jeff (a pseudonym). Jeff is another PhD student in the Computer Science Department with a background in finance; he worked as a quantitative analyst at a hedge fund before coming to graduate school. A frequent trader, Jeff finished with enough tickets to place himself in the top 15 traders. Of his experience, he said:

> *It seemed like every time I would make a trade the value [of the bet] would fall a little bit...it was frustrating, like everything I was doing was wrong.*

Jeff's bets would fall in value because they would create spikes, which speculators like Rob would quickly sell.

4.3.3 Relative smoothing

While spike dampening involves betting *against* spikes, relative smoothing involved betting *for* undervalued spans. Alex, a Mathematics PhD student that was one of the market's top traders, explained his strategy as follows:

> *My main strategy was to purchase time periods that were given <10% opening probability by the market yet I felt had at least a 25% or 30% chance of containing the opening date.*
>
> *Most of my gains came as the market started to believe in the summer opening. After moving up to the top five and then into the top two on the leaderboard, I was actually pretty conservative. I sold many of my successful positions for hefty gains and sat on a lot of tickets. At this point, I didn't feel like I had a lot more information that the market so I made small opportunistic trades but held on to most of the gains.*

Brian, the market's top trader, employed both spike dampening as well as relative smoothing. He said:

> *I was mostly just betting against [spikes], but I'd also bet for anything that was abnormally low probability...I was kind of like a regularizer.*

One example Brian gave of an opportunity like this was a series of three weeks, the first and third of which have higher probability than the middle one. Brian would then buy the middle span, wait for the prices to converge, and then sell at a profit.

The strategy employed by the trading bot that made the majority of the market's trades (further discussed in Section 4.4.4) was also designed to smooth market prices. That bot attempted to fit a mixture of Gaussian distributions to the current prices, and trade off of deviations from the fit.

4.3.4 Information gathering

One strategy that was not identified in the cluster analysis but emerged from interviews with traders was *information gathering*. Consider that both spike dampening and relative smoothing are contrarian strategies—betting against prevailing price trends and waiting for further price movements to validate their trades. Because these strategies are both based around the relative prices of various spans (without regard to what the days actually represented), they can be considered *technical* trading strategies. In contrast, information gathering attempts to find the actual values associated with spans.

Elie, a PhD student in the Computer Science Department, invested a great deal of effort in finding out the real opening day. He became a regular at the construction site, getting updates on progress through discussions with the lead foreman, and even got the cell phone number of the building inspector who issued the temporary occupancy permit. On the morning of August 7th, the correct opening day, he visited the work site and got confirmation that the inspector had indeed signed off on the occupancy permit. He pushed the price of August 7th very high, to about 50% of the probability mass, and tried to get spike-dampening traders to move against him so he could extract additional profit. When, instead, this high price held, he drove the price up to almost

100% of the probability mass. Elie was the trader that ended up with the highest Information Addition Ratio (see Section 4.4.5), a measure of how much traders raised the price of the correct opening day. One of the desirable properties of markets is that they reward traders that acquire good information, and indeed information acquisition was extremely profitable for Elie. In the last several weeks of trading, he moved from being in around 100th place to finish as the fifth-ranked trader in the market.

4.4 Analysis of empirical market performance

The large data set of trades linked to user accounts is a valuable product of the GHPM. In this section, we use this data set to explore questions related to trader behavior and performance, and its impact on prices and information aggregation.

4.4.1 Official communications and the GHPM

The market had a complicated reaction to official communications about the opening day. We provide evidence that the market reacted to official communications, but that it correctly anticipated an officially unexpected delay in opening. (Recall that the GHPM was designed to test existing automated market makers over large event spaces, large numbers of participants, and long market durations. It was *not* designed to out-predict technical project completion forecasts from experts.)

Figure 4.4 shows how the distribution of prices changed over time and Table 4.1 shows the officially communicated moving dates. As we explained earlier, the moving day provides an upper bound on the issuance of the occupancy permit because people are not allowed to move into a building without a permit.

Date of Communication	Moving Day	Medium
October 15, 2008	July	Blog
February 14, 2009	August 3rd	E-mail
July 23, 2009	August 3rd	E-mail
July 28, 2009	Approximately August 10th	E-mail

Table 4.1: Officially communicated moving dates.

We can provide a rough narrative of the market from these two sources. Following some initial skepticism, market prices moved towards the correct prediction, becoming very prescient by the end of November. By then, the exterior framing of the buildings was complete. Over the next several

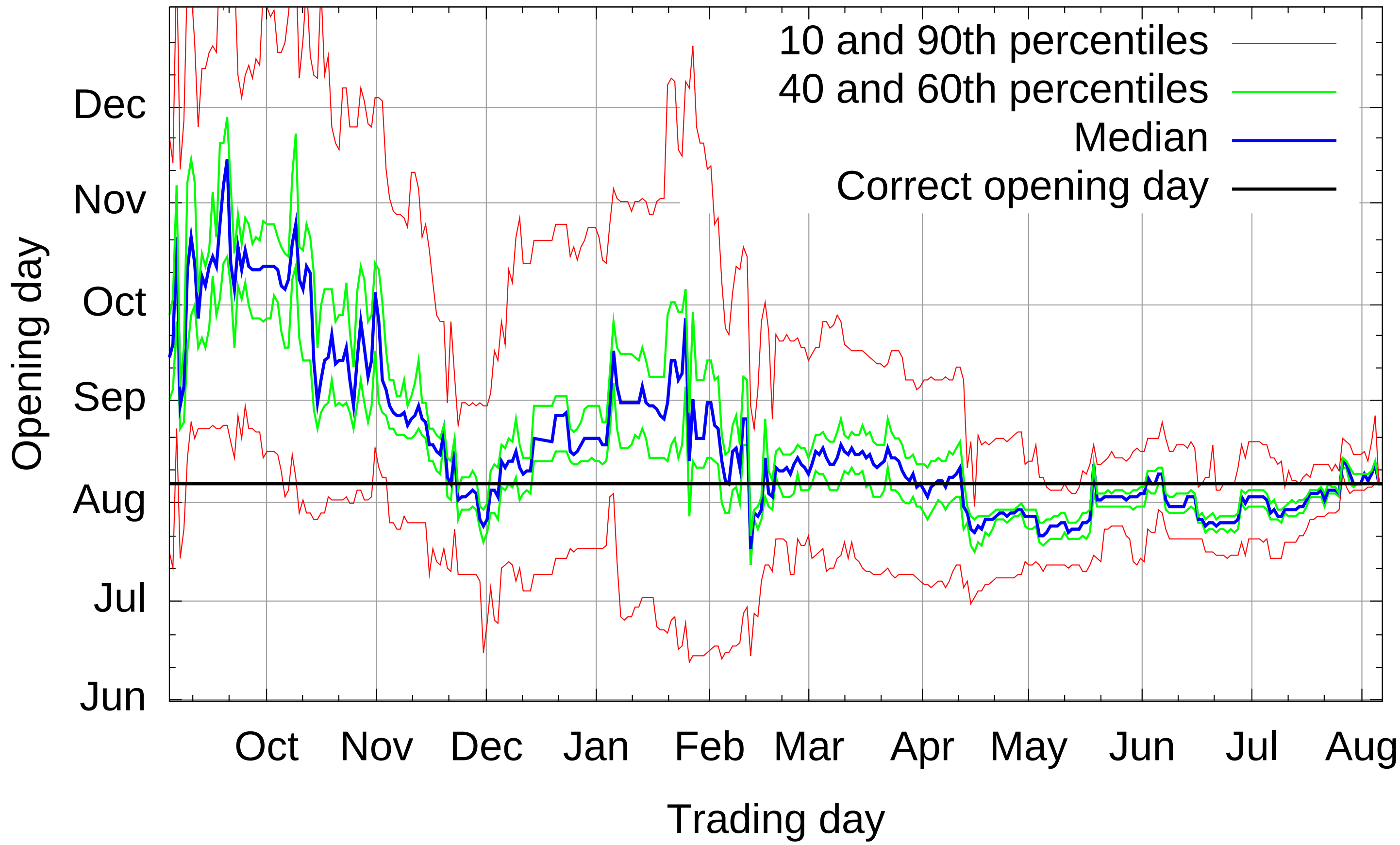


Figure 4.4: The set of prices offered by the market maker corresponds to a probability distribution. This figure shows the percentile curves of the probability distribution. The x-axis ranges over the trading days of the market, while the y-axis ranges over the possible opening days (the contracts in the market). The market actually spanned opening dates from April 1, 2009 to March 31, 2010; the y-axis is truncated here for clarity.

months, the outside appearance of the buildings did not improve measurably, and there were no official communications during this period. Prices reflected this seeming lack of progress.

The weather may have further reinforced traders' beliefs in a delay; the winter of 2008-09 was particularly cold in Pittsburgh and featured the lowest temperatures in fifteen years. Pittsburgh's average temperature in January 2009 was 22 degrees, six degrees colder than the historic average of 28 degrees. As Figure 4.5 shows, the market's probabilities for the building opening in the correct span peaked in late November and fell throughout December and January.

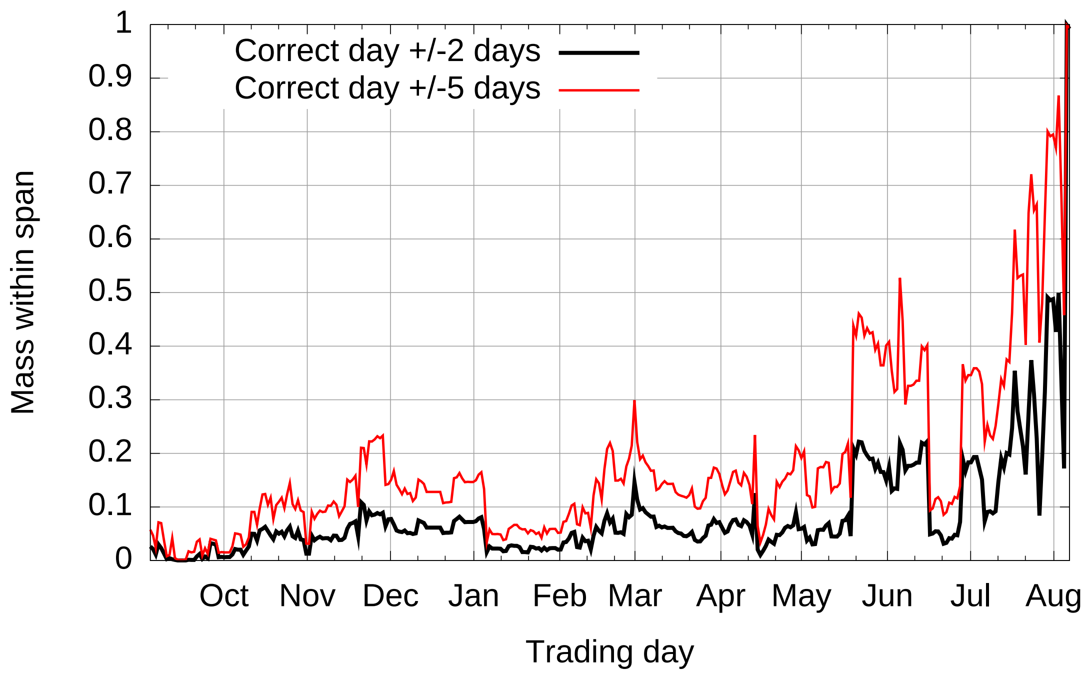


Figure 4.5: The amount of probability mass around the correct opening day. The x-axis ranges over days the market was open. The lines indicate the mass of spans around the opening day on each trading day; upper line is for August 2 to 12, the lower line for August 5 to 9.

Did market prices anticipate or lag public disclosures? To test this, we simulated the performance of a trader with inside information of the public announcements. (Recall that the members of the building committee that made these announcements were not allowed to trade in the GHPM.) How well would the official-information trader have done?

We considered two different schemes for how such an inside trader could operate:

- In the *Sell-quickly* scheme, the official trader spends some fraction of his wealth the day before making a public announcement, and then sells it the day after making the announcement.

- In the *Buy-and-hold* scheme, the official trader spends some fraction of his wealth the day before making a public announcement, and holds that position until the day before the next announcement is made (in the case of the final announcement, we assume he holds the position until the building opens).

There is some ambiguity inherent in the official statements. We interpreted "July" as meaning the whole month of July, and we interpreted "Approximately August 10th" as meaning the range August 7–13th, a one week range centered on August 10th. Observe that this span includes the correct opening day, August 7th.

The returns from both the Buy-and-hold and the Sell-quickly schemes depend on the fraction of wealth invested at each announcement. Figure 4.6 shows the final wealth of the inside-information trader for each scheme. Both schemes finish with more than their 20 initially allotted tickets as long as the fraction of wealth invested with each announcement is positive.

Our simulation suggests that the inside-information trader would have profited from his better information by participating in the GHPM. Qualitatively, the official trader out-performed the market. As a counterfactual, however, our simulation could be wrong quantitatively. Specifically, our simulation is probably optimistically biased towards the returns of the official communications trader. Had the official trader actually participated in the markets, he would have driven up the prices of his desired spans, and so other traders probably would have been less inclined to drive up the prices of those spans even more. However, it is also possible trend-following traders might have driven the prices on those spans up even higher. This would result in higher quantitative returns for the inside-information trader.

Looking at the returns made by the schemes before each announcement helps us contextualize how much each announcement was anticipated. Table 4.2 displays this information, both in terms of total tickets after each action and percent return. The data in the table comes from each trader investing all of their wealth before each announcement, and so the final row of Table 4.2 corresponds to the rightmost values in Figure 4.6.

Communication date	Sell-quickly	% Return	Buy-and-hold	% Return
October 15, 2008	23.83	19	55.40	177
February 14, 2009	43.33	82	81.80	48
July 23, 2009	68.17	57	21.80	-73
July 28, 2009	71.16	4	40.39	85

Table 4.2: Tickets and percent returns from following the two strategies after each announcement, where each agent invests all of his wealth in every action.

Positions in the Sell-quickly scheme are liquidated a day after the official announcement is made. Therefore, the returns from the Sell-quickly scheme indicate how much the market moved

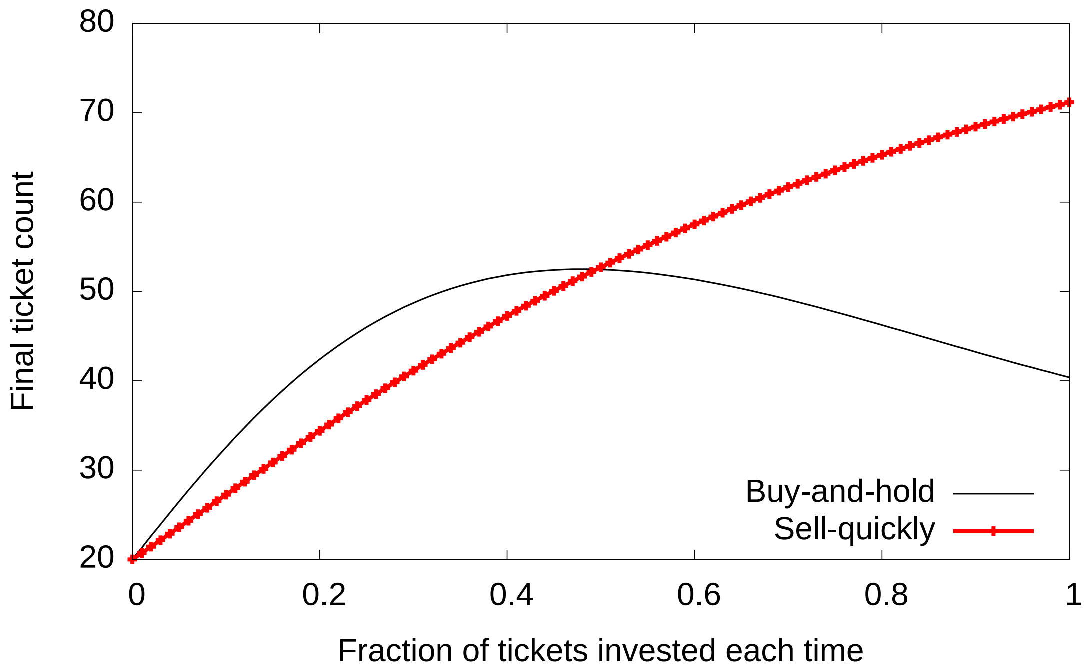


Figure 4.6: The simulated final returns of a trader with inside information of official communications, starting with 20 tickets. The trader either holds their position until immediately before making a new communication ("Buy-and-hold") or closes their position the day after making an official communication ("Sell-quickly").

in the short-term in response to the official information. Positions in the Buy-and-hold scheme are held until immediately before the next public communication. As a result, the returns from the Buy-and-hold scheme indicate how valuable each announcement was over the longer term.

Several of the values in the table merit further discussion.

- The highest-value trade in Table 4.2 is holding the "July" position from October 14th to February 13th. This produced a return of 177% for the Buy-and-hold scheme, nearly tripling the Buy-and-hold trader's tickets. We attribute the success of this position to the announcement being nearly correct at a very early stage of the market.
- The highest return from the Sell-quickly strategy was the 82% return from buying "August 3rd" on February 13th and selling it on February 15th. This exceptionally large return suggests that the market reacted quickly and dramatically to the announcement on February 14th.
- All the values in Table 4.2 are positive with the exception of the Buy-and-hold strategy buying "August 3rd" on July 22nd and selling the position on July 27th. In just five days, this position

> loses almost three-quarters of its value. This result, combined with the very modest returns of only 4% from the Sell-quickly trader's last action, suggests that the market anticipated the building would be delayed beyond August 3rd. However, the strongly positive returns (85%) for the last action of the Buy-and-hold trader suggests that the market was anticipating a much longer delay than actually occurred. So, even though the probability of the building opening on August 3rd had fallen in the five days between announcements, the probability mass did not shift to the correct day, August 7th, but rather to later in August. This is confirmed by observing the skew of the probability distribution in Figure 5.1 at the end of the market.

Both the Sell-quickly and the Buy-and-hold strategies produced positive earnings, which argues against the market fully anticipating every official communication. However, the losses of the Buy-and-hold strategy in the five days between denying and confirming a delayed opening suggest that the market correctly anticipated that the building would be delayed. However, the market appeared to anticipate a significantly longer delay than actually occurred.

4.4.2 Self-declared savviness

When traders signed up, they were asked "How savvy do you think you are relative to the average market participant?". They were given five choices, "Much less savvy", "Less savvy", "About the same" (the default selection), "More savvy", and "Much more savvy". Participants were informed that their answer to this question would not impact their payouts or the way they interacted with the market.

Because people are usually over-confident in various settings—and in prediction markets in particular (Forsythe et al., 1999; Graefe and Armstrong, 2008)—it was our expectation that traders would be over-confident in their own abilities relative to others. Instead, we found the opposite.

Reported under-confidence

Based on prior studies of over-confidence in markets, we would expect to see most traders rate themselves as at least comparable to the average trader in the market. Table 4.3 shows our survey results. 77 traders described themselves as less or much less savvy than average, while only 13 traders described themselves as more savvy than average. Surprisingly, not a single trader listed themselves as much more savvy than the average trader.

Why did we find traders under-confident, instead of over-confident, in their own abilities? Recent research by Moore and Healy (2008) on confidence sheds some light on this issue. They find that

Self-Declared Savviness	Number of Traders
Much Less than Average	30 (17.8%)
Less than Average	47 (27.8%)
Average	79 (46.7%)
More than Average	13 (7.7%)
Much More than Average	0

Table 4.3: Counts of self-assessed savviness.

> *On difficult tasks, people...mistakenly believe that they are worse than others; on easy tasks, people...mistakenly believe they are better than others.*

A novel market setting, such as the web-based automated market maker with span-based elicitation we used in the GHPM, is unfamiliar enough to a new trader as to seem potentially difficult. Prior market studies, because they have used traditional market interfaces that even the most casual participant is familiar with, would seem potentially less difficult and therefore would be susceptible to overconfidence.

Traders poorly predicted their own performance

We found that traders' self-reported savviness relative to other traders had little bearing on their relative performance. Table 4.4 groups traders by self-reported savviness and displays the group medians. The median over all traders was 17.46 tickets, identical to the least-savvy group and within a ticket of the two next-savvy groups. Ironically, traders identifying themselves as more savvy than the average trader had a median return more than 10 tickets lower than any other group.

Self-Declared Savviness	Median Tickets
Much Less than Average	17.46
Less than Average	16.78
Average	18.36
More than Average	6.05
Much More than Average	N/A

Table 4.4: Traders who self-identified as "more savvy than the average participant" in the market had dramatically lower median performance than other traders, while those traders identifying as "much less savvy than the average participant" had the same overall median performance as the trading population as a whole.

4.4.3 Trade frequencies suggest a power law

The numbers of bets made by traders appear to closely fit a power law distribution. Figure 4.7 shows the relationship in terms of the probability of a trader having more than a certain number of trades from our data set, and the best-fitting power law distribution. (We also tried a log-normal distribution and the fit was poor.)

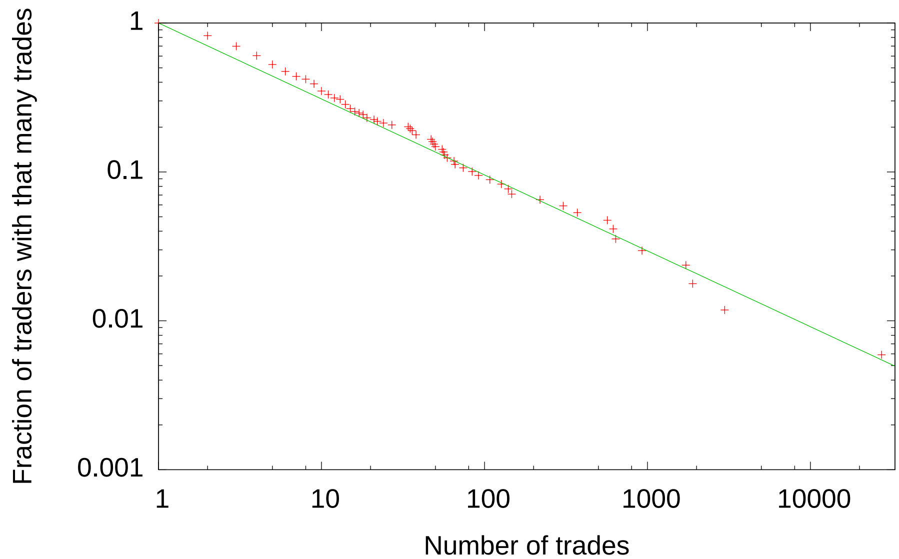


Figure 4.7: A log-log plot showing the relationship between traders and the number of trades they placed. The straight line shows a power law fit for $\alpha = .51$.

Unfortunately, with only 169 traders we can not assert an appropriate level of statistical significance, so we cannot rule out the data being generated by other distributions. However on a log-log plot the data do appear to snugly fit the canonical straight line of a power law distribution.

Why might one expect a power law distribution of trade frequency? It seems reasonable to suggest that a trader both makes new bets and sells old bets in proportion to the number of bets she has currently outstanding, with the constraint that she never go under one bet outstanding (in order to collect her two free tickets each week). As Mitzenmacher (2004) discusses, this type of generative model yields a power law distribution.

4.4.4 Trading by a bot

Conventionally, when we think about prediction markets, we think about a collection of individuals making probability judgments. This is a quality distinct from traditional exchanges, in which automated trading is common and frequent. But as Berg et al. (2001) discuss, trading bots make up a large fraction of the observed volume in the Iowa Electronic Markets (IEM), the longest-running prediction market, and must be considered in any sort of qualitative summary of the properties of prediction markets. We found that trade in the GHPM was also dominated by a bot.

This was surprising because we did not make automated trading easy. The GHPM did not use an API, so any trading bot would have to come up with a way to parse the web page and simulate its user's actions on the page. Jim, a graduate student in the Computer Science Department, took two days to write a trading bot. The bot fit the current prices to a mixture of Gaussian distributions and identified trading opportunities based on deviations from the fit.

The bot made 68.5% of the trades in the market (27,311 of 39,842). The median number of trades placed for all traders was five.

Jim's bot did well in the market; at its peak it was the second-highest-valued trader. Jim turned his bot off after the e-mail of February 14th and began trading manually. He ended up losing the bulk of his tickets by betting on the building opening earlier than it actually did, finishing 158th of the 210 registered users and 117th out of the 169 traders.

4.4.5 Trader-level data is consistent with Marginal Trader Hypothesis

Prediction markets have been demonstrated to be at least as accurate as, and in many cases more accurate than, predictive techniques like polls (Berg et al., 2001; Goel et al., 2010). How do markets incorporate information and generate good prices? A line of research from the administrators of the IEM has suggested that a small group of so-called *marginal traders* are responsible for producing accurate results in prediction markets (Forsythe et al., 1992, 1999; Berg et al., 2001; Oliven and Rietz, 2004). This group, which appears to be about 10% of the traders in the IEM, essentially arbitrage the remainder of the market participants. This theory of how markets function is called the *Marginal Trader Hypothesis*.

In this section, we show that the GHPM is consistent the Marginal Trader Hypothesis. To our knowledge this is the first experimental consistency of the Marginal Trader Hypothesis within a market that used an automated market maker. This finding is significant because using an automated market maker prevents less-sophisticated agents from making many of the errors that could be thought to drive the lopsided distribution of performance that the Marginal Trader Hypothesis predicts. We also argue that the GHPM rejects one of the most intriguing hypotheses of the literature: that marginal traders are disproportionately male.

Why isn't it easy to identify marginal traders?

The techniques from the IEM literature do not apply to our market setting. In Forsythe et al. (1992) marginal traders were identified by the particular kinds of trades they placed. Specifically, marginal traders were those traders that provided liquidity and set prices by placing limit orders (orders with a limiting price, e.g., "I will buy 10 contracts at a price-per-contract of no greater than 60 cents"), rather than market orders ("I will buy 10 contracts"). Put another way, marginal traders served as market makers. In later work, this connection was made more explicit. Oliven and Rietz (2004) classify traders as either marginal, price-setting, market makers, or as non-marginal price takers.

One of the key findings of this literature is that non-marginal traders violated the *law of one price*. To illustrate what the law of one price is, consider the example of the Red Sox and Yankees playing a baseball game with a traditional prediction market not equipped with an automated market maker. Imagine that the current order books on the events are given in Table 4.5.

Contract	Best bid	Best ask
Red Sox	.34	.40
Yankees	.63	.65

Table 4.5: Hypothetical prices in a baseball game prediction market.

Now consider two actions at these prices: buying a share of Red Sox stock, or selling a share of Yankees stock. If a trader buys a share of Red Sox stock, his payoffs are $\{0.6, -0.4\}$, where $\omega_1 =$ Red Sox win and $\omega_2 =$ Yankees win. If a trader sells a share of Yankees stock, his payoffs are $\{0.63, -0.37\}$. Observe that this payoff vector strictly dominates the payoff from buying a share of Red Sox stock—no matter whether the Red Sox or Yankees win the game, the trader receives a higher payoff from the latter action. Consequently, if a trader were to do the former action and buy a share of Red Sox stock, he is said to be violating the law of one price. More formally, violations of the law of one price occur when an agent takes on a strictly-worse payoff vector than one that could be constructed at the current market prices. Interestingly, even though violating the law of one price is a dominated action, the IEM literature has found that non-marginal traders perform these actions frequently (Oliven and Rietz, 2004).

In summary, the literature from the IEM gives us two trade-level ways of distinguishing between marginal and non-marginal traders: marginal traders place limit orders, and non-marginal traders violate the law of one price. Unfortunately, when using the LMSR, neither of these distinguishing characteristics are available to us. First, the LMSR *only* uses market orders, so agents cannot place limit orders. Therefore, traders cannot be price setters, only price takers. Second, the LMSR precludes any violation of the law of one price, because the LMSR maintains a probability distribution over the different states of the world. This probability distribution enforces the condition that equivalent bets for the set of states "A" and against the set of states "not A" have equivalent payoff vectors.

An example will be helpful in seeing why this is the case. Consider our example of the Red Sox-Yankees baseball game again. "Buying the Red Sox" is equivalent to taking on the contract

$\{1, 0\}$. "Selling the Yankees" is equivalent to taking on the contract $\{0, -1\}$. For the payoff vector associated with the former contract, we subtract the amount we must pay for the contract from the vector $\{1, 0\}$. For the payoff vector associated with the latter contract, we add the money we receive from selling the contract to the vector $\{0, -1\}$. We will show that for any initial $\mathbf{x}$ and any translation-invariant cost function C, the payoff vector formed by these two bets is equivalent.

The two bets are equivalent if

$$\{1, 0\} - (C(\mathbf{x} + \{1, 0\}) - C(\mathbf{x}))\,\mathbf{1} = -\,(C(\mathbf{x} + \{0, -1\}) - C(\mathbf{x}))\,\mathbf{1} + \{0, -1\}$$

By re-arranging, we see that this condition is

$$\begin{aligned}
(C(\mathbf{x} + \{1, 0\}) - C(\mathbf{x})) - (C(\mathbf{x} + \{0, -1\}) - C(\mathbf{x})) &= 1 \\
C(\mathbf{x} + \{1, 0\}) - C(\mathbf{x} + \{0, -1\}) &= 1 \\
C(\mathbf{x} + \{1, 0\}) &= C(\mathbf{x} + \{0, -1\}) + 1 \\
C(\mathbf{x} + \{1, 0\}) &= C(\mathbf{x} + \{0, -1\} + \{1, 1\}) \\
C(\mathbf{x} + \{1, 0\}) &= C(\mathbf{x} + \{1, 0\})
\end{aligned}$$

Here the penultimate step relies on the translation invariance of the cost function (i.e., $C(\mathbf{x}) + 1 = C(\mathbf{x} + \mathbf{1})$).

Given that we cannot identify marginal traders in the ways suggested by prior literature, in the next section we discuss how we employed the spirit of the original work of the IEM to select a candidate set of marginal traders.

Identifying marginal traders in the GHPM

The IEM researchers made their criteria for selected marginal traders on the basis of informative trading patterns. They then showed that their set of marginal traders had better performance and more frequent trade than their set of non-marginal traders. Here, we take a similar approach. We start by describing a way of gauging the information content of the bets a trader makes, and how this metric produces an intuitive way to isolate a set of marginal traders.

We dub the metric we used to assess the information content of a trader's bets the *Information Addition Ratio (IAR)*. This measure attempts to separate a trader's return from speculative activities from a trader's return from information-adding activities. It answers the question "If we see a trader making a one-ticket bet, what is her expected return if she were to hold that bet until the market closes?". A return of one ticket on each ticket invested is always available to a trader by betting on the entire range of exhaustive contracts. Traders who inject valuable information into the market will have an IAR greater than one, while traders who have a deleterious impact on information will have an IAR of less than one. Essentially, IAR measures how much each trader increased the price of August 7th, the correct opening day. IAR is an attempt to compress a complex concept into a scalar, and such an enormous dimensionality reduction is inherently lossy. IAR places a focus exclusively on rewarding traders for making bets that raised the price of the correct opening day.

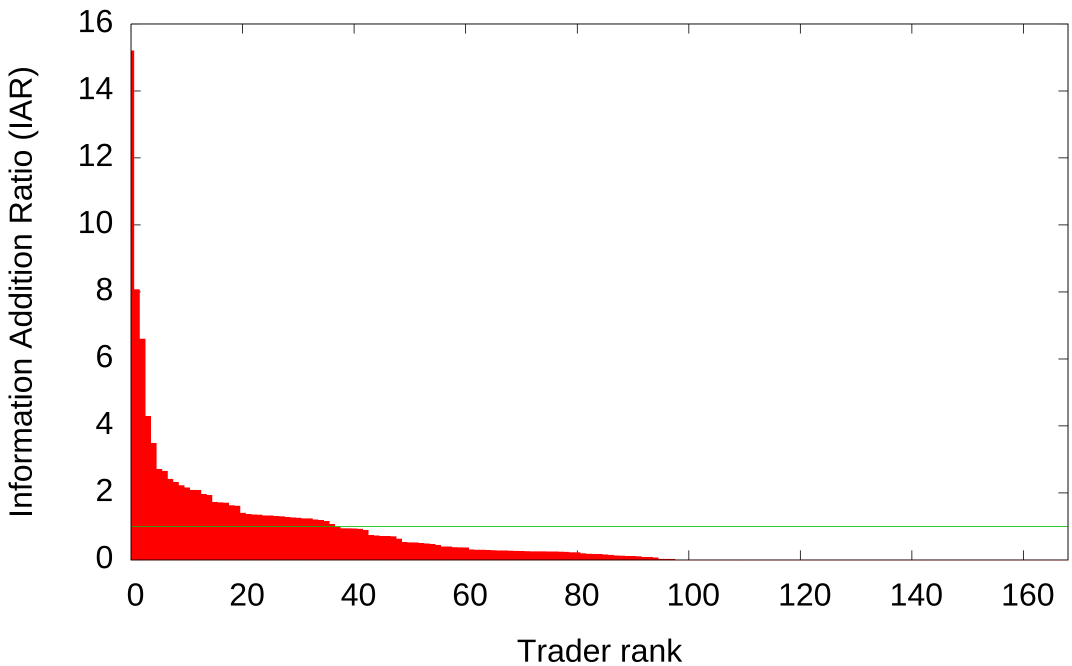


Figure 4.8: The distribution of IAR of traders ordered by rank. The straight line shows an IAR of 1 (one ticket expected per ticket wagered).

Figure 4.8 displays the distribution of IARs. It is heavily skewed and unequal. The median trader had a return of .16 tickets per ticket bet, and 79 traders (47%) placed all their bets on losing intervals, i.e., spans that did not include the correct date, August 7th. This is surprising for several reasons. First, a return of one ticket per ticket bet was *always* available to traders by betting on the entire span. Second, in the ternary elicitation interface, one of the bets offered will always include August 7th, and the other will not, so there was no inherent bias against traders making correct bets. Finally, the median number of bets per trader was five, meaning that the bulk of traders made poor judgements several times, not just once.

We can gauge how skewed a distribution is at a glance by measuring its Gini coefficient, a standard measure of inequality. Assuming we have the data points $x_1 \leq x_2 \leq \ldots \leq x_n$, the Gini coefficient, G, of the sample is given by

$$G = \frac{2\sum_i i x_i}{n\sum_i x_i} - \frac{n+1}{n}$$

The Gini coefficient ranges from zero to one and can be thought of as a measure of how unequal drawn samples are, with particular sensitivity to large outliers. Table 4.6 displays the Gini coefficients for the GHPM in context with other distributions. The Gini coefficients of both final tickets and IARs is very high, which indicates the underlying distributions are skewed. Taken as a whole, these results indicate that the majority of market participants consistently made judgments that hurt the accuracy of the GHPM.

Data Set	Gini Coefficient
Normal Distribution $\mu = 5, \sigma = 1$	.113
Denmark Income	.247
Uniform Distribution	.333
United States Income	.408
Log-normal Distribution $\mu = 5, \sigma = 1$	.521
GHPM Tickets	**.700**
Namibia Income	.743
GHPM IARs	**.762**

Table 4.6: Gini coefficients are a standard measure of the degree of inequality of a distribution. As this table shows, the distribution of both information (IARs) as well as overall performance (tickets) were extremely unequal. For reference, we include country income inequality coefficients from the United Nations (2008); Denmark had the lowest coefficient and Namibia the highest.

On the other side, there was clearly a small and select group of traders responsible for actually making the GHPM produce meaningful prices. Only 37 traders (22%) had an IAR of more than one, and only 13 traders (8%) had an IAR of more than two. A trader with an IAR of greater than two was making sophisticated judgements to bet correctly on less than half of the market's probability mass. A trader could achieve an IAR of more than one by making not-very-nuanced bets with the market maker (e.g., by betting on all but the earliest day), but there is no way a trader could have an IAR of more than two without making nuanced judgements. Consequently, we advance these 13 traders as our candidate set of marginal traders.

Comparison of marginal and non-marginal traders

In this section, we will compare the group we selected as candidate marginal traders to the group we selected as non-marginal traders. We show that, just like in the IEM literature, the performance and level of involvement of the marginal traders was much higher than the non-marginal traders. Thus, we argue that the GHPM is consistent with the Marginal Trader Hypothesis.

Table 4.7, taken from Forsythe et al. (1992), displays the performance of traders in the 1988 IEM presidential election market. Their set of marginal traders were about ten percent of participants, and the marginal traders had a much higher percent return and invested much more in the market than the non-marginal traders.

	IEM marginal	IEM non-marginal
Fraction of traders	.11	.89
Median percent return	9.6	0.0
Average total investment	56	21

Table 4.7: The performance of marginal and non-marginal traders in the 1988 IEM presidential market (Forsythe et al., 1992)).

Table 4.8 displays the performance of our candidate set of marginal traders relative to non-marginal traders in the GHPM. Just like in the IEM, about 10% of traders are classified as marginal. Additionally, the marginal traders in the GHPM were a much more active presence than non-marginal traders. The median marginal trader made almost 30 times more bets than the median non-marginal trader.

Finally, just like in the IEM, the performance of the marginal traders was much better than the performance of the non-marginal traders. However, whereas the median marginal trader in the IEM had a rate of return of 9.6%, the median marginal trader in the GHPM had a rate of return of almost 300%. We attribute this discrepancy to the weekly ticket handout, as well as the subsidy given up by the market maker (recall that the LMSR runs at a loss). We conjecture that had the IEM awarded traders for participation and had the IEM used a market making agent that subsidized traders, then most of these rewards would have gone to the marginal traders because they were the most active participants.

	GHPM marginal	GHPM non-marginal
Fraction of traders	.08	.92
Median tickets	79.2	16.3
Median trades	147	5

Table 4.8: The relative performance of our candidate set of marginal traders in the GHPM.

One possible criticism is that the results of this section were a foregone conclusion, and that our criteria for classifying a trader as marginal necessitated that they end up being an active, profitable participant. But just like in Forsythe et al. (1992), we used trade-level attributes to decide whether or not a trader would be classified as marginal, and there was no guarantee that a trader selected as marginal would end up with the most tickets or the most trades. It is easy to imagine, in fact, traders who could qualify as marginal by our definition that would not be active or involved participants; such a trader could merely have placed a single small bet on the market opening in July, August, or September soon after the market's initiation. While this bet would have been smart, it would have produced an IAR of greater than two for that trader without significant involvement in the market and without that trader amassing a large number of tickets. Furthermore, IAR need not have any correspondence to a trader's actual returns from the bets they placed; if they were to sell their correct bets before the market's expiration, they would earn fewer tickets or, depending on short-term prevailing market prices, could even lose tickets on those bets.

However, our results show that this was not the case, and that traders whose bets raised the price of the correct opening day tended to be by far the most profitable and active traders in the GHPM as a whole. In terms of tickets, our set of marginal traders included the top trader, four of the top five, and six of the top ten. In terms of trades, our set of marginal traders included the top trader, four of the top five, and seven of the top ten.

Are only men marginal traders?

One of the most curious findings of Forsythe et al. (1992) was that their pool of marginal traders was *exclusively male*. In contrast, our pool of marginal traders was not statistically significantly different in gender composition from our pool of traders as a whole. Table 4.9 compares the results of the two prediction markets. (These results are not an artifact of using IAR to select marginal traders, because women were also well-represented under alternative selection criteria. By final ticket count, three of the top ten traders were women. By number of trades, two of the top ten traders were women.)

	IEM marginal	IEM non-marginal	GHPM marginal	GHPM non-marginal
Number	22	170	13	156
Fraction Male	1.00	0.68	0.77	0.74

Table 4.9: The gender composition of the marginal trader pools in our study and in the IEM (as reported in Forsythe et al. (1992)).

For the null hypothesis that the GHPM marginal and non-marginal traders are drawn from the same gender distribution, we get a p-value of 0.60. Since $p > 0.05$, this means it is not statistically significant to reject the null hypothesis.

Consequently, we reject the hypothesis that the gender composition of the marginal traders differs from the gender composition of non-marginal traders. Since there does not seem to be any causal reason that women should be worse traders than men, and Forsythe et al. (1992) do not justify any mechanism for their gender findings, we contend that the hypothesis of women being less likely to be marginal traders should probably be broadly dismissed, or at the least, merits further study before acceptance.

4.5 Extensions

Our experience with the liquidity insensitivity that led to substantial price volatility in the mature stages of the GHPM motivates Chapters 5 and 6. In both of those chapters we develop liquidity-sensitive market makers that make prices "stiffer" (i.e., the price changes less as a function of the amount that is bet) in markets where lots of trade volume has been observed.

The GHPM featured a new interaction interface that enabled participation by both sophisticated and unsophisticated users over a very large event space. Perhaps other interaction interfaces could be developed that lead traders to place bets that might better reflect their beliefs while still being simple enough for unsophisticated users. One suggestion would be to have users wager on payout vectors, with the default being an equal payout on each day for simplicity, as in the GHPM. This could be a way to avoid the spikiness in prices observed in the market, as presumably traders would not consent directly to very spiky prices.

Chapter 5

Extensions to convex risk measures

Recall from Chapter 3 that a convex risk measure is a cost function that satisfies monotonicity, convexity, and translation invariance. In this chapter, we introduce two separate sets of extensions to convex risk measures. First, we relax the setting from a finite discrete event space to either a countably or uncountably infinite event space and describe how constant-utility cost functions can be used to create market makers that satisfy several desirable properties over these much larger event spaces. Second, we describe a new family of market makers (for the standard setting) that qualitatively replicate the behavior of real human-mediated market makers, while still retaining the desirable bounded loss property of the algorithmic agents of the prediction market literature. Recall that convex risk measures do not have the property of positive homogeneity, and lack the simple recourse to expanding liquidity that homogeneity provides. The market makers in Section 5.2 possess alternative means to expand liquidity and are the most realistic trading agents derived from convex risk measures in the literature.

