INTERVIEW

*by **Reuvain Borchardt***

POLLS APART

THOMAS MILLER, faculty director of the data science program at Northwestern University, discusses a data model he came up with for predicting elections, based on election-betting trends.

His model, posted on his website called The Virtual Tout, reviews his success at predicting the outcome of the 2020 presidential election and the Georgia Senate runoff elections in 2021.

Miller correctly predicted the outcome of every state in the 2020 presidential election besides Georgia. As if to prove he could in fact nail Georgia down, Miller refined his Georgia approach for the two January 5, 2021, Senate runoffs in that state. According to an article in Fortune by Shawn Tully, a reporter who has extensively touted Miller's work, "The polls gave Republican David Perdue a wide lead over Democrat Jon Ossoff, and showed the GOP's Kelly Loeffler in a dead heat versus opponent Raphael Warnock. By contrast, Miller's numbers had Loeffler heading for a big loss, and Ossoff en route to a modest victory. Once again, the contrarian academic nailed it: Miller was just 0.2% short on Warnock's 2.0% margin, and precisely on target in forecasting Ossoff's 1.0% final bulge at the ballot box."

Miller is the author of six textbooks about data science. He also owns Research Publishers LLC, a California company that, in addition to publishing books and periodicals, provides research and consulting, measurement services, event forecasting, and natural language processing solutions.

Tell us about your methodology for forecasting elections.

It has varied over time.

With the 2020 presidential election, I used what could be described as a bottom-up approach.

I had access to the 56 Electoral College prediction markets (i.e. types of electoral bets you can make) on the site PredictIt, and I was able to, through the prices in the prediction markets, come up with an estimate of probability that any one of those markets would go for the Democratic or Republican tickets.

With that information in hand, every hour like a crazy person, I ran a million hypothetical elections in the Electoral College, and out of those, I would take the mean number of electoral votes. And that was my forecast for that hour.

And then I tracked those forecasts using the same methodology with new prices every hour. I tracked those forecasts across time, and in the Methods section of The Virtual Tout website, which explains my methodology more in depth. That plot is shown essentially from the end of September until Election Day in 2020.

How do you determine odds from the betting markets?

They have what's called a contract, which essentially is an opportunity to bet on one candidate or another.

For example, to buy a contract on who's going to win the presidential election, the price quoted might be 56 cents for the Democrat and 48 cents for the Republican. This means to win a dollar — it's actually a little less than a dollar because you have to pay fees — you'd have to pay 56 cents if you are buying the contract on the Democrat, and only 48 cents to buy the contract on the Republican.

This essentially means the chances of the Democrat winning are 56% and the chances of the Republican winning are 48%.

In the forecasting model, we normalize these values so they sum to 100%.

You used a different methodology for the Georgia Senate runoff elections in 2021.

Yes, for those I invented a new technology, called prediction surveys, which is marriage of prediction markets and political polls. I paid people money to participate, and then paid them extra if they were correct in predicting the winner of the election.

I haven't been able to do prediction surveys since they're very expensive — you have to pay people when they're right! And surveys are always expensive, especially if you're going to do a representative sample.

Now this year, I'm doing something different, because the prediction market that I have the highest regard for — PredictIt — does not have as many markets or contracts available.

So, I am watching one in particular: the contract for which party you believe is going to win the presidential election.

And I am using a top-down approach this time. My methodology is based upon the history of 60 years of presidential elections, and the current prices in this prediction market on PredictIt.

So it's a combination of historical and prediction-market information.

Why do you use this methodology instead of polls?

Polls are problematic for a number of reasons.

