

Estimating Voter Registration Deadline Effects with Web Search Data

Alex Street

Political Science, Carroll College, Helena, MT 59625
e-mail: street.alex@gmail.com (corresponding author)

Thomas A. Murray

Department of Biostatistics, MD Anderson Cancer Center
e-mail: tamurray@mdanderson.org

John Blitzer

Google, Inc., Mountain View, CA
e-mail: blitzer@google.com

Rajan S. Patel

Google, Inc., Mountain View, CA
e-mail: rajan@google.com

Edited by R. Michael Alvarez

Electoral rules have the potential to affect the size and composition of the voting public. Yet scholars disagree over whether requiring voters to register well in advance of Election Day reduces turnout. We present a new approach, using web searches for “voter registration” to measure interest in registering, both before and after registration deadlines for the 2012 U.S. presidential election. Many Americans sought information on “voter registration” even after the deadline in their state had passed. Combining web search data with evidence on the timing of registration for 80 million Americans, we model the relationship between search and registration. Extrapolating this relationship to the post-deadline period, we estimate that an additional 3–4 million Americans would have registered in time to vote, if deadlines had been extended to Election Day. We test our approach by predicting out of sample and with historical data. Web search data provide new opportunities to measure and study information-seeking behavior.

1 Introduction

One in seven Americans eligible to vote in the 2012 presidential election was not registered, and was thus unable to cast a ballot.¹ Every U.S. state, except North Dakota, requires voters to register. The earliest deadlines are currently 1 month in advance of the election. In some states, however, registration remains open, or re-opens to allow Election-Day Registration (EDR). The United States is unusual among democracies in placing the responsibility to register largely on the voter, and in leaving the administration of elections to state and local officials (Powell and Bingham 1986; Jackman 1987).

Authors' note: The authors thank Mike Alvarez and two anonymous reviewers for valuable comments, and Joshua Dyck, Peter Enns, Matt Filner, Alex Kuo, Renee Liu, Philipp Rehm, Steve Scott, Daniel Smith, Nigel Snoad, Seth Stephens-Davidowitz, Hal Varian and seminar participants at Cornell University and Google for helpful suggestions. Replication data are available in Street et al. (2015). Supplementary materials for this article are available on the *Political Analysis* web site.

¹For details on this calculation, see Supplementary Section S.1.

© The Author 2015. Published by Oxford University Press on behalf of the Society for Political Methodology. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

The effects of requiring early registration are disputed. Some scholars argue that obliging citizens to register well before Election Day reduces turnout, by imposing an additional cost on would-be participants and by preventing the mobilization of new voters in the final days of the campaign, when political interest is most intense (Wolfinger and Rosenstone 1980; Teixeira 1992; Burden et al. 2014). Skeptics contend that most people who fail to register have little interest in politics or political participation (Highton 2004; Hanmer 2009). Existing research on the effects of voter registration laws compares turnout across elections covered by different laws. However, laws that facilitate voter registration are easier to pass in time periods or districts where politicians and voters place greater value on electoral participation (Hanmer 2009). This implies that omitted variables may confound estimates of deadline effects that rely on comparing turnout under different election laws.

Here, we present an alternative approach that directly addresses the question of how many people missed the deadline but were nonetheless interested in registering. Specifically, we use the volume of web searches for “register to vote” and related terms as a measure of interest. We show that daily web search volume is closely correlated with the daily number registering, when registration is open. Using Bayesian models of the relationship between search and registration timing, we calculate counterfactual predictions for the number of additional registrations that would have been observed in 2012 if all U.S. states had extended the deadline to Election Day. The estimates rely on the assumption that web search activity is an equally strong indicator of interest in registering in the pre- and post-deadline periods. We assess the credibility of this assumption, and conclude by discussing other ways in which scholars could use web search data to study elections and information-seeking behavior.

2 The Effects of Electoral Rules

Research on electoral rules is motivated, in part, by the concern that incumbents may manipulate the terms of the contest to their own advantage. U.S. election administration is unusually decentralized, which, historically, has left wide scope for the abuse of power (Tokaji 2008; Keyssar 2009). Changes in electoral rules were central to the post-Reconstruction exclusion of African Americans and other minority groups from the franchise (Key 1949; Kousser 1974, 1999), and some scholars suggest that a similar dynamic is at work again today (Bentele and O’Brien 2013). The effects of registration laws have received sustained scholarly interest. Prior research on the effect on turnout of allowing voters to register up to Election Day yields estimates ranging from -2% to $+14\%$ points (see Supplementary Table S1 for summaries of 15 studies). Estimates of the effect of allowing late registration have fallen over time. The mean estimate in publications through the 1990s was that allowing EDR or keeping registration open through Election Day would produce a 6.4% point increase in turnout, but the mean estimate since 2000 is 3% points.