5.1 Market making over infinite spaces

What makes infinite event spaces challenging is the tension between loss boundedness and sometimes offering traders bets that would be irrational to accept regardless of what event materializes. It is easy to create market makers with bounded loss if they charge agents as much as they could possibly win. On the other hand, it is also easy to create market makers that only offer bets a rational agent could accept, but have unbounded loss. For instance, the most popular automated market maker in practice is the LMSR, which for n events has worst-case loss $\Theta(\log n)$. Over infinite event spaces, however, it has unbounded loss (Gao et al., 2009).

Because of this tension, it has been an open problem whether there can exist market makers with bounded loss over infinite event spaces that never offer agents bets which are immediately bad.

In this section, we construct such market makers, first for countably infinite event spaces, and then for uncountably infinite event spaces. We begin by first discussing prior work on large and infinite event spaces.

5.1.1 Prior work

The difficulty of constructing a market maker with both bounded loss and bets which are not immediately bad is informed by the impossibility result of Gao and Chen (2010). That work concerned automated market makers operating over continuous spaces (i.e., the unit interval). The result of Gao and Chen (2010) is that a convex risk measure over this space that satisfies a certain continuity-monotonicity condition cannot have bounded loss. This result would, at first glance, seem to explicitly preclude the existence of the market makers we construct in this chapter.

However, the market makers in this section subvert this impossibility result in a technical, but natural, way. Consider a bet of arbitrary size over some interval (A, B) with $A < B$, and let $\pi(A, B)$ denote the cost of the bet to the trader. The continuity-monotonicity condition of Gao and Chen (2010) has two parts:

1. That decreasing the size of the betting interval results in a continuous decrease in the cost of the bet.
2. That the limit of this process is a zero-cost bet. Formally, $\lim_{A \to B} \pi(A, B) = 0$.

The market makers we develop in this section satisfy the first property, but not the second, and so circumvent the impossibility result. Observe that the first condition is more natural than the second, because a market maker satisfying the second condition would price arbitrarily large bets over sufficiently small intervals at epsilon cost to a trader.

It is simple to construct a market maker with bounded loss over infinite event spaces. The sup cost function (which can be thought of as an infinite-dimensional max)

$$C(\mathbf{x}) = \sup_{\omega} \mathbf{x}(\omega)$$

accomplishes this goal. This is easy to see because a trader is charged as much as she could possibly gain from the realization of any event. However, sup is one example of a cost function that *offers bets which are immediately bad*, meaning that it offers some bets that could never benefit a rational agent. (We will formalize this property in each of the next two sections.)

Agrawal et al. (2009) present several convex risk measures with a worst-case loss of $\Theta(1 - 1/n)$, which implies bounded loss as n gets large. However, they achieve their bounded worst-case loss by offering good bets only up to a certain amount of uncovered exposure. After reaching this limit (which is controlled by a parameter set *a priori* by the market administrator) the price of some events is set to zero, implying that the other events form a proper subset with unit marginal price. A bet on this subset would then be immediately bad. (Furthermore, the market makers in Agrawal et al. (2009) rely on solving a relatively complex convex optimization, so it is not immediately clear how to generalize their technique over infinite-dimensional spaces, or whether such a generalization is even possible.)

5.1.2 Countable event spaces

In this section, we consider how to take bets on a countably large event space $\{\omega_1, \omega_2, \ldots\}$ with both bounded loss and without offering bets which are immediately bad. We begin by discussing the changes in framework necessary to move from a discrete, finite event space to a discrete, infinite event space.

Translation of qualities to countable spaces

Unlike in the finite case, here we can no longer take *arbitrary* bets over the event space. This is because degeneracies can arise with respect to the sums of infinite series that can lead the market maker to have undefined expected utility from bets. For example, consider a market maker with beliefs $p_i = 1/2^i$ employing a constant-utility cost function to price the bet $x_i = (-3)^i$. The market maker's expected utility for accepting this bet at cost c is

$$\sum_i \frac{1}{2^i} \left(c - (-3)^i\right)$$

which does not converge for any c, and so the payout vector $x_i = (-3)^i$ has undefined cost.

To avoid these infinite summation problems, we allow traders to only make bets that correspond to *valid* payout vectors.

Definition 11. A payout vector $\mathbf{x}$ is *valid* for a constant-utility cost function having utility function u and a belief distribution $\mathbf{p}$ if there exists a cost for $\mathbf{x}$ for all initial utility levels. Formally, for all $x' \in \mathbf{dom}\ u$ there exists a c such that

$$\sum_i p_i u(c - x_i) = u(x')$$

Here, x' is an initial amount of wealth and so $u(x')$ is the market maker's initial (constant) utility level.

We can also formalize the notion of immediately bad bets.

Definition 12. A differentiable convex risk measure C *offers bets which are immediately bad* if there exists an event i with zero marginal price.

The reasoning here is that, because C is a convex risk measure and an event has 0 marginal price, a bet on the rest of the events (a proper subset of the event space) sums to 1. Then, as the name implies, trade with a cost function of this sort can sometimes be a dominated action: a trader pays at least as much as she could possibly make from any contract, but there also exists at least one event for which the trader loses her bet.

Bounded loss and no bad bets over countable spaces

With the theoretical framework in place for fielding bets over a countably large event space, we can present our main result of this section.

Proposition 9. *Let u be a differentiable barrier utility function, let the market maker have belief distribution $\mathbf{p}$ with $p_i > 0$, and let $x^0 > 0$. Then for every valid payout vector $\mathbf{x}$, the constant-utility cost function given by the solution to*

$$\sum_{i \in \{1,2,\ldots\}} p_i u\left(C(\mathbf{x}) - x_i\right) = u(x^0)$$

loses at most x^0 and never offers bets which are immediately bad.

Proof. To prove the specified loss bound, suppose that the market maker lost more than x^0 when event ω_i occurred. Then $x_i + x^0 - C(\mathbf{x}) > x^0$, and so by re-arranging, $C(\mathbf{x}) - x_i < 0$. Therefore since u is a barrier utility function, we have $u(C(\mathbf{x}) - x_i) = -\infty$. But since the $p_i > 0$, this means the expected value of the market maker's utility is $-\infty$, rather than $u(x^0) > -\infty$, a contradiction.

To prove the resulting cost function never offers bets which are immediately bad, we must show that for every valid payout vector $\mathbf{x}$,

$$p_i \nabla_i C(\mathbf{x}) > 0$$

for which (because $p_i > 0$) it suffices to show that $\nabla_i C(\mathbf{x}) > 0$. Recall from Chapter 3 that the equation for the gradient of a constant-utility cost function is

$$\nabla_i C(\mathbf{x}) = \frac{p_i u'(C(\mathbf{x}) - x_i)}{\sum_j p_j u'(C(\mathbf{x}) - x_j)}$$

Because the utility function is strictly increasing and differentiable, its derivative is strictly positive. Coupling this with the fact that $p_i > 0$, and we have that the terms in the integrand of both the numerator and denominator are positive. Thus $\nabla_i C(\mathbf{x}) > 0$ and so the cost function never offers bets which are immediately bad. ■

Proposition 9 relies on the existence of a belief distribution $\mathbf{p}$ such that $p_i > 0$. One distribution of this sort is to assign $p_i = 1/2^i$.

5.1.3 Uncountable event spaces

In this section, we consider how to take bets on the unit interval $[0, 1]$. At first glance this might seem like a very restricted domain of inquiry. However, traditional notions of *dimensionality* do not strictly apply to the spaces where probability measures live. In fact, atomless probability distributions over the unit interval are isomorphic to atomless probability measures over *any* measurable space that satisfies mild technical conditions (Malliavin, 1995, Section 6.4). Consequently, the results of this section also extend to settings like the real number line and geometric shapes, with the caveat that the notion of a *sub-interval* that we use in the results of this section needs to be translated to match the dimension of the space under consideration. For instance, in making markets over a rectangle, a sub-interval could translate to be a proper sub-rectangle with corners (x_1, y_1) and (x_2, y_2), and in making markets over the surface of a sphere, a sub-interval could consist of a rectangular or triangular patch on the surface.

Translation of qualities to the unit interval

On the unit interval the events are $\omega \in [0, 1]$, and bets still map from events to how much the trader gets paid based on what event materializes. However, bets and payout vectors (which were represented as points in $\mathbb{R}^n$ in the finite case) are now functions over the event space; formally, $\mathbf{x} : [0, 1] \mapsto \mathbb{R}$.

For example, we might have a prediction market for forecasting where on Florida's coastline the next hurricane will hit. The events ω could then be represented as points over the unit interval (i.e., with Pensacola to the west and Jacksonville to the east representing the interval's endpoints). An example bet a trader might make is one that demands a uniform payoff over the coastline. Another potential bet could pay off as a triangle distribution centered around Miami. Both of these bets $\mathbf{x}$ can be expressed as functions over the interval.

Cost functions now generalize to be *functionals* which map these functions $\mathbf{x}$ to scalar values, i.e., $C : ([0, 1] \mapsto \mathbb{R}) \mapsto \mathbb{R}$. Furthermore, instead of the market maker having subjective probability distribution $\mathbf{p}$ over the event space, the market maker instead uses some probability distribution F, with some corresponding density f.

Similar to the countable case considered in Section 5.1.2, we cannot field arbitrary bets from traders. Even over the unit interval, degeneracies arise with respect to pricing arbitrary bets that can lead the market maker to have undefined expected utility. For instance, suppose a market maker had utility function $u(x) = \log x$ and uniform beliefs over the interval, and was attempting to calculate the cost for the payout vector $\mathbf{x}(\omega) = \sin(1/\omega)2^{1/\omega}$. Observe that this payout vector oscillates wildly near 0. Since for any c the expected utility integral

$$\int_0^1 \log\left(c - \sin(1/\omega)2^{1/\omega}\right)\, d\omega$$

diverges, there is no cost that can be associated with this payout vector. To ensure that costs are always well defined, we allow traders to only make bets that correspond to *valid* payout vectors.

Definition 13. A payout vector $\mathbf{x}$ is *valid* for a constant-utility cost function using utility function u and a belief distribution F if, at all initial utility levels, there exists a cost for $\mathbf{x}$ that is well defined. Formally, for all $x' \in \mathbf{dom}\, u$ there exists a c such that

$$\int_0^1 u(c - \mathbf{x}(\omega))\, dF(\omega) = u(x')$$

Here, x' is an initial amount of wealth and so $u(x')$ is the market maker's initial (constant) utility level.

Observe that all bounded and measurable (and therefore continuous, because continuous functions over compact intervals using a Borel sigma-algebra are bounded and measurable) payout vectors are valid. This is because for bounded payout vectors there exists a C such that for all $c > C$ and all $\omega \in [0, 1]$, $c - \mathbf{x}(\omega) > 0$. This implies that the integral is finite and can be adjusted to $u(x^0)$ by selecting the appropriate cost.

Recall that over discrete, finite spaces, the market maker solves for the cost $C(\mathbf{x})$ of a payout $\mathbf{x}$ implicitly through

$$\sum_i p_i u(C(\mathbf{x}) - x_i) = u(x^0)$$

Over the unit interval this equation becomes the more general

$$\int_0^1 u\left(C(\mathbf{x}) - \mathbf{x}(\omega)\right)\, dF(\omega) = u(x^0)$$

for every valid payout vector, where once again the terms on the left represent the market maker's (subjective) expected utility, with subjective beliefs here expressed through the probability distribution F, and the term on the right representing the market maker's initial utility.

Observe that monotonicity carried over from the finite case. Let $\mathbf{x}$ and $\mathbf{y}$ be valid and select some initial $u(x^0)$. Then if $\mathbf{x} \geq \mathbf{y}$, we have that $C(\mathbf{x}) \geq C(\mathbf{y})$, because the utility function is strictly increasing and at all $\omega \in [0, 1]$, $\mathbf{x}(\omega) \geq \mathbf{y}(\omega)$.

Translation invariance also holds. Consider the relation between $C(\mathbf{x})$ and $C(\mathbf{x} + \alpha\mathbf{1})$, and in particular the arguments to the utility function inside the integral. Since at any point ω, $C(\mathbf{x}) + \alpha - \mathbf{x}(\omega) - \alpha = C(\mathbf{x}) - \mathbf{x}(\omega)$, we have that $C(\mathbf{x}) + \alpha = C(\mathbf{x} + \alpha\mathbf{1})$.

For continuous **x**, it is easy to see the C is continuous with respect to the sup norm. Let $N_\epsilon(\mathbf{x})$ represent the neighborhood of all continuous functions ϵ-close to **x**, that is,

$$N_\epsilon(\mathbf{x}) = \{\mathbf{y} \mid \mathbf{y} \text{ continuous and } \sup_{\omega\in[0,1]} |\mathbf{x}(\omega) - \mathbf{y}(\omega)| < \epsilon\}$$

but then for all $\mathbf{y} \in N_\epsilon(\mathbf{x})$

$$|C(\mathbf{y}) - C(\mathbf{x})| < \epsilon$$

by translation invariance and monotonicity.

Solving for $C(\mathbf{x})$ requires solving an integral equation, and so the feasibility of our scheme relies on how easy it is to accurately evaluate the integral. While it is possible to describe degenerate cases where the integral cannot be computed effectively, there are many natural domains where the integral can be accurately computed using numerical techniques in a straightforward manner. Consider, for instance, taking bets on the real line using a Gaussian prior distribution. Then if payout vectors are simple polynomials the market maker's cost function can be solved using Gauss-Hermite quadrature (Judd, 1998).

Recall that in the discrete setting, we needed every $p_i > 0$, or else the price of a contract on ω_i occurring would always cost nothing. Consequently, a trader could make an arbitrarily large bet on ω_i for nothing, and then the market maker would have unbounded loss if ω_i were to occur.

This property is not directly meaningful when we talk about probability densities instead of probability masses. Instead, it generalizes to the unit interval through the notion of *strictly positive* density.

Definition 14. A probability density function f is *strictly positive* if $f(\omega) > 0$ for all ω.

The unit interval admits many distributions with strictly positive probability density, the uniform distribution being an example. Of course, the notion of strictly positive density is not just restricted to the unit interval. Over $\mathbb{R}$ an example of a strictly positive density function is a Gaussian distribution.

Bounded loss and no bad bets over the unit interval

Before we can give the main result of this section, we need to rigorously define what bounded loss and bad bets mean on the unit interval. This is far from simple because of the inherent degeneracies of dealing with measure zero sets. To simplify our analysis, we choose to define our relations relative to sub-intervals of the unit interval.

Definition 15. A cost function C on $[0,1]$ *has bounded loss over every interval* if the average loss over any sub-interval is bounded. Formally, if for all valid payout vectors $\mathbf{x}$ and all intervals $[a,b] \subseteq [0,1], a < b$, we have

$$\int_a^b C(\mathbf{x}) - \mathbf{x}(\omega)\, d\omega > -\infty$$

We can also define what an immediately bad bet looks like over an interval.

Definition 16. A cost function C on $[0,1]$ *never offers bets which are immediately bad over any interval* if an agent is always offered a bet over a proper sub-interval for less than unit cost. Formally, if for all valid payout vectors $\mathbf{x}$, all proper intervals $[a,b] \subset [0,1]$, and $\alpha > 0$ we have

$$C(\mathbf{x} + \alpha \mathbf{I}_{a,b}) - C(\mathbf{x}) < \alpha$$

where $\mathbf{I}_{a,b}$ is the indicator function over the interval $[a,b]$.

If this definition is not satisfied, the trader could be offered a bet that pays out α on the selected interval, but costs at least α. Therefore, the trader would pay at least as much as she could possibly make from the bet, but there also exist events not in the interval for which the trader loses her bet.

With our framework in place for fielding bets over the unit interval, we can present our main result of this section.

Proposition 10. *Let u be a differentiable barrier utility function, and let f be a strictly positive probability density function on $[0,1]$ with F its corresponding distribution function, and let $x^0 > 0$. Then for every valid payout vector $\mathbf{x}$, the constant-utility cost function given by the solution to*

$$\int_0^1 u\left(C(\mathbf{x}) - \mathbf{x}(\omega)\right) f(\omega)\, d\omega = u(x^0)$$

has bounded loss over every interval and never offers bets which are immediately bad over any interval.

Proof. To prove the specified loss bound, suppose that the market maker lost more than x^0 on a average on some interval $[a,b]$. Because the market maker is using a barrier utility function, at every point ω at which the market maker loses more than x^0 we must have that the amount paid out minus the amount paid in is greater than x^0, or that

$$\mathbf{x}(\omega) + x^0 - C(\mathbf{x}) > x^0$$

because $\mathbf{x}(\omega)$ is the amount paid out by the market maker and $C(\mathbf{x}) - x^0$ is the amount paid in (recall $x(\mathbf{o}) = 0$). By simplifying, we get that $C(\mathbf{x}) - \mathbf{x}(\omega) < 0$. Consequently, because u is a barrier utility function at those ω

$$u(C(\mathbf{x}) - \mathbf{x}(\omega)) \leq u(0) = -\infty$$

Then because f is strictly positive, the distribution has positive measure over the interval, so

$$\int_a^b u\left(C(\mathbf{x}) - \mathbf{x}(\omega)\right) f(\omega)\, d\omega = -\infty$$

and consequently the market maker's expected utility is $-\infty$ rather than $u(x^0) > -\infty$, a contradiction.

To prove the resulting cost function never offers bets which are immediately bad over any interval, consider some arbitrary $\mathbf{x}$, $\alpha > 0$, and proper sub-interval $[a, b]$. First, because C is translation invariant, we have

$$C(\mathbf{x}) + \alpha\mathbf{1} - C(\mathbf{x}) = \alpha$$

and because C is monotonic,

$$C(\mathbf{x} + \alpha\mathbf{I}_{a,b}) - C(\mathbf{x}) \leq \alpha$$

Now consider that

$$\int_0^1 u(C(\mathbf{x} + \alpha\mathbf{I}_{a,b}) - \mathbf{x}(\omega) - \alpha\mathbf{1}(\omega))\, dF(\omega) < u(x^0)$$

which holds because the utility function is strictly increasing, $\mathbf{x} + \alpha\mathbf{1} \geq \mathbf{x} + \alpha\mathbf{I}_{a,b}$ at every point in $[0, 1]$ and $\mathbf{x} + \alpha\mathbf{1} > \mathbf{x} + \alpha\mathbf{I}_{a,b}$ on $\mathbf{1} - \mathbf{I}_{a,b}$, which is a set of positive measure (recall that f is a strictly positive density function). But because the utility function is strictly increasing, this implies $C(\mathbf{x} + \alpha\mathbf{1}) > C(\mathbf{x} + \alpha\mathbf{I}_{a,b})$.

Bringing together our relations we have

$$\begin{aligned} C(\mathbf{x} + \alpha\mathbf{I}_{a,b}) - C(\mathbf{x}) &< C(\mathbf{x} + \alpha\mathbf{1}) - C(\mathbf{x} + \alpha\mathbf{I}_{a,b}) + C(\mathbf{x} + \alpha\mathbf{I}_{a,b}) - C(\mathbf{x}) \\ &= C(\mathbf{x} + \alpha\mathbf{1}) - C(\mathbf{x}) \\ &= \alpha \end{aligned}$$

which proves the cost function never offers bets which are immediately bad over any interval. ■

5.2 Adding profit and liquidity to convex risk measures

In this section, we develop market making agents that provide a principled way to extend the worst-case results of the literature to incorporate desirable, realistic features in practice. These are the first agents to include three human properties: the ability to make a profit, and (provided sufficient trading volume) unlimited market depth, and a vanishing bid/ask spread. At the same time, our agents retain the bounded loss property of the algorithmic agents in the literature. Furthermore, our agents incentivize myopic traders to directly reveal their private beliefs. In addition to these desirable theoretical properties, our new market making agents have other key practical properties . Just like in real markets, but unlike most agents in the literature, our market making agents are not path independent: a trader that buys from and then sells to our market making agents will incur a small loss. Additionally, our market making agents provide a straightforward way to incorporate the principal's subjective belief over the future into the quoted prices, and those quoted prices can be computed efficiently and simply.

5.2.1 Desirable properties for a market maker

In this section, we use a real example to demonstrate how liquidity is provisioned in real markets. We then use the insights gleaned from the example to motivate four desiderata for automated market makers. Finally, we give an overview of market makers from the literature to show that, while they can achieve every combination of three of the desiderata, no existing approach satisfies all four.

Our study of stock and prediction markets

The word *liquidity* in financial markets is burdened with several connotations. O'Hara (1995) memorably quips: "liquidity, like pornography, is easily recognized but not so easily defined". She goes on to describe several perspectives on what liquidity means: the volume in the market; the size of the marginal bid/ask spread (i.e., the spread for the smallest possible quantity); and the degree to which large bets move prevailing prices (equivalent to the volume of bets placed near the marginal bid/ask spread, the *market depth*). In traditional markets, all the senses of the term liquidity are conflated because these characteristics tend to accompany each other. We proceed to illustrate the conflation of these several views of liquidity using live, current markets as examples.

In real markets as the volume of trade increases, the bid/ask spread falls. This implies both that a fixed size bet moves the market less and that the marginal bid/ask spread decreases. Although there can be exceptions—such as the Flash Crash—this relationship holds broadly and was discussed at least as early as 1968 (Demsetz, 1968).

To provide an example of the effect of volume on bid/ask spreads, we collected values from five live markets. Table 5.2.1 shows the total cost to buy and then sell a thousand dollars of underlying contracts (without regard to any trading fees imposed by the exchange or brokers). We took these values from a snapshot of the relevant order books at 3pm on September 8th, 2011.

Market	Difference ($)	Approximate Daily Volume ($)
NWS stock	0.61	140 million
NYT stock	1.51	18 million
MNI stock	14.41	1.2 million
Obama 2012	31.58	900
Higgs Boson 2011	970	1.20

Table 5.1: The difference in bid and ask prices for $1,000 dollars of the underlying in several markets. "Obama 2012" is the Barack Obama re-elected contract on Intrade, and "Higgs Boson 2011" is the discovery of the Higgs Boson in 2011 contract on Intrade.

To illustrate how we calculated these values, imagine a certain security had the following order book: bid orders for 200 shares at both 2 and 3 dollars, and ask orders for 250 shares at 4 dollars. Then the bid/ask spread would be calculated as 300 dollars: 250 shares would be purchased at 4 dollars, exhausting the thousand dollar budget. Then, 200 shares would be resold at 3 dollars for a 1 dollar loss each, and the remaining 50 shares would be re-sold at 2 dollars for a 2 dollar loss each.

NWS, NYT, and MNI are all equities in news companies. The Obama 2012 Intrade contract, which pays off if the President is re-elected, is one of the most popular contracts on what is probably the most popular Internet prediction market. The Higgs Boson 2011 contract pays out if the Higgs Boson is discovered before the end of 2011 and is a very sparsely traded Intrade contract. The NWS contract is very liquid—well over a hundred million dollars worth is exchanged each day and sizable positions can be exited at almost no spread (the bid/ask spread on each share of NWS is one cent, the smallest possible value). In contrast, the Higgs Boson contract is very illiquid, seeing almost no trade each day. The bid/ask spread on a sizable position is large: buying and then selling a 1,000-dollar position results in an immediate loss of 970 dollars.

5.2.2 Four oppositional desiderata in the literature

The observations of the previous section suggest that real markets have a shrinking bid/ask spread for fixed-size bets as the volume gets large. This, combined with the profit motives of real market makers and the algorithmic worst-case guarantees of the market makers from the literature, yields four desiderata for market making agents:

1. Bounded worst-case loss.

2. The ability to enter a state where the market maker books a profit regardless of which future state of the world is realized.

3. A marginal bid/ask spread that approaches zero in the limit as volume gets large.

4. The price of any fixed-size transaction approaches the marginal bid or ask price, so that there is unlimited depth in the limit as volume gets large.

In addition to their self-evident desirability, *all four combinations of exactly three of these properties already exist in the literature*. Satisfying all four characteristics is challenging because several of the qualities are oppositional; for instance, making a profit involves charging extra, but charging extra means that the bid/ask spread may not vanish. As another example, a market maker that is very deep must not move prices very much in the face of large bets, but if prices are not moved enough then worst-case loss can become unbounded. To illustrate the challenge involved in satisfying all four of these desiderata, we proceed to provide a quick survey of the relevant constructions from the literature. (The next sections will define the desiderata formally.)

Fixed prices

Probably the simplest automated market maker is to determine a probability distribution over the future states of the world, and to offer to make bets directly at those odds. This scheme offers unlimited depth and no marginal bid/ask spread, but has unbounded worst-case loss and no ability to book a profit.

Fixed prices with profit-taking

Adding a profit cut on top of the fixed odds gives the market maker the ability to make a profit and retains the unlimited depth of the fixed pricing scheme. However, this market maker has unbounded worst-case loss and a non-vanishing bid/ask spread.

Fixed prices with shrinking profit cuts

If we allow the profit cut to diminish to zero as trading volume increases, the resulting market maker has three of the four desired properties: the ability to make a profit, a vanishing marginal bid/ask spread, and unbounded depth in limit. However, it still has unbounded worst-case loss because a trader with knowledge of the true future could make an arbitrarily large winning bet with the market maker.

Convex risk measures

Convex risk measures are the general class of market makers (Agrawal et al., 2009; Othman and Sandholm, 2011b) featured in much of the prediction market literature (Chen and Pennock, 2007; Peters et al., 2007; Agrawal et al., 2009; Chen and Vaughan, 2010; Othman and Sandholm, 2010a; Abernethy et al., 2011), including the most widely-used automated market maker in practice, the *Logarithmic Market Scoring Rule (LMSR)* (Hanson, 2003, 2007). These market makers can offer bounded worst-case loss and no marginal bid/ask spread. However, they do not offer the ability for the market maker to book a profit, and they offer a fixed market depth that does not increase with volume. (For instance, in the LMSR, the depth of the market is fixed by the parameter b which is an exogenous constant set *a priori*.)

Convex risk measures with profit-taking

Adding a fixed charge on top of a convex risk measure gives the market maker the ability to make a profit, but the marginal bid/ask spread will not vanish.

Convex risk measures with shrinking profit cuts

If we allow the profit cut to shrink to zero as trading volume increases, the resulting market maker can have bounded loss, the ability to make a profit, and a marginal bid/ask spread that vanishes. However, the depth of the market is still an exogenous constant, and a fixed size bet will always diverge in price from the marginal.

Extended constant-utility cost functions

Othman and Sandholm (2011a) describe a market maker that has unbounded depth in the limit, bounded loss, and vanishing bid/ask spread in the limit. However, it has no ability to book a profit. The market making agents of Othman and Sandholm (2011a) are a restricted special case of the market makers we develop in this work, where the utility function and liquidity function are both logarithmic functions and there is no profit function.

Liquidity-sensitive automated market makers

The market maker described in Othman et al. (2010) has three of the desired properties. While it does have unbounded depth in the limit, bounded loss, and the ability to make a profit, the marginal bid/ask spread never goes to zero.

5.2.3 How bets are taken

In this section, we describe how to calculate the prices an agent sees when she trades with our market makers. We denote a market maker by $\mathcal{M}$ and we denote $\mathcal{M}(\mathbf{x}, \mathbf{y})$ to be the *total* price charged to agents for moving the market maker from payout vector $\mathbf{x}$ to payout vector $\mathbf{y}$.

Let u be a utility function and $x^0 \in \mathbf{dom}\, u$. Let f and g be non-decreasing functions $\mathbb{R}^+ \mapsto \mathbb{R}$ with the property that $f(0) = g(0) = 0$. For reasons that will shortly become clear, we call f the *liquidity function* and g the *profit function*. Let $d : \mathbb{R}^n \times \mathbb{R}^n \mapsto \mathbb{R}^+$ be a distance function (metric). The state s is an internal scalar initialized so that $s = 0$. s can be thought of as a measure of the cumulative volume transacted in the market.

In order to price a bet, the following steps are taken:

1. A trader wishes to place a bet that would move the market maker from payout vector $\mathbf{x}$ to payout vector $\mathbf{y}$. The market maker's current state is s.
2. The cost function $C(\mathbf{y})$ is solved implicitly for

   $$\sum_i p_i u\left(C(\mathbf{y}) - y_i + f(s + d(\mathbf{x}, \mathbf{y}))\right) = u(x^0 + f(s + d(\mathbf{x}, \mathbf{y})))$$

   This equation incorporates the liquidity function f but not the profit function g. Just as in constant-utility cost functions, C will be uniquely defined because u is strictly increasing, and it can be calculated efficiently through a binary search.
3. The total cost quoted to the trader for the bet is the sum of the changes to the cost function, liquidity function, and profit function

   $$\mathcal{M}(\mathbf{x}, \mathbf{y}) \equiv C(\mathbf{y}) - C(\mathbf{x}) + f(s + d(\mathbf{x}, \mathbf{y})) - f(s) + g(s + d(\mathbf{x}, \mathbf{y})) - g(s)$$

4. If the bet is taken, the state changes: $s \leftarrow s + d(\mathbf{x}, \mathbf{y})$, and the new value of the cost function $C(\mathbf{y})$ is saved for the next transaction.

We will call a market making agent $\mathcal{M}$ that prices bets this way a *constant-utility profit-charging market maker*.

The values involved in calculating $\mathcal{M}$ all *telescope* from their initial values. Consider a transaction that first moves from payout vector $\mathbf{x}$ to payout vector $\mathbf{y}$, ending at payout vector $\mathbf{z}$. The total amount paid into the market maker is

$$\mathcal{M}(\mathbf{x}, \mathbf{y}) + \mathcal{M}(\mathbf{y}, \mathbf{z}) = C(\mathbf{z}) - C(\mathbf{x}) + f(s + d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z})) - f(s) + g(s + d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z})) - g(s)$$

Observe that the terms from the intermediate payout vector $\mathbf{y}$ (i.e., $C(\mathbf{y})$, $f(s + d(\mathbf{x}, \mathbf{y}))$, and $g(s + d(\mathbf{x}, \mathbf{y})))$ cancel out and so do not appear in this expression. We will use this telescoping property to simplify some of the proofs in Section 5.2.4.

5.2.4 Theoretical properties

In this section, we show that our market makers satisfy several desirable properties. We begin by showing that they have bounded loss. As a step toward this result, we show that they have no money pumps, so they incentivize agents to directly acquire their desired portfolios (rather than taking a roundabout path). We then show that our agents have unlimited depth and no bid/ask spread in the limit. Finally, we formalize the notion of profit-taking and describe how our agents can enter a state of *unconditional profit*, in which no matter what the realized outcome is or what the future actions of the traders are the market maker will book a profit.

Bounded loss

One particularly undesirable property for a market making agent is to be in possession of a *money pump*.

Definition 17. A market maker $\mathcal{M}$ has a *money pump* if there exists a sequence of payout vectors $\mathbf{x_1}, \ldots, \mathbf{x_n}$ and some terminal state $\mathbf{x_0}$ such that

$$\mathcal{M}(\mathbf{x_0}, \mathbf{x_1}) + \cdots + \mathcal{M}(\mathbf{x_{n-1}}, \mathbf{x_n}) + \mathcal{M}(\mathbf{x_n}, \mathbf{x_0}) < 0$$

When a market maker has a money pump a trader can keep arbitraging the market maker for unbounded riskless profit. Perhaps the most prominent difference between our work and the prior literature is that the total prices charged by our market making agents are not necessarily path independent, that is, different paths through quantity space may correspond to different costs being charged by our market making agents. With a path-independent market maker, buying a contract and then immediately selling it is without cost. This is in contrast to both real-world markets and our market making agents, where buying and then immediately selling is costly to the trader (see, e.g., Table 5.2.1).

Every cost function-based market maker is path independent, and path independence is sufficient for a market maker to have no money pumps (Othman et al., 2010). However, path independence is not *necessary* for a market maker to have no money pumps; having a path-dependent market making agent simply means that it is not immediate that the market maker has no money pumps.

Definition 18. A market maker $\mathcal{M}$ obeys the *triangle inequality* if, for all payout vectors $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$

$$\mathcal{M}(\mathbf{x}, \mathbf{y}) + \mathcal{M}(\mathbf{y}, \mathbf{z}) \geq \mathcal{M}(\mathbf{x}, \mathbf{z})$$

By inductive argument, it is easy to see that a market maker that obeys the triangle inequality cannot have a money pump.

Proposition 11. *Let $\mathcal{M}$ be a constant-utility profit-charging market maker. Then $\mathcal{M}$ obeys the triangle inequality.*

Proof. Consider $\mathcal{M}(\mathbf{x},\mathbf{y}) + \mathcal{M}(\mathbf{y},\mathbf{z}) - \mathcal{M}(\mathbf{x},\mathbf{z})$. In terms of the triangle inequality, this is the difference between taking the "long way" around the triangle versus the direct way; we seek to show that it is non-negative for all $\mathbf{x}, \mathbf{y}$, and $\mathbf{z}$.

Because all the values associated with the calculation of $\mathcal{M}$ telescope, this difference is just

$$C_1(\mathbf{z})+f(s+d(\mathbf{x},\mathbf{y})+d(\mathbf{y},\mathbf{z}))+g(s+d(\mathbf{x},\mathbf{y})+d(\mathbf{y},\mathbf{z}))-(C_2(\mathbf{z})+f(s+d(\mathbf{x},\mathbf{z}))+g(s+d(\mathbf{x},\mathbf{z}))) \tag{5.1}$$

where $C_1(\mathbf{z})$ solves

$$\sum_i p_i u(C_1(\mathbf{z}) - z_i + f(s + d(\mathbf{x},\mathbf{y}) + d(\mathbf{y},\mathbf{z}))) = u(x^0 + s + d(\mathbf{x},\mathbf{y}) + d(\mathbf{y},\mathbf{z}))$$

and C_2 solves

$$\sum_i p_i u(C_2(\mathbf{z}) - z_i + f(s + d(\mathbf{x},\mathbf{z}))) = u(x^0 + s + d(\mathbf{x},\mathbf{z}))$$

We will divide the terms in Equation 5.1 and deal with them each in turn. First, we will show that

$$C_1(\mathbf{z}) + f(s + d(\mathbf{x},\mathbf{y}) + d(\mathbf{y},\mathbf{z})) - (C_2(\mathbf{z}) + f(s + d(\mathbf{x},\mathbf{z}))) \geq 0$$

Since u is strictly increasing and d is a distance function, we have

$$u(x^0 + s + d(\mathbf{x},\mathbf{y}) + d(\mathbf{y},\mathbf{z})) \geq u(x^0 + s + d(\mathbf{x},\mathbf{z}))$$

and so

$$\sum_i p_i u(C_1(\mathbf{z}) - z_i + f(s + d(\mathbf{x},\mathbf{y}) + d(\mathbf{y},\mathbf{z}))) \geq \sum_i p_i u(C_2(\mathbf{z}) - z_i + f(s + d(\mathbf{x},\mathbf{z})))$$

which, because u is strictly increasing and the p_i form a probability distribution, implies

$$C_1(\mathbf{z}) + f(s + d(\mathbf{x},\mathbf{y}) + d(\mathbf{y},\mathbf{z})) \geq C_2(\mathbf{z}) + f(s + d(\mathbf{x},\mathbf{z}))$$

so

$$C_1(\mathbf{z}) + f(s + d(\mathbf{x},\mathbf{y}) + d(\mathbf{y},\mathbf{z})) - (C_2(\mathbf{z}) + f(s + d(\mathbf{x},\mathbf{z}))) \geq 0 \tag{5.2}$$

Now consider the other set of terms in Equation 5.1,

$$g(s + d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z})) - g(s + d(\mathbf{x}, \mathbf{z}))$$

Because d satisfies the triangle inequality, $d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z}) \geq d(\mathbf{x}, \mathbf{z})$. Therefore, the argument to the $g(\cdot)$ function on the left is at least as large as the argument to the $g(\cdot)$ function on the right. Since $g(\cdot)$ is a non-decreasing function in its arguments

$$g(s + d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z})) - g(s + d(\mathbf{x}, \mathbf{z})) \geq 0 \tag{5.3}$$

Putting together Equations 5.1, 5.2, and 5.3, we have that $\mathcal{M}(\mathbf{x}, \mathbf{y}) + \mathcal{M}(\mathbf{y}, \mathbf{z}) - \mathcal{M}(\mathbf{x}, \mathbf{z}) \geq 0$, or rewritten, $\mathcal{M}(\mathbf{x}, \mathbf{y}) + \mathcal{M}(\mathbf{y}, \mathbf{z}) \geq \mathcal{M}(\mathbf{x}, \mathbf{z})$, and so $\mathcal{M}$ obeys the triangle inequality. ■

The proof also provides the intuition that, with an increasing profit function, an agent is charged strictly greater prices when he does not acquire inventory on the path directly to his desired allocation. This is because the market maker will collect more money from the profit function, while not losing any additional money from the potential added liquidity from taking a longer path. Consequently, myopic agents are incentivized to directly acquire their desired allocation.

Because the total prices charged by a constant-utility profit-charging market maker obey the triangle inequality, we also have the following result.

Corollary 1. *Let $\mathcal{M}$ be a constant-utility profit-charging market maker. Then $\mathcal{M}$ has no money pumps.*

Recall that a constant-utility cost function employing a barrier utility function has a worst-case loss of the x^0 originally used to seed the utility function. Because $\mathcal{M}$ has no money pumps, we have the following result.

Corollary 2. *Let u be a barrier utility function. Then if $p_i > 0$ for every i, a constant-utility profit-charging market maker $\mathcal{M}$ that uses u loses no more than the x^0 originally used to seed the utility function, regardless of the trades made by agents or the realized future state of the world. (This result holds even if the liquidity function f and the profit function g are zero.)*

5.2.5 Profit, liquidity, and market depth

We begin this section by showing that, under certain conditions, a constant-utility profit-charging market maker has a vanishing bid/ask spread. Then, we show that under more stringent conditions, constant-utility profit-charging market makers also have unbounded market depth.

Throughout this section, to simplify the proofs, we assume d is a distance function implied by any $\mathcal{L}_p$ norm. However, the results of this section hold for any continuous distance function.

Vanishing bid/ask spread

In this section, we show that if the profit and liquidity functions are diminishing, a constant-utility profit-charging market maker has a vanishing bid/ask spread.

Definition 19. For a market maker $\mathcal{M}$ with a differentiable price response, let $\mathcal{M}_i$ denote the marginal cost of a bet on the i-th event. A market maker has a *vanishing bid/ask spread* if $\sum_i \mathcal{M}_i = 1$.

Proposition 12. *Let the liquidity function f and profit function g have the property that* $\lim_{s\to\infty} f'(s) = \lim_{s\to\infty} g'(s) = 0$. *Then a constant-utility profit charging market maker has a vanishing bid/ask spread as s gets large. Formally,* $\lim_{s\to\infty} \sum_i \mathcal{M}_i = 0$.

Proof. At any $\mathbf{x}$ and s, by definition we have

$$\mathcal{M}_i = \nabla_i C(\mathbf{x}) + \nabla_i f(s) + \nabla_i g(s) = \nabla_i C(\mathbf{x}) + f'(s) + g'(s)$$

Since, by construction, both $f'(s)$ and $g'(s)$ go to zero as s gets large, the interesting term is $\nabla_i C(\mathbf{x})$. From the definition of constant-utility profit-charging market makers we have

$$\nabla_i \left(\sum_j p_j u(C(\mathbf{x}) - x_j + f(s)) \right) = \nabla_i u(x^0 + f(s))$$

$$\nabla_i p_i u(C(\mathbf{x}) - x_i + f(s)) + \nabla_i \sum_{j \neq i} p_j u(C(\mathbf{x}) - x_j + f(s)) = u'(x^0 + f(s)) f'(s)$$

$$(\nabla_i C(\mathbf{x}) + f'(s)) \left(\sum_j p_j u'(C(\mathbf{x}) - x_j + f(s)) \right) = p_i u'(C(\mathbf{x}) - x_i + f(s)) + u'(x^0 + f(s)) f'(s)$$

solving for $\nabla_i C(\mathbf{x})$, we get

$$\nabla_i C(\mathbf{x}) = \frac{p_i u'(C(\mathbf{x}) - x_i + f(s)) + u'(x^0 + f(s)) f'(s)}{\sum_j p_j u'(C(\mathbf{x}) - x_j + f(s))} - f'(s)$$

$$= \frac{p_i u'(C(\mathbf{x}) - x_i + f(s)) + f'(s) \left(u'(x^0 + f(s)) - \sum_j p_j u'(C(\mathbf{x}) - x_j + f(s)) \right)}{\sum_j p_j u'(C(\mathbf{x}) - x_j + f(s))}$$

$$= \frac{p_i u'(C(\mathbf{x}) - x_i + f(s))}{\sum_j p_j u'(C(\mathbf{x}) - x_j + f(s))}$$

because by construction

$$u(x^0 + f(s)) = \sum_j p_j u(C(\mathbf{x}) - x_j + f(s))$$

and so their derivatives are also equal. Therefore

$$\sum_i \nabla_i C(\mathbf{x}) = \frac{\sum_i p_i u'(C(\mathbf{x}) - x_i + f(s))}{\sum_j p_j u'(C(\mathbf{x}) - x_j + f(s))} = 1$$

and so

$$\lim_{s\to\infty} \sum_i \mathcal{M}_i = \sum_i \nabla_i C(\mathbf{x}) + nf'(s) + ng'(s) = 1 + 0 + 0 = 1$$

■

Unbounded depth

In this section, we show that under additional, somewhat stronger, conditions a constant-utility profit-charging market maker also has unbounded depth. This result is similar to the one in Othman and Sandholm (2011a). However, that market maker only showed unbounded depth when both u and f were logarithmic functions. Our result in this section is for a much broader class of utility, liquidity, and profit functions.

The first condition we require is that the utility function employed has *vanishing absolute risk aversion*. To our knowledge this concept was first defined in Caballé and Pomansky (1996).