They use smaller sample sizes — typically 500 to 1,500 people. Their samples vary from one poll to the next — they aren't the same people doing each poll with a particular polling company, so you have variability associated with those different samples. They ask people to respond to a question along the lines of, "If the election were held today, which candidate would you vote for?" It's a personal preference opinion. And because those samples are small, every pollster has to develop a way of weighting those samples so they are more representative
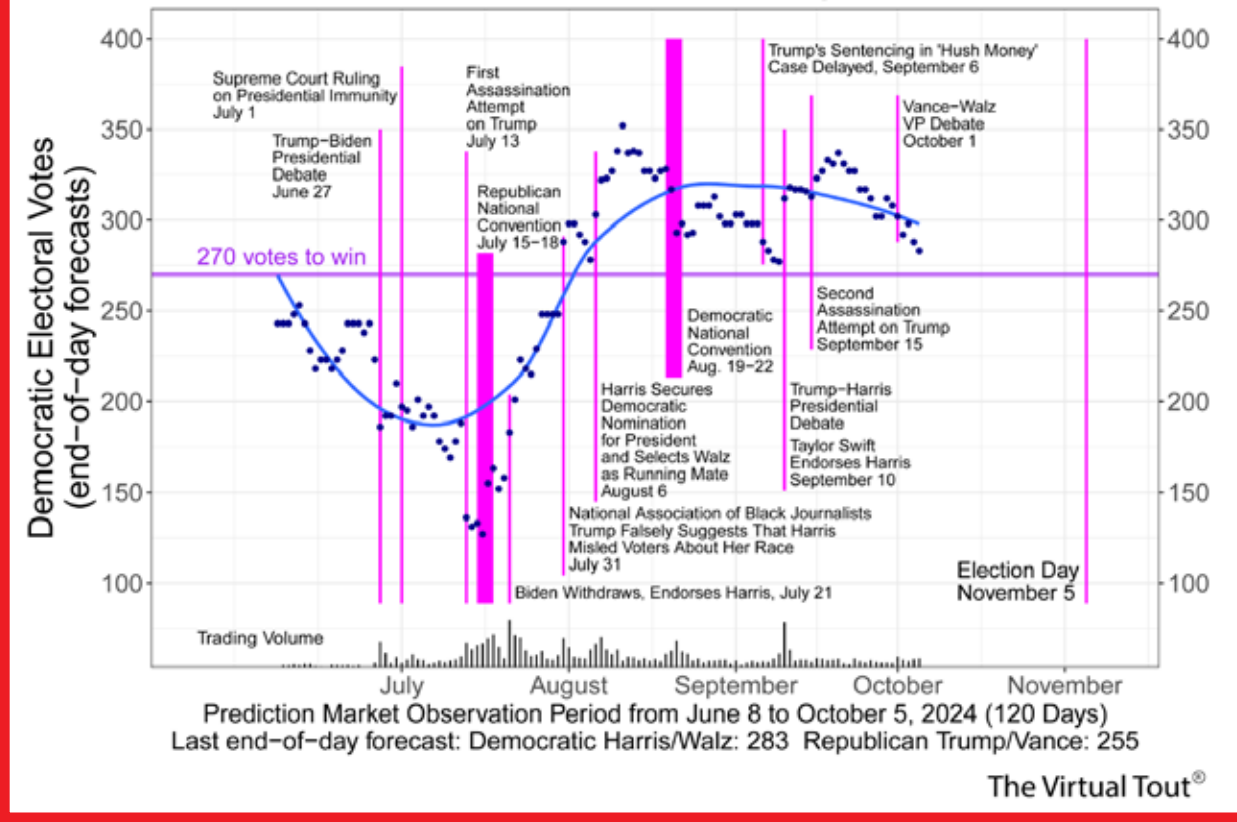
Miller's prediction chart on The Virtual Tout from June 8 to October 5.

of the general voting public. That means every pollster uses his or her own methods for coming up with those percentages that they report. So it is difficult to extract meaning from any individual poll.

If you're being careful about your forecasts, you must take those polls you trust and average them in order to get a forecast for what is going to happen. And even with that, the polls are summarizing what happened at least a few days in the past.

But prediction markets, in contrast to polls, have a fairly constant pool of investors. And on PredictIt, you can buy and sell shares 24/7.

