LANGER RESEARCH ASSOCIATES

SURVEY RESEARCH DESIGN • MANAGEMENT • ANALYSIS

December 2016

MRP Election Projection Modeling

Problematic polling projections in the 2016 election have generated substantial interest in alternative methodologies for increasing the accuracy of election polling and forecasting. Langer Research Associates is offering half-day private workshops on a promising strategy, multilevel regression with poststratification.

MRP has grown in popularity in recent years among researchers seeking to improve survey weighting techniques, particularly in making state-level election forecasts. By combining pre-election polls with Census data and other variables, MRP provides a powerful approach for improving estimates by pooling information across groups. The technique offers a robust alternative to traditional weighting techniques, whose weaknesses may have contributed to polling errors in the 2016 elections.

Our workshops, geared toward applied survey researchers and campaign professionals interested in employing this technique, are based on internal research that employed MRP to predict state-level election results using national tracking poll data. This approach proved highly accurate in predicting state-level results and the national popular vote margin for the 2016 election, exactly predicting the estimated 2-point national margin and correctly forecasting the outcome in 49 of 50 states. In tests of the previous four presidential elections, the model, created by Senior Research Analyst Chad Kiewiet de Jonge, Ph.D, correctly predicts the outcome in 46 states in 2000, 47 in 2004 and 48 in 2008 and 2012 alike.

MRP can be used to generate estimates at the state level (or for other geographical areas) using polling data from individual states, multiple states or national samples. Given sample-size limitations, such samples commonly are weighted using iterative proportional fitting, or raking. MRP improves on this procedure by using additional information across groups to produce more precise estimates. Beyond the modeling applied, data quality is a consideration (our MRP model uses probability-based RDD sampling), as is the availability of empirically based weighting targets.

A description of the method as applied to the 2000-2016 elections follows.

50-State MRP Estimates from National Tracking Polls

In this example, the first stage of the MRP modeling process implements a series of multilevel logistic regression models predicting being a likely voter and supporting an individual candidate.

These models include demographic variables at the respondent level (age, education, race, gender and a series of interactions between these variables) as well as state-level variables to increase precision. For the likely voter models, these state-level variables include battleground state status and previous voter turnout. The vote choice models include the state-level vote in the previous election as well as the size of racial and religious groups in each state.

In the second stage, model estimates from the first stage are used to predict turnout likelihoods and candidate support among the groups defined by the combination of all of the demographic variables selected for inclusion in the model. These predictions are then combined with Census data on the size of each of these groups in each state in order to generate state-level turnout and vote preference estimates.

We have tested this MRP modeling strategy using pre-election national tracking poll data from 2000, 2004, 2008, 2012 and 2016. The few incorrect projections of state winners occur chiefly in states in which the vote was very close or in small states that saw large election-to-election swings; for example, the model predicted victory for Al Gore in Florida and Arkansas in 2000.

Incorrect model projections				
2000	2004	2008	2012	2016
Arkansas	Iowa	Indiana	Florida	Michigan
Florida	New Mexico	Missouri	Ohio	
New Mexico	Wisconsin			
West Virginia				

In terms of electoral vote projections, the model projects the wrong winner in 2000, is extremely accurate in 2004 and 2008, accurate but less precise in 2012, and quite accurate in 2016.

Model vs. Actual Electoral Votes by Election										
	2000		2004		2008		2012		2016	
	Gore	Bush	Kerry	Bush	Obama	McCain	Obama	Romney	Clinton	Trump
Model	297	240	253	284	364	174	285	253	249	289
Actual	266	271	251	286	365	173	332	206	232	306

There's no consistent bias in the estimates across these elections. For 2000 and 2004, the model tilts slightly more toward the Republican candidate than the election results in state level estimates; for 2008 and 2016 toward the Democratic candidate; and there was no partisan lean in 2012.

In terms of correctly predicting the percentage of the vote each candidate received, the model performs best for 2012, with a median error on the Obama-Romney margin of 0.1 points across states. Estimates are somewhat less accurate for the other elections, missing the candidates' final vote shares across states by 1.6 to 3.2 points (median), and missing the Dem.-Rep. margin by 3 to 4.4 points. The root mean squared error (RMSE) on the Dem.-Rep. margin has ranged from 3.5 to 7.2 among all states, and 3.1 to 5.4 among states with sample sizes greater than 100.

	2000	2004	2008	2012	2016
States/D.C. correct (of 51)	47	48	49	49	50
Median D-R bias	-1.2	-0.8	+1.8	+0.1	+1.6
Median absolute errors					
Democratic candidate	2.6	2.2	1.8	1.2	2.0
Republican candidate	1.6	1.8	2.3	1.4	3.2
Dem-Rep margin	3.8	3.0	4.4	2.3	3.3
RMSE					
Democratic candidate	3.1	3.6	3.7	1.9	3.0
Republican candidate	3.3	2.6	3.7	2.0	4.0
Dem-Rep margin	5.6	5.5	7.2	3.5	5.7
RMSE states with n>100					
Democratic candidate	2.8	2.4	2.6	1.8	2.5
Republican candidate	2.5	2.2	3.0	1.8	2.9
Dem-Rep margin	4.4	3.6	5.4	3.1	3.4

For reference, our model's 2016 MRP estimates follow. Contact info@langerresearch.com for additional information and to schedule a customized MRP workshop.

2016 MRP Model Estimates

	Clinton	Trump	Margin		Clinton	Trump	Margin
National	46	44	D +2	Missouri	38	52	R +14
Alabama	36	57	R +21	Montana	36	52	R +16
Alaska	38	47	R +9	Nebraska	33	55	R +22
Arizona	42	48	R +6	Nevada	48	42	D +6
Arkansas	33	60	R +27	New Hampshire	45	43	D +2
California	58	32	D +26	New Jersey	53	37	D +16
Colorado	48	41	D +7	New Mexico	54	38	D +16
Connecticut	52	37	D +15	New York	57	33	D +24
DC	80	13	D +67	North Carolina	43	49	R +6
Delaware	51	40	D +11	North Dakota	33	55	R +22
Florida	45	47	R +2	Ohio	43	47	R +4
Georgia	42	49	R +7	Oklahoma	31	61	R +30
Hawaii	70	21	D +49	Oregon	47	40	D +7
Idaho	28	59	R +31	Pennsylvania	44	46	R +2
Illinois	51	39	D +12	Rhode Island	55	34	D +21
Indiana	37	53	R +16	South Carolina	39	52	R +13
Iowa	43	47	R +4	South Dakota	34	54	R +20
Kansas	34	55	R +21	Tennessee	35	58	R +23
Kentucky	33	59	R +26	Texas	42	50	R +8
Louisiana	38	52	R +14	Utah	31	53	R +22
Maine	47	41	D +6	Vermont	58	30	D +28
Maryland	56	34	D +22	Virginia	47	44	D +3
Massachusetts	55	34	D +21	Washington	50	39	D +11
Michigan	46	45	D +1	West Virginia	29	63	R +34
Minnesota	46	43	D +3	Wisconsin	44	45	R +1
Mississippi	40	52	R +12	Wyoming	25	62	R +37