Definition 20. A utility function has *vanishing absolute risk aversion* if it is twice-differentiable and

$$\lim_{c\to\infty} -\frac{u''(c)}{u'(c)} = 0$$

Most utility functions that are standard in the literature have this property. For instance, every utility function that has Constant Relative Risk Aversion (CRRA) (Mas-Colell et al., 1995) also has vanishing absolute risk aversion. An example of a utility function that does not have vanishing absolute risk aversion is $u(x) = -e^{-x}$.

As we have discussed, the *depth* of a market is the degree to which large bets move marginal prices. A very deep market will clear large orders without moving the marginal price. This leads us to the following definition.

Definition 21. Consider a market maker $\mathcal{M}$ with a twice-differentiable price response, and let $\mathcal{M}_{ii}$ denote the marginal change in the marginal price of a bet on the i-th event. We say that $\mathcal{M}$ has *unlimited depth* if $\mathcal{M}_{ii} = 0$ for all i.

Proposition 13. *Let f be a liquidity function with the property that $\lim_{s\to\infty} f(s) = \infty$. Then in any constant-utility profit-charging market maker, the arguments to the utility function grow arbitrarily large with s. Formally, for any $\mathbf{x}$ and i*

$$\lim_{s\to\infty} C(\mathbf{x}) - x_i + f(s) = \infty$$

Proof. Suppose not. Then there exists some finite bound B such that, for all i

$$\lim_{s\to\infty} C(\mathbf{x}) - x_i + f(s) < B$$

but then, because u is strictly increasing,

$$\lim_{s\to\infty} \sum_i p_i u(C(\mathbf{x}) - x_i + f(s)) < u(B)$$

But by construction,

$$\sum_i p_i u(C(\mathbf{x}) - x_i + f(s)) = u(x^0 + f(s))$$

and $\lim_{s\to\infty} x^0 + f(s) = \infty$. Therefore, there exists some S such that for all $s > S$,

$$x^0 + f(s) > B$$

At such s, we have

$$\sum_i p_i u(C(\mathbf{x}) - x_i + f(s)) < u(B) < u(x^0 + f(s))$$

but this is a contradiction because these values must be equal by construction. ■

Proposition 14. *Let u be a utility function with vanishing absolute risk aversion, and let f be a liquidity function with the property that $\lim_{s\to\infty} f(s) = \infty$, and let the liquidity function f and profit function g have the property that*

$$\lim_{s\to\infty} f'(s) = \lim_{s\to\infty} f''(s) = \lim_{s\to\infty} g''(s) = 0$$

Then the constant-utility profit charging market maker formed by u, f, and g has unlimited depth as s gets large. Formally, $\lim_{s\to\infty} \mathcal{M}_{ii} = 0$.

Proof. By definition $\mathcal{M}_{ii} = \nabla^2_{ii} C(\mathbf{x}) + f''(s) + g''(s)$. Just as in the proof of Proposition 12, since

$$\lim_{s\to\infty} f''(s) = \lim_{s\to\infty} g''(s) = 0$$

the interesting term as s gets large is the cost function term.

For notational simplicity, define $s_i \equiv C(\mathbf{x}) - x_i + f(s)$. Then

$$\nabla_i C(\mathbf{x}) = \frac{p_i u'(s_i)}{\sum_j p_j u'(s_j)}$$

and taking the partial derivative with respect to i of both sides yields

$$\nabla^2_{ii} C = \frac{p_i u''(s_i) s'_i \left(\sum_j p_j u'(s_j)\right)}{\left(\sum_j p_j u'(s_j)\right)^2} - \frac{p_i u'(s_i) \left(\sum_j p_j u''(s_j) s'_j\right)}{\left(\sum_j p_j u'(s_j)\right)^2} \tag{5.4}$$

We need to show that this approaches zero as s gets large.

First, $C(\mathbf{x})$ is a convex function because it is implicitly defined as an argument to equalize a concave utility function (Boyd and Vandenberghe, 2004). Therefore $\nabla^2_{ii} C(\mathbf{x}) \geq 0$.

Now consider the s'_i terms. By construction $s'_{j\neq i} = \nabla_j C(\mathbf{x}) + f'(s)$ and $s'_i = \nabla_i C(\mathbf{x}) - 1 + f'(s)$. Therefore $1 \geq \lim_{s\to\infty} s'_{j\neq i} \geq 0 \geq \lim_{s\to\infty} s'_i \geq -1$. Thus we can bound the second derivative in the limit:

$$\lim_{s\to\infty} \nabla^2_{ii} C(\mathbf{x}) \leq \lim_{s\to\infty} - \left[\frac{p_i u''(s_i) \left(\sum_j p_j u'(s_j)\right)}{\left(\sum_j p_j u'(s_j)\right)^2} + \frac{p_i u'(s_i) \left(\sum_j p_j u''(s_j)\right)}{\left(\sum_j p_j u'(s_j)\right)^2} \right]$$

where we have replaced all the s'_i in the first term of Equation 5.4 with -1 and all the s'_j in the second term of Equation 5.4 with 1, in order to make all of the positive terms as large as possible. (Recall that u is concave, and so $u'' \leq 0$.)

We will deal with the two terms of Equation 5.2.5 in succession, showing that the limit of each as s gets large is 0. We can simplify the first term immediately by canceling out the like term from the numerator and denominator, leaving

$$\lim_{s\to\infty} - \frac{p_i u''(s_i)}{\sum_j p_j u'(s_j)}$$

but then

$$\lim_{s\to\infty} -\frac{p_i u''(s_i)}{\sum_j p_j u'(s_j)} \leq \lim_{s\to\infty} -\frac{p_i u''(s_i)}{p_i u'(s_i)} = 0$$

because u has vanishing absolute risk aversion and because by Proposition 13, we have that $\lim_{s\to\infty} s_i = \infty$.

Now, we address the second term of Equation 5.2.5

$$\lim_{s\to\infty} -\frac{p_i u'(s_i)\left(\sum_j p_j u''(s_j)\right)}{\left(\sum_j p_j u'(s_j)\right)^2}$$

We will split this term into the product of two terms. First

$$\frac{p_i u'(s_i)}{\left(\sum_j p_j u'(s_j)\right)}$$

is just $\nabla_i C(\mathbf{x})$, and is therefore no larger than 1. Then because u has vanishing absolute risk aversion, and because as s gets large the s_j get large

$$\lim_{s\to\infty} -\frac{\left(\sum_j p_j u''(s_j)\right)}{\left(\sum_j p_j u'(s_j)\right)} = 0$$

Consequently,

$$\lim_{s\to\infty} -\frac{p_i u'(s_i)\left(\sum_j p_j u''(s_j)\right)}{\left(\sum_j p_j u'(s_j)\right)^2} = \lim_{s\to\infty} \frac{p_i u'(s_i)}{\left(\sum_j p_j u'(s_j)\right)} \cdot -\frac{\left(\sum_j p_j u''(s_j)\right)}{\left(\sum_j p_j u'(s_j)\right)} \leq 1\cdot 0 = 0$$

Putting together all of our relations, we have

$$0 \leq \lim_{s\to\infty} \nabla^2_{ii} C(\mathbf{x}) \leq 0$$

and so $\lim_{s\to\infty} \nabla^2_{ii} C(\mathbf{x}) = 0$, and therefore

$$\lim_{s\to\infty} \mathcal{M}_{ii} = \lim_{s\to\infty} \nabla^2_{ii} C(\mathbf{x}) + f''(s) + g''(s) = 0 + 0 + 0 = 0$$

■

Revenue bounds

To our knowledge, the only previous formal study of profit-charging behavior within the standard automated market making framework is Othman et al. (2010). The automated market maker in that paper could enter a state of *outcome-independent profit*, which means that regardless of the realized outcome, the agent would book a profit. However, this condition only holds if the market terminates in that state. It is entirely possible, and in some settings virtually assured, that the market will leave the outcome-independent profit state and cause the market maker to run at a net loss.

To explore this notion further, recall that the market maker in Othman et al. (2010) is defined by the cost function

$$C(\mathbf{x}) = b(\mathbf{x}) \log \left(\sum_i \exp(x_i / b(\mathbf{x})) \right)$$

where $b(\mathbf{x}) = \alpha \sum_i x_i$. To see how outcome-independent profit works, consider a simple two-event market where $\alpha = 0.05$ and the market starts from $\mathbf{x^0} = (1, 1)$. Imagine that the market maker takes two bets, the first a 50 dollar payout on the first event and the second a 50 dollar payout on the second event. The market maker is therefore in the state $(51, 51)$ and is in a state of outcome-independent profit, because no matter whether ω_1 or ω_2 is realized, the market maker books a profit of $C(51, 51) - 51 - C(\mathbf{x^0}) \approx 2.5 > 0$. However, imagine that another trader comes along and places an additional 50 dollar payout bet on the first event. Now the market maker is in state $(101, 51)$. If the market terminates and ω_1 is realized the market maker clears $C(101, 51) - 101 - C(\mathbf{x^0}) \approx -1.1 < 0$. So, the final bet has made the market maker exit the state of outcome-independent profit.

This problem is especially likely to arise in practice in settings where more information is revealed over time. For instance, consider a sports game in which betting is left open as the game proceeds. Once a clear winner emerges, traders will likely buy many shares from the market maker in that team, causing the market maker to exit the state of unconditional profit.

There is an important loophole to this argument though. If the automated market maker is not designated with a formal, contractual role that necessitates its activity in a market, e.g., if it is being employed as the automated agent of a trader, then the market maker can circumvent this failure by simply ceasing to trade once it enters a state of outcome-independent profit, or alternately refusing to take any bet that would cause it to exit the state of outcome-independent profit.

In many settings, however, market makers are required to keep a presence in the market, e.g., if they are a designated market maker. For these settings, the notion of *unconditional profit* is more important. We say that a market maker enters a state of unconditional profit if it will make money regardless of the realized outcome *and* the future actions of traders. What is important to note about unconditional profit, as opposed to outcome-independent profit, is that once a market maker enters a state of unconditional profit it can never exit that state.

For example, consider a constant-utility profit-charging market maker using a barrier utility function. If the market maker's prior is strictly positive, and if the profit function has collected more money than the initial value used to seed the liquidity function (i.e., if for all i, $p_i > 0$ and $g(s) > x^0$), then the market maker has entered the state of unconditional profit. This is because the market maker has no money pumps, loses no more than x^0, and has collected a profit of more than this worst-case loss.

Instantiating the theory

In this section we collect the complete set of conditions we require on the various parts of the profit-charging automated market maker. We give several examples of functions that satisfy these conditions; any mix of these choices will produce an automated market maker with desirable qualities, differing only in the specifics of the initial amount of market depth, marginal bid/ask spread, and so on.

The prior distribution should be strictly positive ($p_i > 0$) or the market maker will have unbounded worst-case loss. The uniform distribution is one option, but if the market maker has beliefs over the future these should be used, subject to never setting the probability of any event to zero.

The distance function d should be continuous. This is because a discrete distance function will never be able to achieve a vanishing bid/ask spread. One simple suggestion is to have d be (a positive multiple of) the distance function implied by any $\mathcal{L}_p$ norm.

The utility function u should be a barrier utility function (in order to bound worst-case loss in a simple way), and should have vanishing absolute risk aversion (in order to have unbounded depth). One family of utility functions that satisfy both of these requirements is the standard CRRA utility functions

$$u(x) = \frac{x^{1-\gamma}}{1-\gamma}$$

indexed by $\gamma \geq 1$. (Here, log is the limit case for $\gamma = 1$.)

The liquidity function f should have $f(0) = 0$ and

$$\lim_{s\to\infty} f'(s) = \lim_{s\to\infty} f''(s) = 0$$

and also $\lim_{s\to\infty} f(s) = \infty$, to ensure a vanishing bid/ask spread and unbounded depth. One simple class of functions with both of these properties is $f(s) = \alpha s^{1/\beta}$ for $\beta > 1$ and $\alpha > 0$. Another class of functions that has these properties is $f(s) = \alpha \log(s+1)$, again for $\alpha > 0$.

The profit function g should have $g(0) = 0$ and

$$\lim_{s\to\infty} g'(s) = \lim_{s\to\infty} g''(s) = 0$$

for unbounded depth and vanishing bid/ask spread. Since the conditions on the liquidity function f are a strict superset of the conditions on the profit function g, any liquidity function could also serve as a profit function. As an example of a valid profit function that would not be a valid liquidity function is for the profit function to collect at most a certain amount more than the worst-case loss bound, e.g., setting

$$g(s) = \xi x^0 \left(1 - 1/(s+1)\right)$$

for $\xi > 1$ so that the market maker will collect at most ξ times x^0. Once the market maker has collected more than x^0, it is guaranteed to have entered the unconditional profit state.

To get a perspective on the operation of a representative market maker, consider Figures 5.1 and 5.2, which depict the prices charged by the market maker and the profit made by the market maker, respectively. In this specific example, the number of events is $n = 2$ and the market maker's probability on each event is $p_i = 1/2$. The distance function is implied by the 2 norm of the payout vector. The utility function is $u(x) = -1/x$, the initial wealth is $x^0 = 10$, the liquidity function is $f(s) = 100(\log(s + 10000) - \log(10000))$, and the profit function is $g(s) = 0.6\left(\sqrt{s+100} - 10\right)$. These functions were chosen by experimenting with the classes of functions we suggested previously until we found a combination that made particularly attractive plots.

In both plots, the x axis is the quantity the market maker holds on both events before the interaction of the agents. That is, a value of 10^2 on the x axis indicates the market maker is at the payout vector $(100, 100)$ before the interaction of the trader.

Figure 5.1 shows the marginal price on the first event and the total price charged by the market maker for a one-dollar payout on the first event as the quantity in the market increases. Observe both that the cost of a fixed size bet approaches the marginal price and also that the marginal price approaches a bid/ask spread of 0 (which, because $p_1 = 0.5$ and the two events have equal payouts, corresponds to a marginal price of 0.5).

Figure 5.2 shows the profit made by the market maker as the amount transacted increases. For $s \gtrsim 600$, the market maker enters a state of unconditional profit, because $g(s) > x^0 = 10$. This state is reached at a payout of at most about 430 on each event. If there is additional churn in the market—traders that sell their bets back to the market maker—the state of unconditional profit can be reached at smaller payout vectors. At an extreme, a state of unconditional profit can be made by an agent buying and selling back a single bet to the market maker a large number of times, each time paying a small amount into the profit function.

5.3 Extensions

Both of the extensions to convex risk measures discussed in this chapter merit further experimental study. Measurable spaces with practical implications include notions of location and time. For instance, a prediction market over where and when a hurricane will make landfall. Continuous spaces could also be as tractable relaxations of a discrete space. An example of a setting like this is in options trading, where the space of possible expiration prices for an underlying stock is very

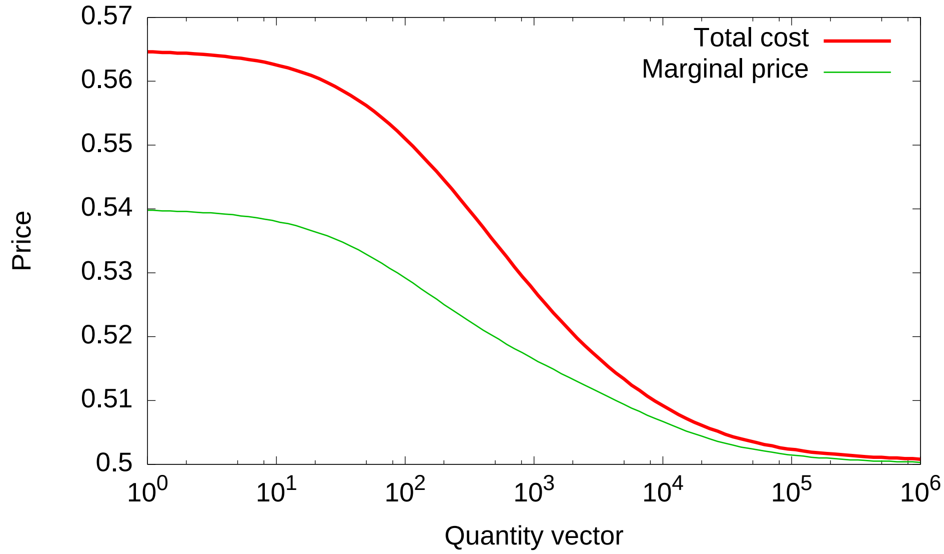


Figure 5.1: Total cost for making a one-ticket bet and marginal prices for the market maker and setting given in the text. Observe that as the quantity transacted increases the bid/ask spread goes to zero (i.e., both lines go to 0.5). The x axis is log-scaled.

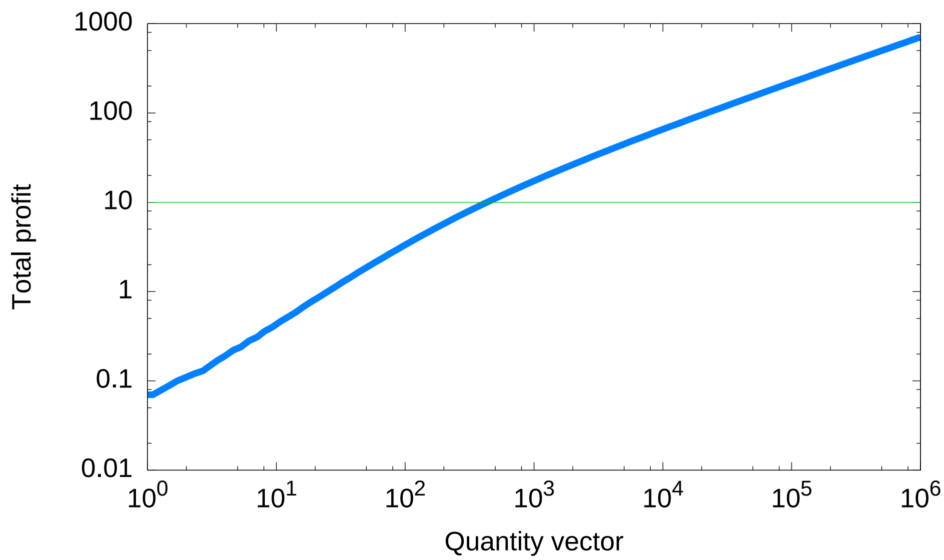


Figure 5.2: The profit collected increases with the quantity transacted in the market. The straight line at 10 represents when the market maker enters a state of unconditional profit. Both axes are log-scaled.

large and discrete (any non-negative one-cent delimited value). We explore this specific example much further in Chapter 7.

The profit-charging constant utility cost functions developed in this chapter seem like a good fit for virtually any prediction market or gambling setting, although care needs to be taken to determine an appropriate combination of utility, profit, and liquidity functions as well as the initial seed value x^0. Finally, it seems straightforward to combine the two extensions discussed in this chapter, creating continuous profit-charging constant-utility cost functions. The risk measure analogues of this class of market maker are likely to be useful in financial applications.

We have heard the argument that it could make more sense to have smaller bid/ask spreads at the market's initiation, to encourage participation, and widen the bid/ask spread as the market matures. This scheme has an appealing intuition: essentially, early participants are being compensated for providing liquidity at the more fragile beginning stages of the market. One way to produce this effect with our market makers is to start s at a large value and decrement it over time (e.g., $s \leftarrow s - d(\mathbf{x}, \mathbf{y})$ when the value is updated). This essentially runs the state variable s in reverse. However, as discussed in Section 5.2.1, this is counter to the way real markets function, and so adopting this methodology may be unintuitive for participants. There may be better ways of doing this, and we believe that our framework of profit and liquidity functions can facilitate such study.

Another extension is to extend our agents to handle not only market orders, but also limit orders—orders that execute only if certain price conditions are reached. Here, a possible direction seems to be to extend the convex optimization framework of Agrawal et al. (2009). That work provided a simple, computationally efficient way to turn a cost function that is capable of only handling market orders into a cost function that can handle both market and limit orders. While adding a profit function to the Agrawal framework is trivial, incorporating a liquidity function appears to be more challenging, because the liquidity function is used in calculating prices. To be explicit, the trading volume affects the liquidity function, which affects the cost function, which affects the number of cleared orders, which affects the trading volume. Resolving this circularity seems like the most direct solution for our market making agents to handle persistent limit orders.

An extension of a different flavor is to consider the values we captured from real markets in Table 5.2.1. To our knowledge, that table represents the first comparison of the prevalent liquidity in prediction markets versus equity markets. That table showed, somewhat surprisingly, that active prediction market contracts have a bid/ask spread roughly comparable to small cap equities, despite the fact that publicly traded equities have several orders of magnitude greater daily volume. This suggests that popular prediction markets may be more robust than would be expected from their small transaction volumes. It would be interesting to study whether this phenomenon holds more broadly, and to understand why.

Chapter 6

Homogeneous risk measures

In this chapter we develop the theory of homogeneous risk measures. Recall that homogeneous risk measures are cost functions that satisfy monotonicity, convexity, and positive homogeneity. As we motivate in this chapter, only homogeneous risk measures provide a *proportional price response* to the actions of traders. Another way of thinking about homogeneous risk measures is that they are *currency independent*, so that they function in a relatively identical way regardless of whether the currency used is thousands of real dollars or dozens of raffle tickets. Practically, this suggests the same homogeneous risk measure will perform adequately regardless of whether the market it is making has a great deal of activity or not. This is in contrast to standard convex risk measures, which have a fixed amount of liquidity set *a priori*.

We begin by illustrating in detail the OPRS market maker, a homogeneous-risk-measure-analogue to the LMSR. We then use the dual space results from Chapter 3 to necessarily and sufficiently classify the set homogeneous risk measures, and finally we use this dual space definition to construct two new families of homogeneous risk measures.

6.1 The OPRS market maker

The OPRS market maker (an acronym of the authors' names) is a cost function that is closely related to the LMSR. It was originally developed in Othman et al. (2010).

6.1.1 Defining the OPRS

The conventional LMSR cost function can be written as

$$C(\mathbf{q}) = b(\mathbf{q}) \log \left(\sum_i \exp(q_i / b(\mathbf{q})) \right)$$

where $b(\mathbf{q}) = b$ is an exogenously set constant. Instead, the OPRS uses the LMSR cost function, but with a variable $b(\mathbf{q})$ that increases with market volume as follows:

$$b(\mathbf{q}) = \alpha \sum_i q_i$$

where $\alpha > 0$ is a constant. The valid region for the OPRS is the set of n-dimensional vectors with all non-negative components (i.e., the positive orthant), omitting the origin. In order to stay in this region we always *move forward in obligation space*.

6.1.2 Moving forward in obligation space

With a market maker using a path-independent cost function, if it costs more than one dollar to acquire a dollar guaranteed payout, a trader could arbitrage the market maker by selling dollar guaranteed payouts to the market maker for more than a dollar.

One way to get around this problem is to only allow the obligation space to move forward. In this section we present two closely related ways to accomplish this goal.

No selling

In this scheme, traders always purchase shares on outcomes from the market maker. Formally, let the market be at state $\mathbf{q^0}$, and let a trader attempt to impose an obligation $\mathbf{q}$ on the market maker, where

$$\min_i q_i < 0.$$

Let

$$\bar{q} \equiv - \min_i q_i$$

Under the usual cost function scheme, that trader would pay

$$C(\mathbf{q^0} + \mathbf{q}) - C(\mathbf{q^0})$$

but instead, in an always moving forward scheme, the trader pays

$$C(\mathbf{q^0} + \mathbf{q} + \bar{q}\mathbb{1}) - \bar{q} - C(\mathbf{q^0})$$

and the market maker moves to the new state

$$C(\mathbf{q^0} + \mathbf{q} + \bar{q}\mathbf{1})$$

noting that the vector $\mathbf{q} + \bar{q}\mathbf{1}$ consists of all non-negative components. This is what we mean by the market maker always moving forward in obligation space.

This scheme is still fully expressive, because with an exhaustive partition over future events the logical condition of betting against an event is equivalent to the logical condition of betting for its complement. Essentially, traders can take on the same obligations as in a traditional scheme, only they will cost more. Furthermore, if

$$\sum_i p_i(\mathbf{q}) > 1$$

then with this scheme when a trader imposes an obligation and then sells it back to the market maker, the trader ends up with a net loss—just like the markets we see in the real world.

Covered short selling

In this scheme, traders are allowed to sell back to the market maker contracts that they have purchased, but are not allowed to directly short sell contracts to the market maker.

Let $\mathbf{q^t}$ represent the vector of payoffs held by trader t, so that q_i^t represents the amount the market maker will pay out to trader t if the i-th event occurs. In a covered short selling scheme, the cost function operates as usual unless trader t suggests a trade that would result in

$$\min_i q_i^t < 0$$

Then, similar to the no selling scheme discussed above, the trader's payoff vector is translated by $\bar{t} \equiv -\min_i q_i^t$, so that instead the trader acquires the vector

$$\mathbf{q^t} + \bar{t}\mathbf{1}$$

noting that for all events i,

$$(\mathbf{q}^t + \bar{t}\mathbf{1})_i \geq 0$$

Discussion

Even though both schemes use the same cost function, they will produce distinct market makers when paired with the homogeneous risk measures we develop in this chapter. A market maker that allows covered short selling permits a trader to buy and then immediately sell at no net cost. With a no selling scheme, that trader will incur a small loss. Which scheme is better depends on the setting; if the set of traders is sophisticated and profitability is a concern, then the no selling scheme is a better choice because it weakly dominates in terms of revenue for the same set of trades. However, if some traders are unsophisticated and user experience is a concern, then the covered short selling scheme could be a better choice because it will not punish users for mistaken bets that they quickly cancel.

In contrast, convex risk measures operating with either scheme or with no scheme at all produce exactly the same quoted costs. Let C be a Hanson market maker. Then because C is translation invariant

$$C(\mathbf{q^0} + \mathbf{q} + \bar{q}\mathbf{1}) - \bar{q} = C(\mathbf{q^0} + \mathbf{q} + \bar{q}\mathbf{1} - \bar{q}\mathbf{1}) = C(\mathbf{q^0} + \mathbf{q})$$

6.1.3 Properties of the OPRS

Even though our modification to the LMSR is simple, it results in a cascade of intriguing properties.

Prices

In a path-independent market maker, the price of state i is given by the partial derivative of the cost function along i. With constant b, this expression is simply

$$p_i(\mathbf{q}) = \frac{\exp(q_i/b)}{\sum_j \exp(q_j/b)}$$

When $b(\mathbf{q}) = \alpha \sum_i q_i$, however, the expression becomes more complex, but still analytically expressible:

$$p_i(\mathbf{q}) = \alpha \log\left(\sum_j \exp(q_j/b(\mathbf{q}))\right) + \frac{\sum_j q_j \exp(q_i/b(\mathbf{q})) - \sum_j q_j \exp(q_j/b(\mathbf{q}))}{\sum_j q_j \sum_j \exp(q_j/b(\mathbf{q}))}$$

Figure 6.1 illustrates the liquidity sensitivity of these prices in a 2-event market. As the number of shares of the complementary event increases, the market's price response for a fixed-size investment becomes less pronounced.

Figures 6.2 and 6.3 show the price of a one-unit bet at various levels of liquidity in a two-event market. Figure 6.2 shows the price of a one-unit bet when the two events have equal quantities

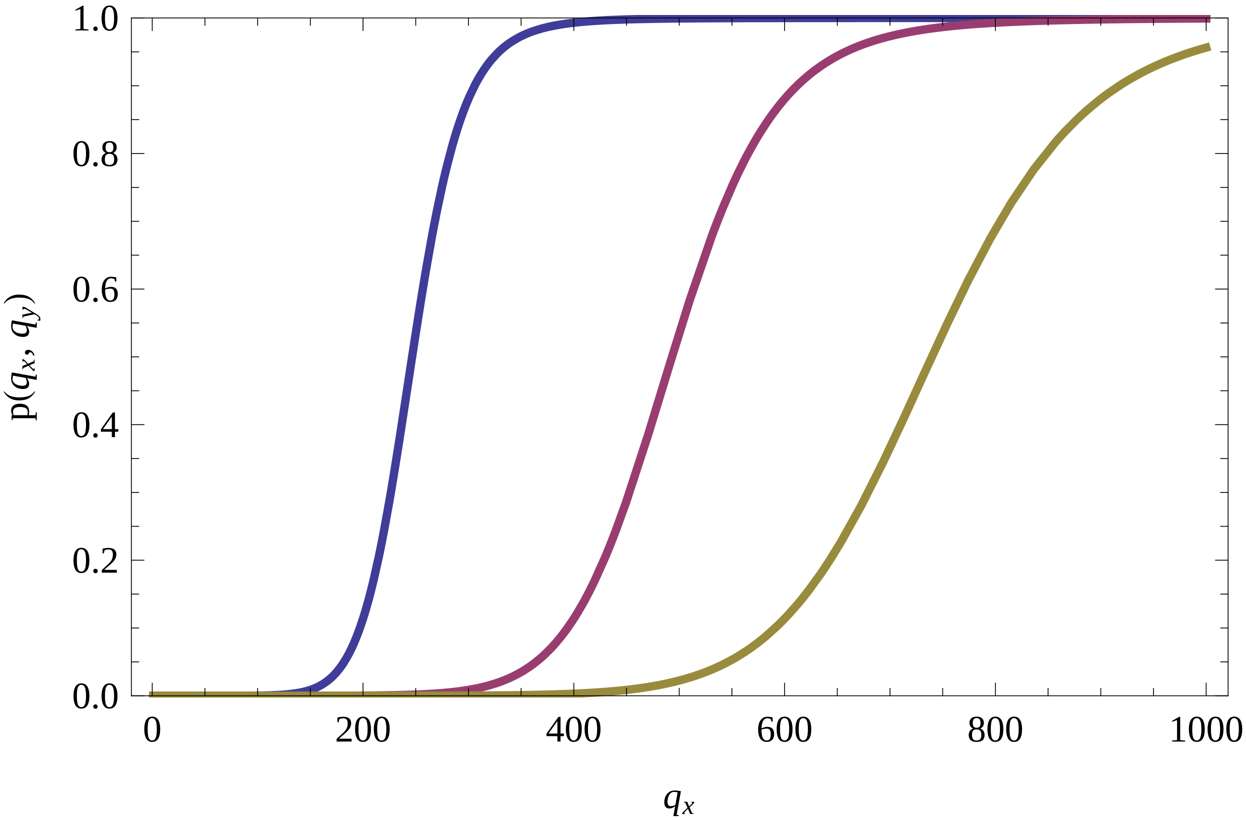


Figure 6.1: In a two-event market with $\alpha = .05$, this plot illustrates the relationship between q_x and p_x for $q_y = 250$, 500, and 750, respectively. The liquidity sensitivity of the OPRS is evident in the decreasing slope of the price response for increasing q_y.

outstanding, while Figure 6.3 has the first event with proportionately higher quantities outstanding. Thus, the unit bet is more expensive in the former than the latter. Though the two figures differ quantitatively, they agree qualitatively: the price of a fixed-size contract shrinks as the level of outstanding quantities increase.

Figures 6.2 and 6.3 also illustrate an important distinction in the OPRS between *instantaneous prices* and *cumulative prices*. Even though, as we show in the next section, the sum of instantaneous prices (i.e., the marginal price for a vanishingly small quantity) is bounded quite modestly for all possible outstanding quantities, at low levels of liquidity these instantaneous prices increase quite quickly. Thus at very small outstanding quantities the cost of a unit bet is more than 90 cents, because the OPRS is very sensitive to bets of large size relative to the quantities outstanding. At higher levels of outstanding quantities, an additional unit bet is relatively small and cumulative prices do not increase much past instantaneous prices.

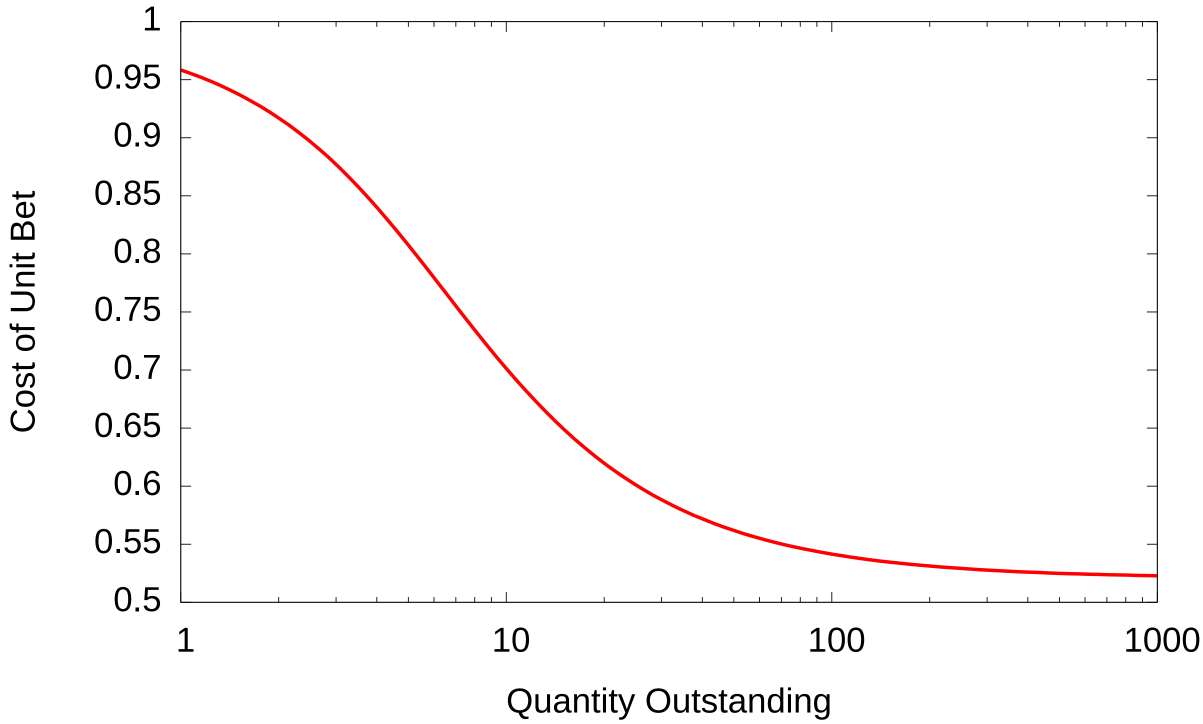


Figure 6.2: In a two-event market with $\alpha = .05$, this plot illustrates the cost of a unit bet on the first outcome when both outcomes have the designated outstanding quantity.

Tight Bounds on the Sum of Prices

In this section, we establish tight bounds on the sum of prices. In particular, we show that

$$1 \approx 1 + n\left[\alpha \log(\exp(1/\alpha) + n - 1) - \frac{\exp(1/\alpha)}{\exp(1/\alpha) + n - 1}\right] \leq \sum_i p_i(\mathbf{q}) \leq 1 + \alpha n \log n$$

Prices achieve their upper bound only when $\mathbf{q} = k\mathbf{1}$ for $k > 0$. Recall that $\mathbf{1}$ is the vector where each element is a 1, so the product $k\mathbf{1}$ yields a vector where each element is a k. Prices achieve the lower bound as $q_i \to \infty$.

Proposition 15. *Prices at $k\mathbf{1}$, for all $k > 0$, sum to $1 + \alpha n \log n$.*

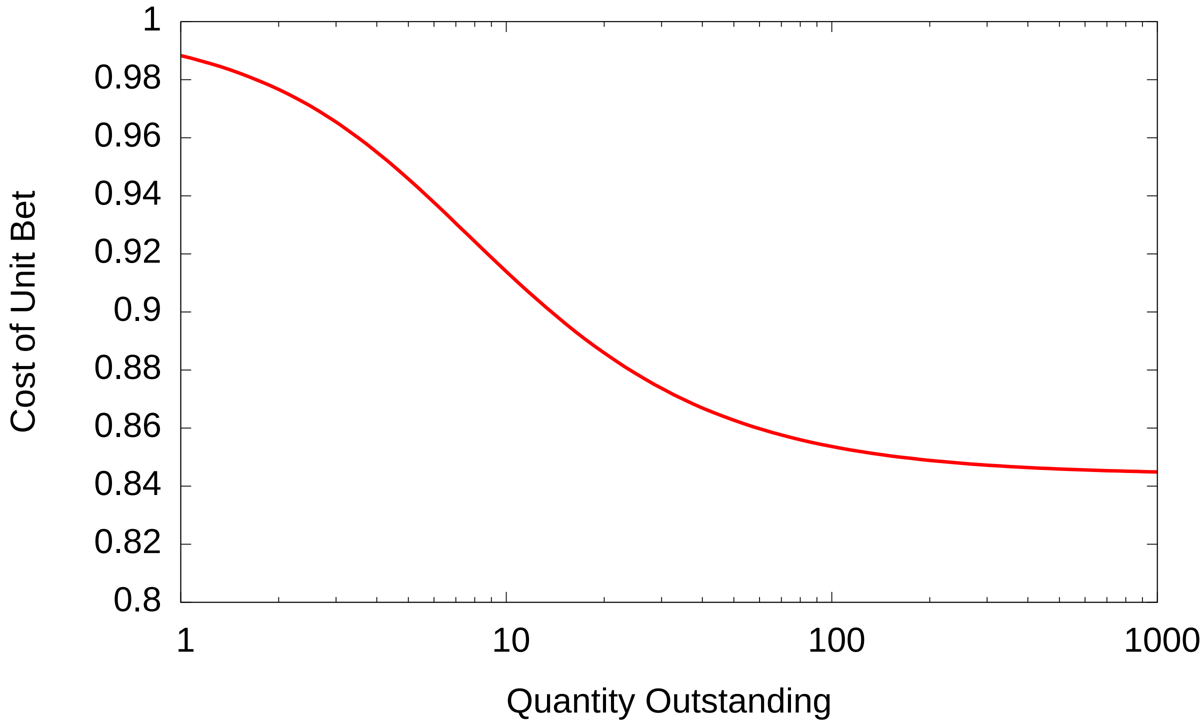


Figure 6.3: In a two-event market with $\alpha = .05$, this plot illustrates the cost of a unit bet on the first outcome when the first outcome has ten percent greater quantity outstanding than the second outcome, where the second outcome's quantity is listed (i.e., a value of 10 corresponds to $(11, 10)$).

Proof. For $\mathbf{q} = k\mathbf{1}$, we have $q_i = q_j$ for all i and j, which allows us to simplify considerably.

$$\begin{aligned}
\sum_i p_i(k\mathbf{1}) &= \sum_i \alpha \log\left(\sum_j \exp(q_j/b(\mathbf{q}))\right) \\
&= n\alpha \log\left(\sum_j \exp(q_j/b(\mathbf{q}))\right) \\
&= n\alpha \log\left(n \exp\left(\frac{1}{\alpha n}\right)\right) \\
&= n\alpha \log\left(\exp\left(\frac{1}{\alpha n}\right)\right) + n\alpha \log n \\
&= 1 + \alpha n \log n
\end{aligned}$$

■

Proposition 16. *The maximum of the sum of prices is obtained at every point of the form $k\mathbb{1}$, where $k > 0$. Furthermore, these are the only points that achieve the maximum.*

Proof. Consider the set of all quantity vectors that sum to $\bar{b} > 0$. We will show that the quantity vector where each event has equal quantity (each one having $\bar{b}/n$) maximizes the sum of prices.

The sum of prices at quantity vector $\mathbf{q}$ is given by

$$\sum_i p_i(\mathbf{q})$$

Without loss of generality, take $\sum_i q_i = 1/\alpha$, so that the space of vectors we consider are those for which $b(\mathbf{q}) = 1$.

So without loss of generality we can rewrite the sum of prices as

$$1 + n\alpha \left[\log\left(\sum_j \exp(q_j)\right) - \frac{\sum_j q_j \exp(q_j)}{\sum_j \exp(q_j)}\right]$$

We will show that

$$\log\left(\sum_j \exp(q_j)\right) - \frac{\sum_j q_j \exp(q_j)}{\sum_j \exp(q_j)} \le \log n,$$

with equality occurring only when $\mathbf{q} = k\mathbb{1}$. We can rewrite the above expression as

$$\sum_j q_j \exp(q_j) \ge \left(\sum_j \exp(q_j)\right) \log\left(\frac{\sum_j \exp(q_j)}{n}\right)$$

Take $p_j \equiv \exp(q_j)$. The expression then becomes

$$\sum_j p_j \log(p_j) \ge \sum_j p_j \log\left(\frac{\sum_j p_j}{n}\right)$$

Without loss of generality, we can scale the p_j to define a probability distribution, to get

$$\begin{aligned}\sum_j p_j \log(p_j) &\ge \log\left(\frac{\sum_j p_j}{n}\right)\\ &\ge -\log(n)\end{aligned}$$

This is a result from basic information theory, which establishes that the uniform distribution has maximum entropy over all possible probability distributions (Cover and Thomas, 2006). Therefore, equality holds only in the case of a uniform distribution, which corresponds to the quantity vector having equal components ($\mathbf{q} = k\mathbf{1}$). ∎

Proposition 17. *At any valid* $\mathbf{q}$, $\sum_i p_i(\mathbf{q}) \geq 1$.

Proof. Define

$$r_i \equiv \frac{q_i}{b(\mathbf{q})}$$

and

$$s_i \equiv \frac{\exp(r_i)}{\sum_j \exp(r_j)}$$

Observe that the s_i form a probability distribution. Then using the entropy operator H:

$$H(\mathbf{x}) = -\sum_i x_i \log x_i$$

we can express prices as

$$p_i(\mathbf{q}) = s_i + \alpha H(\mathbf{s}) \tag{6.1}$$

and therefore the sum of prices as

$$\sum_i p_i(\mathbf{q}) = 1 + \alpha n H(\mathbf{s}) \geq 1.$$

Because the entropy operator is bounded below by zero, the sum of prices is at least 1. ∎

There are two ways to produce a zero entropy distribution of the s_i in the above result.

