Proposal: MS Thesis

Anshul Goel

January 18, 2020

1 Motivation

Betting as such has been considered an evil practise but it is quite interesting to realise its implications in forecasting. The other methods of forecasting relies on application of statistics tools on the data available and is currently more and more following up the same pattern, with the advent of Machine Learning and AI. But, in questions related to our daily lives being as a normal, consumer, a sports lover, a business men, a politician, a stock traders or just a regular citizen, we do realise the vast intellectual capabilities of humans, in terms of interpretation of situations, the possible scenarios, the diverse views and possibilities which is quite difficult and complex to figure out solely with the statistical tools. For example, say a news between comes out from Pakistan Defence Minister related to Indian Army activities, it is the human intellect that cater to the various facts like timings and even the actual words used and their implications in terms of mood and strength of statement, which is quite difficult to capture by statistical teams and would also require huge dataset to just extrapolate the situation from past, which may not always be a good method. But, it would be better to have a group of people, each with some information, own theory formulated from his/her own information (public and non-public), intellect along with own intuition and a mechanism to aggregate all this information to come to a solution, an answer.

This is what the prediction markets promise to deliver. Many studies have found that these markets have provided at par or better forecasts than the expert forecasters be it sports, politics or film industry. Prediction markets collect their power from the idea of wisdom of crowds - that a group of people can forecast better in given conditions than a particular expert. Many corporations like Google & Ford have realised their strength and have created internal prediction markets as a method of information aggregation in the organisation. Other examples include their power in a whole lot of areas, out of which these examples are worth noting.

1. Replication markets by DARPA to predict the replicability of studies in Social Science LINK
2. A decentralised prediction market on almost all walks. be it Politics, Sports, culture. LINK

3. Predicting the timing, nature and impact of scientific and technological advances and breakthroughs LINK

4. Aggregation of people's information and using it in financial markets LINK

And the list continues

The broad question I want to highlight in my work is how and when these markets work, what are the key conditions required for these markets, and how to design a market on own. In the next section, I would highlight the key questions my work would sought to answer.

2 The Big Questions

I have tried to come up with some broad specific questions, which will form the basis of my work and guide me throughout

1. What are the type of problems that these markets can work for?
   It is not sure if the prediction market is applicable for any problem or is there a class of problem for which they work.

2. What are the conditions required for the market to work?
   As we have observed in case of stock markets, they act as a gauge of many economic indicators but there has also been incidents of bubbles and depressions worldwide highlighting certain conditions which are required for prediction markets to work effectively.

3. What are the essential things to focus for designing these markets, or better say, what are the various methods to create this market, which further comprise of following sub questions

   (a) What should be the incentive provided so that people say truth and don't try to manipulate? Should it be monetary or non-monetary?

   (b) How to create 'thick' markets?

   (c) What should be the format of questions? Should we provide possible options, or let them give? On the basis of it, what shoud be the scoring mechanism?

   (d) How to aggregate the information? Should it be similar to methods like voting, stock markets, point spreads, pari-mutuel odds, future contracts?

   (e) How to ensure the conditions for the applicability of prediction markets are met?

   (f) How to reduce affect of various bias and identify or prevent manipulation, misinformation and other determinants to efficient working of markets?

   (g) What should be the target audience/people to involve in these markets?

Please find below the links of the some of literature reviewed:

1. Report 1

2. Report 2

3. Report 3

Week 1-4

Anshul Goel
Research Papers Summary

January 18, 2020

1 Corporate Prediction Markets: Evidence from Google, Ford and Firm X

1.1 Introduction

- Corporates are looking for internal Prediction markets to aggregate information trapped at different levels of organisation
- Issues:
  - Thin Markets: Limited participation on strategic topics $->$ thinner markets $->$ biases (optimism, passion than rational,loss aversion, probability misperception)
- Why these? Biggest Corporate Markets of knowledge
- Markets Covered: Demand forecast, product quality, deadlines being met, external events
- Markets outperformed other forecasting methods
- Though markets have inefficiencies like optimisation bias for Google, but they disappear over time via the 2 way mechanism:
  experienced traders can identify inefficiencies and secondly, the traders (of same experience level) with higher past returns will trade more in future against inefficiencies. Thus, average skill of market improve with time.
- Though insider traders have ulterior motives and are positively biased, the effect is ousted by large group of traders within the company with more experience and fewer reasons to be biased.
- The reason for efficient working of the market: "equilibrium market prices reflect aggregation of information and any subjective biases present", because the possibility of manipulation creates incentives for traders to be informed