Say, for example, there's a debate between Vance and Walz. You hear Walz talking about Minnesota, he's using colloquial terms, he keeps talking over and over about "folks" and Minnesota and it's driving you crazy because you bet on the Democratic ticket, and you're thinking, "Oh no, we've got a problem here." Well, right in the middle of the debate, you could sell your Democratic shares and buy Republican. Not a lot of people did that, by the way. There was a low trading volume during the debate. But you do have that liquidity, just like you would in the stock market, and you can buy and sell at any time.

So that means the prediction markets are much more responsive in real time (to what is happening in the campaigns) than polls are.

And the prediction markets are not talking about what you want to happen or who you plan on voting for. They're asking what you think will happen, and you have to put your money on it. So they're forward looking, not looking at something a few days old.

How can people accurately predict outcomes better than polls — are they not basing their
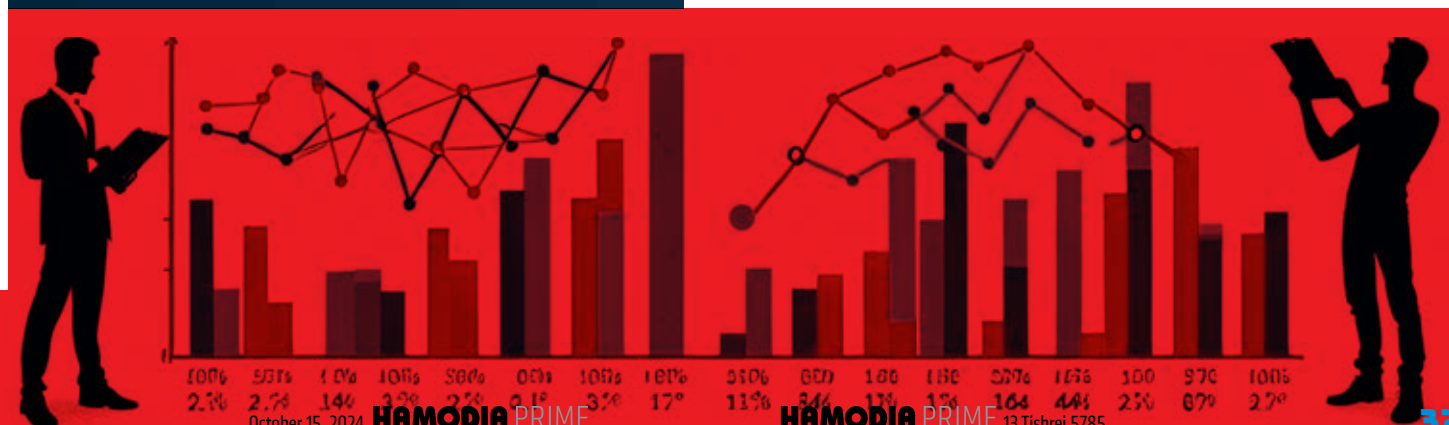
method of predicting political races?

In 2020, I, like everyone else, was stuck inside and had some time on my hands.

Politics and elections have always been interesting to me, and I saw an opportunity to do something different — to use my skills as a data scientist in a way that would be informative and productive, and in a way that other people weren't doing.

I have concerns about the way the media deals with politics. It's not as focused on data and scientific methods as I think it should be.

I'm also planning to introduce a journal called *Data Science Quarterly*. The tagline is, "Following the data, leading with science."

I try to make decisions that make good sense. Utilizing as my primary data source the prediction market, to me, makes sense, because I'm trying to predict the future, not describe the past. But I will be looking at 60 years of presidential election history and be *guided* by the past.

How long have presidential betting markets been around?

Forever. In the Wilson election of 1916, more money was spent on betting than on advertising by the campaigns.

And there's a lot of academic research to indicate that prediction markets are better forecasters than polls.

When Fortune wrote its most recent article about you on September 18, your earlier model showed Harris with 400-plus electoral votes. As we speak today on October 6 — and it might change in the next few days before this interview is published — you have Harris with just 283 votes: still winning, but by a slim margin.

These large swings have happened before: You showed Trump dominating even before his debate with Biden, then

soaring even higher after the debate; when Harris replaced Biden, the Democrats' chances spiked, soaring above 350 votes by the end of the DNC, then dropping to 288 until the debate, when Harris spiked. And now it's close again.