- Were the OPRS defined over all of $\mathbb{R}^n$, we could produce a zero entropy distribution by sending $q_i \to \infty$ and $q_j \to -\infty$ for $i \neq j$. However, the OPRS is not defined over all of $\mathbb{R}^n$, but rather only in the positive orthant.

- As $\alpha \downarrow 0$, the entropy of the distribution of the s_i can approach 0. Letting q_i be positive and $q_j = 0$ for $j \neq i$, we have

  $$r_i = 1/\alpha \qquad \text{and} \qquad r_j = 0$$

  and therefore

  $$s_i = \frac{\exp(1/\alpha)}{\exp(1/\alpha) + n - 1} \qquad s_j = \frac{1}{\exp(1/\alpha) + n - 1}$$

  a distribution which, for fixed n, approaches a unit mass on s_i as $\alpha \downarrow 0$.

Consequently, for fixed positive α, the distribution of the s_i can have nearly zero entropy, but cannot achieve absolutely zero entropy. Thus the minimum sum of prices is not unity but rather very close to it, equal to unity to first order and well within machine precision for small values of α. The following proof formalizes this.

Proposition 18. *The minimum sum of prices is*

$$1+n\left[\alpha\log(\exp(1/\alpha)+n-1)-\frac{\exp(1/\alpha)}{\exp(1/\alpha)+n-1}\right].$$

This minimum is achieved when $q_i>0$ and $q_j=0$ for $i\neq j$. For small $\alpha\gtrsim 0$,

$$1+n\left[\alpha\log(\exp(1/\alpha)+n-1)-\frac{\exp(1/\alpha)}{\exp(1/\alpha)+n-1}\right]=1+O\left(\alpha^2\right).$$

Proof. From our result above, the minimum sum of prices is achieved when the distribution of the s_i has minimum entropy. When restricted to the positive orthant, the corresponding distribution with largest entropy sets one q_i to be positive and the other $q_j=0$ where $j\neq i$.

At these values, we have

$$p_i(\mathbf{q})=\alpha\log(\exp(1/\alpha)+n-1)$$

and

$$p_j(\mathbf{q})=\alpha\log(\exp(1/\alpha)+n-1)+\frac{1-\exp(1/\alpha)}{\exp(1/\alpha)+n-1}$$

Observe that $p_i\approx 1$ and $p_j\approx 0$.

Adding these terms together and simplifying we get that the sum of prices is

$$1+n\left[\alpha\log(\exp(1/\alpha)+n-1)-\frac{\exp(1/\alpha)}{\exp(1/\alpha)+n-1}\right].$$

Within the braces, the left term is larger than unity while the right term is smaller than unity, meaning that the sum of prices as a whole is greater than unity, which is to be expected from our previous result.

As we will discuss, it is natural for α to be set very small. Let

$$f(\alpha)=\alpha\log(\exp(1/\alpha)+n-1)-\frac{\exp(1/\alpha)}{\exp(1/\alpha)+n-1}.$$

Then the Taylor series of the sum of prices on the axes, taken around $\alpha = 0$, is given by

$$\sum_i p_i = 1 + f(0) + \alpha f'(0) + O\left(\alpha^2\right)$$

Since

$$\lim_{\alpha \downarrow 0} \alpha \log(\exp(1/\alpha) + n - 1) - \frac{\exp(1/\alpha)}{\exp(1/\alpha) + n - 1} = 1 - 1 = 0$$

the $f(0)$ term of the expression is zero, meaning that the total deviation away from 1 for small α is given by the term $\alpha f'(0)$. The derivative is a complex expression that we give for completeness:

$$f'(\alpha) = n\left(\frac{e^{1/\alpha}}{\alpha^2(n + e^{1/\alpha} - 1)} - \frac{e^{2/\alpha}}{\alpha^2(n + e^{1/\alpha} - 1)^2} - \frac{e^{1/\alpha}}{\alpha(n + e^{1/\alpha} - 1)} + \log\left(n + e^{1/\alpha} - 1\right)\right)$$

By taking the limit of this expression, we see that

$$\lim_{\alpha \downarrow 0} f'(\alpha) = 0$$

Thus for small α the sum of prices is bounded below by

$$1 + O\left(\alpha^2\right)$$

Put another way, to first order the lower bound of the sum of prices of the OPRS is 1. ■

Figure 6.4 is a plot of the sum of prices in a simple two-quantity market. Prices achieve their highest sum when $q_x = q_y$ and are bounded below by 1.

Selecting α

A possible complaint about our scheme is that we have replaced one *a priori* fixed value, b, of the LMSR with another *a priori* fixed value, our α. In this section, we discuss how the α parameter has a natural interpretation that makes its selection relatively straightforward.

The α parameter can be thought of as the commission taken by the market maker. Higher values of α correspond to larger commissions, which leads to more revenue. At the same time, setting α too large discourages trade.

As we have shown, the sum of prices with the OPRS is bounded by $1 + \alpha n \log n$, and this value is achieved only when all quantities are equal. This bound provides a guide to help set α.

How large should administrators set α within the OPRS? We can look to existing market makers (and bookies) for an answer. Market makers generally operate with a commission of somewhere

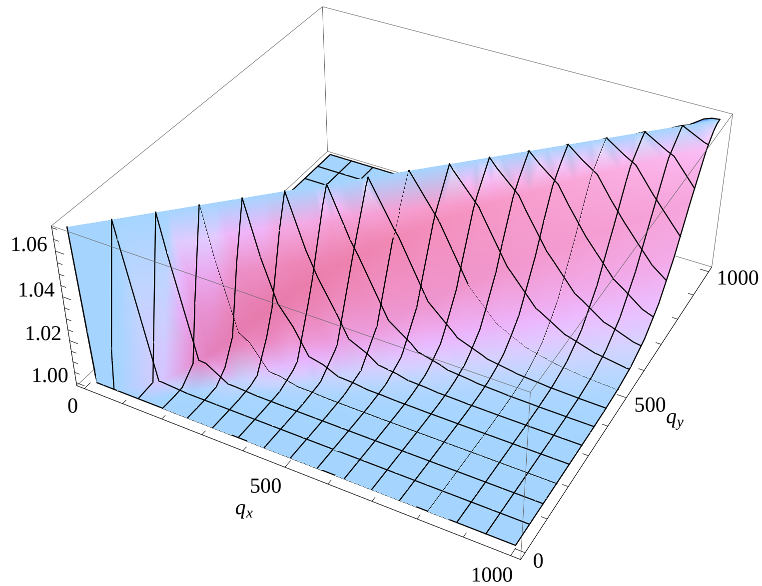


Figure 6.4: Sum of prices where $n = 2$ and $\alpha = 0.05$. The sum is bounded between 1 and $1 + \alpha n \log n \approx 1.07$, achieving its maximum where $q_x = q_y$.

between 2 and 20 percent. To emulate a commission that does not exceed v in the OPRS, the market administrator can simply set

$$\alpha = \frac{v}{n \log n}$$

So, the larger the event space (larger n), the smaller α should be set to maintain a given percentage commission.

Though the sum of prices increases in α, this provides no guidance as to the behavior of the cost function itself—it is not immediate that the cost function increases in α, because it has conflicting effects within the OPRS. Increasing α decreases the terms $q_i/b(\mathbf{q})$ in the cost function, but scales up the output of the log function. However, the following proposition establishes that the OPRS is non-decreasing in α. We are assisted in this result by the following lemma.

Lemma 1. *For the OPRS*

$$C(\mathbf{q}) \geq \max_i q_i$$

Proof. Suppose there exists a valid $\mathbf{q}$ such that

$$C(\mathbf{q}) < \max_i q_i$$

without loss of generality, let

$$q_1 = \max_i q_i$$

and define

$$r_i = \frac{q_i}{b(\mathbf{q})} \geq 0$$

then we have

$$\begin{aligned} \log\left(\sum_i \exp(r_i)\right) &< r_1 \\ \sum_i \exp(r_i) &< \exp(r_1) \\ \sum_{i \neq 1} \exp(r_i) &< 0 \end{aligned}$$

which is a contradiction because $\exp(x)$ is non-negative for all x. ■

Proposition 19. *The OPRS is non-decreasing in α.*

Proof. This result follows if we can show

$$\frac{\partial}{\partial \alpha} C(\mathbf{q}) \geq 0$$

After taking the partial derivative of the OPRS and simplifying, we get

$$\left(\sum_i \exp(q_i/b(\mathbf{q}))\right) C(\mathbf{q}) \geq \sum_i q_i \exp(q_i/b(\mathbf{q}))$$

From Lemma 1 we have

$$C(\mathbf{q}) \geq \max_i q_i$$

and so

$$\left(\sum_i \exp(q_i/b(\mathbf{q}))\right) C(\mathbf{q}) \geq \left(\sum_i \exp(q_i/b(\mathbf{q}))\right)\left(\max_i q_i\right) \geq \sum_i q_i \exp(q_i/b(\mathbf{q}))$$

which completes the proof. ■

Recalling that the cost function defines the amount paid into the market maker, an informal way to interpret this result is that the market maker's revenue increases with the α parameter for any given quantity vector. Of course, increasing α results in higher prices, which can affect trader behavior, so the overall effect in practice might be ambiguous.

Bounded Loss

Like the LMSR, the OPRS has bounded loss.

Proposition 20. *The OPRS has bounded loss. Specifically, its loss is bounded by* $C(\mathbf{q^0})$.

Proof. By Lemma 1

$$C(\mathbf{q}) \geq \max_i q_i$$

and so

$$\max_i q_i - C(\mathbf{q}) \leq 0 \quad \Rightarrow \quad C(\mathbf{q^0}) + \max_i q_i - C(\mathbf{q}) \leq C(\mathbf{q^0})$$

so the OPRS's loss is bounded by

$$C(\mathbf{q^0})$$

■

Since

$$\lim_{\mathbf{q}\to\mathbf{0}} C(\mathbf{q}) = 0,$$

setting the initial market quantities close to **0**, the worst-case loss becomes arbitrarily small. But reducing the initial vector too much comes at a cost, however, because

$$\lim_{\mathbf{q}\to\mathbf{0}} b(\mathbf{q}) = 0$$

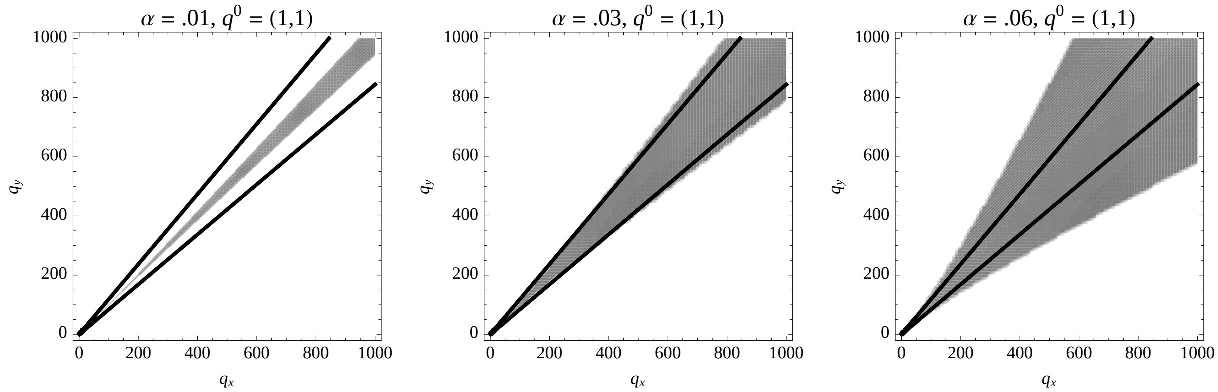
so the market becomes arbitrarily sensitive to small bets in its initial stage.

In contrast, to get near-zero loss in the LMSR, one would have to set b near zero, which would cause arbitrary sensitivity to small bets throughout the duration of the market. Since other convex risk measures are not liquidity sensitive either, they suffer from the same problem. In the OPRS, by setting the initial quantities close to zero, we achieve near-zero loss while containing the high sensitivity to the initial stage only.

Worst-Case Revenue

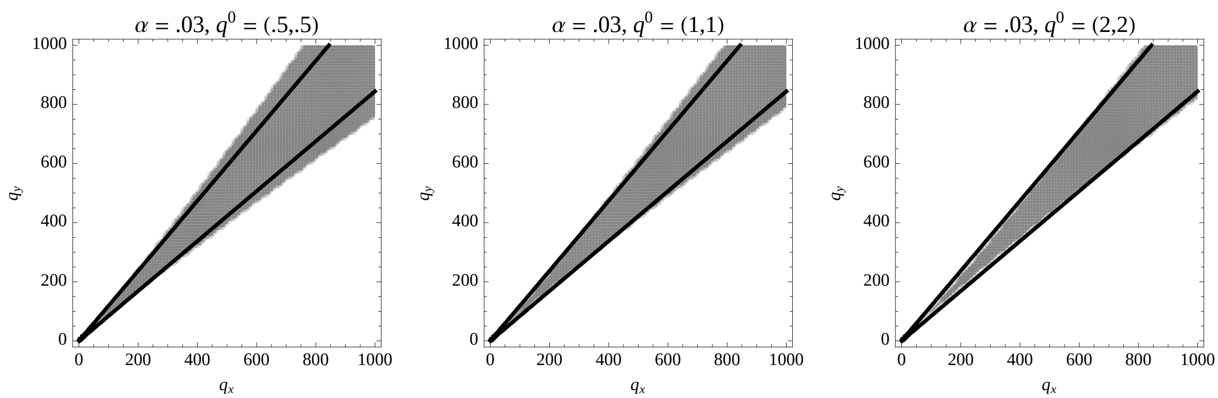
In addition to always having bounded loss (and near-zero loss if desired), under broad conditions on the final quantity vector of the market, we can guarantee that the OPRS actually makes a profit (regardless of which event gets realized). The worst-case revenue is

$$\mathfrak{R}(\mathbf{q}) \equiv C(\mathbf{q}) - \max_i q_i - C(\mathbf{q^0})$$



(a) (b) (c)

Figure 6.5: The shaded regions show where the market maker has outcome-independent profit in a two-outcome market with initial quantity vector $(1, 1)$ and various values of α. Figure (a) sets α equal to $.01$, Figure (b) equal to $.03$, and Figure (c) equal to $.06$. The top black ray represents $p_y = .95$ and the bottom black ray represents $p_x = .95$.



(a) (b) (c)

Figure 6.6: The shaded regions show where the market maker has outcome-independent profit in a two-outcome market with $\alpha = .03$ and various initial quantity vectors. Figure (a) sets $\mathbf{q^0}$ equal to $(.5, .5)$, Figure (b) equal to $(1, 1)$, and Figure (c) equal to $(2, 2)$. The top black ray represents $p_y = .95$ and the bottom black ray represents $p_x = .95$.

If $\mathfrak{R}(\mathbf{q}) > 0$ when the market closes, the market maker will book a profit regardless of the outcome that is realized. We say that in such states the market maker has *outcome-independent profit*. Figures 6.5 and 6.6 show the set of market states for which $\mathfrak{R}(\mathbf{q}) > 0$ for various values of α and initial quantity vectors $\mathbf{q}^0$.

Figure 6.5 shows varying values of α. From Theorem 19, the cost function is non-decreasing in α, which is reflected by the increasing areas of outcome-independent profit as α gets larger. Figure 6.6 shows varying initial quantity vectors. Since revenue is trivially decreasing in the cost of the initial quantity vector, as the cost of our initial quantity vector increases, the area of outcome-independent profit shrinks.

From the figures, it might appear that large portions of the state space will result in the OPRS losing money. However, prices and quantities have a highly non-linear relationship: prices quickly approach 1 as quantities become imbalanced. The straight black rays on the plane represent a price of $.95$ for one of the two events. Therefore, the plots indicate that as long as markets are terminated while events have reasonable levels of uncertainty, the market maker can book a profit regardless of the realized future .

Figure 6.7 contrasts the revenue of the OPRS against the LMSR. In particular, the figure shows the revenue *surplus* of the LMSR relative to the OPRS. Positive values represent how much more the OPRS would collect if the market terminates in the each obligation state. The comparison between the two market makers is valid because both the market makers have the same bound on worst-case loss, set by aligning the α and $\mathbf{q}^0$ parameters in the OPRS with the b parameter in the LMSR. What is especially notable is how large the revenue difference between the two market makers becomes for lopsided obligation vectors, when the market maker has to pay out much more if one event happens than if the other event happens. As Figures 6.5 and 6.6 showed, generally at lopsided obligation vectors the OPRS does not book an outcome-independent profit. However, as Figure 6.7 shows, the OPRS delivers significantly less loss than the LMSR for lopsided obligation vectors.

Homogeneity

Recall that a positive homogeneous function f of degree k has

$$f(\gamma x) = \gamma^k f(x)$$

for $\gamma > 0$. "Positive homogeneous functions of degree one" are often referred to as just "positive homogeneous". As it turns out, the cost function of the OPRS is positive homogeneous, and in this section we prove and explore the implications of that result.

Proposition 21. *The OPRS is positive homogeneous of degree one.*

Proof. Let $\gamma > 0$ be a scalar and $\mathbf{q}$ be some valid quantity vector. Without loss of generality, we

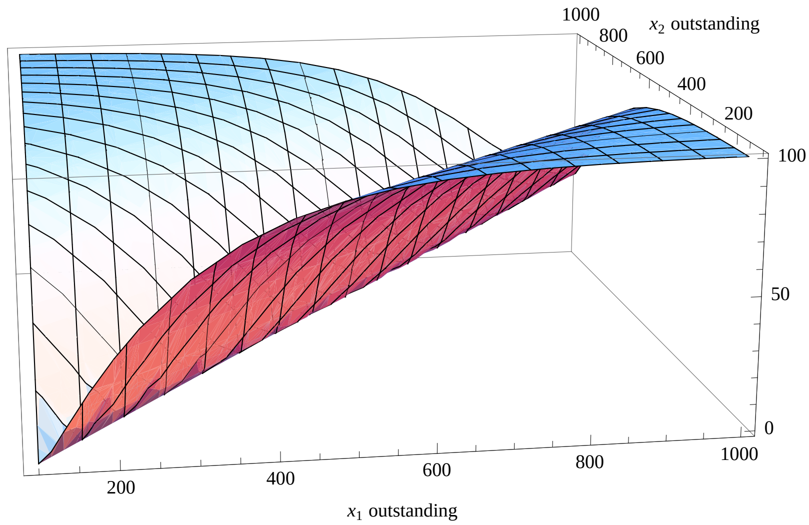


Figure 6.7: Revenue comparison between the OPRS and the LMSR. The z-axis is how much more the OPRS makes than the LMSR. The parameters are aligned so that the two market makers have the same worst-case loss ($\sim$ 104.2), reflected by the zero revenue surplus at $\mathbf{q^0}$. In the OPRS, $\alpha = .03$ and $\mathbf{q^0} = (100, 100)$, and in the LMSR, $b = 150.27$.

can assume $\sum_i q_i = 1$. Then

$$\begin{aligned} C(\gamma \mathbf{q}) &= b(\gamma \mathbf{q}) \log \left(\sum_i \exp(\gamma q_i / b(\gamma \mathbf{q})) \right) \\ &= \gamma \alpha \log \left(\sum_i \exp \left(\frac{\gamma q_i}{\gamma \alpha} \right) \right) \\ &= \gamma C(\mathbf{q}) \end{aligned}$$

■

It is crucial that the cost function be positive homogeneous, because that allows the price response to scale appropriately in response to increased quantities. One of the primary concerns about using the LMSR is the relation of the fraction of wealth invested in the market to the displayed prices. If the b parameter is set too low in the LMSR, that is, if the market is thick but the market maker's price response is too sensitive, then tiny fractions of the overall wealth in the market can move prices a great deal. On the other hand, if the b parameter is set too high all the wealth in the market would be insufficient to move prices significantly enough to reflect this skewed distribution of bets.

A market maker would ideally provide a price response proportional to the amount of wealth in the market. Such a market maker would appropriately scale liquidity, requiring progressively larger trades to achieve the same price response as the market accumulated more and more money. Scaling price responses proportional to the state of the market is the correct liquidity-sensitive behavior because it yields a relative price response that is the same regardless of whether the amount of money in the market is tens, thousands, or millions of dollars. Another way of thinking about this property is that a proportional-scaling market maker is currency independent: without any further adjustment it will function equally as well regardless of whether trading is done in millions of yen or fractions of a dollar, because only the relative, rather than absolute, amounts wagered affect the market maker's price response. This leads us to the following definition.

Definition 22. Prices *scale proportionately* if

$$p_i(\mathbf{q}) = p_i(\gamma \mathbf{q})$$

for all i, $\mathbf{q}$ and scalar $\gamma > 0$.

In fact, only homogeneous cost functions provide this price response.

Proposition 22. *Prices scale proportionately if and only if the cost function is positive homogeneous of degree one.*

Proof. Proportional scaling is equivalent to the price functions being positive homogeneous of degree zero. Since the k-th derivative of a positive homogeneous function of degree d is itself a positive homogeneous function of degree $d - k$, if and only if the cost function is positive homogeneous of degree one will prices scale proportionately. ■

We can now prove that the OPRS is a homogeneous risk measure.

Proposition 23. *The OPRS is a homogeneous risk measure (for vectors in the non-negative orthant).*

Proof. We must prove that the OPRS satisfies positive homogeneity, convexity, and monotonicity. Proposition 21 shows that the OPRS satisfies positive homogeneity.

Monotonicity also holds, because it is possible to write individual prices in the OPRS as the sum of non-negative components (e.g., Equation 6.1), and a function with well-defined positive partial derivatives satisfies monotonicity.

Convexity of the OPRS follows from the relation of the OPRS to the perspective function of the convex *log-sum-exp function*:

$$\log\left(\sum_i \exp x_i\right)$$

The perspective function g of a convex function f is defined as

$$g(\mathbf{z}, t) \equiv t f(\mathbf{z}/t)$$

for $t > 0$. The perspective function is convex in both $\mathbf{z}$ and t (Boyd and Vandenberghe, 2004). Now let f be the log-sum-exp function, which is convex. Then consider the relation between $g(\mathbf{x}, \alpha \sum_i x_i)$ and $g(\mathbf{y}, \alpha \sum_i y_i)$. Since the perspective function is convex in both of its arguments, we have for all $\lambda \in [0, 1]$:

$$\lambda g\left(\mathbf{x}, \alpha \sum_i x_i\right) + (1-\lambda) g\left(\mathbf{y}, \alpha \sum_i y_i\right) \geq g\left(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}, \alpha \sum_i \lambda x_i + (1-\lambda) y_i\right)$$

But observe that

$$C(\mathbf{z}) \equiv g\left(\mathbf{z}, \alpha \sum_i z_i\right)$$

and so

$$\lambda C(\mathbf{x}) + (1-\lambda) C(\mathbf{y}) \geq C(\lambda\mathbf{x} + (1-\lambda)\mathbf{y})$$

which proves convexity of the OPRS. ■

6.2 Dual space results

Recall from Chapter 3 that homogeneous risk measures have a particularly concise representation in dual price space. Specifically, we can represent homogeneous risk measures with a compact convex set $\mathbb{S}$ in the non-negative orthant that provides its support in dual space. The relations between convex and monotonic cost functions, convex and positive homogeneous cost functions, and their respective duals are a consequence of well-known results in the convex analysis literature (Rockafellar, 1966, 1970).

Proposition 24. *A risk measure is convex and monotonic if and only if the set* $\mathbb{S}$ *is exclusively within the non-negative orthant.*

Proposition 25. *A risk measure is convex and positive homogeneous if and only if its convex conjugate has compact* $\mathbb{S}$ *and has* $f(\mathbf{y}) = 0$ *for every* $\mathbf{y} \in \mathbb{S}$.

In the literature this latter result relates *indicator sets* (here, the set $\mathbb{S}$) to *support functions* (here, the cost function). Since $f(\mathbf{y}) = 0$ for all $\mathbf{y} \in \mathbb{S}$, the cost function conjugacy is defined only by the set $\mathbb{S}$. Consequently, we will abuse terminology slightly and refer to the cost functions as conjugate to the convex compact set alone. A necessary and sufficient condition on the set of homogeneous risk measures follows.

Corollary 3. *A cost function is a homogeneous risk measure if and only if it is conjugate to a compact convex set in the non-negative orthant.*

Only the convex conjugate set $\mathbb{S}$ of a homogeneous risk measure is responsible for determining the market maker's behavior, because the conjugate function f takes value zero everywhere in that set. In this section, we explore two features of the conjugate set that produce desirable properties: its *curvature* and its *divergence* from the probability simplex. We then discuss these properties in relation to the OPRS in Section 6.2.3.

6.2.1 Curvature

We would like for the OPRS to always be differentiable (outside of $\mathbf{0}$, where a derivative of a positive homogeneous function will not generally exist). The OPRS is differentiable in the non-negative orthant (again, excepting $\mathbf{0}$) while max is differentiable only when the maximum is unique. In this section, we show that only curved conjugate sets produce homogeneous risk measures that are differentiable. (It might be argued that what we are really interested in, particularly if we claim that curved sets act as a regularizer in the price response, is whether or not curved sets also imply *continuous* differentiability of the cost function. Continuous differentiability would mean that prices both exist and are continuous in the quantity vector. These conditions are in fact the same for convex functions defined over an open interval (such as $\mathbb{R}^n \backslash \mathbf{0}$), because for such functions differentiability implies continuous differentiability (Rockafellar, 1970).)

Definition 23. A closed, convex set $\mathbb{S}$ is *strictly convex* if its boundary does not contain a non-degenerate line segment. Formally, let $\partial\mathbb{S}$ denote the boundary of the set. Let $0 \leq \lambda \leq 1$ and $\mathbf{x}, \mathbf{x}' \in \partial\mathbb{S}$. Then $\lambda\mathbf{x} + (1-\lambda)\mathbf{x}' \in \partial\mathbb{S}$ holds only for $\mathbf{x} = \mathbf{x}'$.

Since strictly convex sets are never linear on their boundary they can be thought of as sets with curved boundaries.

Proposition 26. *A homogeneous risk measure is differentiable on* $\mathbb{R}^n \backslash \mathbf{0}$ *if and only if its conjugate set is strictly convex.*

Proof. First recall that the maximizing argument of the optimization for vector $\mathbf{x} \neq \mathbf{0}$ is given by an extreme hyperplane normal to $\mathbf{x}$ that intersects with the set $\mathbb{S}$. It follows that the maximizing argument is always on the boundary of $\mathbb{S}$.

For the forward direction, consider a set that is convex but not strictly convex. Then there exists a hyperplane H connecting two points $\mathbf{x}$ and $\mathbf{x}'$ on the boundary of the set such that

$$\lambda \mathbf{x} + (1 - \lambda)\mathbf{x}' \in \partial \mathbb{S}$$

Now consider the set of points normal to H which go through $\mathbf{0}$. Observe that these points lie in precisely two orthants. The points in one of those orthants will find that the set of points between $\mathbf{x}$ and $\mathbf{x}'$ are the maximizing arguments in the optimization, which means the subgradient at those points is not unitary, and so the cost function is not differentiable.

Now consider a strictly convex set $\mathbb{S}$. Since the optimization is convex, the maximizing arguments must form a convex set. Therefore if more than one vector is in a maximizing argument, the line connecting those vectors must be on the boundary of $\mathbb{S}$. But then $\mathbb{S}$ is not strictly convex, a contradiction. ∎

6.2.2 Divergence from probability simplex

The amount of divergence from the probability simplex governs the market maker's divergence from translation-invariant prices (i.e., prices that sum to unity). Recall that max is the homogeneous risk measure that is defined only over the probability simplex.

Proposition 27. *Let $\mathbb{S}$ be the dual set of a differentiable homogeneous risk measure. Then the maximum sum of prices (the most a trader would ever need to spend for a unit guaranteed payout) is given by $\max_{\mathbf{y} \in \mathbb{S}} \sum_i y_i$, and the minimum sum of prices (the most the market maker would ever pay for a unit guaranteed payout) is given by $\min_{\mathbf{y} \in \mathbb{S}} \sum_i y_i$.*

Proof. Recall that the maximizing argument in the maximization yields the gradient of the cost function. The point in $\mathbb{S}$ with the largest sum of components (and therefore the largest sum of prices) is selected as the maximizing argument for $\mathbf{x} = k\mathbf{1}$, $k > 0$. The minimum sum of prices result holds by similar logic; the point in $\mathbb{S}$ with minimum sum of components is selected for $\mathbf{x} = -k\mathbf{1}$. ∎

Given any (efficiently representable) convex set corresponding to a differentiable homogeneous risk measure, the extreme price sums can be solved efficiently, since it is a convex optimization over a convex set.

It was shown in Proposition 16 that the OPRS achieved its maximum sum of prices for quantity vectors that are scalar multiples of $\mathbf{1}$. A corollary of the above result is that this property holds for every homogeneous risk measure. (Other vectors may also achieve the same sum of prices.)

Corollary 4. *In a homogeneous risk measure every vector that is a positive multiple of* $\mathbf{1}$ *achieves the maximum sum of prices.*

In addition to maximum prices, the shape of the convex set also determines the worst-case loss of the resulting market maker. The notion of worst-case loss is closely related to our desideratum of bounded loss—a market maker with unbounded worst-case loss does not have bounded loss, and a market maker with finite worst-case loss has bounded loss.

In homogeneous risk measures, the amount of liquidity sensitivity is proportional to the market's state. Since in practice there is some latent level of interest in trading on the event before the market's initiation, it is desirable to seed the market initially to reflect a certain level of liquidity. It is desirable to have a tight bound on that worst-case loss, reflecting that in practice, market administrators are likely to have bounds on how much the market maker could lose in the worst case. Tight bounds on worst-case loss assure the administrator that that bound will be satisfied with maximum liquidity injected at the market's initiation.

Proposition 28. *Let* $\mathbb{S}$ *be a convex set conjugate to a homogeneous risk measure that includes the unit axes but does not exceed the unit hypercube. Then the worst-case loss of the risk measure is* tightly *bounded by the initial cost of the market's starting point.*

Proof. First, note that any convex set that includes the unit axes is conjugate to a cost function that is at least as large as max, because max is conjugate to the minimal convex set that includes the unit axes and increasing the size of the feasible region never decreases the value of a maximization. Therefore, the worst-case loss is bounded from above by $C(\mathbf{x^0})$. To show that this bound is tight, we need to show that there exists a terminal state $C(\mathbf{x})$ where

$$\max_i x_i = C(\mathbf{x})$$

Such a terminal state is given by the axes. ■

By bringing $\mathbf{x^0}$ as close as desired to $\mathbf{0}$, we have the following corollary, which is a generalization of a similar result for the OPRS.

Corollary 5. *Let $\mathbb{S}$ be a convex set conjugate to a homogeneous risk measure that includes the unit axes. Then the worst-case loss of the risk measure can be set arbitrarily small.*

A bound on prices also emerges from this result.

Corollary 6. *Let $\mathbb{S}$ be a convex set conjugate to a homogeneous risk measure that includes the unit axes but does not exceed the unit hypercube. Then the maximum price on any event is 1.*

One of the most powerful features of the OPRS is that it can achieve an outcome-independent profit, so that given a sufficient level of interaction the market maker would never lose money regardless of realized outcome. This is not true generally for translation-invariant risk measures, but it is a feature of homogeneous risk measures defined outside of the probability simplex, as we proceed to show.

Proposition 29. *Let $\mathbb{S}$ be a convex set conjugate to a homogeneous risk measure. Then if $\mathbb{S}$ includes any element not on the probability simplex, the cost function can make an outcome-independent profit.*

Proof. Let $\mathbf{x}$ be a vector for which $\mathbf{y} \notin \Pi$ is selected as a maximizing argument. Since the homogeneous risk measure is defined over all of $\mathbb{R}^n$, such an $\mathbf{x}$ always exists. Then observe

$$C(\mathbf{x}) - \max_i x_i > 0$$

and because the cost function is positive homogeneous this implies that for every $K > 0$ there exists a $\gamma > 0$ such that

$$C(\gamma\mathbf{x}) - \max_i \gamma x_i > K$$

so the gap between $C(\mathbf{x})$ and $\max_i x_i$ can be made arbitrarily large. Since $C(\mathbf{x^0})$ is finite, there exists a $\gamma\mathbf{x}$ such that

$$C(\gamma\mathbf{x}) - \max_i \gamma x_i - C(\mathbf{x^0}) > 0$$

and so the cost function has outcome-independent profit at the payout vector $\gamma\mathbf{x}$. ■

6.2.3 The OPRS and its conjugate set

For the OPRS, however, we already have a homogeneous risk measure, and in this section we will explore how to produce its conjugate convex set.

Recall that the OPRS is given by

$$C(\mathbf{x}) = b(\mathbf{x}) \log \left(\sum_i \exp(x_i / b(\mathbf{x})) \right)$$

where

$$b(\mathbf{x}) = \alpha \sum_i x_i$$

for $\alpha > 0$ and $\mathbf{x} \in \mathbb{R}^n_+$.

Because the OPRS is monotonic and convex, recall that it must be conjugate to a convex set in the non-negative orthant, $\mathbb{S}$, so

$$C(\mathbf{x}) = \max_{\mathbf{y} \in \mathbb{S}} \mathbf{x} \cdot \mathbf{y}$$

Observe that we cannot solve for this convex set directly, because we only know $\mathbf{x}$ and $C(\mathbf{x})$. However, we can solve for the set numerically. For every $\mathbf{x}$, we can find a hyperplane on which at least one point is the outer boundary of the convex set. We define $h(\mathbf{x})$ as

$$h(\mathbf{x}) \equiv \left\{ \mathbf{p} \mid \mathbf{p} \in \mathbb{R}^n_+ \text{ and } \mathbf{x} \cdot \mathbf{p} \leq C(\mathbf{x}) \right\}.$$

For each $\mathbf{x}$, this partition divides the non-negative orthant into two sets—those points that could be part of the convex set (all the points in $h(\mathbf{x})$, under the separating hyperplane), and those points that could not be part of the convex set (or else $C(\mathbf{x})$ would be larger).

In order to fully recover the convex set $\mathbb{S}$, we need to take the intersection of every $h(\mathbf{x})$:

$$\mathbb{S} = \bigcap_{\mathbf{x} \in \mathbb{R}^n_+} h(\mathbf{x})$$

We can simplify this operation considerably, however. Since the OPRS is positive homogeneous, we need only consider the intersection over the $\mathbf{x}$ in the probability simplex. This is because the same point $\mathbf{y} \in \mathbb{S}$ will solve the maximization problem for all $\gamma\mathbf{x}$, $\gamma > 0$. Therefore, it suffices to only consider the intersection of a set of points $\mathbb{X}$ such that for all $\mathbf{x}' \in \mathbb{R}^n_+$, there exists an $\mathbf{x} \in \mathbb{X}$ such that $\gamma\mathbf{x} = \mathbf{x}'$ for some $\gamma > 0$. One such set $\mathbb{X}$ is the probability simplex. Using this result, we proceed to plot the convex conjugate indicator set of the OPRS with $\alpha = .05$ in two dimensions as Figure 6.8. Observe that, just as theory suggests, the conjugate set is curved and bulges slightly away from the probability simplex.

Because the OPRS is only defined in the non-negative orthant, it is only the outer boundary of the convex set that is relevant to the price response. (This is not the case for cost functions defined over all of $\mathbb{R}^n$, because the outer boundary is never selected for vectors in the negative orthant.) In order to show the divergence of the outer boundary, Figure 6.8 displays the inner boundary of the conjugate set as the probability simplex.

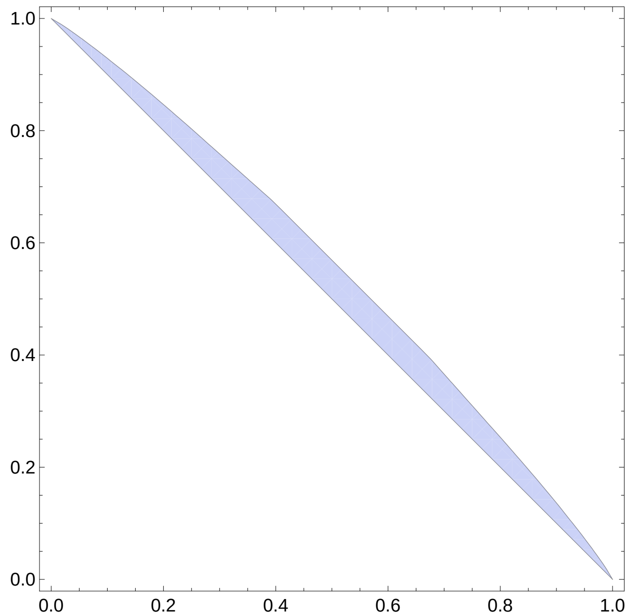


Figure 6.8: The convex set in dual space supported by the OPRS market maker in a simple two-event market. Since that market maker was only defined over the non-negative orthant, it is represented by the curved outer boundary of the set. The inner boundary of the set is drawn here as the probability simplex (i.e., $x + y = 1$) to show how the outer boundary diverges from it.

6.3 Using the dual space

In this section, we proceed to use our theoretical results constructively, to create two families of homogeneous risk measures with desirable properties that the OPRS, the only prior homogeneous risk measure, lacks. These include tight bounds on minimum sum of prices and worst-case losses. Our new families of market makers are parameterized (in much the same way as the OPRS) by the maximum sum of prices. The OPRS is not a member of either family.

6.3.1 Unit ball market makers

One family of homogeneous risk measures is to take as the dual set the intersection of two unit balls in different $\mathcal{L}^p$ norms, one ball at $\mathbf{0}$ and the other ball at $(2/n)\mathbf{1}$. At $n = 2$, the other ball is centered at $\mathbf{1}$.

For $1 < p < \infty$, the intersection of the two balls is a strictly convex set that includes the unit axes but does not exceed the unit hypercube. (At $p = 1$, we get the probability simplex, which is not strictly convex. At $p = \infty$ we get a hypergeometric solid, which is also not strictly convex.)

Let $||\cdot||_p$ denote the $\mathcal{L}^p$ norm. Then we can define the vectors in the intersections of the unit balls, $\mathcal{U}(p)$, as

$$\mathcal{U}(p) \equiv \{\mathbf{y} \mid \mathbf{y} \in \mathbb{R}^n_+, ||\mathbf{y} - (2/n)\mathbf{1}||_p \leq 1, ||\mathbf{y}||_p \leq 1\}$$

Figure 6.9 provides graphical intuition for the set $\mathcal{U}(p)$ in a simple two-event market. This set gives us a cost function

$$C(\mathbf{x}) = \max_{\mathbf{y} \in \mathcal{U}(p)} \mathbf{x} \cdot \mathbf{y}$$

We dub this the *unit ball market maker*. Since we can easily test whether a vector is within both unit balls (i.e., within $\mathcal{U}(p)$), the optimization problem for the cost function can be solved in polynomial time.

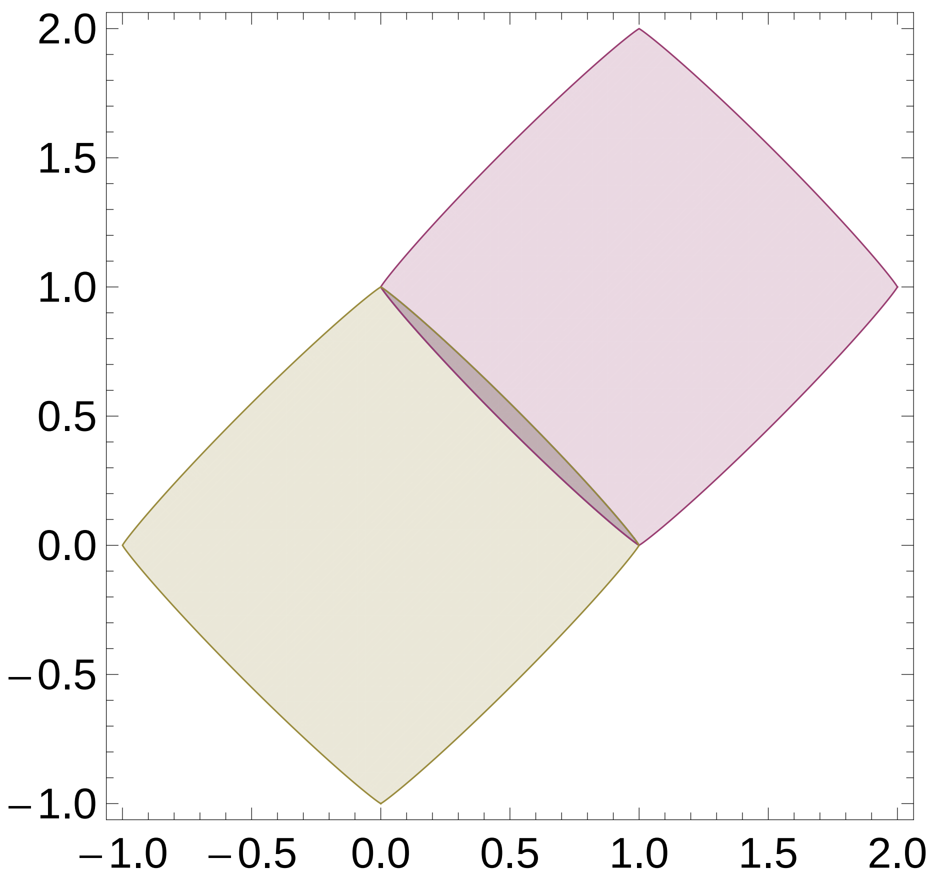


Figure 6.9: The conjugate set of the unit ball market maker in dual space is given by the intersection of two unit balls (the dark region) in a certain $\mathcal{L}^p$ norm. Here, $n = 2$ and $p \approx 1.08$, so by the formula below the maximum sum of prices is 1.05.

