PEORIA Project
2020 Election Predictions Model Results

*Nevada Caucuses - Democratic Party*

Sanders likeliest to win,
Biden, Buttigieg, and Warren in close race to take a distant second place

Innovative Model Incorporates Social Media Variable
of Twitter Mentions to Yield Ranges of Likely Results
We present two models predicting the outcome of the Nevada Caucuses:

The first, or “basic,” model uses our key three variables to predict a Sanders victory, followed distantly by Biden, Warren, and Buttigieg in a close grouping for second place.

The second model captures the “momentum” of the race, incorporating the results of the New Hampshire primary election. This model makes roughly the same vote share predictions, but with greater error when accounting for the surprising results in New Hampshire (with Biden and Warren underperforming and Buttigieg and Klobuchar overperforming), yielding a slightly different ordering of potential second-place finishers.

Presenting these two models tests for how momentum affects predictions. In this case, given the results in New Hampshire, the concept of momentum introduces a great deal of uncertainty, underscoring the volatility of the 2020 race.
The chart and table report the predicted caucus vote share in Nevada for each candidate. For example, Bernie Sanders is predicted to receive 30.6% of the vote share. The bars indicate the upper and lower bounds for the prediction (95% confidence interval).
The chart and table report the predicted caucus vote share in Nevada for each candidate. For example, Bernie Sanders is predicted to receive 29.2% of the vote share. The bars indicate the upper and lower bounds for the prediction (95% confidence interval).
Our Key Three or Four Variables

Our models predict a candidate’s performance based on three or four factors (depending on the model): Twitter mentions, cash on hand, endorsements, and performance in the last nomination contest.

While we are aware that in important ways the Twitter universe does not necessarily reflect the electorate, the quantity of **Twitter Mentions** is a good proxy for the “buzz” a candidate is getting within the wider electorate, and reflects the activity of important opinion leaders.

**Cash on Hand** reflects the strength of the candidate in the “money primary.”

**Endorsements** indicate each candidate’s strength within the party, which speaks to the debate over whether the party decides the outcome of the nomination.

**Performance in the Last Nomination Contest** is the vote share received in the immediately preceding primary or caucus.
Explanation of Models

What Our Models Do
Our models predict the Nevada caucus vote share for each Democratic candidate using three or four predictor variables generated by an equation estimated through an Ordinary Least-Squares (OLS) multiple regression. See the following pages for equations.

How We Predict Vote Share
In order to predict each candidate’s vote share, we input the latest variable data (see below) into the regression model to generate an estimate as well as an upper- and lower-bound for the predicted performance of each candidate.

Twitter Mentions: Measured as the number of mentions on Twitter for each candidate as a percentage share of the total number of mentions for all candidates within the party. The data for these models were tallied through one month leading up to the week prior to the date of the contest. We focused specifically on those Tweets generated from Nevada. Source for data: Crimson Hexagon.

Cash on Hand: Measured as a percentage share of the total cash on hand for all candidates within the party. The most recent data were for Quarter 4 of 2020. Source for data: FEC.gov

Endorsements: Measured as the total number of endorsements for each candidate by US Senators, members of the US House of Representatives, former Presidents and Vice Presidents, former presidential candidates from the current election cycle who had dropped out of the race, elected statewide officials, state legislative leaders, and mayors of large cities. The data for these models were tallied through one month leading up to the week prior to the date of the contest. Source for data: FiveThirtyEight.com

Performance in the Last Nomination Contest: Measured as each candidate’s share of the total vote within the party in the immediately preceding caucus or primary. For the estimates for Nevada, the immediately preceding contest was the New Hampshire primary election.

How We Chose Our Model
To find the best fitting model, we used campaign data from 2012 and 2016 for the three predictor variables above with Nevada vote share for each year as the dependent variable. Several models were created, including OLS, longitudinal (using Q1 through Q4 cash on hand as well as monthly twitter mentions), lasso, ridge, logistic, partial least squares, and principal component regressions. The models with the lowest RMSE while maintaining the highest possible $R^2$ was chosen for this report (in this case, OLS regression).
Descriptive Table of Variables and Regression Model for Nevada Vote Share: The Basic Model

Means, Standard Deviations, and Correlations

<table>
<thead>
<tr>
<th>Variable</th>
<th>M</th>
<th>SD</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Nevada Twitter Mentions</td>
<td>0.231</td>
<td>0.192</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Cash on Hand</td>
<td>0.206</td>
<td>0.220</td>
<td>0.549</td>
<td></td>
</tr>
<tr>
<td>3. Endorsements</td>
<td>258.000</td>
<td>336.011</td>
<td>0.273</td>
<td>0.601</td>
</tr>
</tbody>
</table>

Summary of Regression of the Nevada Caucus Primary Vote Share Prediction

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>SE B</th>
<th>β</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.00175</td>
<td>0.03816</td>
<td></td>
</tr>
<tr>
<td>Twitter Mentions</td>
<td>0.60029**</td>
<td>0.14831</td>
<td>0.58493</td>
</tr>
<tr>
<td>Cash on Hand</td>
<td>0.40080*</td>
<td>0.15632</td>
<td>0.44577</td>
</tr>
<tr>
<td>Endorsements</td>
<td>0.00003</td>
<td>0.00009</td>
<td>0.04705</td>
</tr>
</tbody>
</table>

adj $R^2 = 0.8262, F(3,9) = 20.02$

$p < 0.05, **p < 0.01$

- The equation representing the model is:
  - Predicted Vote Share = $0.00175 + (0.60029 \times \text{Twitter Mentions}) + (0.40080 \times \text{Cash on Hand}) + (0.00003 \times \text{Endorsements})$
- We can interpret the Twitter coefficient as such: As one candidate increases their share of Twitter by 1%, their vote share is predicted to increase by 0.0060029.
The equation representing the model is:
\[
\text{Predicted Vote Share} = -0.00249 + (0.45204 \times \text{Twitter Mentions}) + (0.30706 \times \text{Cash on Hand}) + (0.00006 \times \text{Endorsements}) + (0.24659 \times \text{NH Vote Share})
\]

We can interpret the Twitter coefficient as such: As one candidate increases their share of Twitter by 1%, their vote share is predicted to increase by .0045204.
Questions may be directed to the authors:

**Meg O’Neill**  
megoneill@gwu.edu  
Data Scientist for The PEORIA Project,  
Graduate School of Political Management,  
The George Washington University

**Todd Belt**  
tbelt@gwu.edu  
Professor and Director of the  
Political Management Master’s Program,  
Graduate School of Political Management  
The George Washington University

**Michael Cornfield**  
corn@gwu.edu  
Associate Professor and Director of The PEORIA Project,  
Graduate School of Political Management  
The George Washington University

**Lara Brown**  
larambrown@gwu.edu  
Associate Professor and Director of the  
Graduate School of Political Management,  
The George Washington University
Appendix: Twitter Data

Republican candidates in red, Democratic candidates in blue.
For more information: The PEORIA Project, and the 2020 Weekly Tweeterboard.
For a discussion of the differences between the social media electorate and the broader electorate, click here.