Keele and Minozzi (2013) provide a lucid discussion of the difficulty of estimating the effects of voter registration laws (see also Kousser and Mullin 2007). Early studies used cross-sectional comparisons of state turnout, with the effects of registration laws estimated with a dummy variable for EDR states or a measure of the number of days when registration was closed (Rosenstone and Wolfinger 1978; Nagler 1991). Such estimates identify the effect of requiring early registration only under the assumption that selection into treatment (allowing EDR) is as if random, conditional on observed covariates (Angrist and Pischke 2009, 55). But the selection on observables assumption is dubious, in this case. Other factors, such as norms on the importance of participation, may affect both interest in registering and election laws. Yet these confounders are not observed and can’t be included in the model.

To address this problem, scholars have focused on otherwise similar elections that used different registration laws, such as consecutive elections in a given state (Knack 2001). Difference-in-differences models estimate changes in turnout in states or districts that moved the deadline, while using changes over time in other regions to control for broad temporal trends (Ansolabehere and Konisky 2006; Knee and Green 2011). Keele and Minozzi (2013) employ a regression discontinuity design to compare districts just above and below population thresholds that were used to decide which districts in Minnesota and Wisconsin were obliged to introduce EDR in the 1970s. More

careful research designs may help explain why recent research finds smaller effects of early registration requirements. However, these studies still rest on the assumption that elections held under different rules were otherwise equivalent. As Keele and Minozzi (2013) put it, one's confidence in the results ultimately depends on one's answer to questions such as "How much is Minnesota like Wisconsin?" They argue that more credible research designs yield ever-smaller estimates, and that the most plausible estimate is that EDR has little or no effect. Knee and Green (2011) reach similar conclusions, though they find that EDR has a modest effect on turnout in presidential elections.

Before giving up on the idea that registration laws affect turnout, however, we believe that scholars should draw upon a wider range of measures and methods. We propose to use web search data to measure interest in registering to vote, both before and after the deadline. One advantage of web search data in this context is granularity. These data are available in large quantities across U.S. states and on a daily basis in the period leading up to recent elections. Skeptics have questioned whether the kind of people who miss the deadline are actually interested in registering, but we find substantial last-minute interest.

A growing literature uses web search data to provide up-to-date and localized measures of a range of phenomena, from epidemics of infectious diseases to unemployment claims. Web search data can "predict the present" by providing evidence on time-varying outcomes more quickly than other methods (Ginsberg et al. 2008; Choi and Varian 2012; Varian 2014; but see also Lazer et al. 2014). Scholars have also used web search data to predict consumer behavior into the near future (Goel et al. 2010).

In this article we extend the literature on web search and mass behavior to the electoral domain. We also take on a new methodological task, counterfactual prediction. We ask, how many more people would have registered for the 2012 election, if registration had remained open through Election Day? Since this outcome was not actually observed, there is no definitive answer to this question. Estimates can be made only under certain assumptions. As the literature on web search and mass behavior is quite new, and we are among the first to consider methods for counterfactual predictions in this area (but see Brodersen et al. 2014), it is important for us to be as clear as possible about our approach. To this end, our data and code are published online with the paper (Street et al. 2015).

3 Data

We obtained data on the number of Americans seeking information on voter registration from Google web search logs. These logs are the source of the sample that is publicly available via the Google Trends web site. Using the original source allows us to collect daily data even in small states; not all of these data are available via the Trends web site. Some users issued general queries, while others searched explicitly for voter registration rules in a given state. We therefore chose two generic queries: [voter registration] and [register to vote], and three that referred to state names: [voter registration <state>], [<state>voter registration], and [register to vote <state>]. State names were matched to the state in which the search originated.² Combining several queries yields extra data, while facilitating comparisons with the Google Trends web site, which allows at most five queries (see Supplementary Section S.2).