- Overconfident employees are better for firms as they are emperically found to work harder and take risks beneficial for firm and overvalues their stock options, making the options based compensation cheaper. Thus, employee overconfidence lower the costs for the firm.
- Corporate prediction markets provide tools to measure and potentially correct employees optimism to get better forecats, though it may be in best interest of the firm to correct employees' optimism

1.2 Background

1.2.1 Google

- Prediction market for quarterly OKR (Objectives and Key Results)
- Though not all can be up for prediction market, 60% of company's OKR on prediction market
- Betting comprises of selecting among discrete choices, also binary in many cases

1.2.2 Ford

- Main impetus is on sales forecasting and predicting which features will be popular among customers
- Unlike Google, the payoff is linearly dependent on weekly sales of a particular model for sales forecasting
- Despite the expectation that the market will use long term success as feature for the 'decision market', the feature focus on whether the potential car features would reach a threshold level of interest in market research, if that research were conducted
- For features market, Ford discontinued to share the information and started settling markets based on final trade prices rather than market research outcomes, turning the market into 'beauty contest' markets.

1.2.3 Firm X

- Material and Energy Conglomerate with headquarter in Midwestern USA but global presence
- Most of its business are sensitive to macro-economy and commodity price
-

2 Trading Strategies and Market Microstructure: Evidence from a Prediction Market*

2.1 Introduction

- Starts with the question of how do the asset prices adjust to new information and how do the traders able to find the willing counter part parties.
- **Hypothesis 1:** Able to do due to liquidity needs. Heterogeneous information but form belief based on common prior
- **Hypothesis 2:** Participants have heterogeneous priors and henc different perceptions of same public information, hence if one buys, other is willing to sell.
- These models rely on partitioning the traders into distinctive strategy groups. Several studies have done grouping on capital markets and prediction markets
- Measures derived from Intrade Data: Volume, Transactions, Aggressions, Directional Exposure, Holding Duration, margin profit for each trader's account
- Observations about the traders (Intrade Presidential Election Bets):
  - Small number of highly active traders dominate total volume and transactions
  - 6 categories: Arbitrageurs, others being unidirectional with varying level of bias (extreme, high, moderate, low) on one side
  - Cross market arbitrage on 2 platforms: **BetFair** and **Intrade** where the Romney contract was overpriced on Intrade than BetFair as a single trader had **one-third** of all bets placed on Romney and **one-fifth** of all bets against Obama, providing a support to prices on Intrade. The assumed reason behind this can be to boost fundraising, campaign morale or turnout.
  - 32% volume (87% of traders) never changed the direction of exposure, 37% of volume from traders strongly biased in one direction, a handful of arbitrageurs 16% of volume, 15% volume of unbiased traders

Note: Whenever a trade occurs, it involve a new, incoming order and an older one that is resting in order book. The new order is called ***aggresive*** *or liquidity-taking* order and the old order is called ***passive****/liquidity-providing* order.
Aggression: is the share of total volume that resulted from a marketable order placed by the trader, meaning, these are the orders that trade against resting orders and remove liquidity

2.2 Data

- Positive relationship between volume and number of transactions

- Contains details of the volume, transactions, scope, direction, etc with special emphasis on 10 traders who were top 3 in one of more of the following criteria: volume, trades, frequency, margin, profit/loss. Also curated the trading strategy of 2 traders from their transactions data

2.3 Characterizing Strategies

- Measures used to compute strategies: Aggression, Direction, Exposure, Duration, Margin Profits
- Aggression: Proportion of trader's volume that was initiated by an aggressive order. Arbitrage strategies have intermediate value of aggression measure
- Direction: Size of direction (bias: absolute value of direction) rather than sign matters more in categorising strategies
- Holding Duration: Many have median holding period of zero, sign of arbitrageurs
- Margin: Equal to worst-case loss
- Profit: large positive/negatve profits in case of unidirectional, relatively smaller in arbitrage

2.4 Trading Strategies

- Bias reason: Wishful thinking, i.e, tendency to bend the expectations according to preferences
- 6 trading strategies:
  - Arbitrage: median holding period less than 10 minutes
  - Unidirectional: bias of 1 (non-arbitrage)
  - Extreme bias: $0.9 < \text{bias} < 1$
  - High bias: $0.5 < \text{bias} < 0.9$
  - Moderate bias: $(0.25, 0.5)$
  - Low bias: < 0.25

2.5 Beliefs Objectives

Earlier grouping was based on what they are doing, now question is why they are doing?

