

# Project Proposal

GOVT-653

*Kiernan Nicholls*

*March 7, 2019*

I am hoping to further study the use of virtual prediction markets as tools to predict American Congressional elections. Prediction markets use self-interested traders betting on the outcome of an election to produce a probabilistic view of potential outcomes.

I first began the process of this study last semester in GOVT-696 Programming Applied Political Data Science. During that class, I really only began a rudimentary collection and analysis of the data. For this class, I'm hoping to further explore the economic forces at work in such a system, as well as the fundamentals of the alternative prediction tools (polling and modeling).

## Literature Review

Arrow, Kenneth J., et al. "The Promise of Prediction Markets." *Science*, vol. 320, no. 5878, 2008, pp. 877–878. JSTOR, [www.jstor.org/stable/20054723](http://www.jstor.org/stable/20054723). This article discusses the need for further study on prediction markets in the context of the legal issues surrounding the practice. The United States federal government has legislation which outlaws many types of gambling, especially those around elections. Arrow et al. suggest these laws "create significant barriers to the establishment of vibrant, liquid prediction markets in the United States." The Unlawful Internet Gambling Act of 2006 hinders the potential to utilize the internet as a perfect market to exchange futures contracts. The article proposes two specific steps to facilitate the use and study of prediction markets: (1) the Commodity Futures Trading Commission (CFTC) should create safe-harbor rules for small-stakes prediction markets, and (2) Congress should support prediction markets with legislation to cover the increased spending at the CFTC and secure a legal framework for prediction markets. The date of this paper is noteworthy, as the CFTC does currently extend PredictIt and the Iowa Election Market a "no-action" status due to their academic value. This paper argues more action needs to be taken to truly utilize and explore prediction markets.

Berg, Joyce E., and Thomas A. Rietz. "Market Design, Manipulation, and Accuracy in Political Prediction Markets: Lessons from the Iowa Electronic Markets." *PS: Political Science and Politics*, vol. 47, no. 2, 2014, pp. 293–296. JSTOR, [www.jstor.org/stable/43284536](http://www.jstor.org/stable/43284536). This paper explores the shortcomings of prediction markets, specifically the potential for manipulation. The authors use the Iowa Election Market's 2012 Presidential election. The paper uses the Securities and Exchange Act definition of price manipulation: "To effect, alone or with 1 or more other persons, a series of transactions... raising or depressing the price of (a) security, for the purpose of inducing the purchase or sale of such security by others." The author's extend this definition to political prediction markets by assuming manipulators are using price manipulation not only to profit but as potential means to change the outcome of the election itself. The paper does not weigh in on the debate between "bandwagon theory" and "underdog theory" (which each try and explain the effect perception of victory has on voter behaviour), only going as far as connecting each theory with market price. The paper concludes that manipulation can affect election prediction market forecasting accuracy, although the motives for manipulators is unclear. Market design can mitigate the risk of manipulation. They found little risk that the Iowa Election Market could be successfully manipulated.

Fair, Ray C. "Interpreting the Predictive Uncertainty of Elections." *The Journal of Politics*, vol. 71, no. 2, 2009, pp. 612–626. JSTOR, [www.jstor.org/stable/10.1017/s0022381609090495](http://www.jstor.org/stable/10.1017/s0022381609090495). Uncertainty is a fundamental concept in the prediction of elections. This paper discusses the nature of uncertainty in U.S. elections, the shortcoming of "error" in opinion pollings, and potential role for prediction markets to help both social scientists the the population at large define and understand this uncertainty. The theory of uncertainty assumes that on election day there are  $n$  possible outcomes to an election, each with a probability of  $1/n$ .

If in  $p$  percent of the  $n$  conditions a given candidate wins, then  $p$  is the probability of that candidate winning. Fair explains that “even if we knew the  $n$  possible conditions of nature, the best that one could say at the beginning of election day is that A candidate would win with probability  $p$ ” (612). Weather is given as one example cause of uncertainty on election day; given the truly chaotic nature of weather, and its unknown effect on voter turnout, uncertainty must exist on election day. Prediction markets offer insight into uncertainty due to the balance of interests and the zero to one nature of the futures contracts being trades. Since the outcomes are essentially limited to two (win or lose) and the date of the event is known, the equilibrium price of the prediction market should correlate with the probability of each outcome. Fair contrasts this type of uncertainty with the standard error of opinion polling. Fair explains that if you polled the entire population of voters the day before, the poll would come with a standard error of zero despite uncertainty still existing in nature (thanks to weather, for example). The paper concludes with Fair’s “ranking theory” which provides a role for uncertainty in the strategic behaviour of political operatives in their ranking of campaign priorities.

Lee, David S., and Enrico Moretti. “Bayesian Learning and the Pricing of New Information: Evidence from Prediction Markets.” *The American Economic Review*, vol. 99, no. 2, 2009, pp. 330–336. JSTOR, [www.jstor.org/stable/25592420](http://www.jstor.org/stable/25592420).

This paper by David Lee and Enrico Moretti explored the topic of prediction markets as they relate the Bayesian statistical theory. Since prediction markets allow social scientists to explore the opinion of the market over time in direct response to changes in information. The paper explores the prediction market hosted on the now-defunct Intrade regarding the 2008 Presidential Election. The paper explores the direct relationship opinion polling has on prediction markets. In theory, the market prices should reflect all information available to the traders, with opinion polling being the most popular and useful in 2008. The researchers found that “rather than anticipate significant changes in voter sentiment, the market price appears to be reacting to the release of the polling information” (330). In their model, investors have their own opinion of each candidate’s probability which is updated after receiving the information obtained by a poll.

