

# Final Project Memo

*Kiernan Nicholls*

*October 16, 2018*

## Proposal

I would like to study the effectiveness of prediction markets as a means of forecasting political outcomes. I hypothesize that prediction markets serve as a more useful tool than any given public poll as a means of predicting the outcome of elections in the United States. Prediction markets theoretically encompass polling data in their aggregation, in addition to the benefits derived from the wisdom of the crowd and personal financial stakes.

## Prediction Markets

Prediction markets are exchanges set up to trade “stock” in binary outcomes. For any given event, traders can buy shares of each of the binary outcomes: the event occurring or not. For every market, “contracts” are automatically established between two buyers. Traders buy shares of each outcome based on their individual assessment of likelihoods combined with a financial cost-benefit, risk-reward analysis.

For example, a market might be established that asks the question “Will the Democratic party win a majority in the House of Representatives in the 2018 Midterm Election?” Traders can then buy “Yes” and “No” shares on that market. For every “Yes” buyer, there is a corresponding “No” buyer agreeing to a contract. The price of each share is an analog for probability and must sum to 100. If the “Yes” buyer feels a share is worth \$0.67, then the “No” buyer is agreeing that his shares are worth \$0.33. When the event occurs in reality, the shares of each outcome are automatically sold for \$1 or \$0 depending on the outcome. If the Democratic party does win a majority, the “Yes” shares become worth \$1 each and the “No” shares become worthless. Essentially, the winner of the bet receives the funds wagered by his counterpart.

Theoretically, we can use these markets as a predictive tool. In aggregate, the average trading price for each binary outcome should reflect the probability of that outcome. Both the “Yes” and “No” buyers will want to maximize their financial gain and minimize their loss. Traders will not buy shares at a price higher or their believed worth. You would not buy a “Yes” share of an event at \$0.67 if you believe that event is unlikely to happen. In real time, the average trading price of shares reflects the aggregate belief of the market. Traders will theoretically take into account all available information to make smart trades, including traditional polling. Prediction markets do away with the mathematical nuances of random sampling and polling weights.

As of right now, shares of Democrats winning a majority in the midterms are trading for \$0.67 on PredictIt.org market. This can be interpreted as a 66% probability. Meanwhile, the forecasting method used by FiveThirtyEight.com places that same probability at 84.8%. Traders on the prediction market are less confident than the mathematical model. If my hypothesis is true, prediction markets should correctly predict the outcome of elections more often than polling or modeling.

## PredictIt

PredictIt.org is a prediction market run by non-profit Victoria University of Wellington, New Zealand for educational purposes. Users of this site can legally trade on the exchange with real currency, winning real money if their prediction comes true and their shares sell for \$1. Traders can also sell their shares at any time, provided there is a corresponding buyer; as public opinion shifts, “Yes” or “No” shares may become more or less valuable. The site is exempt from the usual ban on both online and political betting by working with researchers to study prediction markets as a political tool. In October of 2014, the Commodity Futures Trading Commission granted the site a No Action letter allowing them to operate.

The site hosts markets on nearly every conceivable political event, electoral or otherwise. Will Donald Trump be president at year-end 2018? Will the federal government be shut down on February 9? Will Ted Cruz be re-elected to the U.S. Senate in Texas in 2018? Will Facebook's Mark Zuckerberg run for president in 2020? How many tweets will @realDonaldTrump post from noon Oct. 10 to noon Oct. 17?

I am interested in the election markets, comparing them to both individual polling (Sienna College Poll), polling averages (<https://realclearpolitics.com>), and forecast modeling (<https://fivethirtyeight.com>) as a means of predicting the winner of political elections.

## Example Data

```
library(tidyverse)
library(gtrendsR)
library(lubridate)
```

First, we are going to read in FiveThirtyEight's 2018 Midterm forecast data for the House of Representatives. This is a proprietary model used to predict the outcome of the election. The model takes into account adjusted polling, their CANTOR method (polling from nearby districts), and a set of fundamental factors (fundraising, incumbency). Ultimately, we are looking for their predicted probability of the Democratic party winning control of the House of Representatives.

```
forecast <- read_csv("https://goo.gl/E9XJUf")
forecast <- filter(forecast, party == "D")
forecast <- select(forecast, forecastdate, win_probability)
colnames(forecast) <- c("date", "forecast.prob")
```

Then, we are going to do the same for PredictIt's public market data. Academic researchers have access to more detailed data, but this will do for now. Similarly, we are only interested in this market's probability of the Democratic party winning control of the House of Representatives. We will use the closing price of the "Democratic" share for every day as an analog for probability.