- Plain Arbitrage: risk-free profit with minimum capital outlay
- Cross-Market Arbitrage: Betfair(GBP) and Intrade (USD)
- Speculation:

2.5.1 Speculation

- Based on Expectation Maximisation Hypothesis
- Speculation will happen only when your subjective probability > current market price of contract

- Hedging:To hedge asset prices, like market indices
- Manipulation: discussed in next section

2.6 Manipulation

- Though high volume and tight spread of Intrade 2012 prediction market, are there the possibilities of market manipulation?
- Anamoly detected on election date, which calls for 3 possibilities:
  - Romney was underpriced: Not completely rejected as BetFair is pound based
  - Manipulation for financial gain: Not seems the reason from glance as there is a weak positive correlation rather than a negative one
  - Manipulation for political reasons: Most probable as capital outlay for market manipulation is quite low compared to fundraising vcampaigns, and in terms of voters momentum and sentiment (theoritical, not much emperical).
  - *still the forecasts from prediction market matched aptly with forecasters after considering for favorite longshot bias*
  - Though manipulation is possible with comparably small amounts, the predictions are accurate and the anamoly are captured easily

2.7 Discussion

- Seems the market is driven by the hypothesis of "different interpretation of public information"
- contains not just public information but also scattered private information held by traders

Week 5-8

Anshul Goel
Progress Report: Week 3-8

January 18, 2020

Contents

1 *Questions* & Ideas

"The mathematics of martingale provide provides a variety of theoretical predictions. Are they correct in real prediction markets? This makes an interesting topic for student course projects, to gather data and test theory for oneself, or to read academic papers which have done so."

1. Verify, if possible, if price probabilities follow martingale for certain prediction markets?
   - Gather Data & Test the theory
   - ***What is the method to test if a sequence of random variable is Martingale? Is there a genralised method or we just check with empirical data if they satisfy the properties like Serious Candidate principle and we are good to disprove that it is a martingale?***
     [Ref: Slide 22
     "So it seems most appropriate to call the assertion prediction market prices should behave like martingales, a hypothesis, and seek to see if its mathematical consequences are consistent with empirical data."]
   - Look for reasons of why or why not X_n is a martingale? Delve into the Economic theory. (Motivation: Section 2.3 & 3.3.1)
   - Try to create a trading strategy
     (on lines of **Conservation of Fairness Theorem**, fall of which "may" be sufficient to prove that prices don't follow Martingale)
2. **Halftime Price Principle**
   (a) Verify halftime price principle
   (b) Delve into the reasons of its success and failures
3. Test various theoretical predictions derived from Martingale's Mathematics (start with the ones in this chapter, and author's site)
   - ***Out of all contracts starting at price around 40, about 40% do in fact end with the event happening?***
     (a) Number of Crossings in an interval
   - **Serious Candidate Principle:** Can be applied to any competition with single winner
     Motivation:
     Theoretical argument that every prediction market is a martingale should not be affected by fashion or opinion poll results. So, check if prediction markets behaved differently from what theory says which will indicate unusual aspect of 2012 race.
     (a) ICC cricket World Cup and maybe used to check for betting in cricket market [more depth & though required]
     (b) To detect anomalies in the World Cup, etc

(c) **Note:** Keep in mind the martingale theory is applicable for fair odds game, and can't compare some incoming teams like Oman to say, Australia/England/India

- **Down-crossing Inequality:**
  (a) To check if there are too many fluctuations to be a martingale?
  (b) May provide some insight about match fixing or anomalies, a sudden increase in down-crossings?, more down-crossings or breaking the down-crossing limit for some matches and not the others
- One theorem about martingales says that the overall result of any system for deciding how much and when to bet, and when to stop, within this odds" setting, is simply equivalent to a single bet at fair odds. **So one can prove theorems about martingales by inventing hypothetical betting systems and analyzing their possible outcomes.** [Slide 14, David Aldous Lecture 9, STAT150]
  (a) Can be used to create some principles and theorems using martingale's definition, which can be used to detect anamolies and possibly match fixing.
  (b) ***Need to delve deep into martingales?***

2 Thesis Plan

2.1 Spot Fixing/ Detecting Anomaly/Creating Prediction Markets (Theoretical Work)

Timeline: Mostly in next semester, do some data more empirical or "safer" work this semester