This family of market makers is parameterized by the $\mathcal{L}^p$ norm that defines which vectors in dual space are in the convex set. By choosing the value of p correctly, we can engineer a market maker with the desired maximum sum of prices. The outer boundary of the set is defined by the unit ball from **0** in $\mathcal{L}^p$ space. Its boundary along **1** is given by the k that solves $\sqrt[p]{nk^p} = 1$. Solving for k we get

$$k = n^{-1/p}$$

and so the maximum sum of prices is

$$nk = n\left(n^{-1/p}\right) = n^{1-1/p}.$$

For prices that are at most $1 + v$, we can set

$$1 + v = n^{1-1/p}$$

Solving this equation for p yields

$$p = \frac{\log n}{\log n - \log(1 + v)}$$

Given any target maximum level of vigorish, this formula provides the exponent of the unit ball market maker to use. Considering that only small divergences away from unity are natural to the setting, the p we select for our $\mathcal{L}^p$ norm should be quite small. The norm increases in the maximum sum of prices, and for larger n the same norm produces larger sums of prices.

One of the advantages of the unit ball market maker is that it is defined over all of $\mathbb{R}^n$, as opposed to just the non-negative orthant. Its behavior in the positive orthant is to charge agents more than a dollar for a dollar guaranteed payout, because the outer boundary diverges outwards from the probability simplex. Its behavior in the negative orthant, where its points on the inner boundary are selected in the maximization, is to pay less than a dollar for a dollar guaranteed payout. Its behavior in all other orthants is equivalent to max, as the unit axes are selected as maximizing arguments. Finally, if we restrict the unit ball market maker to only the non-negative orthant (like the OPRS), the sum of prices is tightly bounded between 1 and $n^{1-1/p}$.

6.3.2 Optimal homogeneous risk measures

Now that we have given a necessary and sufficient characterization of homogeneous risk measures, we might wonder which homogeneous risk measure is the *best*. In this section, we use the dual space to construct the homogeneous risk measure that is optimal in a specific sense—it has the minimum amount of price deviation as traders change the angle of the the market maker's payout vector. The resulting market maker has a shallower, more consistently curved support set than other homogeneous risk measures.

Parameterization

We assume that we are using an always-moving-forward scheme, as in the OPRS, and that so the price response of the corresponding homogeneous risk measure is determined only by the outer boundary of the convex set, as in Figure 6.8 of the OPRS. We denote by $\partial\mathbb{S}$ the outer boundary of the convex set.

For simplicity, we assume in this section that $n = 2$. Consequently, the set of all valid payout vectors is the positive quadrant. There are two natural coordinate systems on the space. One is to parameterize the space by $x \geq 0$ (the position on the first axis) and $y \geq 0$ (the position on the second axis). The second is to parameterize the space by $r \geq 0$ (the distance from the origin) and $\theta \in [0, \pi/2]$ (the angle from the first axis). The latter parameterization is more natural for homogeneous risk measures, because in a homogeneous risk measure prices are constant for constant θ (they do not depend on r). In contrast, prices in a homogeneous risk measure for the former parameterization would depend on both x and y.

Properties of optimality

There are two important properties an *optimal* homogeneous risk measure should have. First, it should have bounded loss, and second, it should not feature marginal prices on any event that exceed unity. In dual price space, this means the convex set $\mathbb{S}$ should include both $(1, 0)$ and $(0, 1)$, but not exceed 1 along either dimension.

Now parameterize by $x(\theta)$ and $y(\theta)$ the point on $\partial\mathbb{S}$ selected by the price vector selected at angle θ,

$$(x(\theta), y(\theta)) = \arg\max_{(x,y)\in\partial\mathbb{S}} (x, y) \cdot (\cos\theta, \sin\theta) \tag{6.2}$$

If $\partial\mathbb{S}$ is strictly convex, then x and y are differentiable. Furthermore, $x'(\theta) \leq 0$ and $y'(\theta) \geq 0$.

Now we can turn to the definition of optimality. Recall that the fundamental problem with the max cost function was that its prices very quickly changed from 0 to 1. For instance, in two dimensions, the rate of change at $\theta = \pi/4$ is undefined (and unboundedly large in the limit as $\theta \to \pi/4$).

Considering this argument, what we would like is for prices to change as little as possible uniformly over the relevant θ.

$$\partial\mathbb{S} = \arg\min_{\partial\mathbb{S}} \max_{\substack{\theta \in (0,\pi/2) \\ x,y\in\partial\mathbb{S}}} |y'(\theta)| + |x'(\theta)| = \arg\min_{\partial\mathbb{S}} \max_{\substack{\theta \in (0,\pi/2) \\ x,y\in\partial\mathbb{S}}} y'(\theta) - x'(\theta) \tag{6.3}$$

Observe that in this equation, we have begged the question that the optimizing $x(\theta)$ and $y(\theta)$ are differentiable on $\theta \in (0, \pi/2)$. However, it is easy to see that if this is not the case, then the value of the sup operation is undefined, as prices would change infinitely fast around the θ at which $\partial\mathbb{S}$ is not differentiable.

Simplifying the optimization equation

The functions x' and y' cannot be arbitrary; we must ensure that a given x' and y' are actually the solution to the optimization in Equation 6.2. For any point $(x, y) \propto (\cos\theta, \sin\theta)$, the argument on $\partial\mathbb{S}$ that is optimal is the point of intersection with the line normal to $\tan\theta$ farthest from the origin.

Observe that since $\tan\theta = \frac{\sin\theta}{\cos\theta}$, the line normal to it has a slope of $-\frac{\cos\theta}{\sin\theta} = -\cot(\theta)$. Consequently, if $\partial\mathbb{S}$ is strictly convex the maximizing argument at θ is given by $(x(\theta), y(\theta))$ when

$$\frac{y'(\theta)}{x'(\theta)} = -\cot(\theta) \tag{6.4}$$

so that the slope of the curve at that point matches the slope of the line normal to the tangent.

A further condition we impose is that the solution x' and y' should be symmetric. In particular,

$$x'(\theta) = -y'(\pi/2 - \theta) \tag{6.5}$$

Combined with the boundary condition that $x(0) = y(\pi/2) = 1$, the symmetry condition implies $x(\pi/4) = y(\pi/4)$. Intuitively, we should expect a minimum curvature solution to be symmetric, because in an asymmetric solution "too fast". (Furthermore, both the OPRS and the unit ball market maker are symmetric.)

Equation 6.4, combined with symmetry between x and y, allows us to dramatically simplify the optimization. In particular, we can focus only on a single function, x', over half the range of the variable θ, $\theta \in (0, \pi/4)$. Let the maximum sum of prices of the homogeneous risk measure be $s > 1$. Then three equations describe the optimization

- The first condition is that

  $$\int_0^{\pi/4} x'(\theta)\, d\theta = s/2 - 1$$

  which ensures that $x(\pi/4) = y(\pi/4) = s/2$.

- The second condition is that

  $$\int_0^{\pi/4} y'(\theta)\, d\theta = \int_0^{\pi/4} \cot(\theta) x'(\theta)\, d\theta = -s/2$$

which ensures that the total change for both x' and y' is 1, and also embeds the tangency relationship of Equation 6.4. Observe that $\cot(\theta) > 1$ for $\theta \in (0, \pi/4)$.

- The optimization is to find x' that minimizes the largest instantaneous total curvature:

$$\min_{x,y \in \partial\mathbb{S}} \max_{\theta \in (0,\pi/4)} y'(\theta) - x'(\theta) = \min_{x \in \partial\mathbb{S}} \min_{\theta \in (0,\pi/4)} x'(\theta) + x'(\theta)\cot(\theta)$$

Once we have solved for x' over $\theta \in (0, \pi/4)$, we can recover y' over $\theta \in (\pi/4, \pi/2)$ through the symmetry relation, Equation 6.5. Consequently, we have either x' and y' defined over the entire interval $\theta \in (0, \pi/2)$. If we have either x' or y' at some θ we can recover the other value through the tangent relationship, Equation 6.4. Finally, once we have the derivatives x' and y', we can recover the original parametric functions x and y from the boundary conditions $x(0) = 1$ and $y(0) = 0$.

Solving the optimization equation

As is standard, we solve the integral constraints by approximating the continuous function x' at a set of nodes $\{k_1, k_2, \ldots, k_n\}$ and associated weights $\{w_1, w_2, \ldots, w_n\}$, such that

$$\int_0^{\pi/4} f(\theta)\, d\theta \approx \sum_i w_i f(k_i)$$

The optimization is to specify the function values $x'(k_i) \equiv f_i$ at the nodes k_i by solving the following optimization for a given maximum sum of prices s:

$$\begin{aligned} &\max_{\mathbf{f}} \left(\min_i f_i + f_i \cot(k_i) \right) \\ &\qquad\qquad \text{such that} \\ f_i &\leq 0 \\ \sum_i w_i f_i &= s/2 - 1 \\ \sum_i w_i \cot(k_i) f_i &= -s/2 \end{aligned}$$

When we solved the optimization equation numerically for small s, we found that it had a very specific, three-part structure:

- For $\theta \in [0, t]$, $x(\theta) = 1$ and $y(\theta) = 0$. That is, the maximal point along the x axis is selected, so that the price on the first event is 1 and the price on the second event is 0. The rate of change of prices in this region is zero.

- For $\theta \in (t, \pi/2 - t)$,
$$y'(\theta) - x'(\theta) = l \tag{6.6}$$
where $l > 0$ is a constant rate of curvature.
- For $\theta \in [\pi/2 - t, \pi/2]$, $x(\theta) = 0$ and $y(\theta) = 1$. That is, the maximal point along the y axis is selected, so that the price on the first event is 0 and the price on the second event is 1. The rate of change of prices in this region is zero.

Because of this specific structure, we can solve for the optimal x and y analytically, instead of numerically. Equations 6.6 and 6.4 set up a system of two differential equations with two unknowns, x and y. Solving these differential equations yields

$$\begin{aligned} x(\theta) &= \frac{l}{2}(\log(\sin\theta + \cos\theta) - \theta) + C_1 \\ y(\theta) &= \frac{l}{2}(\log(\sin\theta + \cos\theta) + \theta) + C_2 \end{aligned}$$

By setting the boundary conditions on the parameters l, t, C_1, C_2 we will produce the desired solution to Equation 6.3. Furthermore, just like the OPRS and unit ball market makers, we can parameterize this family by the maximum sum of prices s.

First, we will use symmetry to eliminate the constant term C_1. By Corollary 4, we know that prices reach their maximum at $\pi/4$. By the symmetry constraint Equation 6.5

$$x(\pi/4) = y(\pi/4)$$

which means

$$l(\log(2)/2 - \pi/4) + C_1 = l(\log(2)/2 + \pi/4) + C_2$$

this yields

$$C_1 = C_2 + \frac{l\pi}{4}$$

Now we will impose additional boundary constraints to solve for l and C_2 (and therefore, C_1) as a function of t. The selection of any $t \in [0, \pi/4)$ imposes two boundary conditions:

$$x(t) = y(\pi/2 - t) = 1$$

and

$$x(\pi/2 - t) = y(t) = 0$$

These two sets of equalities produce two independent equations

$$\frac{l}{2}(\log(\sin t + \cos t) - t) + C_2 + \frac{l\pi}{4} = 1$$

and

$$\frac{l}{2}(\log(\sin t + \cos t) + t) + C_2 = 0$$

And this system of two equations and two unknowns (l and C_2) solves to

$$\begin{aligned} l &= \frac{1}{\pi/4 - t} \\ C_2 &= -\frac{t + \log(\cos(t) + \sin(t))}{2(\pi/4 - t)} \end{aligned}$$

Again by Corollary 4, the maximum sum of prices is achieved at $\theta = \pi/4$. Call that sum s

$$s \equiv x(\pi/4) + y(\pi/4) = \frac{l \log 2}{2} + C_1 + C_2 = \frac{\log(2)/2 + \pi/4 - t - \log(\cos t + \sin t)}{\pi/4 - t}$$

Figure 6.10 is a plot of this relationship. Observe that $s \to 1$ as $t \to \pi/4$.

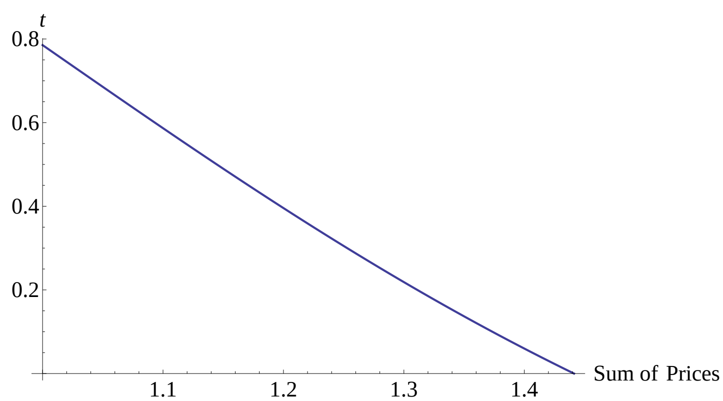


Figure 6.10: The offset angle t required for a specified maximum sum of prices in the optimal homogeneous risk measure.

Figure 6.11 is a graphical depiction of the boundary of the shell $\partial\mathbb{S}$ of the homogeneous risk measure, selecting t so that the maximum sum of prices is 1.05. For comparison, Figure 6.11 also features the probability simplex (i.e., the line $y = 1 - x$ in the positive orthant).

Figure 6.12 is a graphical comparison of the probability simplex, the unit ball market maker with maximum price sum 1.05, and the optimal homogeneous risk measure with maximum price sum

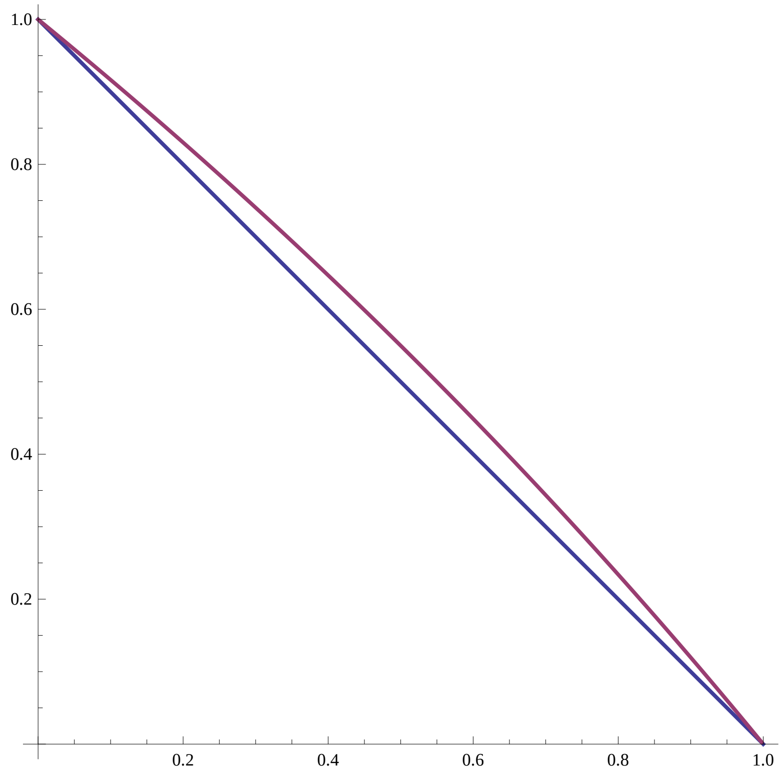


Figure 6.11: The outer shell of the convex set which supports the optimal homogeneous risk measure with maximum sum of prices 1.05.

1.05. The plot is zoomed to the lower-righthand corner of the price space in order to more clearly see the differences between the convex sets.

Intuitively, we would expect the optimal homogeneous risk measure to be as close to the probability simplex as possible while still retaining strict convexity and including the point corresponding to the maximum sum of prices. It is evident from inspection that the unit ball market maker achieves larger prices faster than the optimal homogeneous risk measure. But there is a limit to how large the sum of prices can get (in this case, the line $y = 1.05 - x$). Consequently, the boundary of the set must necessarily be less curved for larger values of θ, and so the prices around those angles must change faster.

6.4 Extensions

Because they automatically expand the depth of the market as more trades are made, homogeneous risk measures seem like they would be a good fit for settings in which the level of interest is un-

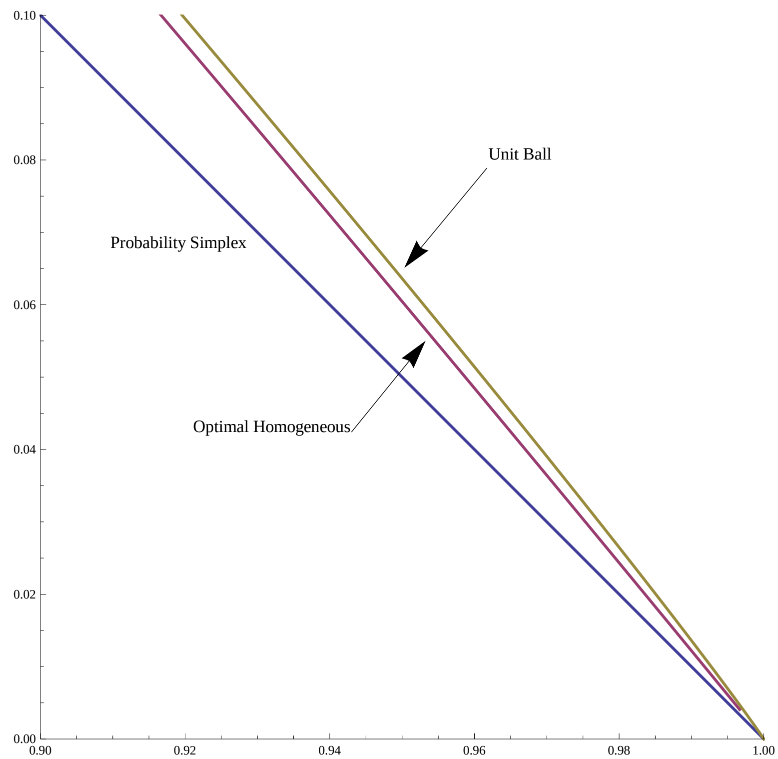


Figure 6.12: A comparison of the convex sets of the unit ball market maker and the optimal homogeneous risk measure.

known *a priori*. Based on our experiences with the GHPM (detailed in Chapter 4), most Internet prediction markets would seem to be characterized by this property. From a risk-measurement context, homogeneous risk measures would be most appropriate for circumstances in which the *ratio* of various payoffs is significant in determining the acceptability of a vector of risks.

In this chapter, we motivated and developed the optimal homogeneous risk measure for two events. Recall that this cost function was optimal in the sense that it has the smallest uniform change of prices while still retaining bounded loss. Under similar precepts, one extension would be to develop optimal homogeneous risk measures for more than two events.

A further extension would be to add limit orders to homogeneous risk measures. Agrawal et al. (2009) provide a framework to simply add functionality to handle limit orders (orders of the form "I will pay no more than p for the payout vector $\mathbf{x}$") into a cost function market maker. That framework relies on convex optimization and so would also be able to run in polynomial time, a significant gain over naïve implementations of limit orders within cost function market makers. However, that work relied heavily on simplifications to the optimization that could be made because of translation invariance, so it is unclear how to embed a market maker whose convex conjugate is defined over more than the probability simplex into a limit order framework.

Chapter 7

Option trading

Automated market makers are almost always used in fake-money applications, not with actual money on the line. In a prediction market, the goal of a market maker is often to assist in information elicitation, and so the losses that inevitably result from these market makers are viewed as subsidies to encourage traders to participate and reveal their information (Hanson, 2003; Pennock and Sami, 2007; Chen and Pennock, 2010). This reasoning provides a contrast between a market maker in a prediction market, which has a designated role to provide liquidity and can lose money acceptably, and a trading agent in a financial market which speculates for its own account. We specifically study the latter in this chapter. A trading agent will make or lose money based on whether its distribution over futures states of the world is better or worse than that of its counterparties, and an outcome-agnostic agent, the norm for making prediction markets, generally has too much entropy in its prior over outcomes to profit.

The finance literature, and particularly the finance literature as it pertains to derivatives, has taken a different modeling approach than the prediction market literature. The modern derivatives literature started with the seminal work of Black-Scholes and Merton (BSM) (Black and Scholes, 1973; Merton, 1973). By providing a practical way to price options contracts effectively, BSM led to the options markets that are the precursors of modern derivatives trading. Additionally, because the formula for calculating prices is entirely self-contained, it provided the constructive groundwork for the notion of *autarky* within finance theory. The BSM formula takes only three inputs—the current price, volatility, and the risk-free rate of return—to produce an ensemble of options prices. Although it is problematic to generalize over the entire finance literature, speaking broadly, most models of asset pricing operate using autarky as a guiding principle. In an autarky model, prices are philosophically prior to the agents that trade on them, so these models have no reliance on an agent's past actions in determining prices.

In this chapter, we synthesize ideas from the prediction markets and finance literatures to create more successful trading agents than either literature on its own. The key insight in this chapter is that these two notions—having good priors, and learning from inventories—are not oppositional. We combine them to create a trading agent that develops actionable prices based on both factors. Interestingly, the combination of the two ideas is not straightforward, and there are significant theoretical hurdles that serve to restrict the valid combinations of prior distributions and the utility models that determine how an agent reacts to its inventory.

The principal contribution of our work is the experimental simulation of five different trading strategies on a large body of recent options data:

1. A zero-intelligence agent, added as an experimental control, that trades randomly.
2. An orthodox BSM Log-normal distribution trader.
3. A Normal distribution trader, with mean and variance matched to the Log-normal trader.
4. The *Logarithmic Market Scoring Rule (LMSR)*, the most popular automated market maker in Internet prediction markets (Hanson, 2003, 2007; Pennock and Sami, 2007) which is equivalent to an exponential utility trading agent with constant uniform priors.
5. A hybrid agent that combines exponential utility with normal distribution priors.

We find that by many different measures, including expected return and worst-observed performance, the hybrid trader outperforms the other traders. Consequently, our results support the hypothesis that a trader's current exposure can be a profitable influence on future actions, and that a trader can learn from their past actions to create a more accurate estimate of the future.

Even though the hybrid trader performs the best on several key metrics, it does not stochastically dominate the performance of the parametric traders from the finance literature. This means it is possible to construct a coherent utility function that would prefer the performance of the parametric traders. However, the relative performance of the hybrid trader relative to the parametric traders provides insights into the larger qualitative question of *how* the hybrid trader is able to perform well. We believe our results are best interpreted by the hybrid trader profitably insuring against the risk that its model of the future is inaccurate, a claim which we justify in detail.

It is important to clarify what we believe is significant about our results. We do not believe that our hybrid trader is the best options trading agent that could be devised. Certainly, BSM can be considered a theoretical model, rather than a practical trading agent, and there are likely other agents that could be constructed that would have better performance on our dataset. These agents could employ more sophisticated models about how prices move through time, detailed order book information, or outside information like press releases and macroeconomic forecasts. But what is significant about our work is this: by incorporating inventories into a Normal distribution trader, we increase performance *without incorporating any new or more sophisticated information into the trading process*. Instead, we achieve these performance gains simply by paying attention to information that autarky models discard.

We are particularly keen on interpreting our results within the intellectual framework of artificial intelligence. Much of modern, optimization-based AI is grounded in the idea of modeling (or actually constructing) a robot that can observe its environment, effect an action, and receive a reward that depends on its state and the actions it has chosen (see Russell and Norvig, 2003, Chapter 2). This model fits naturally into trading options, where the environment is the market, the actions are the contracts to trade, and the rewards are literal monetary profits and losses. History-keeping is vital to what constitutes a *rational agent* in the AI literature. In a general environment an agent that does not retain its history *cannot be rational*. So in this context, our results are not surprising: agent performance is enhanced by both built-in knowledge about the future and the ability to learn from one's past actions.

7.1 Options

In this section we introduce options and their associated terminology, as well as our dataset.

7.1.1 What are options?

Options contracts involve the future opportunity to buy or sell some kind of underlying instrument (*the underlying*) at a set price (the *strike price*).

There are two types of option contracts. A *call* option gives its holder the right to buy an obligation at a specified price, and a *put* option gives its holder the right to sell an obligation at a specified price. These contracts have a hinged form of payouts.

Definition 24. Let the underlying expire at price π. A (European) *call option* with strike price s has value $\max(\pi - s, 0)$. A (European) *put option* with strike price s has value $\max(s - \pi, 0)$.

Options that strike close to the current price of the underlying are known as *at the money*. Options that are valuable at the current price of the underlying (high-strike puts or low-strike calls) are known as *in the money*. Options that are worthless at the current price of the underlying (low-strike puts or high-strike calls) are known as *out of the money*.

There are two principal ways that govern the exercise of options. European options expire in cash at a certain date, the *strike date*. In contrast, American options can be exercised for delivery of the underlying at any point before the strike date. The additional optionality of American options makes them necessarily at least as expensive as European options. However, this additional optionality is rarely exercised, making American options essentially European in practice. Varian (1987) provides a theoretical argument based on no-arbitrage principles for European and American options having exactly the same prices. We examine both European and American options in our dataset.

Options with the same underlying and strike date form what is called an *options chain*. Chains consist of an ensemble of contracts along with their associated prices. For instance, a chain might consist of calls and puts with strikes of 900, 1000, and 1100 for the ^SPX underlying expiring on December 22, 2007. Each of these contracts has a price at which they can be bought or sold (the *ask* and *bid* prices, respectively), and theoretically all of these prices are based on some underlying distribution over the expiration price. This distribution changes over time as the price of the underlying and the time until expiration changes. When we perform our experiments, we step through simulating trading agents on snapshots of each options chain as it evolves over time, from initiation until expiration.

7.1.2 Historical dataset

The dataset we use covers almost seven years of data on eleven underlyings, taken at 15-minute intervals. It is comprised of nearly 300 million {underlying, expiration date, strike price, datetime, best bid, best ask} tuples, and the corresponding {underlying, datetime, underlying best bid, underlying best ask} tuples for the underlying. The data spans from January 2004 through September 2010. To our knowledge this is the one of the more-detailed datasets used in an academic study on options—studies generally use data from daily closing prices, which is much less detailed (and one of the most widely-cited papers on empirical option pricing, Dumas et al. (1998), uses *weekly* data). Table 7.1 gives an overview of the dataset.

In order to provide a clean train/test separation, we divide the first two years (January 2004 through December 2005) to learn the relevant parameters for our simulation, and only test on the remaining data. A naïve split between training and testing data that randomly placed chains or days into different partitions would be tainted, because the training and test sets would overlap temporally. Consequently, only options chains that expire in 2006 and beyond are in our testing set. The number of complete chains in the testing set for each underlying is noted in the "Testing chains" column of Table 7.1.

7.2 Agents based on Log-Normal and Normal distributions

In this section we introduce the first two traders we used in our experiments, the Log-normal and Normal distribution traders. These traders are derived from existing work in the finance literature.

Underlying	Description	Clearing	Number of records	Testing chains
^SPX	S&P 500 Index	European	53.6 million	6
^DJX	Dow-Jones Index	European	43.7 million	9
^NDX	NASDAQ Index	European	30.1 million	3
^IRX	Thirteen-week treasury bill	European	3.4 million	19
^FVX	Five-year treasury yield	European	6.2 million	19
^TNX	Ten-year treasury yield	European	6.9 million	16
^XAU	Gold-Dollar Index	European	10.1 million	16
X	US Steel	American	9.3 million	5
C	Citigroup	American	7.5 million	6
GE	General Electric	American	7.7 million	5
MSFT	Microsoft	American	8.5 million	5
XOM	Exxon-Mobil	American	7.7 million	5

Table 7.1: Our options dataset includes a diverse mix of underlyings, including indices, bonds, commodities, and equities of very liquid to moderately liquid contracts cleared both in American and European ways.

7.2.1 The BSM model

Option pricing was revolutionized and popularized by the work of Black and Scholes (1973) and Merton (1973) (BSM). Those authors described a parametric framework under which prices on the underlying change according to a log-normal distribution, which was the solution to a differential equation.

This framework was essentially unchallenged until "Black Monday" of 1987, where stock prices dropped precipitously in a single day. After this, options now show a persistent "volatility smile" or "volatility skew", where out-of-the-money options are overpriced relative to BSM (MacKenzie, 2006). An interpretation of this phenomenon is that the log-normal distribution of future prices as predicted by BSM is inaccurate, and the skew represents an effort to make the predicted distribution heavier-tailed. Another interpretation is that investors use options to provide insurance against the state of the world in which extremely low values of the underlying are realized.

7.2.2 Calculating contract values when the underlying is log-normally distributed

We use a constant (daily) volatility parameter σ for each underlying. These values are learned from our training data in order to assure a clean train/test separation. Table 7.2 shows the values we used in our experiments.

In addition to the current price and the volatility parameter σ, the BSM model takes an additional input, the so-called *risk-free rate of return*. This value reflects the time-cost of money. In our exploratory data analysis over our training data we did not see significant changes in performance for different realistic values of the risk-free rate (between zero and five percent annualized). For consistency, in our tests we set this value equal to zero for all the trading agents. In practice, banks, market makers, and large hedge funds will have small risk-free rates over short time horizons. Furthermore, setting this rate equal to zero means that any returns generated are exclusively from options trading, and not from interest on passive income.

Because of the popularity of the BSM model, the formulas to calculate prices from a log-normal distribution are well-known. The so-called *partial expectation* of the log-normal distribution has an analytic expression. Let f denote the density function and F the distribution function of a log-normal distribution parametrized by μ and σ. Then the partial expectation formula is:

$$\int_s^\infty x f(x)\, dx = e^{\mu+\sigma^2/2} \Phi\left(\frac{\mu + \sigma - \log s}{\sigma}\right)$$

where $\Phi(\cdot)$ is the distribution function of the standard normal distribution.

Underlying	σ
^SPX	.006814
^DJX	.006781
^NDX	.010448
^IRX	.014504
^FVX	.016970
^TNX	.012679
^XAU	.018846
X	.027520
C	.008436
GE	.009329
MSFT	.010200
XOM	.012430

Table 7.2: The (daily) volatility parameters σ are learned by taking the MLE of the daily changes of each underlying in the training set.

When this value is known, the price of a call option can be calculated, because the price of a call option with strike s is

$$\begin{aligned}\int_s^\infty (x-s)f(x)\,dx &= \int_s^\infty xf(x)\,dx - \int_s^\infty sf(x)\,dx \\ &= e^{\mu+\sigma^2/2}\Phi\left(\frac{\mu+\sigma-\log s}{\sigma}\right) - s(1-F(s)).\end{aligned}$$

By the use of the well-known *put-call parity* we can calculate the price of a put option at the same strike. The parity says that the price of a call at a strike, plus that strike, equals the price of a put at that strike, plus the value of the underlying. The justification for this formula is that buying a call and selling a put at a strike is equal in payout to buying the underlying and holding the amount in cash. (Throughout this work, when we buy or sell underlyings it is assumed that the position will be closed when the options expire.) Once we have calculated the value of the call, the only unknown value in the put-call parity equation is the value of the put.

7.2.3 Calculating contract values when the underlying is normally distributed

We also implemented a trader that models the underlying's expiration price as a normal distribution, instead of a log-normal. This trader sets the mean and variance of the normal distribution to match that of the log-normal distribution.

Normal distributions are a feature of much of the literature on market making in both prediction markets and finance. As O'Hara (1995) discusses, theoretical frameworks often model underlying prices as normal distributions because the conjugate prior to a normal distribution (with known variance) is another normal distribution. Consequently, it is common for agents to have a normal prior distribution and then update that distribution to a posterior normal distribution as more (normally-distributed) information arrives. This allows agents in models to act rationally while still retaining closed-form expressions for analytical tractability. Examples of models using normal distributions to project the future price of an asset include the classic models of Glosten and Milgrom (1985) and Kyle (1985), as well as more recent models such as that of Das and Magdon-Ismail (2009).

In the context of options, however, normal distributions are conceptually much more problematic than log-normal distributions. This is because they have support over the whole real line, rather than just over positive values. Since negative values cannot exist as termination prices (shareholders are not personally liable for the debts of a company), this makes the normal distribution inherently unrealistic[1]. Recognition of this problem dates back to some of the earliest work on asset pricing (Merton, 1971). Reflecting the intuition that a normal distribution is a worse match for the setting than the log-normal distribution, our results showed that the normal distribution trader generally performed worse than the log-normal distribution trader, even though the two traders matched the mean and variance of their distributions.

To solve for prices using a normal prior, we used *Gauss-Hermite quadrature* (Judd, 1998). Let $\mathbb{E}_{\mu,\sigma}(f)$ denote the expectation of the function f under a normal distribution with mean μ and standard deviation σ. Gauss-Hermite quadrature provides a set of nodes x_i and weights w_i such that

$$\mathbb{E}_{\mu,\sigma}(f) \approx \sum_i w_i f(x_i)$$

by implicitly converting the function f to an approximation by orthonormal polynomials.

[1]One way to overcome this limitation is to instead deal with truncated normal distributions, where the probability mass that exists below zero is re-distributed over the positive values (e.g., Chang and Shanker (1986)). However, these concerns are more theoretical than practical, because in our simulations the total probability mass for negative realization values appeared to be small enough that truncating the distribution would not have changed agent actions.

7.3 The LMSR as an option trading agent

In this section we discuss how we implemented the LMSR, the *de facto* automated market maker used in Internet prediction markets, as an options trading agent. Recall that the LMSR (which we first introduced in Chapter 3) is given by the cost function

$$C(\mathbf{x}) = b \log \left(\sum_i \exp(x_i/b) \right)$$

Where $b > 0$ is an exogenous constant known as the *liquidity parameter*. Larger values of b correspond to larger worst-case losses by the LMSR, which loses at most $b \log n$. On the other hand, larger values of b produce tighter bid/ask spreads.

7.3.1 Compressing the state space

For the log-normal and normal distributions, we took the view that the underlying would expire as a continuous process. In reality, the space of expirations is countably infinite, delimited by one-cent intervals.

It is natural, however, to reduce this infinite space to a finite range of possibilities. This is a lossy operation; there is perhaps the chance the final outcome will fall outside the range we specify. Therefore, we should take a wide range. Consider the ^SPX underlying, which tracks the S&P 500; it has a value of around 1000. It is reasonable to assume that for near-term options, its expiration price will be between 100 and 10,000. With one-cent discretization, this implies a space of about $n = 1$ million events.

Very large event spaces like this pose two problems for the LMSR; one practical, and the other theoretical. First, the LMSR is numerically unstable over large event spaces; we observed this problem in the Gates Hillman Prediction Market, and numerical stability was also a problem in Yahoo's Predictalot (which ran over a combinatorially large event space). This numerical instability makes implementing the LMSR over very large event spaces challenging, unwieldy, and potentially inaccurate. The second problem with large event spaces is that they correspond to larger worst-case losses; in order to maintain realistic worst-case losses the b parameter would need to be much smaller, leading to a large bid/ask spread that would not facilitate much trade.

In full form, the size of the event space we would need to consider makes applying the LMSR to options markets extremely challenging. Fortunately, we can achieve a significant lossless dimensionality reduction that makes automated market making for options feasible. The key to compressing the state space is to focus only on the strike prices, not the expiration prices. Let the set of ordered strike prices be given by $\mathbf{s} = \{s_1, s_2, \ldots, s_n\}$ and the market maker have corresponding payout vector $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$. This idea, of compressing the state space only to relevant strike prices, is a feature of many options trading models (e.g., Varian (1987)).

Lemma 2. *Let the expiration price be $s_i < s < s_{i+1}$, such that $s = \alpha s_i + (1-\alpha)s_{i+1}$. Then if there exist no contracts written for a strike price between s_i and s_{i+1}, the market maker must pay out $\alpha x_i + (1-\alpha)x_{i+1}$.*

Proof. An options contract is piecewise linear with a joint at its strike price, underlyings are linear, and our portfolio consists only of these contracts. Since a combination of linear functions is also linear, if no contract has a strike price between s_i and s_{i+1}, our payoffs will move linearly with the realized price between s_i and s_{i+1}. ■

This linearity result allows us to, in effect, collapse the continuous state space of possible prices $[s_1, s_n]$ into the set of discrete prices $\{s_1, \ldots, s_n\}$ by bounding our realized loss.

Proposition 30. *Let $\overline{x} = \max_i x_i$ represent the maximum value the market maker must pay out if $s \in \{s_1, \ldots, s_n\}$ is realized. Then $\overline{x}$ is also the maximum value the market maker must pay out if $s \in [s_1, s_n]$ is realized.*

If a value outside of the strike price range is realized, we might lose more than what would be suggested by the x_i. One way to solve this problem is to include two dummy strike prices: one at zero, and another at an arbitrarily large value. This essentially prevents the final state from falling outside of the span of the strike prices. While this is appropriate for more general settings (e.g., sparsely-traded underlyings) we did not implement these dummy strikes into our LMSR trading agent. We found that for our underlyings the extreme strike prices of our option chains generally gave reasonable bounds on the expiration price.

7.3.2 Implementation details

Recall that trades are priced in the LMSR based on the vector of payouts currently held by the market maker. In this section we describe how to go from an inventory of options contracts to a payout vector. Let $\{s_1, \ldots, s_n\}$ be the ordered set of strikes we are considering, and $\mathbf{x} = \{x_1, \ldots, x_n\}$ be the payout vector corresponding to the realization of each strike. Formally, selling a call at strike s corresponds to the payout vector $\mathbf{x} = \{x_i\}_{i=1}^n$ where

$$x_i = \begin{cases} s_i - s & \text{if } s_i \geq s \\ 0 & \text{if } s_i < s \end{cases}$$

Selling a put at strike s corresponds to the payout vector

$$x_i = \begin{cases} 0 & \text{if } s_i > s \\ s - s_i & \text{if } s_i \leq s \end{cases}$$

Selling the underlying corresponds to a payout vector of

$$x_i = s_i$$

The payout associated with an event (i.e., x_i) depends directly on the strike price associated with that event (i.e., s_i). For example, selling a contract of the underlying corresponds to a payout of 30 dollars if the underlying expires at 30, and a payout of 50 dollars if the underlying expires at 50. Selling a call at a strike of 20 corresponds to a payout of 0 if the underlying expires at 20 (or below), but a payout of 30 if the underlying expires at 50.

Buying any contract induces the negative payout vector of selling that contract. Observe that when we quote the price to buy a contract the value will be negative, suggesting that we need to compensate our counterparty (i.e., pay out money for) the contract in question.

The set of strike prices we model for the LMSR trader is all the currently-offered strike prices, plus any strike prices corresponding to contracts we traded in the past. Given an inventory $\mathfrak{I}$ of bought and sold contracts, we can calculate the payout p_i at any strike s_i by doing an element-wise sum for the payout vector $\mathbf{x}$ of each accumulated contract:

$$p_i = \sum_{\mathbf{x} \in \mathfrak{I}} x_i$$

This cumulative payout vector $\mathbf{p}$ is then used to price the available options in the chain; if a prospective contract induces a payout vector $\mathbf{y}$ the LMSR trader prices the contract at

$$C(\mathbf{p} + \mathbf{y}) - C(\mathbf{p})$$

The final issue is how to set the liquidity parameter b. For our experiments we set b equal to 2500 times the initial underlying price, a value that yielded good performance over the training data. Values much smaller than this resulted in sharply diminished trade because the bid/ask spread was too large. Values much larger than this resulted in marginal prices which stayed close to a uniform distribution for the entire trading period.

7.4 Trading agents based on constant exponential utility

We have presented two different ways of thinking about how to price options. The first is the traditional approach from finance, derived from projecting a distribution over the future and pricing obligations based on this projection. The second approach is derived from automated market making in prediction markets. It involves pricing obligations based only on trades previously made. In this section, we explore how to achieve a synthesis between these two ideas, creating a market maker with a good prior that also responds to past trades.

This effort is immediately complicated by the fact that models from finance typically involve generating continuous distributions over the final strike price, while the LMSR, as we discussed in the previous section, is for discrete distributions. In order to synthesize these two models one needs to either discretize a continuous distribution, or develop a continuous analogue to the LMSR. Since the existing literature in financial options is based around continuous distributions, we choose to do the latter in order to better align our work with that literature. (We study a version of the LMSR with a good discrete prior in Section 7.9.2. We found that it performs better than the maximum-entropy LMSR, but much worse than the trader developed in this section.)

We synthesize the two methodologies by developing a version of the LMSR over continuous spaces by viewing it as a constant-utility cost function. Recall that we developed the theory of cost functions over continuous spaces in Chapter 5. In the continuous setting, our payout vectors are functions $\mathbf{x} : \mathbb{R} \mapsto \mathbb{R}$, and cost functions become functionals that map these functions to scalars, $C : (\mathbb{R} \mapsto \mathbb{R}) \mapsto \mathbb{R}$.

Definition 25. Let μ be a probability distribution over possible expiration prices, $u : \mathbb{R} \mapsto \mathbb{R}$ be an increasing concave function, and $x^0 \in \mathbf{dom}\ u$. A *continuous constant-utility cost function* $C(\mathbf{x})$ is given implicitly by the solution to

$$\int_0^\infty \mu(t) u(C(\mathbf{x}(t)) - \mathbf{x}(t))\, dt = u(x^0)$$

Recall from Chapter 3 that the LMSR is equivalent to an agent with the exponential utility function $u(x) = -\exp(-x/b)$. Given this framework, it seems like it would be straightforward to combine the LMSR with the orthodox BSM forecasting model: simply set μ to be equal to the appropriate log-normal and set u equal to exponential utility. Surprisingly, this approach does not work as planned and instead produces undefined prices for simple actions.

7.4.1 The undefined prices phenomenon

In a nutshell, what produces undefined prices is the tradeoff between how quickly the tails of the agent's prior distribution fall off, and how severely the trading agent's risk aversion reacts to extreme losses. If the aversion to large losses is strong enough, it can outweigh the very small probabilities associated with those large losses. The resulting trading agent would not offer to trade a contract that could produce those losses at any price.