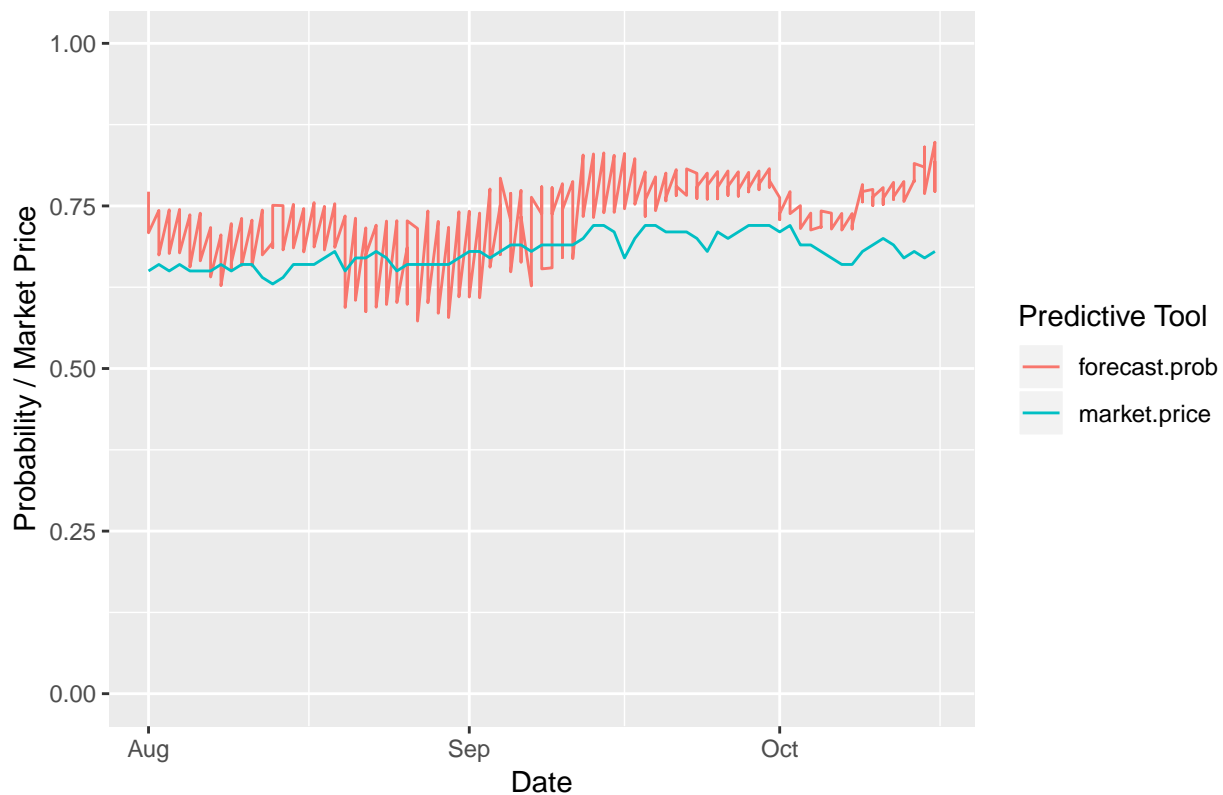
```
market <- read_csv("https://goo.gl/w4K8pP")
market <- filter(market, ContractName == "Democratic")
market <- select(market, DateString, CloseSharePrice)
colnames(market) <- c("date", "market.price")
```

Finally, we can merge these two data frames by a shared date to see how the market and forecast change over time.

```
predict <- as_tibble(merge(forecast, market))
predict <- predict %>%
  gather("market.price",
        "forecast.prob",
        key = "tool",
        value = "prob") %>%
  arrange(date)

ggplot(data = predict) +
  geom_line(mapping = aes(x = date,
                        y = prob,
                        color = tool)) +
  coord_cartesian(ylim = c(0, 1)) +
  labs(title = "Market and Forecast Probabilities of Democratic Midterm Victory",
       x = "Date",
       y = "Probability / Market Price",
       color = "Predictive Tool")
```

## Market and Forecast Probabilities of Democratic Midterm Victory



Here we can see the market price continually underestimate the Democrat's chances compared to the forecasting model used by FiveThirtyEight. Still, the trends between the two predictions are fairly closely linked; when one drops, so does the other, and vice versa.

## Further Questions

I would be interested in seeing if we can statistically establish if the FiveThirtyEight forecast or RealClearPolitics polling *causes* changes in market price, or if they move independently.

Once I partner with PredictIt.org as a research partner, I can gain access to their closed markets. A powerful test in predictive capabilities would be to compare the success rate of the three predictive tools (market, model, poll) in, say, the 2014 Midterms or the 2016 Election.

I would also be interested in using Google Trends data to see how the news affects the prediction market (and how Google searches affect both polling and market price).

