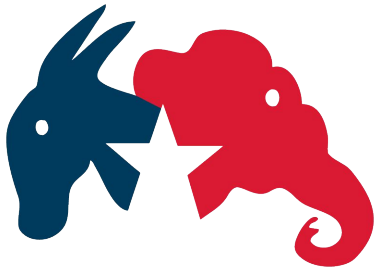
PEORIA Project
2020 Election Predictions Model Results

South Carolina Primary - Democratic Party

Sanders, then Biden likeliest to win

Innovative Model Incorporates Social Media Variable

of Twitter Mentions to Yield Ranges of Likely Results



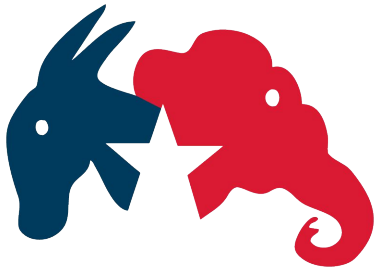
Two Models that Depict the Uncertainty of the SC Primary

GW | GSPM

We present two models predicting the outcome of the South Carolina Primary:

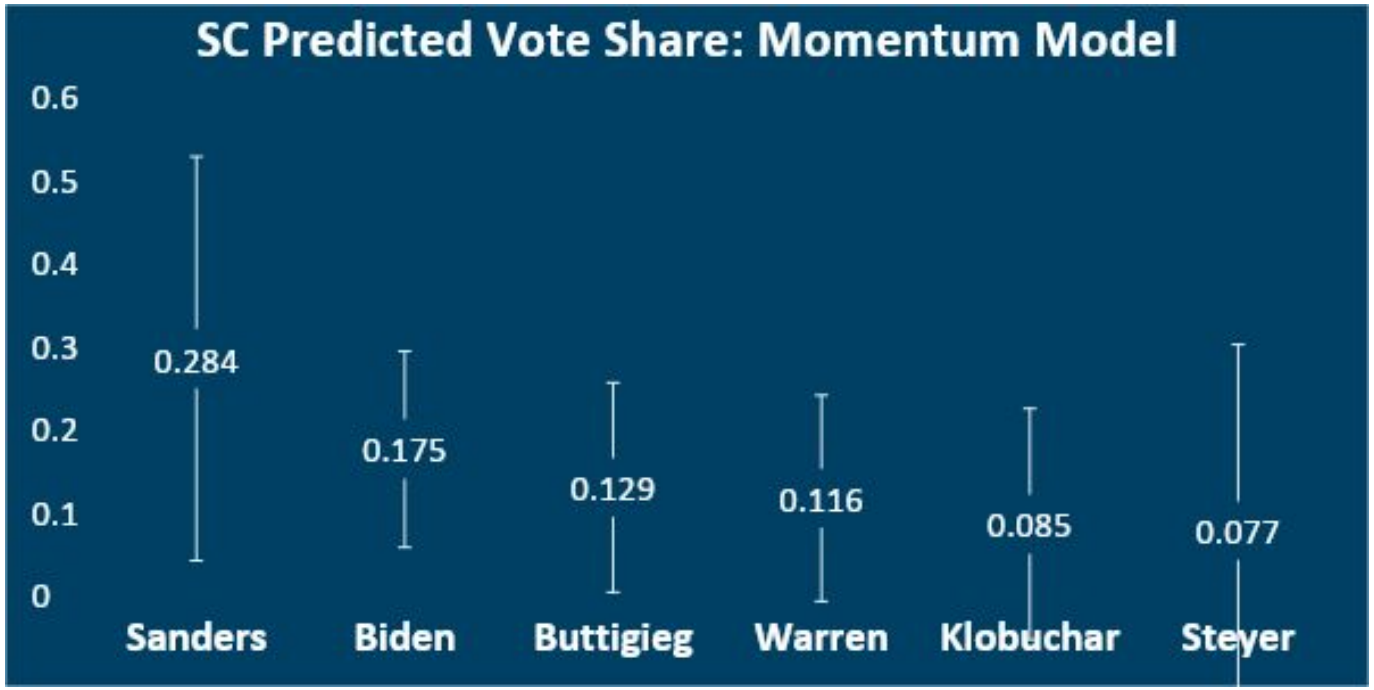
The first model captures the “momentum” of the race, incorporating the results of the most previous primary election. This model predicts a large victory for Sanders over Biden and the rest of the field given Sanders’ strong performance in Nevada. However, there is more uncertainty in the predictions made by this model given how South Carolina often diverges from earlier contests.

The second basic model, without accounting for momentum, demonstrates a closer race based on fundamental factors alone. This model predicts a tighter race between Sanders and Biden, with Warren and Buttigieg close behind.