The specific failure of exponential utility and log-normal priors to produce always-defined prices is known in the finance literature (Henderson, 2002), but working through a realistic example will shed light on how and why this pairing fails. We will then generalize the intuition gained from the example to multiple trades, distributions, and utilities, with a particular focus on what happens with exponential utility.

Example

For this example, we will assume a log-normal prior with $\mu = 5$ and $\sigma = 0.5$. Now consider the calculation involved in pricing the sale of the underlying. The sale of the underlying is given by the payout vector (function) $\mathbf{x}(t) = t$. With exponential utility, recall that the cost function solves, for some $v < 0$:

$$\int_0^\infty -e^{-\left(\frac{C(\mathbf{x})-\mathbf{x}(t)}{b}\right)}\mu(t)\,dt = v$$

Because of the form of the utility function, we can uncouple the cost, which does not feature the dummy integrating variable t, from the vector of payouts

$$\int_0^\infty -e^{\mathbf{x}(t)/b}\mu(t)\,dt = ve^{C(\mathbf{x})/b}$$

which shows that the cost function is defined if and only if

$$\int_0^\infty -e^{\mathbf{x}(t)/b}\mu(t)\,dt$$

converges. For our specific example, with the sale of the underlying and the log-normal distribution, this integral is

$$\int_0^\infty -e^{t/b}\left(\frac{e^{(\log t-\mu)^2/2\sigma^2}}{t\sqrt{2\pi\sigma^2}}\right)\,dt$$

Figure 7.1 shows a plot of the integrand. The x-axis is log-scaled. The integrand tends towards zero for large expiration values (in the thousands, the median of the prior distribution is $e^5 \approx 148$). However, for unrealistically extreme expiration values (more than 100 times the fictional current price) the integrand explodes negatively, and the integral diverges.

Why does this occur? It is because for extremely large realization values our sensitivity to the prospect of extreme loss (since we are selling the underlying we lose when it expires high) outweighs the extremely small probabilities that the log-normal distribution produces for those values. We can show that this behavior holds regardless of the particular b, μ, σ parameterization chosen. Again, consider the pricing integral corresponding to the sale of the underlying. With a log-normal distribution and exponential utility, we have

$$\mu(t) \in \Theta(e^{-\log^2 t}) \text{ and } u(\mathbf{x}(t)) \in \Theta(e^t),$$

so

$$u(\mathbf{x}(t))\mu(t) \in \Theta(e^{t-\log^2 t}).$$

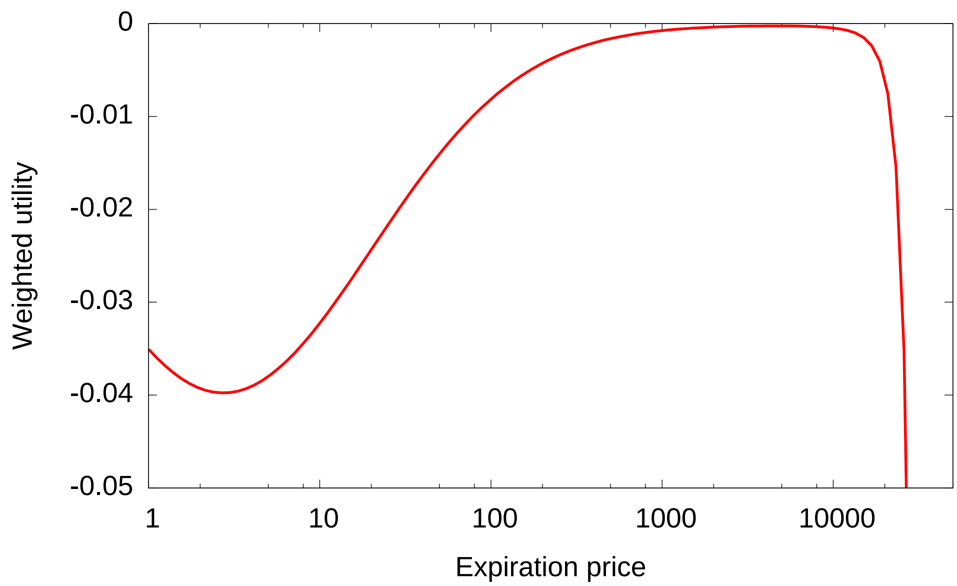


Figure 7.1: Because exponential utility increases faster than log-normal probability falls off, our sensitivity to large losses grows unboundedly.

Since

$$\lim_{t \to \infty} e^{t - \log^2 t} \neq 0,$$

the pricing integral diverges. It is easy to see that this asymptotic analysis also applies to selling a call, since when we sell a call with strike price s,

$$u(\mathbf{x}(t)) \in \Theta(e^{t-s}) \in \Theta(e^t).$$

Consequently, *there exists no amount of money an exponential utility cost function with a log-normal prior would sell an underlying or a call for.*

The theoretical basis of undefined prices

For any probability distribution, the density at extremely large realizations is very small, but our utility function could react to these realizations in a pronounced way. Hence, we have a tension between the density of the distribution at its tails and the response of the utility function. In this section, we provide theoretical bounds for this phenomenon, motivated by demonstrating why using exponential utility with normal priors succeeds, while with log-normal (and many other) priors it fails. Geweke (2001) also explores diverging (undefined) prices, given certain priors and utilities, but only for a specific family of utility functions.

The following result rules out using utility functions like log and $-1/x$ with infinite-domain probability distributions.

Proposition 31. *In a continuous constant-utility cost function, if the prior probability distribution is positive over* $(0, \infty)$ *but the utility function is not defined over all of* $\mathbb{R}$*, then there exists a transaction with undefined price.*

Proof. We will show that the utility function will be evaluated at an undefined value. Let $\mathbf{x}(t) = t$ denote the function corresponding to selling an underlying. By assumption, the utility function is not defined at $v \in \mathbb{R}$. Now, consider $C(\mathbf{o}) > v$. Because the utility function is increasing, the cost function is increasing, and therefore $C(\mathbf{x}) > C(\mathbf{o})$. But because $C(\mathbf{x})$ is finite and the prior distribution is positive at every $t \in (0, \infty)$, there exists some t for which $C(\mathbf{x}) - \mathbf{x}(t) = C(\mathbf{x}) - t = v$, so the utility function will be evaluated at v, producing an undefined value. This argument holds by symmetry in the case where $C(\mathbf{o}) < v$, in which case we would buy the underlying instead to force the undefined evaluation. ■

The following result shows that one way to achieve well-defined costs is to only consider finite probability distributions.

Proposition 32. *In a continuous constant-utility cost function, if the utility function is defined and finite over all of* $\mathbb{R}$ *and the prior distribution with density* μ *is bounded, so that there exists a* T *such that for all* $t > T$, $\mu(t) = 0$, *then the cost function is well-defined.*

Proof. The limits of integration are finite and the utility function is defined and finite for any argument. It follows that the integral to determine the cost function is always well-defined. ■

One application of this result to the option trading problem is to consider a trading agent with a log-normal prior distribution that is truncated above some upper boundary T. (The remaining probability mass above the upper bound could be re-distributed below the bound.) Proposition 32 suggests then that the cost function would always be defined when using such a distribution.

However, this truncation scheme is numerically hazardous because it provides no guidance for *which* upper bound of T to select. Consider again Figure 7.1, which shows the weighted utility of a contract. Since the utility of the contract calculated without an upper bound on the prior diverges, where the upper bound T is set will have a great effect on the calculated utility of taking on the contract. To be concrete, it is clear from inspection that setting the upper bound at 1,000, 10,000, or 100,000 will produce vastly different calculations for the utility of the trader, and therefore for the fair price of the contract.

Now specifically focusing on exponential utility, it turns out we must have very light tails for the cost function to be well-defined.

Proposition 33. *A continuous exponential-utility cost function is defined for every set of options transactions if and only if*

$$\int_0^\infty e^{c \cdot t} \mu(t)\, dt$$

is bounded for every $c \geq 0$.

Proof. Recall we have defined prices if and only if

$$\int_0^\infty -e^{\frac{\mathbf{x}(t)}{b}} \mu(t)\, dt$$

is well-defined. By removing constant factors, we see that

$$\int_0^\infty e^{\mathbf{x}(t)} \mu(t)\, dt$$

must be well-defined. Options contracts are continuous, piecewise-linear functions, so therefore $\mathbf{x}(t)$ is a continuous, piecewise-linear function. Therefore there exists some c such that $\mathbf{x}(t) \leq c \cdot t$, where the bound is tight by selling c underlyings. Now since e^x is an increasing function, this implies

$$\int_0^\infty e^{\mathbf{x}(t)} \mu(t)\, dt \leq \int_0^\infty e^{c \cdot t} \mu(t)\, dt$$

so if the right-hand equation is finite, the left-hand equation is finite also. ∎

Recall that the expression

$$\int_0^\infty e^{c \cdot t} \mu(t)\, dt$$

is also known as the *moment generating function* of the distribution corresponding to the density μ. However, moment generating functions are generally used only for the values of their derivatives at the argument $c = 0$, whereas for our result here the function must be defined for any positive c.

This result has the effect of significantly limiting what distributions we can use with exponential utility. We have already discussed how we cannot use the log-normal distribution, but other distributions, like the chi-square, exponential, and Weibull are also ruled out. (This begs the question of whether there is a theoretical or empirical reason to use these distributions.)

Now, consider that for the normal distribution,

$$\mu(t) \in \Theta(e^{-t^2})$$

This means that for arbitrary c the integrand is

$$\Theta(e^{c \cdot t - t^2})$$

which results in a well-defined integral for any c. This means we can combine a normal distribution prior with exponential utility and still get defined prices.

7.4.2 Our Exponential utility trader

For our Exponential utility trading agent, we use the same normal distribution prior as the Normal trader and the same b parameter as the LMSR trader. This is to better facilitate comparisons between the traders.

7.5 Random agent as a control

In order to provide a performance benchmark for the effect of trading in the market, we introduce a random trading agent as a control in our experiments. *Random* is a trading agent that performs a uniform random action over the set of possible actions at each time step. That is, for all the bids and asks on the relevant calls, puts, and the affiliated underlying, Random selects a bid or ask to trade uniformly at random. Random can be thought of as a *zero-intelligence agent* (Gode and Sunder, 1993) that does not learn or optimize.

We would expect Random to consistently produce slightly negative returns. Consider that the bid and ask prices are spread, so we should expect that an agent that takes either side of the market at random should consistently "eat" this spread. That is, we can roughly think of the Random trader as buying an option at price $p + \epsilon$ and then selling it at $p - \epsilon$, resulting in a guaranteed loss of 2ϵ.

7.6 Experimental setup

As we discussed in Section 7.1, we simulated the performance of the trading agents on each options chain in our testing set. In our simulations, we step through 15-minute increments on each chain from initiation to expiration. Figure 7.2 shows a flowchart of the simulation steps over each testing chain (recall the number of testing chains for each underlying is listed in Table 7.1). Section 7.7 features a worked example of the simulation process with each trading agent on one snapshot of a single option chain.

Any simulation on historical data is fraught with the risk of overfitting, producing an unreasonably rosy picture of real-world performance. We took several steps to combat overfitting. These

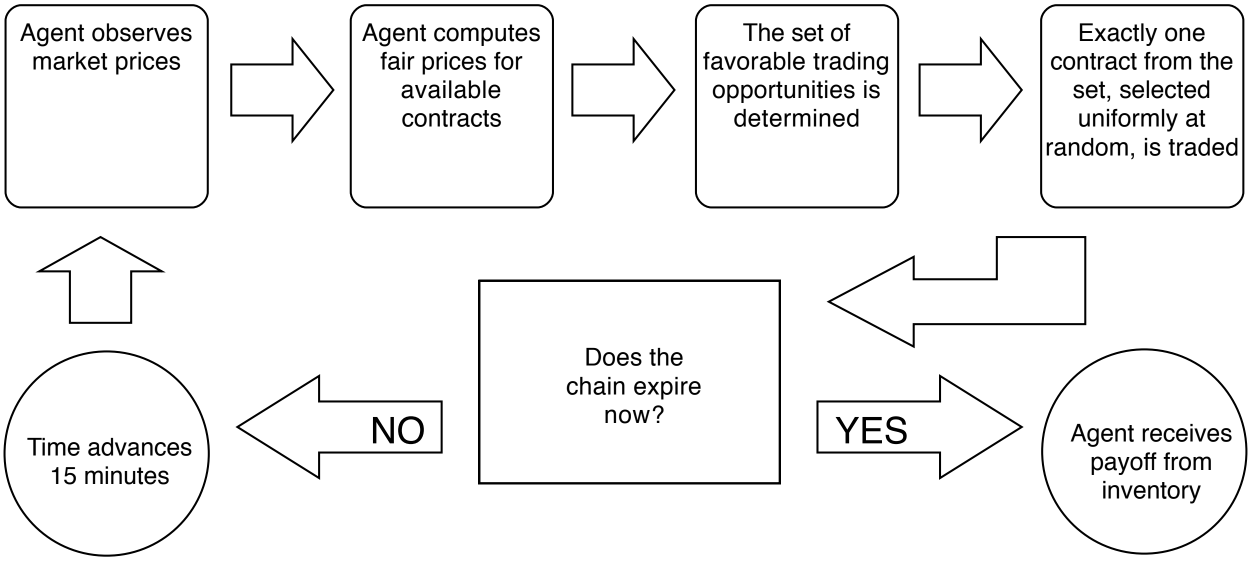


Figure 7.2: The steps taken by an agent over each option chain. Agents differ in the second step, the way they determine the fair prices for the offered contracts.

limits are intended to give a more accurate picture of live performance than a naïve optimization over the dataset.

It is, of course, impossible in hindsight to accurately produce a counterfactual answer to the question of how well a trading bot would have performed. The bids and asks we fill could cause changes in the behavior of other traders, which is an effect we cannot judge from past data. However, Even-Dar et al. (2006) suggest that, as long as agents in the market trade on *absolute* values (fixed prices) rather than *relative* values (based on the current state of the order books), overall price series will be resistant to the small changes produced by simulation. Options markets are driven mainly by BSM-style models and often feature relatively large (compared to the underlying) bid/ask spreads. Therefore, we believe that these markets are populated mostly by traders of the absolute, rather than the relative type, making simulation by an agent trading a small amount of total volume appropriate.

Another concern related to the difference between real trading and simulation involves the effect of adding orders. In the real world, we would be able to add our own limit orders to the order book so that we would both provide as well as consume liquidity. However, for robustness we do not model this property within our data; we only take prices and do not simulate the effect of adding limit orders to the dataset. Presumably, adding the ability to place limit orders at auspicious prices would only help the performance of these trading agents in the real world, as well as limiting slippage by contributing to, instead of subtracting from, market liquidity.

We take two more steps to handicap the performance of our trading agents to avoid overfitting to the historical data:

- We limit the frequency of trading to only a single contract every 15 minutes. We are concerned about *slippage*—the tendency of prices to move against a trader's actions as the trader absorbs liquidity. Trading a single contract is a conservative measure of performance that avoids slippage, because when a counterparty sets a price, the counterparty must sell at least one contract at that price.

- We choose *which* contract to trade uniformly at random from the set of desirable contracts. Say that our algorithm has identified purchasing an underlying, selling a call at 20, and buying a put at 30 to be beneficial trades given the current market prices. Then we will do only one of these, each with probability one third. This is the case even if, say, our trading algorithm thinks that buying the underlying would be better than selling the call at 20. This restriction is designed so as to not overfit on our static snapshots of prices. In a real setting, trading opportunities may arise and disappear quickly and we may not be able to trade the best opportunity that exists at a given moment.

  We contrast the results of trading a random contract with trading only the contract an agent thinks is best in detail in Section 7.9.1. Our results show the former model disadvantages the inventory-based traders more than the finance literature traders, so this restriction is conservative, as desired, for the conclusions we will draw in Section 7.8.

  Furthermore, when we examined the trading behavior that arose from only trading the best contracts, we saw that it was biased towards trading the same contracts again and again. This is unrealistic behavior, because presumably the liquidity associated with those contracts will be exhausted if they are traded so frequently, and the prices would slip. Since a trading agent generally has several contracts it is interested in trading at a given time step, trading uniformly at random produced more realistic behavior in the simulation by spreading trading activity among several contracts.

In the next section we work through the pricing behavior of each agent on a single snapshot of an option chain, with the goal of giving the reader a better sense of what our experiments entailed and how our trading agents worked.

7.7 Snapshot of a single option chain

We will consider the snapshot of the ^TNX option chain expiring December 19th, 2009, taken at 9:45 AM on October 22nd, 2009. This particular snapshot was chosen arbitrarily, but with attention to having a relatively small number of relevant strike prices. Table 7.3 shows the option chain.

The actions of the Random trader are the simplest to describe. At this snapshot the Random trader picks one of the possible actions to perform uniformly at random. Observe that not all

Strike	Call Bid	Call Ask	Put Bid	Put Ask
17.5	15.6	20.4	X	1.5
20.0	13.1	17.9	X	1.5
22.5	10.6	15.4	X	1.5
25.0	9.0	12.0	X	1.5
27.5	6.6	9.6	X	1.5
30.0	4.6	7.0	X	1.5
32.5	2.6	5.0	X	1.5
Underlying: Current Price 34.34				
35.0	1.45	2.95	0.9	2.4
37.5	0.35	1.85	2.15	3.9
40.0	X	1.5	3.7	6.1
42.5	X	1.5	5.6	8.6
45.0	X	1.5	8.0	11.0
47.5	X	1.5	10.5	13.5
50.0	X	1.5	12.1	16.9
52.5	X	1.5	14.6	19.4
55.0	X	1.5	17.1	21.9
57.5	X	1.5	19.6	24.4
60.0	X	1.5	21.9	26.9

Table 7.3: The example option chain at our snapshot. Contracts that do not have open interest are designated by an “X”.

actions associated with the options chain are available. For instance, the Random trader cannot sell the call at strike 60, because there is nobody in the market who is offering to buy that contract. The randomization is only over the set of contracts with open interest.

Now consider the Log-normal and Normal trading agents. Based on the historical volatility learned from the training set, the time until expiration, and the current price, the Log-normal trader sets its $\mu = 3.54$ and $\sigma = .0959$. Matching the mean and variance of this distribution, the Normal trader sets its $\mu = 34.5$ and $\sigma = 3.32$. From these values, by using the techniques we described in Section 7.2, we generate the prices in Table 7.4.

By comparing the two tables, we see that the set of positive-expectation actions for both traders is buying the underlying and selling the call at 32.5. As per our rules, only one of these actions is

Strike	Log-normal Put	Log-normal Call	Normal Put	Normal Call
17.5	0.00	17.00	0.00	17.00
20.0	0.00	14.50	0.00	14.50
22.5	0.00	12.00	0.00	12.00
25.0	0.00	9.50	0.00	9.50
27.5	0.01	7.01	0.02	7.02
30.0	0.10	4.60	0.14	4.64
32.5	0.52	2.52	0.57	2.57
Underlying: Both value at 34.50				
35.0	1.60	1.09	1.60	1.10
37.5	3.37	0.37	3.33	0.33
40.0	5.60	0.09	5.57	0.07
42.5	8.02	0.02	8.01	0.01
45.0	10.50	0.00	10.50	0.00
47.5	13.00	0.00	13.00	0.00
50.0	15.50	0.00	15.50	0.00
52.5	18.00	0.00	18.00	0.00
55.0	20.50	0.00	20.50	0.00
57.5	23.00	0.00	23.00	0.00
60.0	25.50	0.00	25.50	0.00

Table 7.4: The prices generated by the Log-normal and Normal traders for our example.

chosen at random to be performed in this time step.

The LMSR and Exponential utility traders both depend on the set of trades we have made in the past. For a tractable exposition that still exhibits realistic quantities, imagine that we currently hold the following portfolio:[2]

1. Long 200 underlyings

2. Short 200 puts at 50

[2]If we ran our trading agent on the chain over time, it is highly unlikely we would make 200 of the same trade, but we simulate prices with such a portfolio here to balance the competing desires of having a small number of distinct contracts with having a realistic-size portfolio.

3. Short 200 puts at 45
4. Long 200 calls at 25
5. Short 200 calls at 35

This portfolio corresponds to the payout vector given in Table 7.5.

Strike	Payout
17.5	8500.0
20.0	7000.0
22.5	5500.0
25.0	4000.0
27.5	2000.0
30.0	0.0
32.5	-2000.0
35.0	-4000.0
37.5	-5500.0
40.0	-7000.0
42.5	-8500.0
45.0	-10000.0
47.5	-11000.0
50.0	-12000.0
52.5	-12500.0
55.0	-13000.0
57.5	-13500.0
60.0	-14000.0

Table 7.5: The payout vector used in the LMSR for our example.

The b parameter used in the LMSR and in the Exponential utility agents is 52,875, which is 2500 times the initial underlying price at the first instance of the chain, 21.15. Starting from our set of holdings, we can calculate the fair prices for the LMSR trader by incorporating a prospective contract into our holdings and calculating the difference in cost. Unlike in the Log-normal and Normal distribution traders, the LMSR prices the bid and ask of each contract separately, with a spread. (Due to decimal truncation, the spread is not always visible in the displayed prices.)

Strike	Call Bid	Call Ask	Put Bid	Put Ask
17.5	19.42	19.42	0.0	0.0
20.0	17.1	17.1	0.18	0.18
22.5	14.95	14.95	0.53	0.53
25.0	12.96	12.97	1.04	1.04
27.5	11.14	11.14	1.72	1.72
30.0	9.47	9.48	2.55	2.55
32.5	7.96	7.96	3.54	3.54
35.0	6.59	6.59	4.67	4.67
Underlying: Bid 36.92 Ask 36.92				
37.5	5.36	5.36	5.94	5.94
40.0	4.26	4.26	7.34	7.34
42.5	3.3	3.3	8.88	8.88
45.0	2.46	2.47	10.54	10.54
47.5	1.75	1.75	12.33	12.33
50.0	1.17	1.17	14.24	14.24
52.5	0.7	0.7	16.27	16.28
55.0	0.35	0.35	18.42	18.43
57.5	0.12	0.12	20.69	20.69
60.0	0.0	0.0	23.08	23.08

Table 7.6: The option prices for the LMSR trader.

By comparing Tables 7.3 and 7.6, we find that the set of positive-expectation actions for the LMSR trader is to buy the call at between 25 and 47.5 inclusive, buy the underlying, and buy the puts between 27.5 and 42.5, inclusive.

For the Exponential utility trader, we use the same b parameter as the LMSR and the same μ, σ tuple as the Normal distribution. Just like in the LMSR, but unlike in the Normal distribution, we have separate bid and ask prices for contracts (though in this example the prices are close enough that they are equal when truncated). Table 7.7 displays the option chain prices for the Exponential utility trader.

The set of positive-expectation actions for the Exponential utility trader is to sell the call at between 30 and 37.5 inclusive, and to buy the underlying.

Now that we have demonstrated the experimental setup, we proceed to present the results of

Strike	Call Bid	Call Ask	Put Bid	Put Ask
17.5	16.85	16.85	0.0	0.0
20.0	14.35	14.35	0.0	0.0
22.5	11.85	11.85	0.0	0.0
25.0	9.35	9.35	0.0	0.0
27.5	6.88	6.88	0.03	0.03
30.0	4.51	4.51	0.16	0.16
32.5	2.47	2.47	0.62	0.62
Underlying: Bid 34.35 Ask 34.35				
35.0	1.04	1.04	1.69	1.69
37.5	0.31	0.31	3.46	3.46
40.0	0.06	0.06	5.71	5.71
42.5	0.01	0.01	8.16	8.16
45.0	0.0	0.0	10.65	10.65
47.5	0.0	0.0	13.15	13.15
50.0	0.0	0.0	15.65	15.65
52.5	0.0	0.0	18.15	18.15
55.0	0.0	0.0	20.65	20.65
57.5	0.0	0.0	23.15	23.15
60.0	0.0	0.0	25.65	25.65

Table 7.7: The option prices for the Exponential utility trader.

our experiments.

7.8 Results

We begin by providing the numerical results from our experiments and then discussing those results qualitatively.

7.8.1 Quantitative results

In terms of real-world trading performance, it would be most appropriate to quantify performance of trading agents in terms of net annualized return (e.g., "10% a year"). Of course, net return is a function of both value generated as well as value risked. Because of the form of the options contracts, determining the value risked is not straightforward. Single contracts, like selling a call, could lose an unbounded amount of money in the worst case. Combinations of options could amplify or hedge these losses. In practice, traders need to put up a certain fraction of their positions with the exchange (the margin) in order to maintain those positions. The precise margin amount depends on the rules of the particular exchange the options are traded on and is generally based around historical models of how prices move over time.

While percent return is difficult to calculate and depends on a host of practical matters, net performance (gain or loss) is simple to calculate. Therefore, we use net performance as a measure of trading agent performance. To normalize net performance, we divide a trading agent's gain or loss for a chain by the initial underlying price and by the number of days the chain is active. The resulting figure gives a meaningful way to compare chains where the underlying is in the thousands (like ^SPX) or the ones (like C), over differing numbers of days. We refer to this normalized value as *net underlyings per day (NUPD)*.

Table 7.8 provides summary statistics for the performance of each trading agent over the 114 testing chains in terms of NUPD. The Exponential utility agent had the most positive instances, highest mean, and best worst-observed performance. The Log-normal agent had the highest median performance. The Random trading agent had the worst performance along each of these dimensions.

Agent	Frac. positive	Mean	Median	Worst
Log-normal	.54	.15	**.89**	-6.6
Normal	.54	.12	.85	-6.7
LMSR	.46	-1.57	-.76	-37.8
Exponential utility	**.55**	**.32**	.70	**-3.6**
Random	.08	-1.58	-.82	-38.5

Table 7.8: A summary of comparing each trading agent along a number of dimensions. Mean, median, and worst-observed trials are measured in terms of NUPD. Higher values are better.

To assess the significance of the values in Table 7.8, we performed a bootstrap analysis that simulated running our experiments 10,000 times. Bootstrap analysis was a natural choice for this setting because of the complexity of estimating e.g., worst-observed loss using standard techniques (Davison and Hinkley, 2006). The head-to-head results of this analysis is given in Table 7.9. For instance, the first value in the table indicates that in 68% of our bootstrapped experiments the Exponential utility trader had a higher fraction of positive runs than the Log-normal trader.

Trader	Frac. positive	Mean	Median	Worst	Opponent
Exp. utility	68	93	26	99	Log-normal
Exp. utility	60	95	25	100	Normal
Exp. utilty	100	100	100	100	LMSR
Exp. utility	100	100	100	100	Random
Log-normal	48	57	56	98	Normal
Log-normal	100	100	100	100	LMSR
Log-normal	100	100	100	100	Random
Normal	100	100	99	100	LMSR
Normal	100	100	100	100	Random
LMSR	100	57	42	98	Random

Table 7.9: Percent of the bootstrapped experiments in which the trader on the left had a higher number of positive instances, mean, median, and worst-observed loss relative to the trader on the right.

Table 7.10 shows the relative performance of each trading agent against the others. The values are the number of testing chains in which the agent in the row beat the agent in the column. Our results establish the strict transitive ordering Exponential utility > Log-normal > Normal > LMSR > Random, where "$a > b$" means that trading agent a beat trading agent b in a majority of our testing chains.

	Log-normal	Normal	LMSR	Exp. utility	Random
Log-normal	X	**66**	**69**	52	**75**
Normal	48	X	**70**	49	**77**
LMSR	45	44	X	41	**62**
Exp. utility	**62**	**65**	**73**	X	**82**
Random	39	37	52	32	X

Table 7.10: The number of times the agent in the row beat the agent in the column in our 114 testing chains. Majority winners are denoted in bold.

Table 7.11 adds statistical significance context to the results observed in Table 7.10. Given the produced data, the hypothesis "Trader A has a higher NUPD than Trader B on this chain" was tested for all traders and chains. If the probability that a trader outperforms the other was greater than 0.99, it was recorded as a win for that trader and a loss for the other. All the chains in which the probability lied between 1% and 99% are recorded as ties. Table 7.11 shows the resulting counts. All of the conclusions from Table 7.10 still hold; in particular, the Exponential utility trader is still the Condorcet winner.

Trader	Wins	Ties	Losses	Opponent
Exp. utility	41	40	33	Log-normal
Exp. utility	36	44	34	Normal
Exp. utility	71	5	38	LMSR
Exp. utility	75	12	27	Random
Log-normal	20	80	14	Normal
Log-normal	68	7	39	LMSR
Log-normal	68	13	33	Random
Normal	66	5	43	LMSR
Normal	71	8	35	Random
LMSR	57	9	48	Random

Table 7.11: Counts of head-to-head performance of traders taking into account statistical significance. If a trader had $>$ 99% chance of out-performing its opponent on a chain, it is recorded as a "Win". If it had $<$ 1% chance, it is recorded as a "Loss". All other significance levels are recorded as "Ties".

Table 7.12 compares the performance of the Exponential utility trader against its "parents", the Normal distribution trader and the LMSR trader. The Exponential utility agent beats the performance of both of these traders in 42 of the testing chains (37%) while losing to both in only 18 of the chains (16%). It outperformed a blend composed of equal parts of Normal and LMSR traders in 76 testing chains (67%).

The NUPD of each trading agent over our testing chains can be viewed as noisy realizations of a continuous random variable. We can recover this variable by smoothing the realizations with a kernel. Figure 7.3 shows the kernel-smoothed CDFs of these random variables for a likelihood-maximizing Gaussian kernel. The figure plots the fraction of instances that had net performance no better than the given value. Both the LMSR and Random trading agents had chains on which they performed worse than -10 NUPD, and so the CDFs for those traders do not start 0 on the plot. The Normal and Log-normal CDFs are indistinguishably close together for much of the plot.

Our 114 chains include a mix of bonds, indices, equities, and commodities. We did not ob-

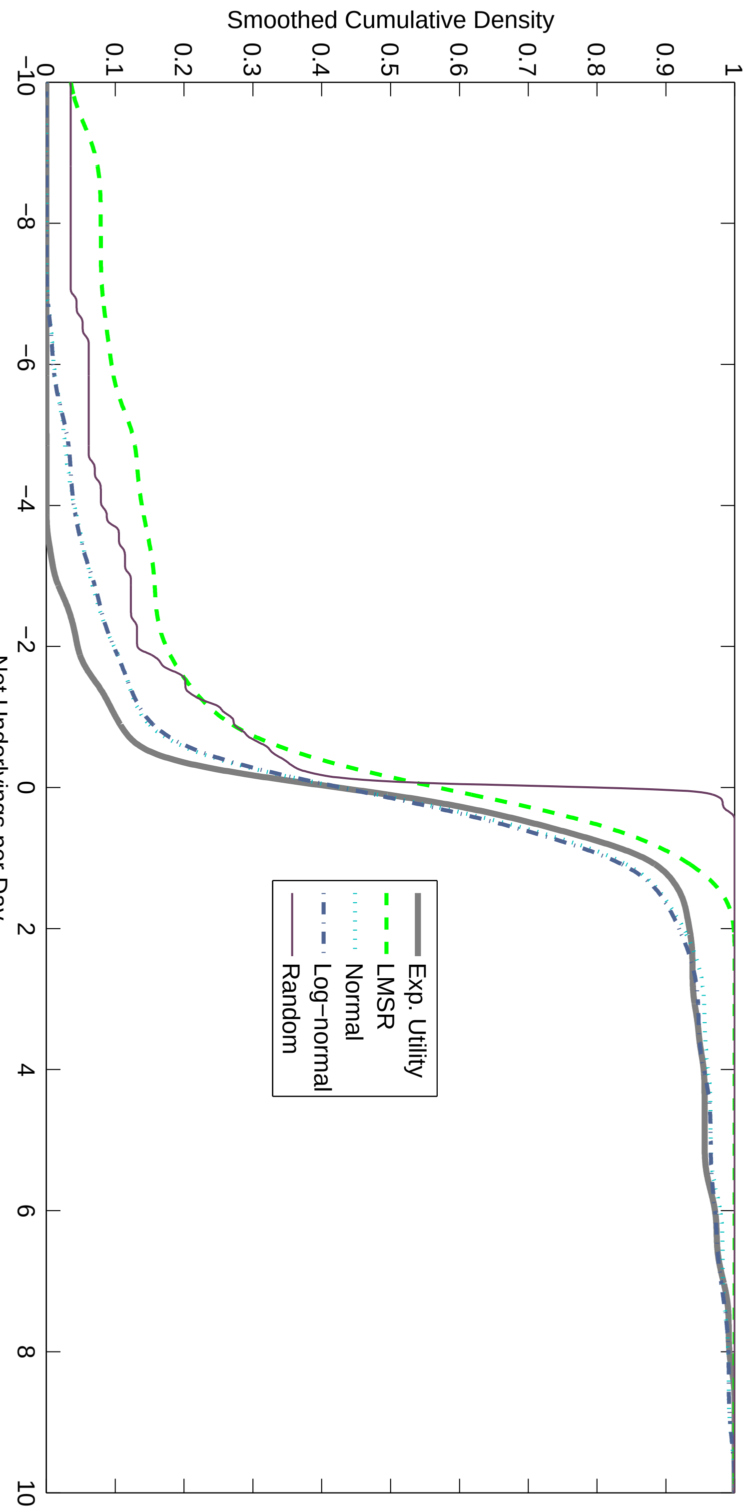


Figure 7.3: The kernel-smoothed CDFs of the NUPD for each of our traders. The Log-normal and Normal CDFs coincide for much of the plot. Lower function values are better.

Relative Exponential utility performance	Frequency
Beats both	.37
Beats one, loses to one	.47
Loses to both	.16
Beats blended	.67

Table 7.12: Distribution of the performance of the Exponential utility trader against the Normal and LMSR traders. "Blended" refers to an equal parts mix of the Normal and LMSR traders.

serve a substantial difference between the relative performance of our trading agents in any of the underlyings. This may be due to the fact that the volatility parameter σ is fit differently for each underlying, allowing for appropriate responses to both volatile and stable underlyings.

One notable feature of the financial markets captured in our dataset was the financial collapse of late 2008. Chains that expired in late 2008 and early 2009 showed the worst performance for our parametric traders. In particular, the Log-normal and Normal traders delivered their worst performances over the ^FVX chain that expired December 20th, 2008. The performance of the underlying from the expiration of the prior chain on September 22nd to expiry is plotted in Figure 7.4. There are several days in which the underlying moved down or up more than 10%. The collapse was not an "anomaly" in our dataset. It was a real event that our trading agents would have been involved in and must be considered when evaluating trading agents on real data.

Finally, we did not see any significant difference in the total volume traded by each agent. We attribute this to the diversity of contracts offered to the trading agents at each time step, all of which generally have tight bid/ask spreads. In our simulation, an agent needs to find only one of the dozens of possible actions desirable at each time step in order to trade. This result could be seen as a consequence of selecting a b parameter for the inventory-based traders large enough to result in small bid/ask spreads (e.g., the snapshot example in Section 7.7); substantially smaller b values would have resulted in larger bid/ask spreads in the agents' prices and consequently less trading activity.

7.8.2 Qualitative results

In this section we attempt to distill our quantitative findings into qualitative facts about the performance of our automated traders.

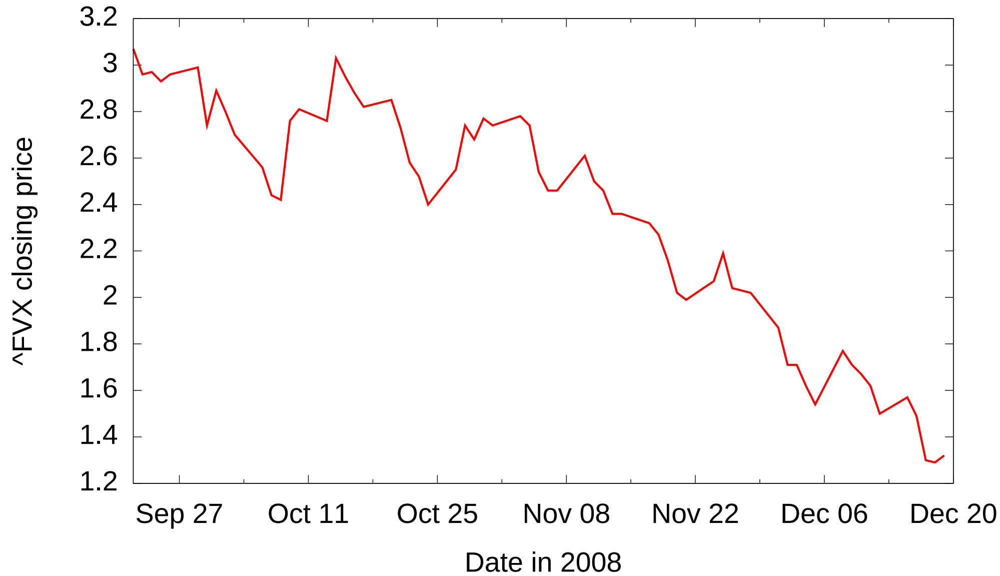


Figure 7.4: The option chain expiring December 20th, 2008 was particularly volatile, leading to poor performance by the parametric traders.

The Random and LMSR traders had the worst performance

Both the Random and LMSR traders were characterized by low mean and median performance and terrifically bad worst-case losses. As Figure 7.3 shows, the LMSR was much more volatile than Random. Random had its performance on the vast majority of the testing chains (about 80%) fall between -2 and 0 NUPD, while the LMSR had many testing chains do better or worse.

As we have discussed, the LMSR is equivalent to an agent with exponential utility and a uniform prior over the strikes. One interpretation of this uniform prior is that the LMSR neither has nor relies on any domain knowledge. Our quantitative results with the LMSR are in line with the recent findings of Brahma et al. (2010) that suggest the LMSR struggles in comparison to trading agents with domain knowledge, and of Chakraborty et al. (2011), who compare the LMSR to a Bayesian market maker that relies on both priors and inventory. Their lab experiments showed that the Bayesian market maker was generally much more profitable than the LMSR. Furthermore, the LMSR's results are not unexpected; it is traditionally used to provide liquidity and subsidize a set of traders for their information in Internet prediction markets. With its losses here, the LMSR did the same thing in our experiments.

One dimension along which the LMSR was able to out-perform Random significantly was the fraction of positive instances over the testing chains. Random recorded positive NUPD in fewer than eight percent of our trials, while the LMSR was positive on about 47% of the trials. Table 7.9

shows that in 100% of our bootstrapped experiments the LMSR trader had a higher fraction of positive instances than the Random trader. This confirms our intuition that the Random trader would "eat the spread" and lock in small losses. Interestingly, the performance of Random was the best relative to the other traders on the highly volatile chains at the end of 2008. For instance, the Random trader had the best performance of all the traders on the ^FVX chain that expired December 20th, 2008, losing about 1.2 NUPD (better than its mean performance over the testing set as a whole). We credit this to the fact that the Random trader is highly non-parametric, and so its performance is not affected by the relative volatility of the underlying.

The LMSR learned plausible distributions

While the LMSR lagged in quantitative performance, that does not mean the concepts behind it are unsound. Figure 7.5 shows an in-progress run of the LMSR (on an ^IRX chain). The implicit probability distribution over strike prices in the LMSR closely matches the fit produced by the Normal distribution trader. This is significant because the Normal trader knows the historical volatility and the current underlying price, while the LMSR trader only knows the trades it has made. This is made more remarkable by the fact that the trading agent has only six crudely-shaped tools (buying or selling calls, puts, or the underlying) to create this distribution.

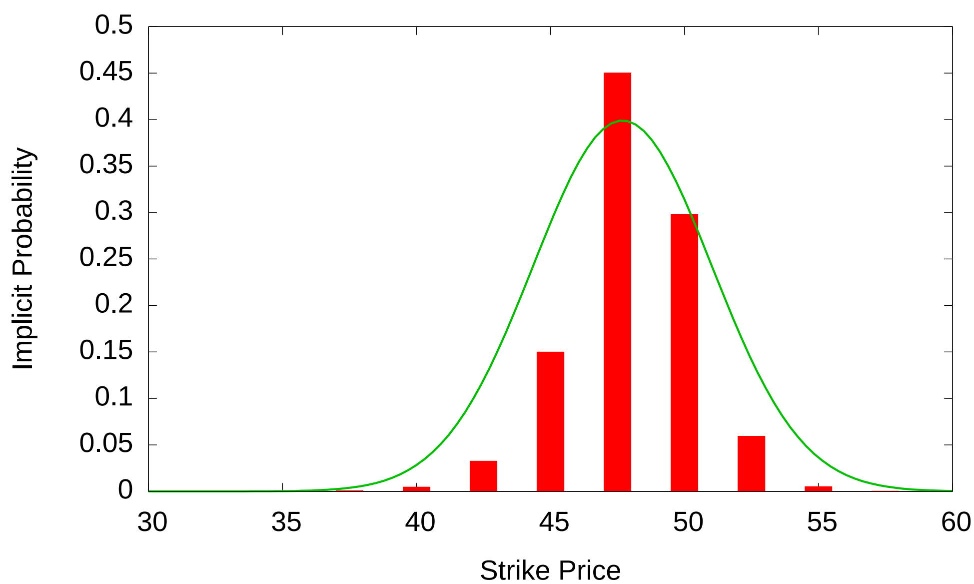


Figure 7.5: In this capture of an in-progress run, the LMSR trader's implied probability distribution (red bars) closely matches the projection of the Normal distribution trader (green curve).

One perspective on what is happening is that the LMSR learns the correct distribution of prices because it is equivalent to a no-regret learning algorithm (Chen et al., 2008; Chen and Vaughan, 2010). Essentially, with each time step through the options chain the LMSR trader makes a small correction to get its implicit probability distribution closer to the market's distribution. This also implies the similarity between the LMSR's probabilities and the Normal distribution trader in Figure 7.5 is partly fallacious, because the LMSR has not been learning from that specific snapshot of the chain but rather making a series of small adjustments in probabilities over time.