My chart, in addition to showing how many electoral votes each candidate was predicted to receive on a particular date, also shows key events that occurred on those dates.

One key event obviously was the Trump-Biden debate on June 27, which most would agree was a disaster for Biden.

Trump did not do much better, by the way, but Biden just did so poorly, and that led to his withdrawal.

Until Biden dropped out July 21, it was really looking like a Republican landslide. At its low point in mid-July — it was during the RNC and just after the first assassination attempt on Trump — the Democratic candidate was on track to get just 125 electoral votes.

But then Biden withdrew, and things turned around for Democrats, and you can see on my chart when those key events occur. At the very bottom of the chart, you can see the spikes in trading volume. That means the traders who are in this market said, "Oh, I bet on the wrong side, I'm going to sell my Republican shares and buy Democratic."

So it starts moving up after Biden withdraws, and there continues to be high trading.

And then something very important happens that a lot of people haven't paid that much attention to. In my mind, it's the turning point in the campaign: Trump's visit to the National Association of Black Journalists on July 31. The way he behaved in that meeting — the things he said in that meeting, falsely suggesting Harris misled voters about her race, and being insulting to the very capable journalists who were in that meeting — in a sense signaled that this was going to be a campaign about identity and race. And in many ways, it has been. But that signal was very important. There was a notable jump in Harris' favor between July 30 and July 31.

After that date, although there have been ups and downs, everything looks like a Democratic win. That

was the turning point in the campaign.

But there are other things that happened. When Harris officially secured the nomination and chose Tim Walz as her running mate on August 6 there was lot of enthusiasm for her, and you can see the Democratic ticket moving higher. You could even say it was becoming a landslide. Curiously, she did not see the "convention bump" the media often predicts.

Then things started moving down for Democrats. On September 9, the Democratic ticket had just 277 votes on our model — only seven more than needed to win. On September 10, you had the presidential debate and Harris was endorsed by Taylor Swift. Trump blew it in the debate. And you again see a significant jump to over 300 for Harris.

And since then, we've seen some ups and downs, but since mid-September the general trend has been in favor of the Republican ticket. As of now we have Harris with 283.

Why do you think Republicans have been gaining recently?

I wish I knew.

I think underlying all this is that we have a lot of division in this country, and it could well be a toss-up election.

What do you think each candidate can do to improve his/her chances?

Trump needs to be less focused on Harris's identity. To make this election about race or gender is a mistake.

He needs to broaden his base, move beyond white males, and be more open to the diversity of America. He hasn't shown a lot of indication that he will. But if he wants to win, I think he needs to do that.

Harris needs to continue doing what she's been doing: welcoming all citizens regardless of ethnic background and previous political affiliation. And she needs to continue talking about the policies, focusing on the working class. To a large extent, I think she has been successful in executing a winning strategy.

Trump has hurt himself in his messaging, and it shows in the forecasts.

It doesn't seem likely there will be another debate, but if there were one, who would benefit more?

From the one debate they had, it seems Harris is clearly positioned to outperform Trump. She's an excellent communicator in a confrontational, prosecutorial environment, and she can keep the facts straight and present them effectively. So if there is another debate, I would expect it to be a repeat of the first. ∎

---

opinions on polls?

They're basing it on polls and on everything else they know: the war in Ukraine; the situation in the Middle East; candidates' stumbles, mistakes and errors; Trump's court cases; and other prediction markets, if they're watching them. All of that information is going to be part of each investor's individual feeling about what's going to happen.

When you have tens of thousands of people watching everything that's going to happen and putting their money on it, you have a market that's anticipating the future. That's the beauty of it.

Did you ever conduct traditional polls?

Only in 2021 with the Georgia senatorial runoff elections. There I conducted traditional candidate preference polls as well as prediction surveys. I had not conducted polls prior to that, although I was a reviewer of polls going way back. And my doctorate is in psychometrics, so I do know a little about measurement.

What made you look for a new