1. Try to derive some theoretical results or principles on prediction markets like done by Aldous for spot fixing or anomaly detection for suggesting a trading strategy
2. Can go into the theoretical framework in either case of prediction market coming out to be/ not be martingale? (*Just presenting my crude idea*)
3. May work to create a real world prediction market of our own which may focus on
   (a) Startups to set prices of their products (*A bit sceptic as pricing is a bit subjective)*
   (b) Prediction markets on more objective stuff like number of app downloads and conversion rate from free to premium subscription, etc. (didn't find any literature on combination)
   (c) Motivated by Wisdom of Crowds (in progress), my submissions for business oriented role in Flipkart & discussion with peers with entrepreneurial endeavours

2.2 A backup project if things don't work out

Timeline: Perform the tasks which can also be clubbed as part of work in above tasks this semester. Possible to devote much more time in next semester

1. Test various theoretical predictions on cricket betting markets discussed by David (section 1.2)

2. Verify if the markets follow martingale

3 Problem Setup

3.1 Theoretical Setup

X_n = probability team A wins, given the information available at time n,and this should be a martingale ($E[X_{n+1}|X_n = x_n, ..., X_1 = x_1] = x_n$)

3.2 Problems

Problems in applying this to actual games/markets:

- Case 1: Setup is correct
  No automatic procedure exists which assigns numerical values to these probabilities (We don't observe probabilities)

- Case 2: Setup is conceptually doubtful
  Reasoning: Say looking at baseball/election market and information we have till time n is of some past events whose conceptual utility in our current setup is doubtful.

In the chapter, this conceptual doubts is not covered and the concept of "consensus probabilities" is introduced

3.3 Consensus Probabilities

Prices in prediction markets are consensus opinions as one can buy if s/he thinks prices are low and vice-versa
Author provided an approximate identification:

prices in prediction market $< - >$
consensus probabilities $< - >$
probabilities within axiomatic setup suggests as plausible hypothesis that
"(Hypothesis:) prices in prediction market should behave,i.e.fluctuate in time approximately as martingale"

4 Summary: Chapter 4 (Book & Slides)

Some of the chapter summary is incorporated into first 2 sections and not repeated for brevity

4.1 Definition: Prediction Market

4.2 Some Data: Baseball & Elections

4.3 Fair Games & Martingale

1. "When X_n denotes some real-world quantity at time n, we could model (X_n) as some unspecified martingale. In some contexts there are compelling reasons why this is reasonable; in some other contexts there are plausible but vague arguments why this is reasonable. In contexts where one has enough data, one can check empirically whether quantitative predictions from martingale theory are correct"

2. Single Fair Bet: Start with known fortune x_0 and play a single bet fair game with your fortune be X_1 gain be $X_1 - x_0$, then the gain has mean zero, i.e. $E[X_1] = x_0$

3. Martingale property gives that whatever strategy in **fair game** you choose which may also depend on what has happened so far $(X_n = x_n, X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, ..., , X_1 = x_1)$, X_{n+1} is a martingale implying your expected wealth will be equal to initial fortune. (follows from Conservation of Fairness Theorem)

4. **Conservation of Fairness Theorem:** In case of betting on a fair game using any arbitrary strategy and deciding time to leave the game T by any strategy, fortune X_T will be such EX_T is equal to initial fortune, same as single fair bet

5. **Basic Formula for Fair Games:** Continue betting on fair game with initial fortune x_0 until it either reaches 0 or $B > x_0$, probability of reaching $B = \frac{x_0}{B}$
   Derivation: $E[X] = x_0 => p * B + (1 - p) * 0 = x_0 => p = \frac{x_0}{B}$

"The mathematics of martingale provide provides a variety of theoretical predictions. Are they correct in real prediction markets? This makes an interesting topic for student course projects, to gather data and test theory for oneself, or to read academic papers which have done so."

4.4 Prediction Markets & Martingales: Conceptual Issues:

Covered under Problem Setup

4.5 Theoretical predictions for the behavior of prediction markets

Theoretical Predictions which can be checked:

1. **Out of all contracts starting at price around 40, about 40% do in fact end with the event happening?**

2. 60% is same as around 40% for opposite event & can make efficient use of data by considering contracts that ever crosses 40 or 60, taking first such crossing as initial time.