The five queries were chosen for construct validity, to measure interest in registering. We confirmed that people who entered these queries were more likely to click on official sources of information on how to register (typically web sites run by the state Secretary of State) than on any other link that Google supplied in response to the query. Our measure of search volume is the daily number of times the five queries were issued in each state, which ranged into the millions. To avoid revealing proprietary information, the data were standardized by subtracting the grand mean and dividing by the standard deviation. We truncated the data by setting the lowest 5% of values to zero (with very little effect on our results).

²Google uses proprietary methods for ascertaining the location from which searches originate. The details of such methods are beyond the scope of this article.

We focus on the 67 days leading up to the 2012 election, from 9/1 to 11/6. This period was chosen because prior research shows that many people register in the final weeks before presidential elections (Cain and McCue 1985; Gimpel, Dyck, and Shaw 2007), and also to facilitate replication with data from the Google Trends web site. In almost all states, voter registration closes for some period before the election. In 2012, the median length of time for which registration was closed was 3 weeks, and 11 states allowed EDR. Most states show two peaks in search activity: at the time of the registration deadline, and on the Monday before the election and Election Day itself. States that allowed EDR in 2012 showed a single peak in search activity, on and shortly before Election Day (see Supplementary Figs. S1–S3 for search data in all states).

We also collected voter files from 16 states, yielding the date of registration for 80 million Americans. Our sample was limited by the fact that some states prohibit research with voter files, while others charge high fees (see Supplementary Section S.3). Note that voter files contain the effective date of registration, even for applications that were mailed on the deadline but processed thereafter. Validation against other sources shows that, while they do contain some errors, voter files are accurate sources of evidence on political behavior (McDonald 2007; Ansolabehere and Hersh 2012).

Figure 1 shows search and registration timing in the 16 states for which both kinds of data were available. In each of the panels, the left axis and the black line show the daily number of registrations, in thousands. The right axis and the dashed gray line show standardized search volume. The horizontal axis shows the date, with D standing for the mail and in-person deadlines.³ States that allowed EDR saw high numbers registering on Election Day. In states that did not allow EDR, the highest number of registrations was observed on the day of the deadline. Although some applications for voter registration were processed after the deadline (e.g., for people registering a vehicle), those who registered after the deadline were not eligible to vote in the coming election, and registration rates were thus much lower.

As Fig. 1 reveals, daily web search volume in the weeks leading up to the 2012 election was closely related to the daily number registering. The Spearman's rank correlation between search and registration during the registration period was 0.85 ($n = 742$, $p < 0.01$; see Supplementary Table S2 for details of each state). In the states in our sample that allowed EDR in 2012, at least for the presidential ticket—Alaska, Idaho, Maine, Rhode Island and Wyoming—much of the search and registration activity occurred on Election Day. In other states we also see a spike in search activity on and immediately before Election Day. This suggests that many Americans were interested in registering at the last minute, but were unable to do so.

4 Our Critical Assumption

The strong correlation between web search volume and voter registration totals, when registration was open, as well as the spikes in search activity around Election Day, suggest that web search data can be used to measure the post-deadline potential for extra registrations. To do this, we model the pre-deadline relationship between daily web search and voter registration totals, and use the resulting coefficients and the data on post-deadline searches to create counterfactual predictions of the number of people who would have registered if deadlines had been extended to Election Day. We create prediction intervals (PIs) around these estimates; these differ from confidence intervals in that they account not only for the uncertainty around the values of parameters in the model, but also for the range of outcomes that are consistent with these values.

Following Angrist and Pischke (2009, 14), in observational studies one can think of differences between people affected by a policy, and those not affected, as the sum of the average treatment effect on the treated and selection bias. This bias arises when the units that select into treatment differ from those that do not. One way to remove selection bias is to conduct a randomized experiment, and use the control group to estimate counterfactual outcomes for the treated. In our case, one could compare turnout in states randomly assigned to allow or forbid EDR. But true experiments with election laws are not feasible, and the natural experiments that have been

³Typically, the mail and in-person registration deadlines fall on the same day. A few states allow online registration, with the same deadline as registration by mail.