Northcott, Robert. “Opinion Polling and Election Predictions.” *Philosophy of Science*, vol. 82, no. 5, 2015, pp. 1260–1271. JSTOR, [www.jstor.org/stable/10.1086/683651](http://www.jstor.org/stable/10.1086/683651).

In this paper, Robert Northcott says “election prediction by means of opinion polling is a rare empirical success story for social science” (1260). In so much of social science, predicting future outcomes proves difficult due to the open system nature of human elections. Northcott cites many before him who have claimed there are simply too many variables acting on too many unique individuals to establish a model. Northcott takes a philosophical angle in his analysis of how opinion polling has proven the exception to this understanding, providing social scientists with a tool to predict the actions of a population in advance. Northcott concludes that the success of opinion polling is no mystery, but a result of scientifically rigorous methodology. The accuracy of forecasting in the 2012 Presidential election by Huffington Post, FiveThirtyEight, and Princeton Elections is, according to Northcott, “a stunning success story, arguably with few equals in social science” (1261). Northcott explores the two levels of opinion polling, the first level demographic balancing and the potential second level aggregation that has recently become popular. He addresses sampling errors and election system idiosyncrasies as sources of errors in polling’s ability to translate to the true population. In this paper, the difference between explaining an election and predicting an election is discussed, an importance difference when expressing the need for prediction in social science.

Silver, Nate. *The Signal and the Noise: Why Most Predictions Fail but Some Don’t*. Penguin Press, 2012.

In the preface to the print edition, Silver discusses the nature of “big data” and its place in Gartner’s hype cycle. When Silver’s data journalism venture FiveThirtyEight used a forecasting model to accurately predicted all 50 states in the 2012 Presidential election, big data potentially at the peak of the hype cycle. In the preface, written in 2014, Silver worries data-based predictions might soon enter the second phase of the hype cycle, the “trough of disillusionment.” This insight is particularly interesting in the wake of the 2016 Presidential election only two years later, where Silver’s model was potentially the most accurate of the prediction models, but still gave the eventual loser of the election a probability of victory greater than 70% on the eve of the election. While not a scholarly work, *The Signal and the Noise* is useful in the context of this paper because FiveThirtyEight provides the probabilistic prediction to which markets would be compared.

In the book, Silver explores the nature of predictions and uncertainty, especially as they relate to elections. Prediction markets are brought up throughout. In one instance; Silver recalls being challenged to “put his money where his mouth is” (a key concept in prediction market theory) when his model-based prediction for the 2012 Republican Iowa Caucus contradicted the market price at the time. Ultimately Silver won that bet, and he continues to challenge reliance on prediction markets due to their fundamentally human nature. The fallibility of human judgement is a key concept in this book. This book provides crucial insight into Silver’s FiveThirtyEight model and the nature of probability and uncertainty in predictions.

## Theoretical Question

Throughout the campaign, with varying levels of information available to traders and models, can prediction markets offer a more accurate view of probabilities compared to mathematical forecasting models? Data

I would use market data from PredictIt.com, which provides their data for free for academic use. I have partnered with the service to obtain data on the 2018 midterm markets. Below is a random sample of market data with selection of variables:

Table 1: 44,711 observations of 6 variables

ID	Market	Contract	Date	Low	High	Close	Volume
3883	TX29.2018	DEM.TX29.2018	2018-08-06	0.96	0.96	0.96	1
3503	KING.MESENATE.2018	n/a	2017-09-05	0.83	0.83	0.83	0
3737	WA08.2018	DEM.WA08.2018	2018-02-07	0.77	0.81	0.77	3
3489	STAB.MISENATE.2018	n/a	2018-04-30	0.85	0.85	0.85	0
3863	TX05.2018	GOP.TX05.2018	2018-01-20	0.86	0.86	0.86	0
4263	NC09.2018	DEM.NC09.2018	2018-11-06	0.25	0.99	0.25	1243
4016	PA09.2018	DEM.PA09.2018	2018-06-21	0.11	0.11	0.11	0
2940	SANDERS.VTSENATE.2018	n/a	2018-07-14	0.96	0.96	0.96	0
3480	HEIT.NDSENATE.2018	n/a	2017-09-07	0.56	0.60	0.60	30
2941	MANCHIN.WVSENATE.2018	n/a	2017-06-04	0.62	0.62	0.62	0

I would use model data from FiveThirtyEight, who provide top-level outputs of their model for free to the public. Below is a selection of variables from the 2018 House model:

Table 2: 299,760 observations of 7 variables

Date	State	District	Special	Party	Incumbent	Model	Win Prob.	Avg Share
2018-09-02	WV	2	NA	D	FALSE	lite	0.087	41.43
2018-08-06	VA	5	NA	D	FALSE	deluxe	0.411	49.12
2018-09-29	TN	5	NA	R	FALSE	deluxe	0.000	32.33
2018-10-31	IL	10	NA	R	FALSE	lite	0.000	27.13
2018-08-15	WY	1	NA	D	FALSE	deluxe	0.000	29.28
2018-10-30	NY	1	NA	D	FALSE	classic	0.071	43.70
2018-08-31	NY	24	NA	R	TRUE	deluxe	0.875	53.76
2018-10-19	KY	6	NA	NA	FALSE	classic	0.000	2.94
2018-10-21	CA	43	NA	D	TRUE	classic	1.000	78.98
2018-08-12	OH	6	NA	D	FALSE	lite	0.010	35.37

Ultimately, I would like to explore the accuracy of each tool in their ability to predict elections. Furthermore, I would like to explore how changes in each tool’s inputs affect their accuracy (e.g., total trade volume, market manipulation, opinion polling, Google search trends, etc).

Preliminary R code to collect and format this data for comparison can be found on my github repository for this project: <https://github.com/kiernann/predictr>.