3. Probability that price reaches b before it reaches a is:
   $p_x(a,b) = \frac{x-a}{b-a}$
   Derivation: Let probability to go from x to b first be p and to a first be 1-p. From conservation of fairness theorem:
   $E[X] = x_0$ which gives:
   $p * b + (1-p) * a = x_0$, which gives: $p = \frac{x_0-a}{b-a}$
   , leading to these 2 formulas:

   - **Crossing of an interval**
     - Consider an interval, say [40,60] & if the price is ever in this interval, it will have to cross 40 or 60, say, start at 40
     - x = 40, a = 0, b = 60, $p_{40}(0,60) = \frac{2}{3}$, that is with probability of $\frac{2}{3}$, price can reach 60. Call this **First Crossing**
     - Let C be number of such crossings, C = 0,1,2,... At point b, with prob of $\frac{1}{3}$, it will go to 100 and with $\frac{2}{3}$ to a(=40).
       Therefore, $P(C = i) = \frac{1}{3}(\frac{2}{3})^i$
     - When applied on 103 baseball matches, the predicted and observed crossings were quite close to lesser number of crossings but discrepancies for larger counts because near the end of the game there can be a huge impact of single hit and there may be no trades between hits
   - Maximum & Minimum Prices
     For a contract starting at x (in favor of A), either
     - team A wins, the contract expires at 100 and there is some overall minimum price L_x such that $L_x \leq x$
     - team A loses, the contract expires at 0 and there is some overall maximum price L_x such that $L_x \geq x$
     - The formulae for distribution of L_x is:
       * $\Pr(L_x < a) = \frac{a(100-x)}{100(100-a)}, 0 < a < x$
       * $\Pr(L_x > b) = \frac{X(100-b)}{100(b)}, x < b < 100$
     - Derivation:
       $\Pr(L_x < a, Awins) = \Pr(\text{price hits a before 100}|\text{initial price x}) \times \Pr(\text{A wins}|\text{current price=a})$
       $\Pr(L_x < a, Awins) = (1 - p_x(a,100)) \times p_a(0,100)$
       $\Pr(L_x < a, Awins) = (1 - \frac{x-a}{100-a}) \times \frac{a}{100} = \frac{a(100-x)}{100(100-a)}$
       Similarly, can do for losing case

4.6 Were there improbably many candidates for the 2012 Republican nomination whose fortune rose or fell? A case of Serious Candidate Principle

Comparatively large number of candidates with rising & falling popularity in 2012 nominations. Many discussions that number of such candidates is higher than usual.

Need to clarify these 2 things:

1. Mathematics doesn't say nothing about how people's opinion would fluctuate. One can't create statistics for comparison with past data but there are no mathematical predictions on fluctuations in opinions
2. **Theoretical argument that every prediction market is a martingale should not be affected by fashion or opinion poll results**

So, check if prediction markets behaved differently from what theory says which will indicate unusual aspect of 2012 race.

4.6.1 Serious Candidate Principle

Use **Serious Candidate Principle** for that.

Definition:"Consider an upcoming election with several candidates, and a (prediction market) price for each candidate, and suppose initially all these prices are below b, for given 0 ¡ b ¡ 100. Theory says that the mean number of candidates whose price ever exceeds b equals $\frac{100}{b}$ "

Derivation: Buy a contract for every candidate if and when their price reaches b.
Total cost = bN_b, Gain = 100, Net Gain = 100 - bN_b
Conservation of Fairness Theorem: $E[100 - bN_b] = 0$ which gives $E[N_b] = \frac{100}{b}$

Expected Number and observed number of candidates exceeding threshold price (b) were quite close:

	expected	observed
b = 33.3	3	3
b = 20	5	5
b = 16.6	6	9
b = 12.5	8	10

Technical Considerations:

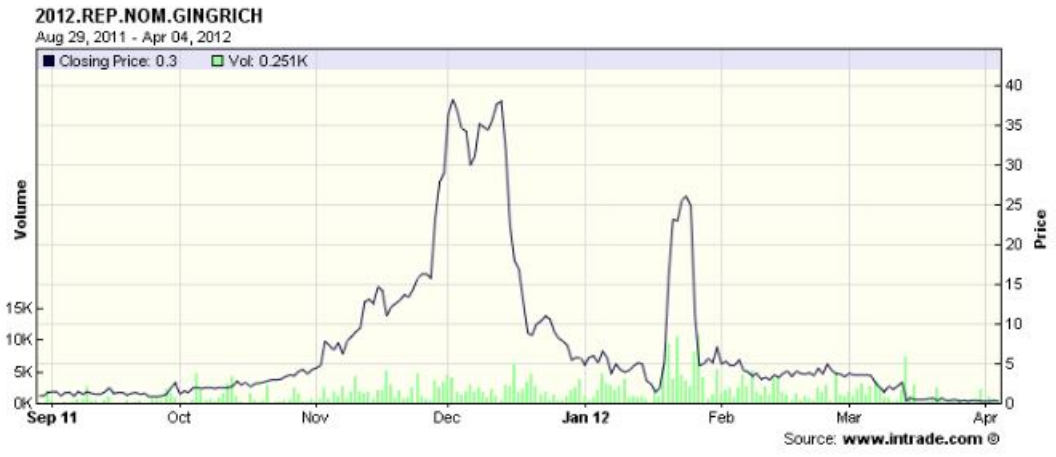
1. Data on initial prices is somewhat reliable because of thinly traded markets in the start
2. **Covering your position Argument:** For long duration contract, low price contracts overstates the consensus probability because of the requirement of margins to cover their positions (opportunity cost of holding the money for margin vs investing elsewhere). Correcting for this would make the expected numbers larger than that shown