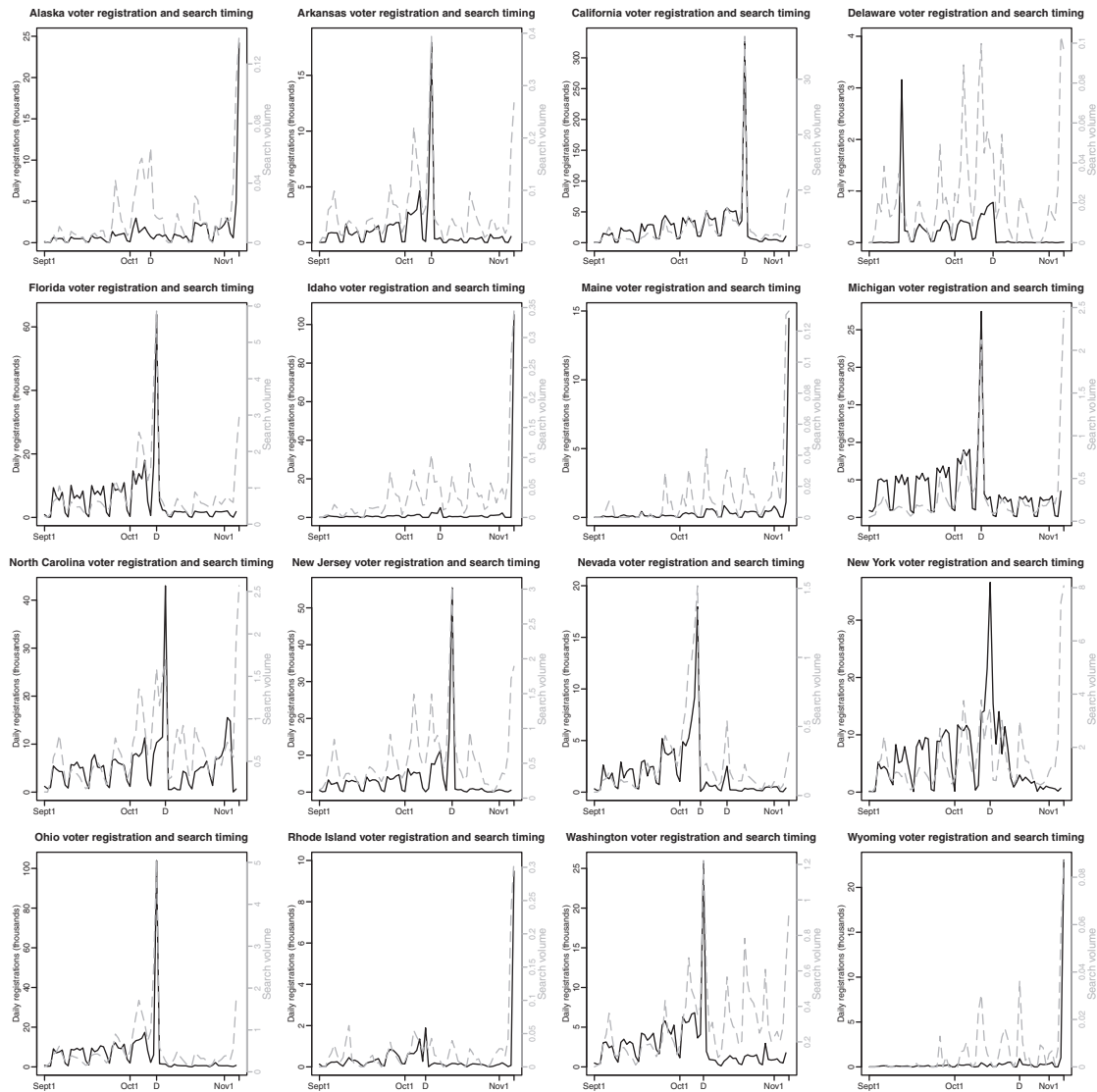


Fig. 1 Web searches for “voter registration” and observed registration numbers, September to November 2012. Black lines and left axes show daily registrations, in thousands. Dashed gray lines and right axes show standardized search volume. Horizontal axes show dates; D marks the mail and in-person registration deadlines (the same day in most states).

suggested in this area are questionable (Keele and Minozzi 2013). States select which election laws to apply, and comparisons across states (or even over time in a given state) risk mistaking the effects of the laws for the reasons behind choices of (or changes in) the laws. Our new measure of post-deadline interest in registering does not preclude other, unobserved differences across states with different deadlines.