Predicting SC Vote Share: The Momentum Model

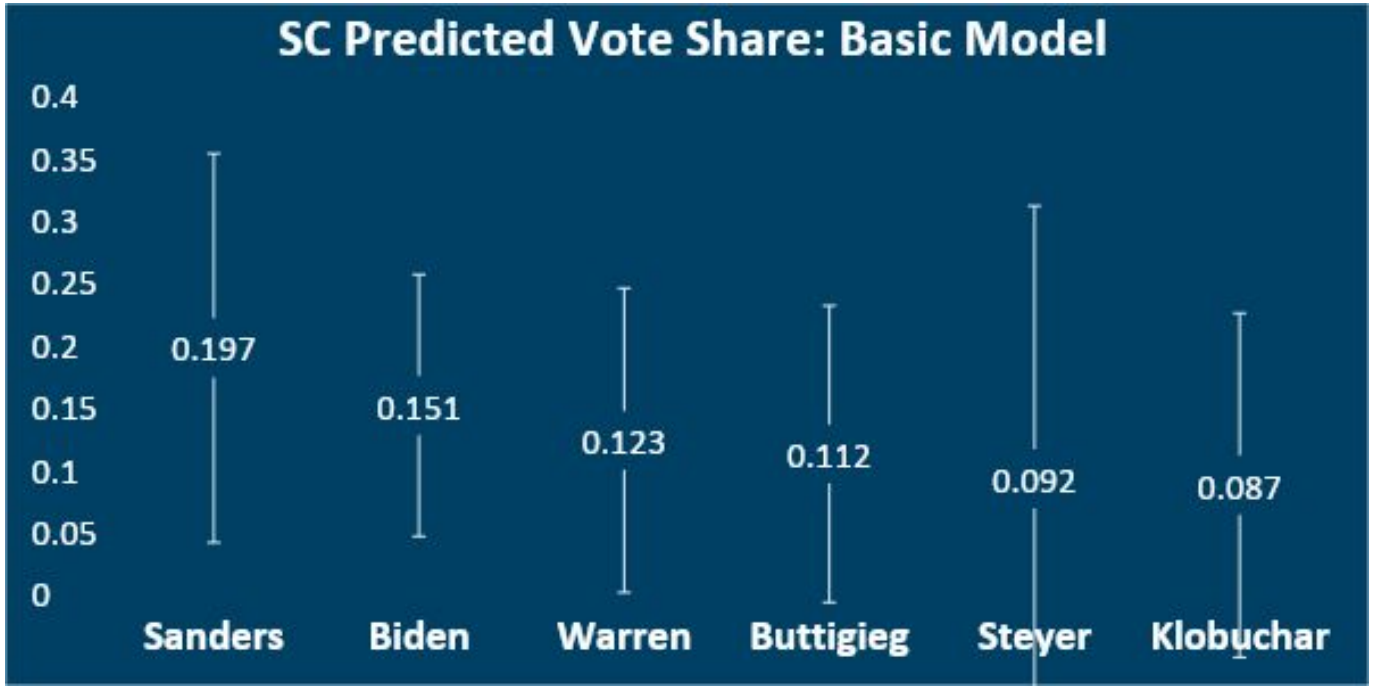
GW | GSPM



Candidate	Average Predicted Vote Share	Lower Bound	Upper Bound
Sanders	**0.284**	0.041	0.528
Biden	**0.175**	0.057	0.292
Buttigieg	**0.129**	0.003	0.255
Warren	**0.116**	-0.008	0.24
Klobuchar	**0.085**	-0.054	0.224
Steyer	**0.077**	-0.147	0.301

The chart and table report the predicted vote share in South Carolina for each candidate. For example, Bernie Sanders is predicted to receive 28.4% of the vote share. The bars indicate the upper and lower bounds for the prediction (95% confidence interval).

Predicting SC Vote Share: The Basic Model



Candidate	Average Predicted Vote Share	Lower Bound	Upper Bound
Sanders	**0.197**	0.041	0.354
Biden	**0.151**	0.046	0.256
Warren	**0.123**	0.001	0.245
Buttigieg	**0.112**	-0.007	0.232
Steyer	**0.092**	-0.127	0.312
Klobuchar	**0.087**	-0.051	0.224

The chart and table report the predicted vote share in South Carolina for each candidate. For example, Bernie Sanders is predicted to receive 19.7% of the vote share. The bars indicate the upper and lower bounds for the prediction (95% confidence interval).

Our Key Three or Four Variables

GW | GSPM

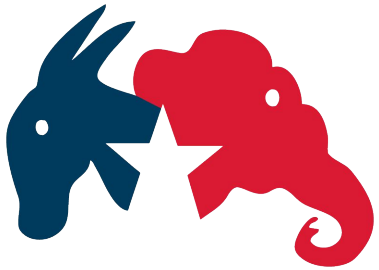
Our models predict a candidate's performance based on three or four factors (depending on the model): Twitter mentions, cash on hand, endorsements, and performance in the last nomination contest.

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

GW | GSPM

What Our Models Do

Our models predicts the South Carolina 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 Tweets originating from South Carolina and tallied through one month leading up to the week prior to the date of the contest. Source for data: Crimson Hexagon.