Conclusion: "even if it is actually true that the month-to-month fluctuations in opinion poll standings were greater than usual, we can see no sign that this unduly influenced the smart money being wagered on the prediction market"

4.6.2 Upcrossing Inequality

Too many fluctuations to be possibly a martingale?

Figure 4.3: Intrade price for Gingrich nomination.



Use the concept of upcrossing inequality.

Definition: "Consider a price interval 0¡ a ¡ b ¡100, and consider an upcoming election with several candidates, and a (prediction market) price for each candidate, where initially all these prices are below b. Theory says that the expected total number of down-crossings of prices (sum the numbers for each candidate) over the interval [a,b] equals $\frac{100-b}{ba}$ "
Expected down-crossing count came out to be 8 while data gave 7 down-crossings for interval [10,20]
Derivation: For each candidate, buy a contract if price reaches b. If price fall to a, sell buy again if it reaches b. Exactly 1 will expire at 100.
Others will be sold at price a, the number $D_{a,b}$ is number of downcrossings in [a,b].
Gain = 100 - b - $D_{a,b}(b-a)$
Conservation of Fairness Theorem: E[100-b-$D_{a,b}(b-a)$] = 0 which gives the required formulae: $D_{a,b} = \frac{100-b}{b-a}$

4.7 The Halftime Principle

4.7.1 Not Martingale!

Here we detour away from martingales to see a more elementary pice of theory which can be checked against data

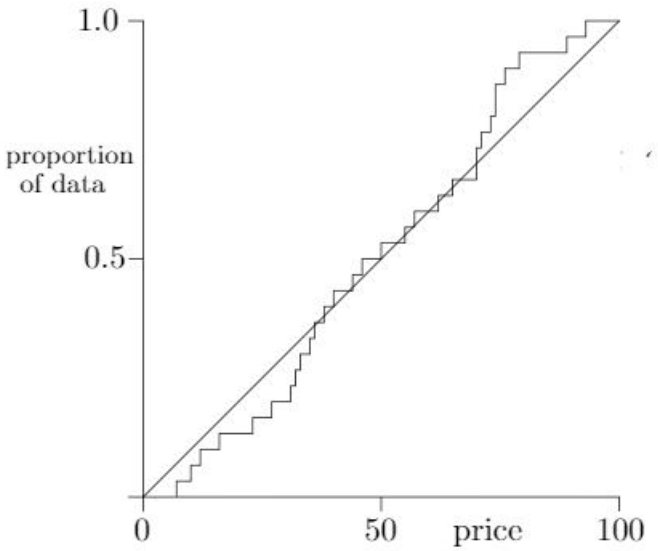
4.7.2 Halftime Principle

"In a sports match between equally good teams, at halftime there is some (prediction market) price for the home team winning.This price varies from match to match, depending largely on the scoring in the first half of the match. Theory says its distribution should be approximately

uniform on [0,100]"
$\Pr(z_1 + z_2|z_1) = \Pr(z_2 > -z_1) = F(z_1)$

4.7.3 Setup

- Assume teams are equally good(50%) possibility of each team winning
- z_1 is point difference in 1^{st} half, z_2 in 2^{nd} half
- Assumptions:
  - Realistic:
    1. z_1&z_2 are independent random variables with same distribution
    2. Distribution is symmetric about 0, i.e. $F(z) = 1 - F(-a)$
  - Unrealistic:
    1. Distribution is continuous (for mathematical simplicity)
- **Probability home team wins, given 1^{st} point diff. is z_1 is**
  $\Pr(z_1 + z_2|z_1) = \Pr(z_2 > -z_1) = F(z_1)$
  a continuous distribution it is always true that F(z_1) has uniform distribution on (0%,100%)
- An example of prediction market prices at halftime for which the initial prices were close to 50 from 30 baseball matches:



- For unequal time, the distribution of half time price depends on initial price distribution & z_1

4.8 Other Martingale Calculations

Not important for our discussion

4.9 Stock Markets & Prediction Markets

1. Interpretation of price in stock market may be subjected to debate but in prediction markets, it tells the probability of event happening is more definite
2. Conceptually simpler than stock market as it is based on outcome of specific external event and will expire at either 0 or 100
3. Need no empirical data for making theoritical predictions whereas in stock markets, one needs an estimate of variance
4. Thinly traded and less martingale
5. Prediction market is zero sum game, even negative considering transaction costs while stock markets can be considered as they will exceed long term risk free rewards (seems positive sum game)

4.10 Wrap-Up

Prediction markets can serve as hedge against real-world risks but currently are thinly traded.