A deeper perspective on this learning process is visible in Figure 7.6. This figure shows the ˆIRX chain used above and three other (arbitrarily chosen) chains with nearby expiration dates, and measures the K-L divergence of the LMSR trader's marginal prices and the Log-normal trader's probability distribution.

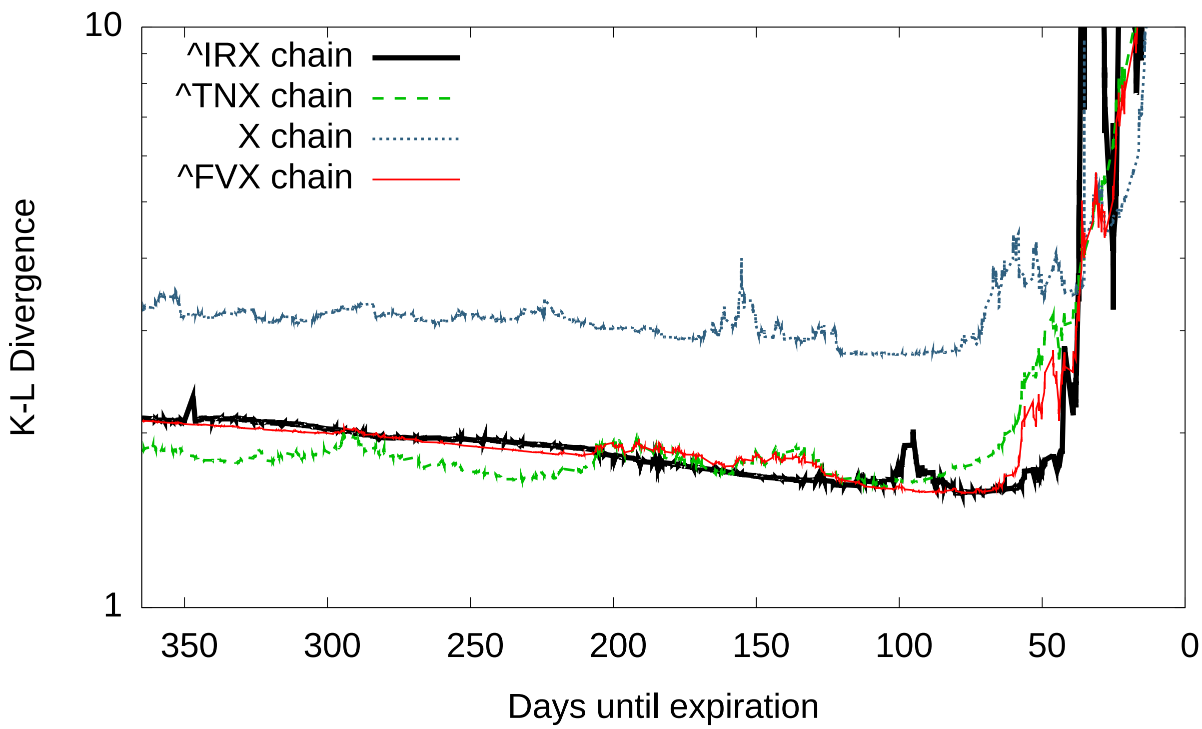


Figure 7.6: The K-L divergence of the LMSR from the log-normal distribution taken over four chains.

Recall that K-L divergence is a measure of how dissimilar two distributions are. (Decreasing K-L divergence means the distributions are more similar.) Formally, let LN^t denote the log-normal probability density function at time t, and let π_i^t be the marginal probability of the LMSR trader at strike price s_i. Then the K-L divergence at time t, KL^t, on the above plot is calculated as

$$KL^t \equiv \sum_i \pi_i^t \log_2 \left(\frac{\pi_i^t}{LN^t(i)} \right)$$

One interpretation of the K-L divergence is the number of extra bits required for the log-normal distribution to encode the LMSR marginal distribution.

Although the trend is not consistent, the K-L divergence between the two distributions seems to decrease slowly until roughly 75 days before expiry, when it begins to increase significantly. We have truncated the plot at an upper boundary of 10, but in the final days before expiry the K-L divergence increases without bound. It appears to be accurate, then, to divide the LMSR's behavior into two regimes: the last 75 days, in which the K-L divergence becomes arbitrarily large, and the time previous, in which the K-L divergence is stable or slightly falls. This earlier period represents the LMSR's "learning" process, as the marginal prices begin to approach the log-normal distribution.

The reason the K-L divergence increases so dramatically immediately before the market expires is that the log-normal distribution becomes tighter and tighter around a single value (the current price), eventually converging to a single unit mass at the expiration price in the final time step. These progressively tighter distributions require huge numbers of additional bits to encode the more spread out LMSR distribution, because the extreme strike prices have such a low likelihood when the log-normal distribution is tight. Consequently, the K-L divergence between a tight, late-stage log-normal distribution and a more diffuse LMSR distribution is large, and as the log-normal distribution approaches a unit mass, it goes to infinity.

Log-normal slightly outperformed Normal

A log-normal distribution is intuitively a better and more-realistic fit for stock prices than a normal distribution, because the former reflects that the stock price cannot go below zero. Reflecting this intuition, the Log-normal trader performed better than the Normal trader in our experiments. The Log-normal trader had a slightly higher mean and median performance than the Normal trader and won the majority of head-to-head comparisons between the traders. However, the performance characteristics on the whole were close, as can be seen visually by the overlapping density lines in Figure 7.3. Table 7.9 shows that in our bootstrap analysis, neither trader had a higher fraction of positive instances, mean NUPD, or median NUPD in more than 57% of the experiments. Finally, Table 7.11 also shows that for the bulk of option chains (80 out of 114, or 70%), neither the Log-normal or Normal trader produced higher values with 99% confidence. This similarity could be considered as a likely consequence of the design of the two traders, because the Normal trader matches the mean and standard deviation of the Log-normal trader at each timestep.

Exponential utility won but did not stochastically dominate

The Exponential utility trader was the winner in our trials by most of the measures we used. It had the highest mean and fraction of positive testing instances, a much better worst-case loss, and only a slightly lower median than the Log-normal and Normal traders. The bootstrap analysis in Table 7.9 suggests that the Exponential utility trader consistently outperformed the Log-normal and Normal traders in terms of mean performance and worst-case loss, but that neither the Exponential utility trader's higher fraction of positive trials nor the Normal and Log-normal traders' higher median performance were observed in more than 75% of the bootstrapped experiments. The Exponential utility trader was also the Condorcet winner in head-to-head comparisons against the other traders, beating each of them over a majority of the testing chains. Furthermore, the Exponential utility trader outperformed a mix of the Normal and LMSR traders in two-thirds of our testing chains, indicating that it is more sophisticated than a mere combination of the two techniques. However, the Exponential utility agent did not stochastically dominate the finance literature agents, and so it is possible for some utility functions to prefer the returns of the standard BSM model instead. These utility functions would weight average- and better-case performance and discount worst-case losses.

Exponential utility had more accurate actionable beliefs

One perspective on how the Exponential utility agent performed so well can be found by considering the areas of Figure 7.3 where its CDF diverges from the CDF of the Normal and Log-normal traders. There is a large gap between the lines for negative NUPD, where the Exponential utility trader out-performs the Log-normal and Normal traders, and a smaller gap between 1 and 2 NUPD, where the Log-normal and Normal outperform Exponential utility. What this implies is that the Exponential utility trader is practicing a form of insurance against bad outcomes. It transfers wealth between states of the world in which good things happen (the positive net return realizations) into states of the world in which bad things happen (negative net return realizations). As a result, the good cases become slightly worse (the gap between Normal/Log-normal and Exponential utility on the positive side) but the bad cases are severely reduced (the gap on the negative side).

This interpretation makes sense when we consider that the Exponential utility trader is a risk-averse analogue of the Normal trader. As a risk-averse trader the Exponential utility agent is more willing to hedge future risk, trading off future profits to avoid large losses, and will not increase its exposure to existing risks unless at the offered prices doing so seems exceptionally profitable.

One of the ways the Exponential utility trader accomplished this insurance is by having effectively heavier tails (more probability mass) on extreme cases than its corresponding normal prior. These tails make actions like buying a low-strike put or buying a high-strike call more desirable. These contracts are out of the money, and so they would require significant price movement to not expire worthless; consequently they are also priced cheaply. When a trading agent purchases one of these contracts, a small amount of wealth is transferred from states where extreme events are not realized to become a larger amount of wealth in states in which those extreme events are realized.

This insurance allowed the Exponential utility trader to outperform the Normal and Log-normal traders, because those traders' models were not correct but those agents traded as if they

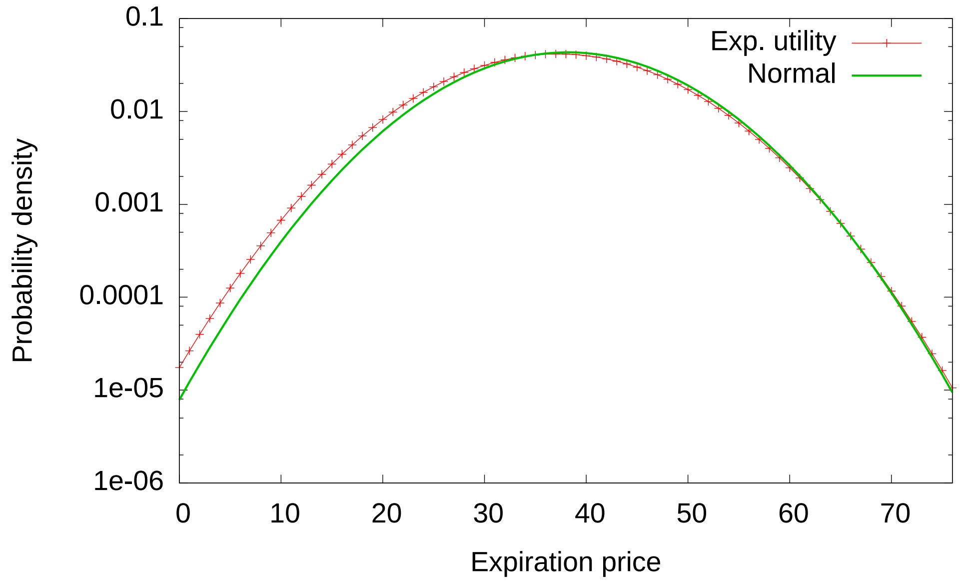


Figure 7.7: In this capture of an in-progress run, the heavier tails of the Exponential utility trader relative to the Normal trader are evident. The y-axis is log-scaled.

were. Consider that, if a trader's beliefs are indeed the correct model of the world, then a risk-neutral agent trading on those beliefs will have a higher expected return than a risk-averse agent trading on those beliefs. (This is because a risk-neutral agent maximizes his expected return by definition.) Put another way, taking insurance should not increase a risk-neutral agent's payout if that agent was acting on correct beliefs. Since, in our experiments, the risk-averse Exponential utility agent had higher expected returns than the risk-neutral parametric traders from the finance literature, the latter traders' models of the world were incorrect. Specifically, the heavier tails of the Exponential utility agent could be a more accurate distribution over the expiration price if there is a chance of large downward shocks to the price (e.g., in the financial crisis).

Figure 7.7 is an in-progress shot of heavy tails in a representative run (in this case, for a GE chain). Here, the underlying price is about 38. The implied probability distribution of the Exponential utility trader has about twice the density at low underlying realizations (the plot is log-scaled to make this more clear). For a continuous constant-utility cost function with belief density μ, we can calculate the implied probability density at t for vector $\mathbf{x}$ by

$$\nabla C(\mathbf{x})(t) = \frac{\mu(t)u'(C(\mathbf{x}) - x(t))}{\int_0^\infty \mu(k)u'(C(\mathbf{x}) - x(k))\, dk}$$

7.9 Modifying the experiments

In this section we examine two changes to the experiments we ran in the previous section. First, we relax the random uniform trading restriction, allowing trading agents to trade only the contract they think is most beneficial at each time step. Second, we examine the performance of an LMSR trader that incorporates a better discrete prior.

7.9.1 On the random uniform trading restriction

Recall that our second constraint on trading simulations was to select a contract to trade uniformly at random from the set of contracts identified as favorable. This was done to avoid overfitting on the fact that our data comes in the form of static snapshots over the option chains. In a real trading environment, favorable trades will come and go, possibly quickly enough to preclude our trading on them. Therefore, selecting uniformly at random provides a more conservative measure of how each trading agent would perform in real settings.

With this scheme, the performance of each trading agent on each testing chain becomes a random variable. When the trading agent keeps state (as in the LMSR and Exponential utility agents), then this random variable becomes quite complex as future actions depend on the actions selected randomly in the past.

This randomization gives rise to two concerns: First, that the variance between different runs (realizations of this random variable) is large enough to mitigate the significance of the differences between agent performance. Second, that uniformly trading, rather than trading the best contract from the set of actions, unfairly penalizes some agents over others. In this section, we study each of these concerns in turn.

The difference between runs was generally small

We completed four runs over each agent over each testing chain. To measure volatility, we took the maximum difference in NUPD between the four runs. This can be considered an adversarial measure that is particularly sensitive to the outliers from each run. The resulting differences appear in Table 7.13.

These results indicate that the runs were fairly robust over different realizations of trading strategies. For all but the Random agent, about two-thirds of all runs had all four experimental runs fall in a range less than 1/16 NUPD wide, and those agents had all of their runs fall within a 0.5 NUPD band in more than 95% of the chains.

As could be expected, the Random agent showed particularly large volatility between different runs. This is because the Random agent had the largest set of possible actions in each time step (i.e., all the available contracts). The Random agent over the ^XAU chain expiring on 2009-09-30

	Percent with max difference smaller than					
Agent	0.0625	0.125	0.25	0.5	1.0	2.0
Log-normal	64	83	96	99	99	99
Normal	66	83	92	97	97	98
LMSR	67	83	92	96	97	98
Exponential utility	68	82	92	97	97	98
Random	46	65	83	92	96	96

Table 7.13: We performed four runs over each trading agent in the dataset. The percent of chains with difference between maximum and minimum NUPD is noted for each agent.

had the largest difference in trials overall, with NUPD of -0.10, -0.11, -10.3, and -10.4. This appears to be due to a possible anomaly within the dataset: the best ask on a put option at 105 on September 28th, 2009 is listed in our data set at 200,000 dollars. When the Random agent buys this contract, it induces a massive loss for the chain as a whole. It is unclear whether this contract was actually listed at this price on the exchange, or if it is an error in the data set. Regardless, because this price was so uncompetitive the other, intelligent, trading agents were able to avoid it.

Trading only the best contracts produces no qualitative changes

To see whether our results were significantly affected by trading uniformly at random from the set of contracts they deemed favorable, we also simulated our trading agents trading only the best (highest expected profit) contract from the set of contracts they deemed favorable. Observe that this produces a deterministic trading agent. (Consequently, the Random trading agent is no longer relevant in this setting and is omitted.) We are interested in testing whether our qualitative results are merely an artifact of the restriction to trading uniformly or not. We want to examine how the agents from the finance literature perform relative to the Exponential utility agent when only the best contract is traded.

Table 7.14 shows the mean and median NUPD of the deterministic agents relative to trading uniformly. They are phrased in terms of *surplus NUPD*, subtracting the deterministic NUPD for each chain from the average NUPD of that chain when trading occurs uniformly at random.

As might be expected, trading only the best contracts produces slightly higher median NUPD for all the deterministic agents. The LMSR agent is notable for having significantly higher mean surplus NUPD than the other agents. This appears to be due to a combination of two factors: trading only the best contracts allows the LMSR agent to avoid making some of the worst trades, and furthermore, the LMSR agent's performance was already poor enough to allow for a large

	Surplus NUPD	
Agent	Mean	Median
Log-normal	0.00	0.04
Normal	0.04	0.09
LMSR	1.13	0.10
Exp. utility	0.26	0.13

Table 7.14: The difference in NUPD between deterministic agents that only trade the contracts they think will deliver the best return versus trading uniformly at random from the set of agreeable trades, over the 114 testing chains.

boost. These results suggest that trading only the best contracts does not change the qualitative performance ordering of the agents: the Exponential utility trader outperforms the Log-normal trader, which does about as well as the Normal trader, which in turn outperforms the LMSR.

One feature of simulating only the best contract that we observed, especially for the parametric finance literature traders, was the tendency to buy (or sell) exactly the same contract over and over again. It is easy to see why this would be the case if there is not a tremendous amount of movement in the market in-between 15 minute intervals; in this case, the most agreeable contract at the current time step is likely to be the most agreeable contract 15 minutes later, as well. Particularly for contracts at extreme strike prices, this kind of behavior is inherently unrealistic in a simulation because sustained trade in a single contract is likely to move market prices. The Exponential utility and LMSR agents are able to mitigate this phenomenon because as they trade a contract, further expansion of that position becomes less desirable. However, our principal way of simulating in the paper—taking a uniform sample from the set of favorable trades—is a more conservative simulation. As we have shown here though focusing on only the best contracts does not change our qualitative conclusions.

7.9.2 The LMSR with a better discrete prior

Recall that there were two differences between the Exponential utility trader and the LMSR trader: a better prior, and a continuous relaxation of the event space. To investigate which of these changes was responsible for the improvement in performance, in this section we examine incorporating a good discrete prior into the traditional LMSR.

Recall from Chapter 3 that the LMSR with a discrete prior π corresponds to the cost function

$$C(\mathbf{x}) = b \log \left(\sum_j \pi_j \exp(x_j/b) \right)$$

We form the *Discrete-prior LMSR trader* by setting

$$\pi_i \propto LN(x_i)$$

where LN is the density function of the Log-normal trader for the same option chain at the same snapshot in time. To facilitate comparison, we keep the b parameter the same as in the LMSR and Exponential utility trading agents, and we use the same compression of the state space to strike prices that we developed and motivated in Section 7.3.1.

Results

We replicated our experiments from Section 7.6 with the Discrete-prior LMSR trader. Table 7.15 compares the Discrete-prior LMSR trader against the LMSR, Log-normal, and Exponential utility traders. The Discrete-prior LMSR trader outperforms the LMSR trader over 63% of the option chains, but loses a similar frequency of head-to-head matchups against the Log-normal and Exponential utility trader.

Comparison Trader	Fraction
LMSR	.37
Log-normal	.61
Exponential utility	.64

Table 7.15: Fraction of options chains on which the LMSR, Log-normal, and Exponential utility traders out-performed the Discrete-prior LMSR.

Table 7.16 compares the performance of the Discrete-prior LMSR with the Log-normal, LMSR, and Exponential utility traders. The Discrete-prior LMSR had performance that closely matches the original LMSR, although with slightly better mean, median, and worst-case loss.

Discussion

While the Discrete-prior LMSR trader outperformed the LMSR trader, the difference was slight. The common factor between the two agents was the compression of the state space to just the relevant strike prices, and our results suggest this compression was the dominant factor in determining

Agent	Frac. positive	Mean	Median	Worst
Discrete-prior LMSR	.46	-1.54	-.73	-37.7
Log-normal	.54	.15	.89	-6.6
LMSR	.46	-1.57	-.76	-37.8
Exponential utility	.55	.32	.70	-3.6

Table 7.16: Comparison of the Discrete-prior LMSR with other trading agents.

the performance of the traders. Since both the Log-normal and Exponential utility traders have much better performance than the Discrete-prior LMSR, this suggests that as the number of events (future expirations) considered increases, the performance of the discrete-prior LMSR would improve. This is because as the number of events increases, the discrete market maker becomes a closer approximation to these continuous market makers. However, as we have discussed, an increasing number of events makes the computation of values much more challenging, and compressing the state space to only the traded strike prices made a huge state space tractable without opening the trader up to unbounded loss. Still, it seems that compressing the state space to strike prices, while not affecting worst-case loss, does come with the hidden cost of a much less expressive prior.

One issue that should not be lost in this discussion is that using a discrete market maker with a fine-grained prior does not provide a way to circumvent the impossibility results of Section 7.4. Consider using the LMSR with a large number of expiration prices (events) with a prior that matches the distribution of the Log-normal trader. The prior distribution that is actually being generated is a log-normal distribution truncated at the upper bound of the event space (the largest expiration price considered). This truncation scheme is covered by Proposition 32, which shows that exponential utility will produce meaningful prices if the prior distribution is identically zero above some upper bound. However, as we discussed in Section 7.4, this truncation is numerically hazardous because its effect on prices is unpredictable and highly sensitive to the upper bound.

7.10 Extensions

In terms of practical impact, we acknowledge that there are considerable gaps between our experiments and the actual implementation of a trading strategy. These include features like margin rules, which dictate how much of our currently-held position we are required to stake, and trading costs, which determine how expensive it would be to take on new positions. Of course, once these practical hurdles are known they could be incorporated into the decision logic of a trading agent, but regardless, they are likely to affect performance. However, we do believe that our experiments are robust in terms of implementation, particularly because we handicapped the trading volumes and frequencies of our agents. This was done with the intention of producing experimental results that would be more accurate reflections of the actual performance of those strategies in real markets. In summary, we are guardedly optimistic about the efficacy of this line of research in practice, although implementing these strategies in a real market poses obstacles that could deleteriously impact returns.

The framework we used to construct the Exponential utility trader, combining a utility function with a probability distribution, is very general. We believe there are significant opportunities for expanding and broadening the model in terms of each component.

Probability distribution In the time since the BSM model was first promulgated, other ways of analyzing how prices move through time have also been proposed that better fit actual performance. These include GARCH (Engle and Ng, 1993) and jump models (Kou, 2002), and they imply probability distributions over future realizations that are not log-normal. Another intriguing possibility for a prior distribution is the technique of Ait-Sahalia and Lo (1998) that uses non-parametric techniques to fit a probability distribution to the prices in an option chain.

Utility function On the other side, it would be interesting to explore other utility functions besides exponential utility. We are particularly concerned that exponential utility's extreme weighting at large values has a tendency to produce undefined prices for many distributions. Perhaps a utility function that had polynomial risk aversion, rather than exponential, would provide a more intuitive price response and would allow for the use of a broader range of prior distributions. Regardless of the utility function-prior distribution pairing, they must still obey our results in Section 7.4.1. In particular, power-law utilities (including log utility, which could be considered a standard choice) do not correspond to traders that produce meaningful prices.

Most well-established option trading strategies[3] in the literature operate by trading multiple contracts in a single chain. We modeled our experimental methodology after these approaches, and in our experiments we treated each testing chain independently. But in practice, there is a correlation between chains with different expiration dates for the same underlying. For instance, if an underlying expires at a low realization, it is more likely that the chain expiring three months later will also expire at a low realization. Another extension to our model is to explore how to add a time dimension to incorporate information between the chains of different expiration dates for the same underlying. In this view, the prior distribution would be over *paths* of prices over time, and when contracts are traded at expirations it affects the probability distribution over those paths. Furthermore, different expiration dates are linked through the underlying; a trading agent may make plans to purchase an underlying now to sell at a specific date in the future, or to sell conditional on how prices move over time.

[3]The Chicago Board Options Exchange (CBOE) operates `http://www.cboe.com/Strategies`, which features popular options trading strategies.

Finally, we have applied our techniques to options markets but the experimental success of the Exponential utility trader suggests traders that take into account both a good prior distribution as well as their previous trades could be successful in other real applications, too. There are many settings where we have good priors over the future, and they often involve considerable financial risk and reward; examples include a casino handling sports betting or a proprietary trading desk at a bank. We are interested in adapting the synthesized model to other trading activities where we have reasonable priors over the future and the ability to trade through time.

Chapter 8

Rational market making with probabilistic knowledge

The previous chapters of this thesis have taken a quasi-adversarial view of a market maker's counterparties and over the events themselves. The market maker's goal is not to minimize worst-case loss, and despite possibly having a subjective prior over the events, he is sensitive to the bets made by counterparties although he is oblivious to what those bets will be.

In this chapter we consider a different setting in which the market maker has a good prior on the future state of the world, and on how traders will bet for different prices the market maker offers. This may be a more realistic setting for market making settings like Las Vegas sports betting or a proprietary trading desk at a bank, where significant investment has been made in developing knowledge over the future states of the world.

A complication arises when traders and the market maker have substantially different beliefs. Then, the market maker must balance two competing factors: the desire to hedge bets for a certain profit, and the desire to profit in expectation from wagers made at favorable prices. For instance, a market maker could find itself in a situation where it could either increase its exposure to an event it thinks will probably occur at a bargain price, or hedge out its current risk on that event in order to guarantee a small but certain profit.

In this chapter, we compute the policy of a *Kelly criterion* market maker over a series of interactions with traders. The Kelly criterion (Kelly Jr, 1956) is a way to make bets that mandates maximizing the expected log utility of a setting. While simple as a guiding precept, the Kelly criterion accomplishes a broad range of objectives: over a series of bets, it is the fastest way to double an initial investment, produces the highest median wealth, and produces the highest mode wealth (Poundstone, 2006).

Our experimental results show that a Kelly criterion market maker follows a complex time-dependent strategy. In the early stages of wagering, the market maker will attempt to match orders to profit from the bid/ask spread. Towards the end of trading, the policy gradually shifts to myopic optimization on the market maker's private beliefs. Perhaps surprisingly, we show that in the early stages of the market, profiting from the bid/ask spread dominates the desire to sell inventory at agreeable prices, that is, if it facilitates more trade, a Kelly criterion market maker should buy obligations at a price higher than, or sell obligations at a price lower than, its private beliefs. Moreover, because the inventory a risk-averse market maker accumulates affects the prices it offers, the market maker could offer bets that are myopically irrational for the entire trading period. This is in contrast to a risk-neutral market maker that would never offer a myopically irrational bet.

8.1 Model

Following Glosten and Milgrom (1985), the setting is a repeated sequential interaction between the market maker and a set of traders. In each period, the market maker sets prices for a finite set of bets, and then a trader is drawn randomly from a large pool of potential traders. That trader enters the market and selects one of the offered bets to make with the market maker (or none at all). After a finite number of periods the process halts, one of the n events is realized, and the bets are settled with the traders.

8.1.1 Traders

The traders have the following features:

- Traders are *anonymous*, so there is no way for the market maker to distinguish between traders. Anonymity is a standard component of many models in the literature (for example, Feigenbaum et al. (2003) and Das (2008)), because it is natural for settings where prices are posted publicly, as is the standard in electronic markets.
- Traders are *myopic*, not strategic. They exist for only a single period: they enter the market, perceive the prices offered by the market making agent, select a bet to take (or no bet), and then exit. The traders do not learn from historical prices or strategize about their behavior. Myopic traders (also known as *zero-intelligence agents* or sometimes *noise traders*) are a feature of much of the literature (Glosten and Milgrom, 1985; Kyle, 1985; Othman and Sandholm, 2010b). Empirical studies of market microstructure have shown that the behavior of these agents is qualitatively very similar to behavior observed in real markets with human traders (Gode and Sunder, 1993; Othman, 2008). However, in some settings the simple behavior of these agents may be an unrealistic model (Chen et al., 2007; Dimitrov and Sami, 2008; Chen et al., 2010).

- The number of trading periods is drawn independently of the market maker's policy. Since traders have the ability to decline to place a bet with the market maker if they do not find the offered bets agreeable, this condition means that the number of traders placing bets with the market maker is *not* a constant—instead, it will depend on the market maker's policy. We assume the market maker knows the true distribution of the number of trading periods.

8.1.2 Utility and the Bellman equation

The market maker's *state* can be represented by a tuple $(t, \mathbf{w})$ of the index of the participating agent $t \in \{1, 2, \ldots\}$, and the *wealth vector* $\mathbf{w}$, where w_i is the market maker's wealth (payoff) if state of the world $\omega_i \in \Omega$ is realized. (Since exactly one trader appears in each period, the variable t can be thought of as an index over discrete time.) There is a termination state $(\bar{t}, \mathbf{w})$, where the market maker gets an expected utility payout based on his subjective beliefs $\hat{\mathbf{p}}$, which he believes to be the correct distribution over the possible futures:

$$V(\bar{t}, \mathbf{w}) \equiv \sum_{i=1}^{n} \hat{p}_i u(w_i)$$

Without loss of generality, a risk-neutral market maker receives its expected linear utility on termination:

$$V(\bar{t}, \mathbf{w}) \equiv \sum_{i=1}^{n} \hat{p}_i w_i$$

A Kelly criterion market maker receives its expected log utility on termination:

$$V(\bar{t}, \mathbf{w}) \equiv \sum_{i=1}^{n} \hat{p}_i \log(w_i)$$

The *bets* a market maker offers can be expressed by vectors in payout space $\mathbf{x} \in \mathbb{R}^n$, so that x_i is the *trader's* payoff (that is, the market maker's loss) if ω_i is realized. For instance, imagine that the market maker is fielding bets on which of three horses will win a horse race. A bet that pays the trader 10 dollars if the first horse wins, 5 dollars if the second horse wins, and nothing if the third horse wins, is represented by the vector $(10, 5, 0)$.

The market maker's *policy* when interacting with trader t, $\pi(t, \cdot) : \mathbb{R}^n \mapsto \mathbb{R}$, maps these vectors to the amount the market maker would charge the agent for each bet. We denote by the zero-vector bet $\mathbf{0}$ an agent declining to make a bet with the market maker, and set $\pi(\mathbf{0}) = 0$. (This can be interpreted as the intersection of the individual rationality constraint of the traders (who would want $\pi(\mathbf{0}) \leq 0$) and of the market maker (who would want $\pi(\mathbf{0}) \geq 0$).) The market maker knows the probability that an agent will accept a bet given the prices. Because traders are anonymous, the market maker has no way to distinguish between traders and so these probabilities are the same for all traders.

In full generality, there is a chance $\delta(t)$ of the interaction terminating immediately before the t-th trader participates. Consequently, the value of being in state $(t, \mathbf{w})$ is

$$\begin{aligned} V(t, \mathbf{w}) = (1 - \delta(t)) \sum_{\mathbf{x}} & \mathbb{P}\left(\text{Trader takes bet } \mathbf{x} \text{ at price } \pi(\mathbf{x})\right) \\ & \cdot V(t+1, \mathbf{w} - \mathbf{x} + \pi(\mathbf{x})) \\ + \delta(t) V(\bar{t}, \mathbf{w}) & \end{aligned}$$

In every state $(t, \mathbf{w})$, a utility-maximizing market maker employs the optimal policy π^* defined by the Bellman equation

$$\begin{aligned} \pi^*(t, \mathbf{w}) = \arg\max_{\pi} (1 - \delta(t)) \sum_{\mathbf{x}} & \mathbb{P}\left(\text{Trader takes bet } \mathbf{x} \text{ at price } \pi(\mathbf{x})\right) \\ & \cdot V(t+1, \mathbf{w} - \mathbf{x} + \pi(\mathbf{x})) \\ + \delta(t) V(\bar{t}, \mathbf{w}) & \end{aligned}$$

with respective values V^* defined by

$$\begin{aligned} V^*(t, \mathbf{w}) = (1 - \delta(t)) \sum_{\mathbf{x}} & \mathbb{P}\left(\text{Trader takes bet } \mathbf{x} \text{ at price } \pi^*(\mathbf{x})\right) \\ & \cdot V(t+1, \mathbf{w} - \mathbf{x} + \pi^*(\mathbf{x})) \\ + \delta(t) V(\bar{t}, \mathbf{w}) & \end{aligned}$$

Solving these equations when the market maker has log utility is very challenging. We proceed to discuss how we solve for the optimal policy and values in this case.

8.2 Computation of the policy of a Kelly criterion market maker

When given a specification of the value function $V^*(t+1, \mathbf{w})$, it is simple to calculate the optimal value V^* and policy π^* of any state in the previous time step t. Thus, backward induction from the termination state is a straightforward way to solve for optimal values and policy across every time step. A complication arises from the difficulty in representing arbitrary $V^*(t+1, \mathbf{w})$. While the termination state is closed form, the previous time steps will generally not have closed form representations. In order to solve a Kelly criterion market maker's problem with backward induction, we must find a way to approximately represent the value function concisely.

8.2.1 Shape-preserving interpolation

While the value function for an arbitrary time step may have a complex, non-analytic form, we know a great deal about its *shape* from the properties it inherits from the log utility of the terminating state (Stokey et al., 1989). In particular: (1) it is increasing in wealth, (2) it is concave, and (3) it goes to minus infinity as the wealth in any state goes to zero.

Since these properties are intrinsically linked to the logarithmic utility of the Kelly criterion market maker, we choose to adopt an approximation technique that preserves these properties, *shape-preserving interpolation*. Specifically, we employ the shape-preserving interpolation developed theoretically in Constantini and Fontanella (1990). By shape-preserving, we mean that the technique retains the partial derivatives, concavity, and monotonicity of the original function, and by interpolation, we mean that the approximated function precisely matches the actual function at a set of interpolating points. While shape-preserving interpolation is well-known in the scientific computing literature (Judd, 1998), this specific technique has been featured rarely. Perhaps the most practical example is Wang and Judd (2000), who study a tax planning problem with stochastic stocks and bonds.

Because the theory of shape-preserving interpolation developed in Constantini and Fontanella (1990) is complete only for two dimensions, we focus only on settings with two events for the rest of the paper. While it does appear possible to extend the interpolation into n dimensions, it would suffer from the curse of dimensionality and take significantly longer to compute the approximate value function. The restriction to two events is not as limiting as it might first appear, because many realistic and popular settings involve wagers on binary events. An example from sports betting is whether the Red Sox or Yankees will win their upcoming match. An example from finance is credit-default swaps, where a bond either does or does not experience a default event.

In order to properly preserve the shape of the function, shape-preserving interpolation requires computing the partial derivatives with respect to the wealth in each state at the interpolating points. We compute these values by using the *envelope theorem*; since $V^*(t, \mathbf{w})$ is given by the maximizing policy π^*, we calculate the partial derivatives with respect to wealth by numerically differentiating the value function when the maximizing policy is followed (Mas-Colell et al., 1995; Wang and Judd, 2000).

We proceed to describe the interpolation procedure at a high level, first on a single rectangle and then over the whole positive orthant. Figure 8.1 shows a sample expected utility function over a single rectangle.

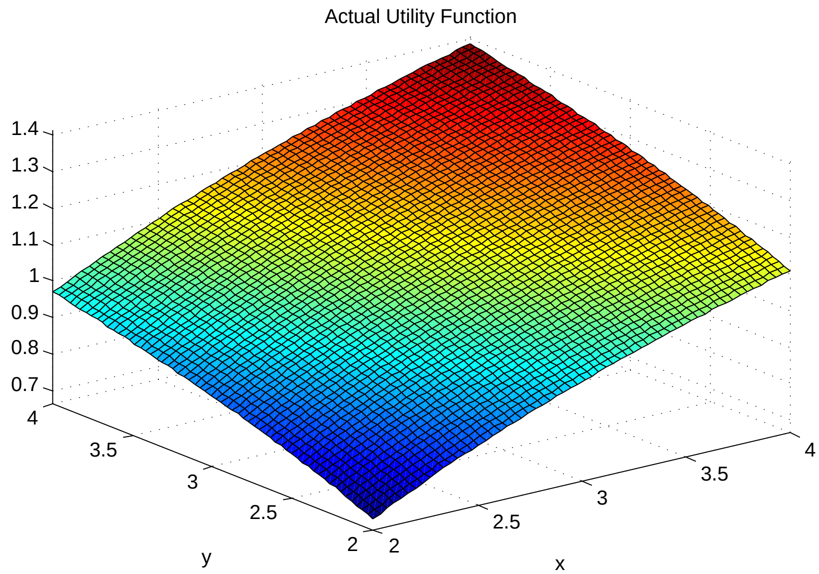


Figure 8.1: The utility function $u(x, y) = .6 \log x + .4 \log y$ on the rectangle $[2, 4]^2$.

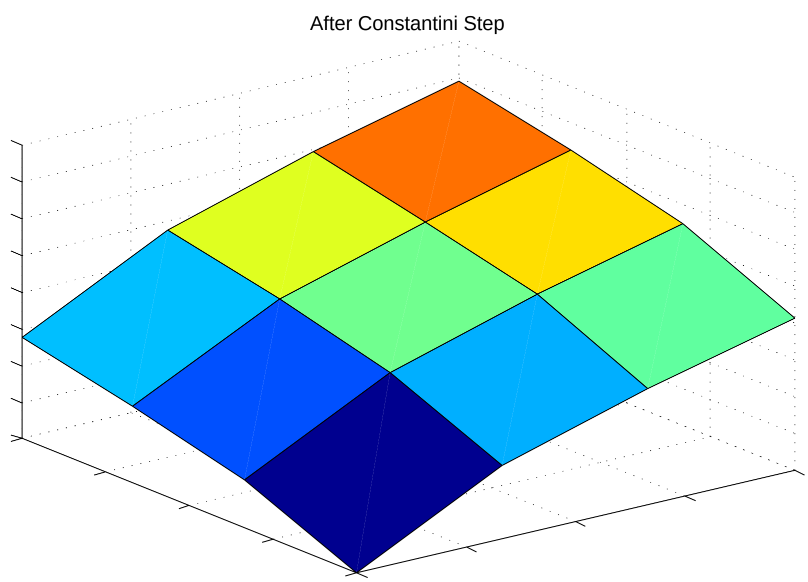


Figure 8.2: The quilt which matches the function values and partial derivatives.

The first step to creating an approximate interpolating function on this rectangle is to generate a three-by-three *quilt* (continuous, piecewise-linear approximation) of the function by matching the function values and partial derivatives at the vertices of the rectangle. Figure 8.2 shows the quilt that results from the utility function in Figure 8.1. This quilt retains the monotonicity, concavity,

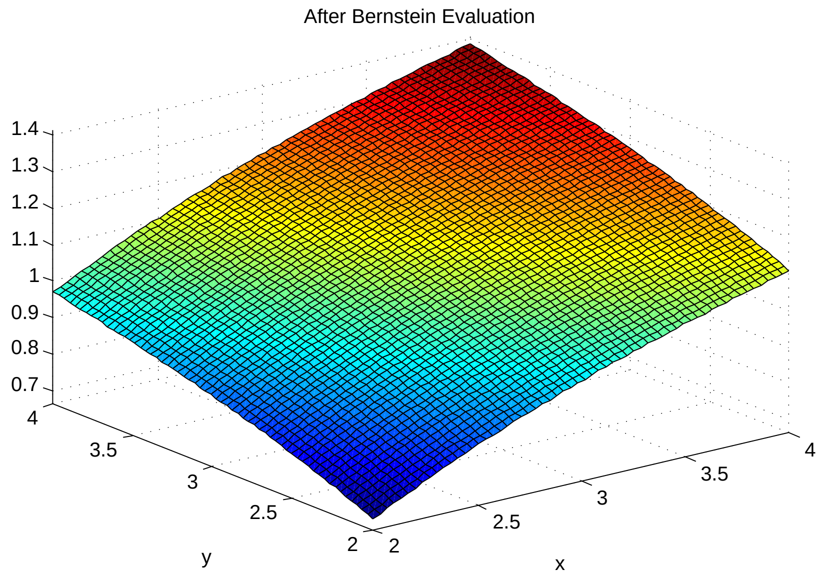


Figure 8.3: Evaluating the quilt using Bernstein bases produces a good approximation.

and partial derivatives at the vertices of the original function.

The final step is to evaluate the quilt using *bivariate Bernstein basis functions*. These are a variation-minimizing set of functions that retain the monotonicity, concavity, and partial derivatives at the vertices of the quilt. Of course, the quilt retained these properties from the original function itself, and so the interpolation is shape preserving. By variation minimizing, we mean that the bases are weighted to produce a polynomial that minimizes the sup ($\mathcal{L}_\infty$) norm error over the quilt. It is therefore accurate to think of the Bernstein bases as smoothing the piecewise linear quilt (Judd, 1998). Figure 8.3 shows the interpolated function that results from the process. Since the Bernstein evaluation step works by directly interpolating on the quilt, the function is approximated concisely: for each interpolating rectangle we only need to store the sixteen values that create the quilt.

Computing the shape-preserving interpolated function is more involved than a simple linear interpolation (table lookup). However, the benefit of these extra steps is the dramatically improved accuracy of the evaluated function or, put another way, a substantial decrease in the degree of grid fineness required to compute the value function to the same level of accuracy. Table 8.1 compares the accuracy of the shape-preserving interpolation versus a simple linear interpolation at an arbitrary collection of wealth vectors for the representative utility function used in Figure 8.1.

The shape-preserving interpolation is between 72 and 895 times more accurate than a linear grid at the example points. Perhaps unsurprisingly, we found that the inverse of this relation also appeared to hold—to achieve the same level of accuracy as shape-preserving interpolation, the grid used in linear interpolation would need to be roughly one thousand times finer. We estimate that the

Wealth vector	Shape-preserving error	Linear error	Ratio
$(2, 2)$	$.0020$	$.14$	72
$(5, 1.1)$	$.0026$	$.36$	135
$(20, 25)$	1.2×10^{-6}	$.0011$	895
$(50, 10)$	1.0×10^{-5}	$.0021$	210

Table 8.1: Relative errors for shape-preserving interpolation versus linear interpolation on identical rectangles. At each wealth vector, the interpolating rectangle is $(w_1 \pm 1, w_2 \pm 1)$, i.e., a square with side length 2 centered at the wealth vector.

running time of our experiments on a commodity PC using linear interpolation would take about a week; in contrast, solving the dynamic program took about ten minutes using shape-preserving interpolation.

8.2.2 Extending the technique

We have described how shape-preserving interpolation works on a single rectangle over which the function to be approximated is finite. It is straightforward to extend this technique from a single rectangle to a finite grid of rectangles over which the function to be approximated is finite. (In this case, care must to be taken to ensure that the function approximation is continuous at the boundaries of the individual interpolating rectangles, but this can be accommodated without too much additional complexity, see Constantini and Fontanella (1990) for details.)