Rather than using state-level variation in turnout to estimate the effects of registration laws, we take a different approach. Our identification strategy is to make assumptions that allow us to model the counterfactual outcomes.⁴ Crucially, we assume that queries for “voter registration,” which

⁴In this respect, our approach is similar to recent studies that create “synthetic controls” to estimate counterfactual outcomes for units affected by a given intervention. See Abadie, Diamond, and Hainmueller (2010); Brodersen et al. (2014).

were generated after the deadline, are equally strong evidence of potential for actual registrations as the same queries in the pre-deadline period. Since we fit models on aggregate web search and registration totals on a given day, our key assumption pertains to the conditional distribution of the daily number of registrations (we allow for lower registration volume on weekends, and for deadline effects). More formally, we assume:

$$\begin{aligned} &P(\text{registration volume} \mid \text{search} = \text{srch}, \text{weekend} = w, \text{deadline} = d, \text{before deadline}) \\ &= P(\text{registration volume} \mid \text{search} = \text{srch}, \text{weekend} = w, \text{deadline} = d, \text{after deadline}). \end{aligned} \quad (1)$$

Assumption (1) cannot be directly tested. But we can use indirect evidence to assess whether the assumption is plausible. Post-deadline search activity might actually be *stronger* evidence of an intent to register. Many people participate in electoral politics because they are asked to do so, and voter mobilization is easier when an election is imminent (Rosenstone and Hansen 1993; Verba, Scholzman, and Brady 1995). Alternatively, the relationship between web search and true registration potential in the post-deadline period might be weaker. This would be especially problematic, since it would lead us to over-estimate the effect of requiring early registration.

The core assumption could be violated in two main ways. The first is if the kind of people who sought information on “voter registration,” after the deadline, were systematically different from those who sought the same information beforehand. For example, conscientious citizens may be more likely both to search before the deadline and to actually register (although recent research suggests that conscientiousness is a weak predictor of electoral participation; see Mondak et al. 2010; Gerber et al. 2011; Gallego and Oberski 2012). In order to ensure privacy, Google does not provide evidence based on multiple pieces of data generated by the same user. We thus lack individual-level data on user characteristics. But we can use aggregate data to test for certain patterns that would arise if our key assumption is misguided. One possibility is that people who searched after the deadline were less interested in registering. Perhaps the late searches were generated by people seeking news on the election process, rather than by citizens interested in voting. Or perhaps the queries were generated by people who were already registered, and were trying to find their polling places. We guard against these possibilities by using data on the links most often chosen, after the relevant queries had been entered. Specifically, we test whether people who searched after the deadline were any less likely to click through to the official (Secretary of State) web sites with information on how to register.⁵

The second main way in which assumption (1) could be violated is due to contextual differences. The states in our sample were selected based on the availability of registration data. Note that the sample in which we observed both search and registration data includes one state where in-person registration did not close (Maine), two states with in-person registration deadlines only a few days before the election (North Carolina and Washington), and several states that allowed EDR. As such, we are able to use observed outcomes from the final days of the campaign, as well as the data from September and early October. The sample is diverse in terms of population size, the tendency to support Democrats or Republicans, and the competitiveness of the 2012 presidential race. Nonetheless, the sample may not be representative of the entire country. To test our ability to predict beyond this sample, we successively hold out each state for which we have data, and re-estimate the model. We use the resulting coefficients, along with search data from the state that was held out, to “predict” voter registration numbers in the held-out state. Finally, we test whether the observed number of registered voters in the held-out state, during the period when registration was open, is within the PI. This cross-validation exercise tests our ability to predict out of sample, and allows us to confirm that no single state is driving our results. Of course, we can only check our predictions against observed outcomes. We have no such baseline with which to compare our counterfactual, post-deadline predictions.

⁵Comparing click-through rates before and after the deadline is a hard test. Some users may have seen, from the brief description that Google provides with each suggested link, that the deadline had already passed, without having to click on the link. For further discussion of state-level correlates of the search data, see Supplementary Section S.6.

Another concern is the activities of other groups, besides the people who generated the web search data. One can think of this as a “general equilibrium” issue. Political parties and civic groups, such as churches, the League of Women Voters, or the NAACP, play an important role in registering voters (Green and Gerber 2008; Herron and Smith 2012, 2013). If the deadline shifted to Election Day, we expect that third-party registration drives, which climax in the period immediately before the deadline, would shift to the later date. Our response to this issue is to include in our models not only web search volume but also indicator variables to capture deadline effects. The models include at most one indicator for each kind of deadline (mail, in-person, or online) for each state, on the day when the deadline fell in 2012. Our counterfactual post-deadline predictions are therefore based only on the raw relationship between search activity and registrations, with no additional deadline effects. We see this as a conservative approach, since mobilization efforts on or immediately before Election Day, when media coverage and interest in the election peaks, might be even more effective than similar efforts 3 or 4 weeks earlier. In addition, we compare search activity in “safe” and “battleground” states. If third parties focus their registration efforts on the more competitive states, and succeed in registering most people interested in voting in those states, then it would be inappropriate to treat post-deadline searches in less competitive states as strong signals of extra registration potential.