Week 9-12

Anshul Goel
Write Up: Week 9-12

January 18, 2020

Contents

1 Data

1. Data for football matches (English Premier League) has been scrapped from bet365. Sample data can be viewed here:
   Link

2. Data for various sports including football and cricket is obtained from betfair. Past data goes to year 2015. Also, data for various types of betting markets is available. Please see the link below for match odds market data:
   Original Data
   A bit formatted Link
   Need some thought because the 3 data points: Home Team winning, Away Team Winning and Draws are updated at different times, as expected in an exchange. Maybe we can create 3 different martingales for each type of data point, say 1 market for probability of Home team winning and one for Away team winning.

3. Difference between bet365 and betfair data:

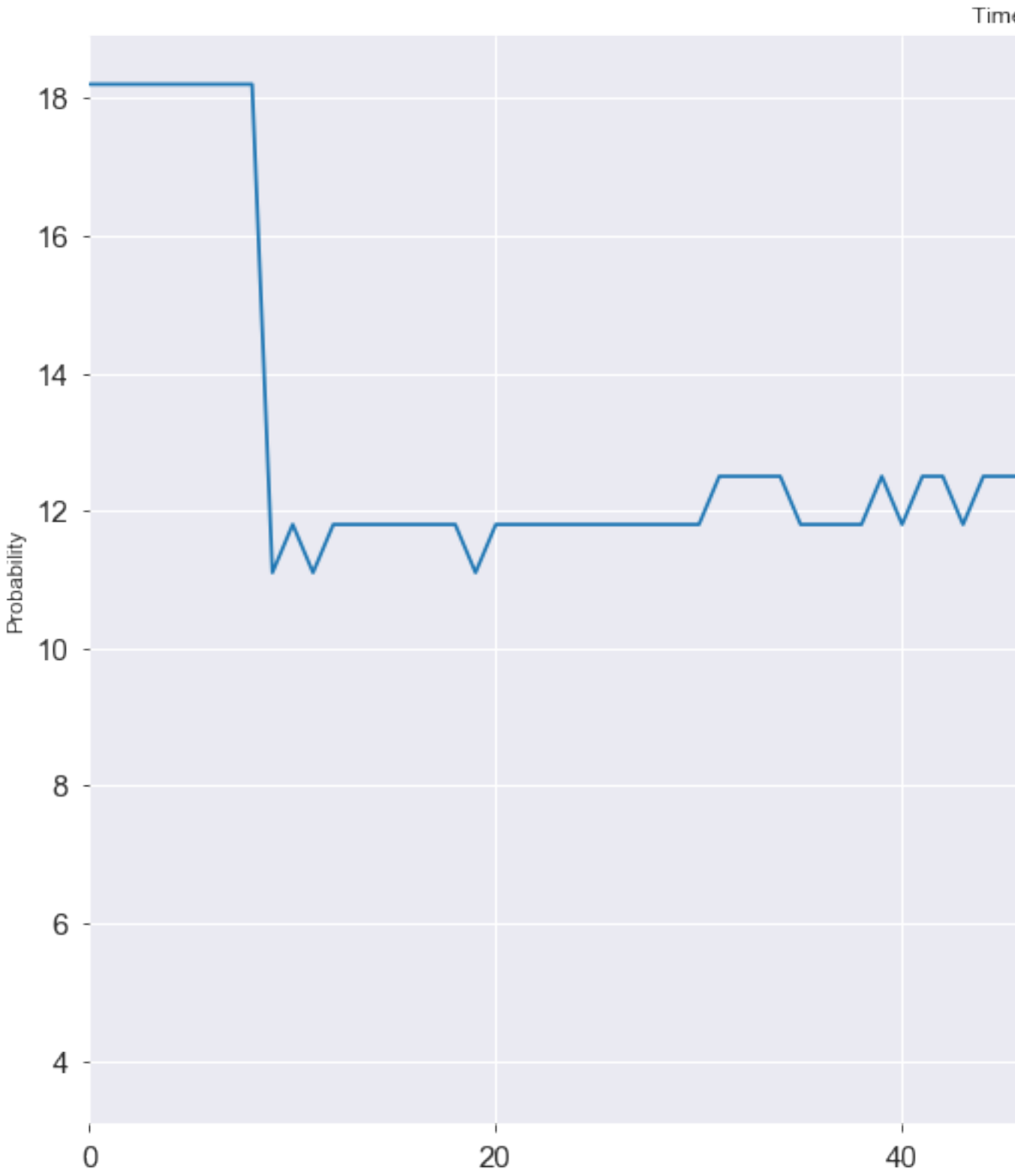
   (a) Bet365 is online bookmaker, hence the odds are set by bookmaker according to bookmaker and are not accurate, the reason being to keep the odds unequal to give incentive to people to bet. Also, the probabilities sum to ¿ 1 and anything additional to 1 is bookmaker's margin, which is subtracted from payoff of winning party. **Hence, the data from bet365 may not reflect the true wisdom of crowds**, and the bookmakers' estimates of implied probability are derived from some of the statistical techniques like this: Link
   By the way, I also connected with the person mentioned in the article and he is quite interested about application of neural networks to create the strategies and advised to focus on smaller elements of game, like focusing on the strategy to predict a players performance than predicting outcome of IND-AUS match