However, the value function we are approximating is not just a finite function over a finite grid: it fails this in two separate ways. First, since $\lim_{x \downarrow 0} \log x = -\infty$, we have that at the lower boundary of the positive orthant (i.e., values close to zero along either dimension) the value function goes to $-\infty$. Second, the value function has no finite upper bound on its input—it is defined over the entire positive orthant. Consequently, we must extend the interpolation technique from the literature to accommodate the specific properties of a Kelly criterion market maker. Our solution is to have a large finite grid of interpolating rectangles on which we can apply the standard shape-preserving technique, and then to employ custom extensions to approximate below the lower boundary and above the upper boundary of the grid.

Beyond the lower boundary of the grid

We interpolate beyond the lower boundary of the grid as if the value were given by setting the value of a state equal to its termination value plus a constant that ensures continuity at the boundary of the grid. Formally, to approximate the value of state $\mathbf{w}$, with nearest point on the interpolating grid $\mathbf{w_g}$, we set

$$V(t, \mathbf{w}) \approx V(\bar{t}, \mathbf{w}) + (V(t, \mathbf{w_g}) - V(\bar{t}, \mathbf{w_g}))$$

(Observe that as $\mathbf{w} \to \mathbf{w_g}$, $V(t, \mathbf{w}) \to V(t, \mathbf{w_g})$). This approximation ensures the monotonicity of the value function and that it goes to negative infinity as the wealth of either state goes to zero, but, it is only an exact approximation for the termination function itself. To ensure that this extension does not change the overall value function substantially, in our experiments we start the interpolating grid at a small value, so the additional interpolation is only relevant over a small fraction of the state space. In our exploratory data analysis, we experimented with different lower bounds for the interpolating grid and found that different small values did not noticeably affect calculated optimal policies. We attribute this to states at the lower boundary of the grid having such low utility that they will be avoided, and are therefore largely irrelevant to the optimization problem as a whole.

Beyond the upper boundary of the grid

Consider the market maker's pricing problem at the upper boundary of the grid at time t. If the size of the trader's bet is bounded (say, to be no larger than c), then the market maker can approximately compute the optimal pricing policy by using an interpolating grid at time step $t+1$ whose upper boundary is larger than the grid at time t by at least c. Using this insight, we eliminate the need to calculate a value beyond the upper boundary of the grid by increasing the upper boundary of the grid as time proceeds. (In fact, recalling that we solve the dynamic program through backward induction, from an algorithmic perspective we are actually reducing the upper boundary of the grid as we solve backwards through time.) In contrast to our extension to compute values below the lower boundary of the interpolating grid that we discussed above, this extension uses the same mechanics as the rest of the shape-preserving interpolation process and so suffers from no additional loss of accuracy.

8.2.3 Alternative approaches

As an alternative to the gridded approach here, we also considered but rejected a global shape-preserving approximation technique along the lines of De Farias and Van Roy (2003). This would involve selecting basis functions ϕ_i that are each monotonic and concave, and representing the value function in each time step as a conical combination of these functions:

$$V^*(t, \mathbf{w}) \approx \sum_i \gamma_i^t \phi_i(\mathbf{w}), \quad \gamma_i^t \geq 0.$$

Such a representation retains the monotonicity and concavity properties of the value function and is concise. We rejected this approach for two reasons. First, the heuristic selection of the basis provides little guidance. Which set of functions is a good choice, and why? Observe that many standard basis function selections, such as radial basis functions, will not in general preserve the monotonicity or concavity of the value function and so could lead to nonsensical policies.

The second reason we chose to reject this technique is the difficult optimization to select the weights γ^t. In particular, a standard linear regression that maximizes the deviation from sum of squares at a set of relevant nodes can create aberrant behavior and an approximation that deviates significantly from the actual value function (Gordon, 1995; Guestrin et al., 2001; Stachurski, 2008). The correct optimization to use to determine the weights is to minimize the sup norm (that is, $\mathcal{L}_\infty$, rather than $\mathcal{L}_2$), which is a significantly more challenging problem to solve numerically (Judd, 1998).

8.3 Experiments

With only two possible events, it is possible to characterize bets in terms of a single event. In particular, setting p to be the probability that the first event occurs implies $1 - p$ is the probability that the second event occurs. Applying this logic to the market maker's policy, we can without loss of generality have the market maker buy and sell contracts on the first event only, because buying (selling) a contract on the first event implicitly yields the sale (purchase) of a contract on the second.

The *ask* is the price at which the market maker will sell a contract, and the *bid* is the price at which the market maker will buy a contract. For non-degenerate settings, ask prices will always be higher than bid prices. In this section, we describe the optimal ask and bid prices for two different settings.

8.3.1 Parameterization

Following Das (2008), in our experiments, traders have a belief drawn from a Gaussian with a mean belief of $p = 0.5$ and standard deviation 0.05. The traders are zero-intelligence agents; a trader visits the market maker exactly once and behaves myopically. They purchase a unit contract if they see an ask price lower than their belief, sell a unit contract if they see a bid price higher than their belief, and do not transact with the market maker otherwise.

In our experiments, we set 50 trading periods (that is, $\delta(t = 51) = 1, \delta(t < 50) = 0$), although we found our results hold qualitatively for other distributions of traders. Recalling from Section 8.2.2 that the upper boundary of the interpolating grid increases in each trading period, we set the interpolating grid for trader t to $[1, 1.5, 2, 3, \ldots, 250, 250 + t]^2$.

In our experiments with Kelly criterion market makers, we consider only relatively small levels of wealth (alternatively, large bets relative to the amount of wealth). This is because for bets with large levels of wealth, a market maker maximizing expected log utility can be well-approximated by a risk-neutral, linear utility agent. To see why, consider the Taylor expansion of log utility at wealth x:

$$\log(x+\epsilon) = \log(x) + \frac{\epsilon}{x} - \frac{\epsilon^2}{x^2} + \Theta\left(\epsilon^3\right)$$

If x is large enough that $x^2 \gg x$, then $1/x^2 \ll 1/x$. Consequently, at large wealths, the impact of small bets on the utility function can be well-approximated by the linear function $\log(x) + \left(\frac{1}{x}\right)\epsilon$, with negligible higher-order effects.

We now turn our attention to how to calculate the optimal policy for a risk-neutral market maker, and the qualitative properties of that policy.

8.3.2 Optimal risk-neutral policy

For this setting, a risk-neutral market maker's optimization problem is significantly simpler than the general case. Recall that in the two-event case, the market maker's knowledge of the future, the vector $\hat{\mathbf{p}}$, can be represented by a single scalar $\hat{p}$ (e.g., "Team A has a 50 percent chance of winning the game"). Then the termination state $V(\bar{t}, \mathbf{w})$ is

$$V(\bar{t}, \mathbf{w}) \equiv \hat{p}w_1 + (1-\hat{p})w_2$$

Let the agents have beliefs on the first event distributed according to the cumulative density function F with probability density function f. In the penultimate step $\bar{t}-1$, a risk-neutral market maker sets their bid and ask price to maximize their utility in the termination state, conditioning on three cases: the bid being taken, the ask being taken, and neither offer being taken. Formally,

$$\begin{aligned} V^*(\bar{t}-1, \mathbf{w}) &= \max_{b,a} F(b)V(\bar{t}, (w_1 - b + 1, w_2 - b)) \\ &\quad + (1 - F(a))V(\bar{t}, (w_1 + a - 1, w_2 + a)) \\ &\quad + (F(a) - F(b))V(\bar{t}, \mathbf{w}) \end{aligned}$$

and since $V(\bar{t}, \mathbf{w}) = \hat{p}w_1 + (1-\hat{p})w_2$ the right-hand side optimization simplifies to

$$\begin{aligned} \max_{b,a} \quad & F(b)(\hat{p}(w_1 - b + 1) + (1-\hat{p})(w_2 - b)) \\ & + (1 - F(a))(\hat{p}(w_1 + a - 1) + (1-\hat{p})(w_2 + a)) \\ & + (F(a) - F(b))(\hat{p}w_1 + (1-\hat{p})w_2) \end{aligned}$$

which further simplifies to

$$\max_{b,a} V(\bar{t}, \mathbf{w}) + F(b)(\hat{p} - b) + (1 - F(a))(a - \hat{p}) \tag{8.1}$$

which implies

$$V^*(\bar{t} - 1, \mathbf{w}) + C = V(\bar{t}, \mathbf{w})$$

where C is a constant that does not depend on t or $\mathbf{w}$. Consequently, by inductive argument working back from the terminal state the optimal policy for a risk-neutral market maker does not depend on t or $\mathbf{w}$. Equation 8.1 also makes it easy to see that the optimal arguments (b^*, a^*) have $b^* \leq \hat{p} \leq a^*$, because if not changing to a policy satisfying that inequality would yield a higher value. Thus, a globally optimal risk-neutral market maker is always myopically rational. (This argument also applies to the general setting discussed in Section 8.1.2 with more than two bets and events. In that case, by similar reasoning, the result is that a risk-neutral market maker will always price a bet $\mathbf{x}$ such that $\pi(\mathbf{x}) \geq \hat{\mathbf{p}} \cdot \mathbf{x}$.)

We have shown that the optimal policy of a risk-neutral market maker is constant and invariant to time and wealth. To actually compute the optimal b^* and a^*, we can take the first-order condition of the optimization in Equation 8.1 to get

$$\begin{aligned} F(b^*)(\hat{p} - 1) + f(b^*)(\hat{p} - b^*) &= 0 \\ (1 - F(a^*))(1 - \hat{p}) - f(a^*)(a^* - \hat{p}) &= 0 \end{aligned}$$

If $\hat{p} \in (0, 1)$ and $f(x) > 0$ for all $x \in (0, 1)$, the existence and uniqueness of optimal $b^* < \hat{p} < a^*$ are guaranteed. When F and f are well-behaved smooth functions (as is the case for our experiments where F is a normal distribution), the optimal values can be solved quickly by numerical root-finding techniques.

8.3.3 Optimal $\log$-utility policy

Following the procedure outlined in Section 8.2, we computed the optimal value and policy functions for several different parameterizations of wealths and beliefs for both Kelly and risk-neutral market makers.

We begin by considering the case where the market maker's private belief aligns with the beliefs of the traders. Figure 8.4 shows the optimal bid and ask prices over the series of traders when the market maker has wealth $(100, 100)$ (thickest line), $(50, 50)$ (medium line), and $(25, 25)$ (thinnest line). That is, the plot shows $\pi(t, (w, w))$ for $t \in \{1, \ldots, 50\}$ and $w \in (25, 50, 100)$.

Here, prices throughout the interaction are very close to the myopic optimization for the last trader, and the prices are very similar for all of the sampled wealths. In this scenario, the prices are also essentially equivalent to the optimal policy of a risk-neutral market maker.

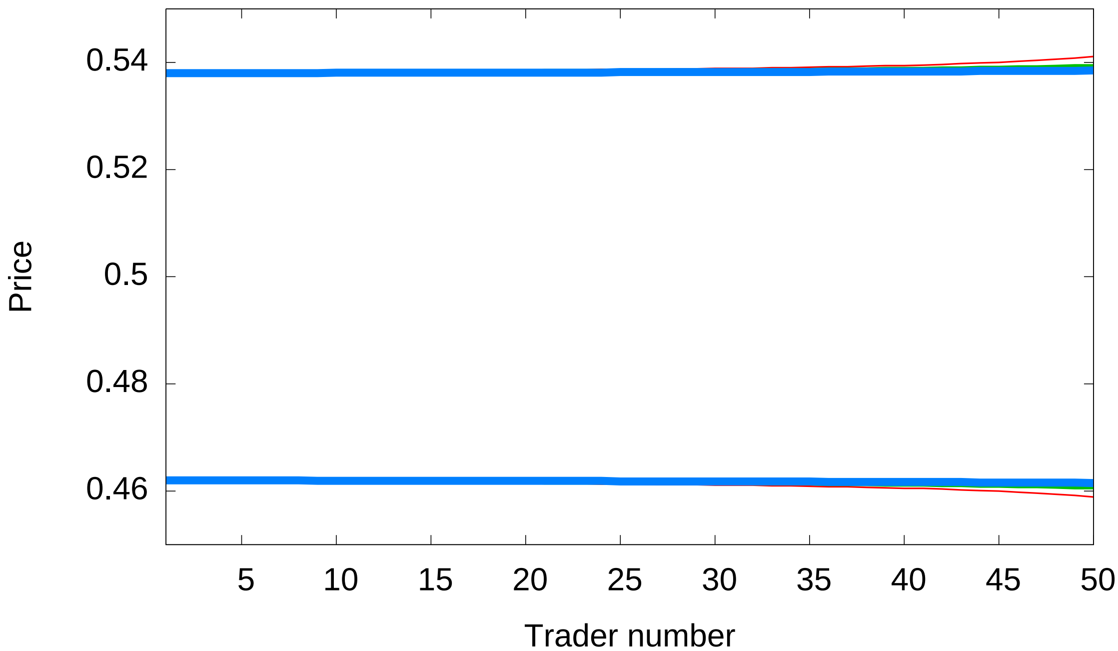


Figure 8.4: When the market maker's private beliefs align with those of the traders, the optimal ask prices (top lines) and bid prices (bottom lines) do not change significantly over the course of the interaction period.

In contrast, Figure 8.5 shows the optimal policies when the market maker's belief is $p = 0.6$ (shown by the cross-hatched line). This value is two standard deviations higher than the mean of the traders' beliefs. The policies are calculated at the same wealths as in Figure 8.4, that is, the policy of a market maker with wealth of 25, 50, and 100 in both states at every time step.

Unlike in the previous figure, the optimal policies change over time and are wealth-dependent. In this scenario, the optimal risk-neutral policy is a bid of 0.52 and an ask of 0.62. Because with large wealth a logarithmic utility market maker making small bets can be approximated well by linear utility, we know that as wealth increases, the market maker's optimal policy throughout the trading period will converge to be the optimal linear utility policy. However, at the smaller levels of wealth in our experiments, for all except the last few traders, the asking price for a unit contract is below the market maker's belief that the event will occur. Thus, for much of the trading period, from a myopic perspective the Kelly criterion market maker offers irrational bets.[1]

On the surface, this result seems confusing and even paradoxical. To see why it is the optimal

[1]One might think that this phenomenon could be explained by the market makers accumulating wealth from spread profits, and therefore becoming absolutely less risk-averse over time. However, even market makers with considerably larger endowments than could possibly be made through spread profits still display the same qualitative behavior.

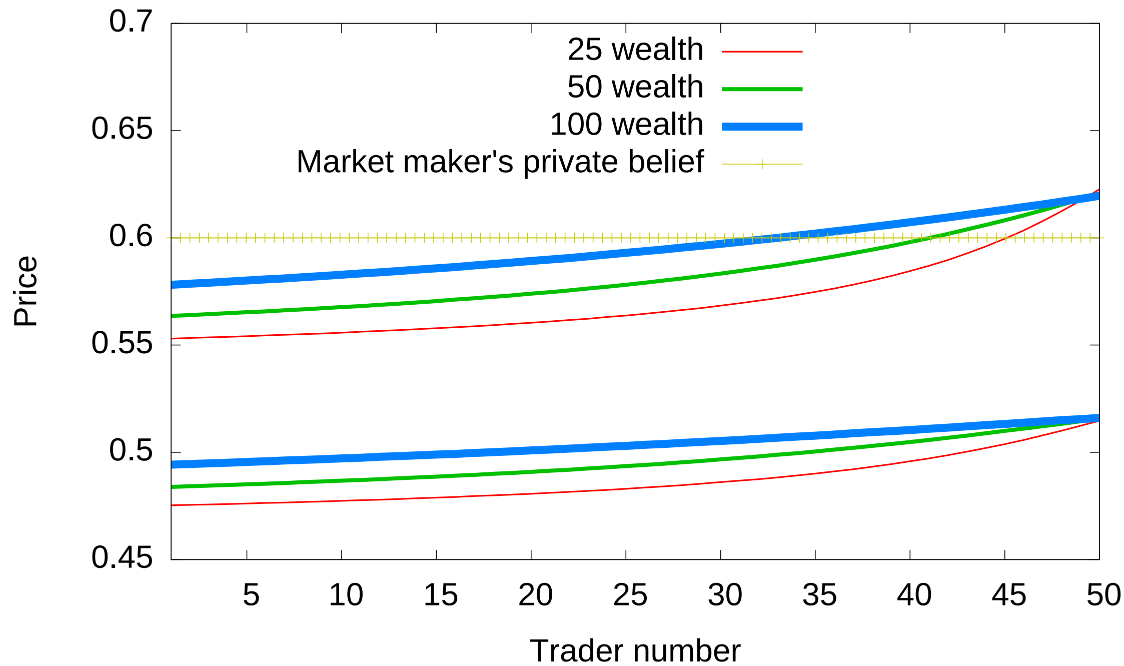


Figure 8.5: When the market maker's private beliefs do not align with those of the traders, the optimal policy is highly time and wealth dependent.

policy, consider Figure 8.6, which displays the probability of each trader taking the bets offered by a market maker with a (constant) wealth of 25 in both states. It shows how the probability of a trader selling at the bid price rises over time, while the probability of a trader buying at the ask price falls. The first trader is about twice as likely to sell at the bid price than to buy at the ask price, while the last trader is about *87 times* more likely to sell than to buy.

For early traders, the market maker's bid and ask prices are roughly centered around 0.5, just like the distribution of agent beliefs. Consequently, the market maker has a reasonable chance of matching traders' bids and asks and thus profiting off the bid/ask spread. A market maker that successfully matches the bids and asks of traders books a profit regardless of the personal beliefs of the market maker, even if those beliefs are, as in this scenario, very different from the prices in question. Of course, as fewer traders remain the setting more and more resembles a myopic optimization where, with equivalent wealths in both states, the market maker will employ a myopically rational strategy. This sophisticated policy emerges solely from the introduction of risk-aversion to the termination state, because a risk-neutral market maker in an identical setting displays none of this behavior.

Once the optimal policy is computed, we can simulate the behavior of the market maker against the pool of traders. Figure 8.7 shows the simulated prices of wealth 25 market maker in a sample

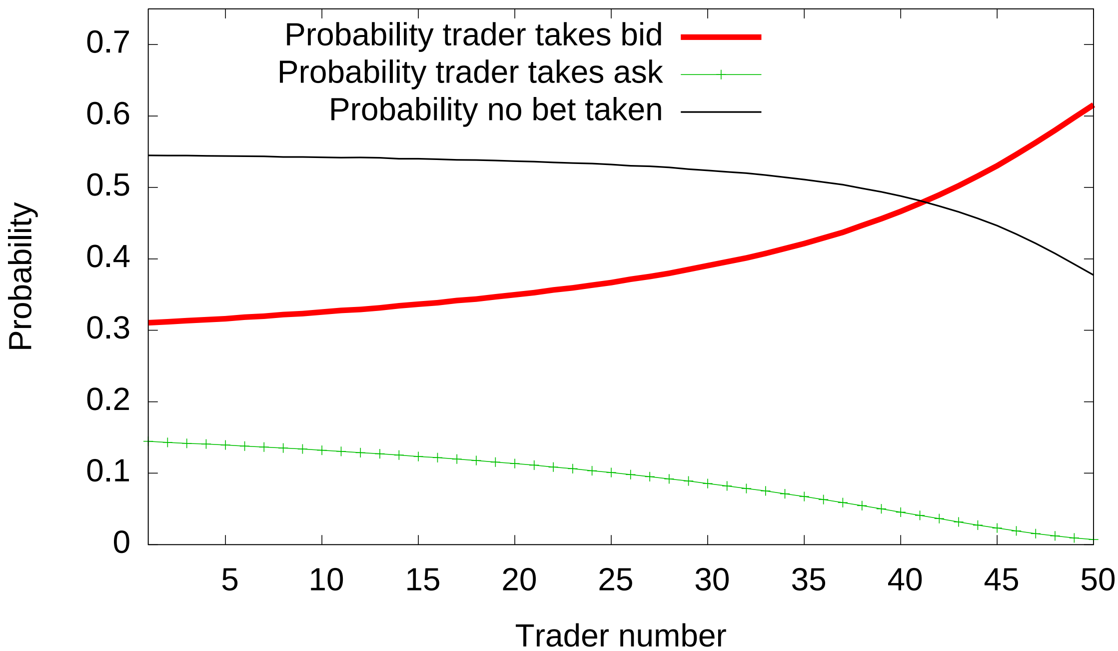


Figure 8.6: The probability of a trader taking each offered bet from a market maker with 25 wealth in each state over the entire trading period.

interaction for the case where the the market maker has belief 0.6 and agents have belief mean of 0.5 (i.e., the setting for Figures 8.5 and 8.6). The thin lines, with values marked by the left axis, show the ask (upper line) and bid (lower line) prices of the market maker. The thick black line, with values market by the right axis, shows the market maker's net inventory, i.e., the market maker's payoff if the event occurs. The values on Figure 8.7 show the prices faced by trader i but the inventory *after* the participation of the trader. (The inventory line starts at 1 in this case because the first trader took the market maker's bid.) In this simulation, the market maker's expected utility from their wealth vector increases from 3.22 before any traders participate to 3.27 after all 50 traders participate.

Recall that in this setting the market maker has a significantly higher belief that the event will occur than does the pool of agents, so it is natural for the market maker to accumulate inventory. As the market maker accumulates inventory, its prices fall. This is because a risk-averse market maker prefers to take a small sure profit (the bid/ask spread from matching orders) over a somewhat larger speculative gain (from holding inventory). Consequently, the prices from the simulation are very different than the prices in Figure 8.5, because the prices in Figure 8.5 captured the prices of a market maker with constant wealth in both states over time. If the market maker were not taking on inventory, its prices would rise, as in Figure 8.5, but because the market maker in our simulation takes on inventory, that price rise is effectively dampened. Observe that in this simulation, because

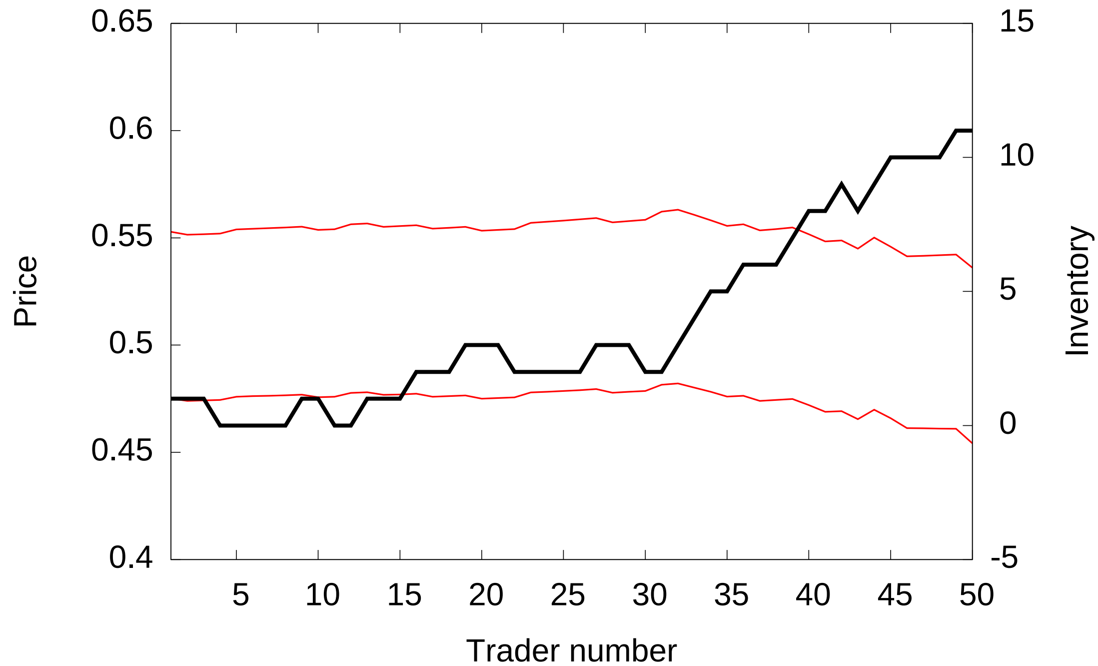


Figure 8.7: Simulating the prices (left axis; thin lines) and net inventory (right axis; thick black line) that result from the interaction of an optimal log utility market maker starting with 25 wealth in both states. In this figure, both the inventory and prices change over time.

the price rise is dampened, the market maker's asking price is always less than 0.6 and therefore is myopically irrational for the entire trading period!

8.4 Extensions

There are several extensions to consider to our framework. Our model assumed the market maker was monopolistic, so that it could maximize profit without fear of competition. One extension could be to examine a competitive setting between several risk-averse market makers.

Another extension would be to incorporate informed traders into the pool of trading agents. These agents could have correct knowledge about the future, but, more importantly, the market maker could know of and react to their existence. The presence of informed traders that influence the market maker in this way would make our setting much closer to the Bayesian market maker setting explored in Das and Magdon-Ismail (2009), but would be even more complex because of the risk aversion of the market maker.

While our framework applies to any number of events and bets, our computational experiments focused on the binary case. An extension would be to develop algorithms to solve for optimal policy with multiple events. The Constantini shape-preserving technique used in this paper could presumably be applied to more than two events, although it will suffer from the curse of dimensionality. Perhaps an alternate approach to approximating the value function could be used in this case, such as a spline of radial basis functions, although this would be unlikely to preserve the monotonicity and concavity of the value function and so could lead to poor or unrealistic policies.

Bibliography

J. Abernethy, Y. Chen, and J. W. Vaughan. An optimization-based framework for automated market-making. In *ACM Conference on Electronic Commerce (EC)*, 2011.

M. Abramowicz. *Predictocracy: market mechanisms for public and private decision making*. Yale University Press, 2008.

S. Agarwal, Z. Wang, and Y. Ye. Parimutuel betting on permutations. In *International Workshop on Internet and Network Economics (WINE)*, 2008.

S. Agrawal, E. Delage, M. Peters, Z. Wang, and Y. Ye. A unified framework for dynamic parimutuel information market design. In *ACM Conference on Electronic Commerce (EC)*, pages 255–264, 2009.

Y. Ait-Sahalia and A. Lo. Nonparametric estimation of state-price densities implicit in financial asset prices. *The Journal of Finance*, 53(2):499–547, 1998.

M. Ali. Probability and Utility Estimates for Racetrack Bettors. *The Journal of Political Economy*, 85(4):803, 1977.

K. Arrow. *Social choice and individual values*. New Haven: Cowles Foundation, 2nd edition, 1963. 1st edition 1951.

P. Artzner, F. Delbaen, J. Eber, and D. Heath. Coherent measures of risk. *Mathematical finance*, 9 (3):203–228, 1999.

P. Barrieu and N. E. Karoui. Pricing, hedging, and designing derivatives with risk measures. In R. Carmona, editor, *Indifference Pricing*, chapter 3, pages 77–146. Princeton University Press, 2009.

A. Ben-Tal and M. Teboulle. An old-new concept of convex risk measures: The optimized certainty equivalent. *Mathematical Finance*, 17(3):449–476, 2007.

J. Berg and T. Rietz. Prediction markets as decision support systems. *Information Systems Frontiers*, 5(1):79–93, 2003.

J. Berg, R. Forsythe, F. Nelson, and T. Rietz. Results from a Dozen Years of Election Futures Markets Research. *Handbook of Experimental Economics Results*, pages 486–515, 2001.

F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of political economy*, 81(3), 1973.

P. Bossaerts, L. Fine, and J. Ledyard. Inducing liquidity in thin financial markets through combined-value trading mechanisms. *European Economic Review*, 46(9):1671–1695, 2002.

S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

A. Brahma, S. Das, and M. Magdon-Ismail. Comparing Prediction Market Structures, With an Application to Market Making. Technical report, Dept. of Computer Science, Rensselaer Polytechnic Institute, 2010.

J. Caballé and A. Pomansky. Mixed risk aversion. *Journal of Economic Theory*, 71(2):485–513, 1996.

R. Carmona. *Indifference Pricing: Theory and Applications*. Princeton University Press, 2009.

P. Carr, September 2010. Personal communication.

P. Carr, H. Geman, and D. Madan. Pricing and hedging in incomplete markets. *Journal of Financial Economics*, 62(1):131–167, 2001.

M. Chakraborty, S. Das, A. Lavoie, M. Magdon-Ismail, and Y. Naamad. Instructor rating markets. In *Conference on Auctions, Market Mechanisms and Their Applications (AMMA)*, 2011.

J. Chang and L. Shanker. Hedging effectiveness of currency options and currency futures. *Journal of Futures Markets*, 6(2):289–305, 1986. ISSN 1096-9934.

Y. Chen and D. M. Pennock. A utility framework for bounded-loss market makers. In *Proceedings of the 23rd Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 49–56, 2007.

Y. Chen and D. M. Pennock. Designing markets for prediction. *AI Magazine*, 31(4), 2010.

Y. Chen and J. W. Vaughan. A new understanding of prediction markets via no-regret learning. In *ACM Conference on Electronic Commerce (EC)*, pages 189–198, 2010.

Y. Chen, D. Reeves, D. Pennock, R. Hanson, L. Fortnow, and R. Gonen. Bluffing and Strategic Reticence in Prediction Markets. In *WINE*, 2007.

Y. Chen, L. Fortnow, N. Lambert, D. M. Pennock, and J. Wortman. Complexity of combinatorial market makers. In *ACM Conference on Electronic Commerce (EC)*, pages 190–199, 2008.

Y. Chen, S. Dimitrov, R. Sami, D. M. Reeves, D. M. Pennock, R. D. Hanson, L. Fortnow, and R. Gonen. Gaming prediction markets: Equilibrium strategies with a market maker. *Algorithmica*, 58:930–969, December 2010.

P. Constantini and F. Fontanella. Shape-preserving bivariate interpolation. *SIAM J. Numer. Anal.*, 27:488–506, April 1990. ISSN 0036-1429.

T. Cover and J. Thomas. *Elements of Information Theory*. John Wiley & Sons, 2006.

S. Das. The effects of market-making on price dynamics. In *International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 887–894, 2008.

S. Das and M. Magdon-Ismail. Adapting to a market shock: Optimal sequential market-making. In *Advances in Neural Information Processing Systems 21 (NIPS)*, pages 361–368, 2009.

A. Davison and D. Hinkley. *Bootstrap Methods and Their Application*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 8th edition, 2006.

D. De Farias and B. Van Roy. The linear programming approach to approximate dynamic programming. *Operations Research*, pages 850–865, 2003.

F. Delbaen, P. Grandits, T. Rheinländer, D. Samperi, M. Schweizer, and C. Stricker. Exponential hedging and entropic penalties. *Mathematical Finance*, 12(2):99–123, 2002.

H. Demsetz. The cost of transacting. *The Quarterly Journal of Economics*, 82(1):33–53, 1968.

S. Dimitrov and R. Sami. Non-myopic strategies in prediction markets. In *ACM Conference on Electronic Commerce (EC)*, pages 200–209, 2008.

B. Dumas, J. Fleming, and R. Whaley. Implied volatility functions: Empirical tests. *The Journal of Finance*, 53(6):2059–2106, 1998.

R. J. Elliott and J. van der Hoek. Duality methods. In R. Carmona, editor, *Indifference Pricing*, chapter 9, pages 321–386. Princeton University Press, 2009.

R. Engle and V. Ng. Measuring and testing the impact of news on volatility. *Journal of finance*, pages 1749–1778, 1993.

E. Even-Dar, S. M. Kakade, M. Kearns, and Y. Mansour. (In)Stability properties of limit order dynamics. In *ACM Conference on Electronic Commerce (EC)*, pages 120–129, 2006.

J. Feigenbaum, L. Fortnow, D. Pennock, and R. Sami. Computation in a distributed information market. In *ACM Conference on Electronic Commerce (EC)*, pages 156–165, 2003.

H. Föllmer and A. Schied. Convex measures of risk and trading constraints. *Finance and Stochastics*, 6(4):429–447, 2002a.

H. Föllmer and A. Schied. *Stochastic Finance (Studies in Mathematics 27)*. De Gruyter, 2002b.

R. Forsythe, F. Nelson, G. Neumann, and J. Wright. Anatomy of an experimental political stock market. *The American Economic Review*, 82(5):1142–1161, 1992.

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

L. Fortnow, J. Kilian, D. M. Pennock, and M. P. Wellman. Betting boolean-style: a framework for trading in securities based on logical formulas. In *ACM Conference on Electronic Commerce (EC)*, pages 144–155, 2003.

X. Gao and Y. Chen. An axiomatic characterization of continuous-outcome market makers. In *International Workshop On Internet And Network Economics (WINE)*, 2010.

X. Gao, Y. Chen, and D. M. Pennock. Betting on the real line. In *International Workshop On Internet And Network Economics (WINE)*, pages 553–560, 2009.

J. Geweke. A note on some limitations of CRRA utility. *Economics Letters*, 71(3):341–345, 2001.

L. R. Glosten and P. R. Milgrom. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of financial economics*, 14(1):71–100, 1985.

T. Gneiting and A. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

D. Gode and S. Sunder. Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality. *Journal of Political Economy*, pages 119–137, 1993.

S. Goel, D. Pennock, D. Reeves, and C. Yu. Yoopick: a combinatorial sports prediction market. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, pages 1880–1881, 2008.

S. Goel, D. Reeves, D. Watts, and D. Pennock. Prediction without markets. In *ACM Conference on Electronic Commerce (EC)*, pages 357–366, 2010.

G. J. Gordon. Stable function approximation in dynamic programming. In *International Conference on Machine Learning (ICML)*, pages 261–268, 1995.

A. Graefe and J. S. Armstrong. Can you beat the market? Accuracy of Individual and Group Post-Prediction Market Judgments. In *Third workshop on prediction markets, part of the ACM conference on Electronic Commerce (EC)*, 2008.

C. Guestrin, D. Koller, and R. Parr. Max-norm projections for factored MDPs. In *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 673–682, 2001.

R. Hanson. Combinatorial information market design. *Information Systems Frontiers*, 5(1):107–119, 2003.

R. Hanson. Logarithmic market scoring rules for modular combinatorial information aggregation. *Journal of Prediction Markets*, 1(1):1–15, 2007.

P. Healy, S. Linardi, J. Lowery, and J. Ledyard. Prediction markets: Alternative mechanisms for complex environments with few traders. *Management Science*, 56(11):1977–1996, 2010.

V. Henderson. Valuation of claims on nontraded assets using utility maximization. *Mathematical Finance*, 12(4):351–373, 2002.

V. Henderson and D. Hobson. Utility indifference pricing. In R. Carmona, editor, *Indifference Pricing*, chapter 2, pages 44–73. Princeton University Press, 2009.

T. Ho and H. Stoll. Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial economics*, 9(1):47–73, 1981.

J. Jackwerth. Recovering risk aversion from option prices and realized returns. *Review of Financial Studies*, 13(2):433, 2000.

J. Jackwerth and M. Rubinstein. Recovering probability distributions from option prices. *Journal of Finance*, 51(5):1611–1631, 1996.

K. Judd. *Numerical methods in economics*. The MIT Press, 1998.

J. Kelly Jr. A new interpretation of information rate. *IRE Transactions on Information Theory*, 2(3): 185–189, 1956.

S. Kou. A Jump-Diffusion Model for Option Pricing. *Management Science*, 48(8):1086–1101, 2002.

A. S. Kyle. Continuous Auctions and Insider Trading. *Econometrica*, 53(6):1315–1335, 1985.

N. Lambert, J. Langford, J. Wortman, Y. Chen, D. Reeves, Y. Shoham, and D. Pennock. Self-Financed Wagering Mechanisms for Forecasting. In *ACM Conference on Electronic Commerce (EC)*, pages 170–179, 2008a.

N. S. Lambert, D. M. Pennock, and Y. Shoham. Eliciting properties of probability distributions. In *Proceedings of the 9th ACM conference on Electronic Commerce (EC)*, 2008b.

J. Ledyard, R. Hanson, and T. Ishikida. An experimental test of combinatorial information markets. *Journal of Economic Behavior & Organization*, 69(2):182–189, 2009.

D. MacKenzie. *An engine, not a camera: How financial models shape markets*. The MIT Press, 2006.

D. Madan, August 2010. Personal communication.

P. Malliavin. *Integration and probability*, volume 157. Springer, 1995.

M. Mania and M. Schweizer. Dynamic exponential utility indifference valuation. *The Annals of Applied Probability*, 15(3):2113–2143, 2005.

H. Markowitz. Portfolio Selection. *The Journal of Finance*, 7(1):77–91, 1952.

A. Mas-Colell, M. Whinston, and J. R. Green. *Microeconomic Theory*. Oxford University Press, New York, NY, 1995.

R. Merton. Theory of rational option pricing. *The Bell Journal of Economics and Management Science*, 4(1):141–183, 1973.

R. C. Merton. Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, 3(4):373–413, 1971.

M. Mitzenmacher. A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, 1(2):226–251, 2004.

D. Moore and P. Healy. The trouble with overconfidence. *Psychological review*, 115(2):502–517, 2008.

M. Musiela and T. Zariphopoulou. A valuation algorithm for indifference prices in incomplete markets. *Finance and Stochastics*, 8(3):399–414, 2004.

J. Nocera. Risk mismanagement. New York Times Magazine, January 2 2009.

M. O'Hara. *Market Microstructure Theory*. Blackwell Publishing, 1995.

K. Oliven and T. Rietz. Suckers are born but markets are made: Individual rationality, arbitrage, and market efficiency on an electronic futures market. *Management Science*, 50(3):336–351, 2004.

A. Othman. Zero-intelligence agents in prediction markets. In *International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 879–886, 2008.

A. Othman and T. Sandholm. Automated market-making in the large: the Gates Hillman prediction market. In *ACM Conference on Electronic Commerce (EC)*, pages 367–376, 2010a.

A. Othman and T. Sandholm. When Do Markets with Simple Agents Fail? In *International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 865–872, Toronto, Canada, 2010b.

A. Othman and T. Sandholm. Automated market makers that enable new settings: Extending constant-utility cost functions. In *Conference on Auctions, Market Mechanisms and Their Applications (AMMA)*, 2011a.

A. Othman and T. Sandholm. Liquidity-sensitive market makers via homogeneous risk measures.

In *International Workshop On Internet And Network Economics (WINE)*, 2011b.

A. Othman, D. M. Pennock, D. M. Reeves, and T. Sandholm. A practical liquidity-sensitive automated market maker. In *ACM Conference on Electronic Commerce (EC)*, pages 377–386, 2010.

D. Pennock and R. Sami. Computational Aspects of Prediction Markets. In *Algorithmic Game Theory*, chapter 26, pages 651–674. Cambridge University Press, 2007.

D. M. Pennock. A dynamic pari-mutuel market for hedging, wagering, and information aggregation. In *ACM Conference on Electronic Commerce (EC)*, pages 170–179, 2004.

M. Peters, A. M.-C. So, and Y. Ye. Pari-mutuel markets: Mechanisms and performance. In *International Workshop On Internet And Network Economics (WINE)*, pages 82–95, 2007.

W. Poundstone. *Fortune's Formula: The Untold Story of the Scientific Betting System That Beat the Casinos and Wall Street*. Hill and Wang, 2006.

K. Rock. The specialist's order book and price anomalies. *Review of Financial Studies*, 1996.

R. T. Rockafellar. Level sets and continuity of conjugate convex functions. *Transactions of the American Mathematical Society*, 123(1):46–63, 1966.

R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall, 2nd edition, 2003.

L. Savage. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, pages 783–801, 1971.

G. Schwert. Stock volatility and the crash of '87. *Review of Financial Studies*, 3(1):77–102, 1990.

S. Shalev-Shwartz and Y. Singer. A primal-dual perspective of online learning algorithms. *Machine Learning*, 69:115–142, December 2007.

J. Stachurski. Continuous state dynamic programming via nonexpansive approximation. *Computational Economics*, 31(2):141–160, 2008.

N. Stokey, R. Lucas, and E. Prescott. *Recursive methods in economic dynamics*. Harvard University Press, 1989.

H. Stoll. The supply of dealer services in securities markets. *The Journal of Finance*, 33(4):1133–1151, 1978.

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

N. Taleb. *The black swan: The impact of the highly improbable*. Random House Inc, 2007.

G. Tziralis and I. Tatsiopoulos. Prediction markets: An extended literature review. *The journal of prediction markets*, 1(1):75–91, 2007.

United Nations. Human Development Report 2007/08, 2008.

U.S. Commodity and Futures Trading Commission and U.S. Securities & Exchange Commission. Findings regarding the market events of May 6, 2010. Technical report, Joint Advisory Committee on Emerging Regulatory Issues, September 2010.

H. R. Varian. The arbitrage principle in financial economics. *J. Economic Perspectives*, 1(2):55–72, 1987.

H. R. Varian. *Microeconomic analysis*. New York: W. W. Norton, 1992.

S.-P. Wang and K. L. Judd. Solving a savings allocation problem by numerical dynamic programming with shape-preserving interpolation. *Comput. Oper. Res.*, 27:399–408, April 2000. ISSN 0305-0548.

O. M. Williams. Regulatory oversight and recent initiatives to address risk posed by credit default swaps. Testimony Before the Subcommittee on Capital Markets, Insurance, and Government Sponsored Enterprises, March 2009.

R. Winkler. Scoring rules and the evaluation of probability assessors. *Journal of the American Statistical Association*, pages 1073–1078, 1969.

R. Winkler. Evaluating probabilities: Asymmetric scoring rules. *Management Science*, pages 1395–1405, 1994.

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

J. Wolfers and E. Zitzewitz. Five Open Questions About Prediction Markets. Technical Report 1975, Institute for the Study of Labor (IZA), Feb. 2006.

BIBLIOGRAPHY