

Assessment of Wang et al. (2015)

In this paper, the authors base their election forecasts analysis on a non-representative survey dataset. It was an opt-in poll available continuously on the Xbox gaming platform during the 45 days preceding the 2012 US presidential election (Wang et al., 2015, p981). Of the eight variables reported in this investigation, Sex, Age and Education are the three least representative ones of the data. This is because the Xbox is a gaming platform, which is more popular among young men. Women or people over forty years old are less likely to be interested. Besides, college graduates tend to spend more time on research or working and they probably have less leisure time for game. Therefore, there is a huge discrepancy of the ratio of college graduates between the Xbox dataset and the broader voting population. Among the rest five variables, Race, State and 2008 Vote are the most representative.

To perform a post-stratification re-weighting of the respondents, the authors utilize two data sources, which are exit poll data and Xbox survey data. The commonly used Current Population Survey (CPS) is missing some crucial key variables. Thus, exit poll data from the 2008 presidential election is taken advantage of. "Exit polls are conducted outside voting stations on election day, and record the choices of exiting voters." (Wang et al., 2015, p984) One of the disadvantages of using exit poll dataset is that the demographic shifts of the intervening four years between 2008 and 2012 cannot be captured. However, for the sake of simplicity and transparency, the authors did not try to combine CPS data and exit poll data to make analysis (Wang et al., 2015, p984).

The final prediction based on three datasets could be detected from Fig.2 and Fig.3 in the paper. In Fig.2, we can see that during the last three weeks, the daily unadjusted Xbox estimates of the two-party Obama support never exceed 50% line. Thus, the raw Xbox data would definitely suggest a victory for Mitt Romney. However, the line

of the daily aggregated polling results from Pollster.com was really close to the 50% for the last three weeks. During the last several days, there was an increase in the support rate of Obama. Something might have happened on Nov. 1st. Thus, although with huge uncertainty, Pollster.com would have predicted the success of Obama. The line representing post-stratified Xbox data manifests that the final three weeks' estimation was close to the actual outcome, 52%. With the procedures mentioned in the fourth part of the paper for generating election day forecasts, the adjusted Xbox data would also predict Obama's winning.

References

Wang, Wei, David Rothschild, Sharad Goel, and Andrew Gelman, "Forecasting Elections with Non-Representative Polls," *International Journal of Forecasting*, 2015, 31(3), 980-991.