   (b) Betfair is an exchange, though it also offers bookmakers features but that is a completely different market, independent of it. It works like conventional stock market with option of going long or short (back or lay respectively). It may offer better insights into our analysis.

4. Because we have bet365 data too and converting the bet365 data will take some time, I will also implement the tests on Bet365 data too, some of which I implmented for one of the match, discussed in next section.

2 Tests & Analysis

1. Below are the snapshots of price (probabilities) with match time:


Time
Probability
18
16
14
12
10
8
6
4
0
20
40

Literature Review

Anshul Goel
Various Papers

January 18, 2020

Contents

1 Prediction Markets: Justin Wolfers & Eric Zitzewitz

1.1 Introduction

1. DARPA in 2003 tried to start prediction markets in - Economics, health, Political and events. But, was scraped due to political issues. The goal was to get probability of events & connection between events.

2. Looney provides details of proposal and aftermath

3. Robin Hanson has archive of related news story, etc.

1.2 (Looney 2003)DARPA's Policy Analysis Market for Intelligence: Outside the box or Off the Wall?

1.2.1 Motivation/Goal

Motivated from futures market like petroleum futures where future prices provide information about the events to happen in Middle East. There were supposed to be 2 type of markets:

1. Quarterly contracts based on (i) specific events; (ii) global economic and conflict indicators such as likely occurence of regime change in Syria; (iii) data indices that track economic health, civic stability, etc.

2. Positions on interrelated issues (Motivation: Bayesian Probabilistic forecast used by CIA) ***"Idea is to create chain of events leading to activity of main concern"***

   Some points from Bayesian Forecast Papers

   (a) Objective Probability - prior probability, supplemented with other data to gain better idea - posterior probability

   (b) Application in political forecasts as priors are even more difficult to get (historical examples maybe inadequate or inapplicable), like interpretation of statements

   (c) **Delphi Technique**

2 Discussion: 9 December 2019

Economics behind it: market design -¿ credits? incentives for participation, truth telling? How should people be rewarded? Book: Inefficient markerts - Schleifer (chapter 1) 1st para: broad questions & motivations 2nd para: Very specific questions I intend to answer in 4 months Reference: 1. **Wisdom of Crowds** 2. How to design option markets, how to design matching markets, [**Who gets what - and why**] **& blog**, Micro II

3 Links

Look for market design elements.

1. Some example of prediction markets:
   (a) www.estimize.com : Prediction markets over Stocks and Economic Indicators, analysts and amateur provide estimates (players)
   (b)
2. tradesports : Yes/No based betting/prediction markets on sports.
3. intrade : Explore for future and how to use its design
4. Augur : A method to execute prediction markets
5. **Very Important** Metaculus : Best method/mechanism for creating markets
6.

4 Creating game/market

1. **Questions & Question type** : Numerical, Ordinal, Categorical, Binary
2. **Response** : The form of response, say, to include lower & upper margin, probability of various outcomes, etc. (give numberMetaculus or probabilitiesReplication Markets)
3. **Scoring Rule** : Metaculus provides various possible scoring rules, along with their own

Some comments:

You should be influenced by what others think (since in aggregate they are likely to be reasonably right), but not slavishly follow the crowd, since if you have better insight, that pays off handsomely in your score.