Here I am plotting the Google Trends "hits" value for "Bill Nelson" against the market price for Senator Nelson's election on PredictIt.

```
nelson.trend <- gtrends(keyword = "Bill Nelson",
  time = "today 3-m",
  gprop = "web",
  geo = "US")
nelson.trend <- nelson.trend$interest_over_time
nelson.trend <- as_tibble(select(nelson.trend, date, hits))
nelson.trend$date <- lubridate::as_date(nelson.trend$date)

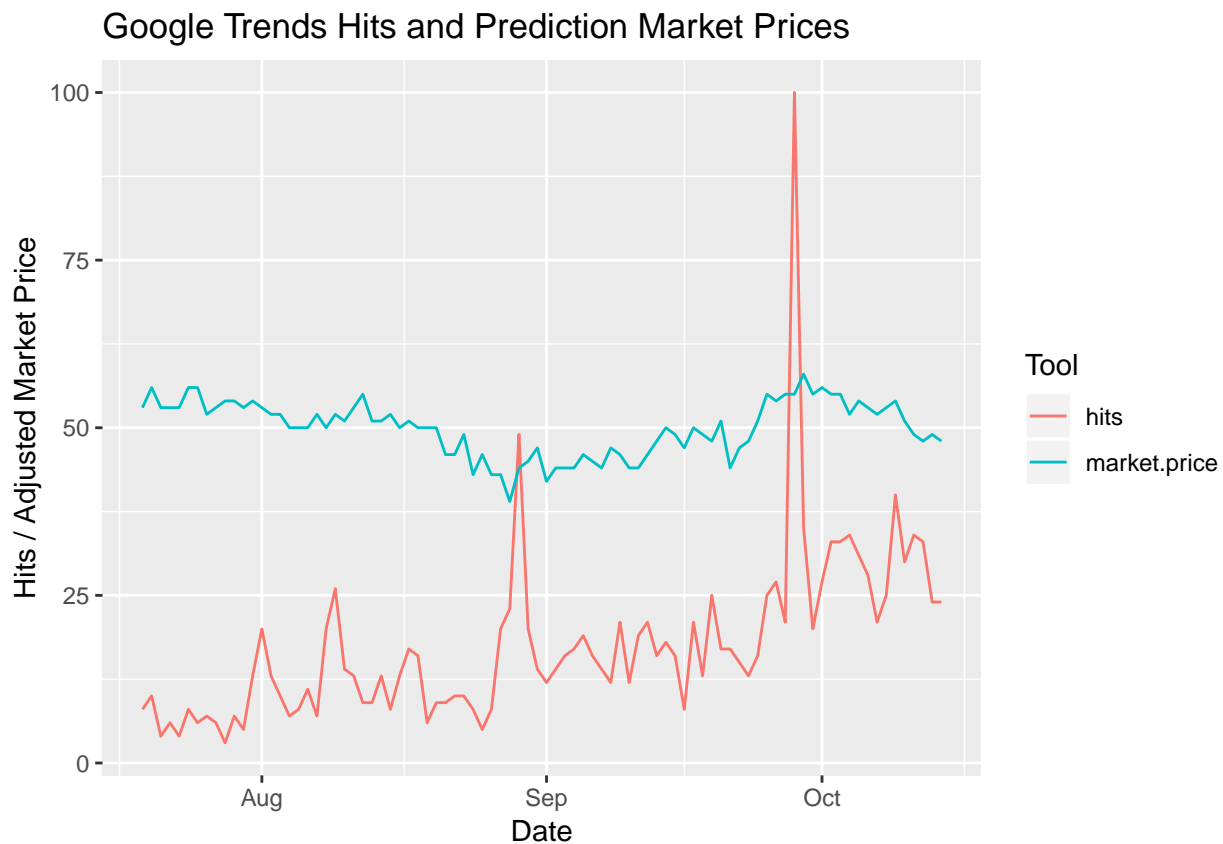
nelson.market <- read_csv("https://www.predictit.org/Resource/DownloadMarketChartData?marketid=2999&time=3m")
nelson.market <- select(nelson.market, DateString, CloseSharePrice)
```

```
colnames(nelson.market) <- c("date", "market.price")

nelson <- as_tibble(merge(nelson.trend, nelson.market))
nelson$market.price <- nelson$market.price * 100

nelson <- nelson %>%
  gather("hits", "market.price",
        key = "tool",
        value = "value") %>%
  arrange(date)

ggplot(data = nelson) +
  geom_line(mapping = aes(x = date, y = value, color = tool)) +
  labs(title = "Google Trends Hits and Prediction Market Prices",
       x = "Date",
       y = "Hits / Adjusted Market Price",
       color = "Tool")
```



We can see that the two spikes in search volume match with a drop in his market price (him doing poorly) and a rise in his market price (doing well).

## Conclusion

I think there is serious value in prediction markets as a tool in political science. Polling can inaccurately capture a sample or incorrectly weigh the responses. Forecasting relies on complex and proprietary modeling, necessitating trust in the modelers.