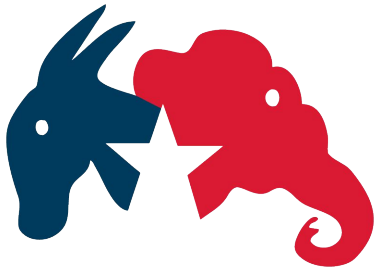
Cash on Hand: Measured as a percentage share of the total cash on hand for all candidates within the party. The most recent data were for January, 2020. Source for data: FEC.gov

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 February 24. 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 South Carolina, the immediately preceding contest was the Nevada caucus for the Democrats and the New Hampshire primary election for the Republicans.

How We Chose Our Model

To find the best fitting model, we used campaign data from 2012 and 2016 for the predictor variables above with South Carolina vote share for each year as the dependent variable. Several models were created, including OLS, longitudinal (using Q1 through Jan 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 were chosen for this report (in this case, OLS regression).



Descriptive Table of Variables and Regression Model for SC Vote Share: The Momentum Model

Means, Standard Deviations, and Correlations

Variable	*M*	*SD*	*1*	*2*	*3*
1. SC Twitter Mentions	0.250	0.194			
2. Cash on Hand	0.250	0.233	0.489		
3. Endorsements	265.583	349.788	0.330	0.739	
4. Last Primary Vote	0.214	0.179	0.802	0.610	0.468

Summary of Regression of the SC Primary Vote Share Prediction with Most Previous Results

Variable	*Estimate*	*SE B*	*β*
Intercept	0.0605	0.0670	
SC Twitter Mentions	0.0218	0.3462	0.0219
Cash on Hand	-0.0165	0.2836	-0.0198
Endorsements	0.0003	0.0002	0.5532
Last Primary Vote Share	0.4597	0.4128	0.4251

adj $R^2 = 0.534$, $F(4,7) = 4.15$

$^*p < 0.05$. $^{**}p < 0.01$.

- The equation representing the model is:
- Predicted Vote Share = 0.0605 + (0.0218 * Twitter Mentions) + (-0.0165 * Cash on Hand) + (0.0003 * Endorsements) + (0.4597 * Most Previous Results)
- 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 .000218.

Descriptive Table of Variables and Regression Model for SC Vote Share: The Basic Model

Means, Standard Deviations, and Correlations

Variable	*M*	*SD*	*1*	*2*
1. SC Twitter Mentions	0.250	0.194		
2. Cash on Hand	0.250	0.233	0.489	
3. Endorsements	265.583	349.788	0.330	0.739

Summary of Regression of the SC Primary Vote Share Prediction

Variable	*Estimate*	*SE B*	*β*
Intercept	0.0634	0.0680	
SC Twitter Mentions	0.3043	0.2391	0.3051
Cash on Hand	0.0619	0.2789	0.0745
Endorsements	0.0003	0.0002	0.5889

adj $R^2 = 0.5201$, $F(3,8) = 4.97$

$^*p < 0.05$. $^{**}p < 0.01$.

- The equation representing the model is:
- Predicted Vote Share = 0.0634+ (0.3043 * Twitter Mentions) + (0.0619 * Cash on Hand) + (0.0003 * 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 0.003043.

Thanks for reading!
Come back each week for new predictions!

GW | GSPM

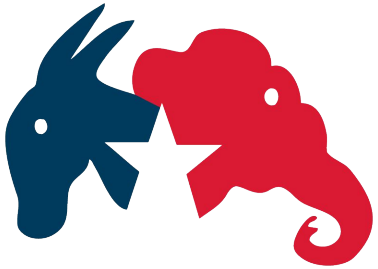
Questions may be directed to the authors:

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

Michael Cornfield
corn@gwu.edu
Associate Professor and Director of 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

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



Twitter Data

2012 Candidate	SC Mentions	Share within Party
Ron Paul	2571	0.480
Mitt Romney	1393	0.260
Rick Santorum	761	0.142
Newt Gingrich	628	0.117
Fred Karger	3	0.001

2016 Candidate	SC Mentions	Share within Party
Donald Trump	8799	0.432
Bernie Sanders	6263	0.525
Hillary Clinton	5660	0.475
Ted Cruz	1737	0.350
Marco Rubio	2469	0.121
Ben Carson	1377	0.068
John Kasich	584	0.029

2020 Candidate	SC Mentions	Share within Party
Donald Trump	22954	0.997
Bernie Sanders	13574	0.356
Joe Biden	7590	0.199
Elizabeth Warren	6164	0.162
Pete Buttigieg	4640	0.122
Michael Bloomberg	2326	0.061
Amy Klobuchar	1852	0.049
Tom Steyer	1092	0.029
Tulsi Gabbard	863	0.023
Bill Weld	65	0.003
Rocky De La Fuente	0	0.000