Besides this evidence on the credibility of our key assumption, we also conduct a general test of our ability to predict post-deadline behavior based solely on the pre-deadline relationship between search and registration timing. We use coefficients from a model of the relationship between web search and registration numbers in Iowa in 2004, when registration closed 10 days before the election, to predict registrations in the state in the final weeks of the election campaign in 2008 and 2012, when Iowa allowed EDR. We compare the predictions (and PIs) to the outcomes actually observed in 2008 and 2012. Finally, we conduct a sensitivity analysis to show how violations of our key assumption would affect our results.

5 Estimation

We model the number of people who registered as voters in each state on each day as a function of daily web search activity in that state, in the 67-day period leading up to the 2012 election. Voter registration was restricted in the period after the deadline, so we treat these observations as missing. Although we have searched data from all states for the entire period, we rely on registration data from a sample of states. Thus, the outcome variable is also missing for the entire period in many states. In order to handle these missing data, we estimate the relationship between daily search activity and daily registrations using fully Bayesian models. These models allow us to calculate posterior predictive distributions for every unobserved value, based on the observed relationships. The predictive distributions reflect the uncertainty around each parameter, given the variation in the observed data.

The patterns in search and registration timing, evident in Fig. 1, support the suspicion of Briars and Grofman (2001) that the final days of the campaign are especially important. One-quarter of all search activity in the 10-week period leading up to the 2012 election was observed in the final 2 days. We therefore estimate the total number of post-deadline registrations through Election Day, rather than some subset of this period. Our counterfactual estimate of additional registrations applies only to states that did not allow EDR in 2012.

Bayesian estimation proceeds in two steps (Carlin and Louis 2009). First, we posit models for the likelihood of the data. Second, we specify prior distributions for the unknown parameters in the models. Our models of the likelihood are designed to reflect various features of the data. The outcome of interest is measured as count data. To allow for over-dispersion, we use a Poisson-gamma mixture formulation of the negative binomial distribution (Zeger 1988). Formally, we model

$$Y_{s,t} | \lambda_{s,t} \sim \mathcal{P}(\lambda_{s,t}), \quad \lambda_{s,t} | \mu_{s,t}, \kappa \sim \mathcal{G}\left(\kappa, \frac{\kappa}{\mu_{s,t}}\right), \quad (2)$$

where $\mathcal{P}(\lambda)$ denotes a Poisson distribution with mean λ , $\mathcal{G}(\alpha, \beta)$ denotes a gamma distribution with

mean α/β and variance α/β^2 , and $Y_{s,t}$ denotes the registration count in state $s = 1, \dots, 50$ on day $t = 1, \dots, 67$. Integrating over $\lambda_{s,t}$, we see that this hierarchical structure implies $E[Y_{s,t}|\mu_{s,t}, \kappa] = \mu_{s,t}$ and $\text{Var}[Y_{s,t}|\mu_{s,t}, \kappa] = \mu_{s,t} + \mu_{s,t}^2/\kappa$. Thus, $\mu_{s,t}$ denotes the expected number of registrations in state s on day t , and κ measures over-dispersion.

Figure 1 shows less web search activity on weekends, and more on registration deadlines. Nonetheless, it is possible that the variation in search activity does not fully reflect the impact of these temporal features of the data. To allow for this possibility, the regression components of our models contain indicators for whether day t was a weekend (w_t), and whether in state s day t was the mail ($md_{s,t}$), in-person ($pd_{s,t}$), or online ($od_{s,t}$) registration deadline, plus an indicator for Election Day in states that allowed EDR ($ed_{s,t}$).

Another concern is autocorrelation. Even after accounting for search volume and the other covariates, outcomes on successive days in a given state may be correlated due to unobserved time-dependent variables. The Durbin–Watson test in a linear model yields evidence of moderate first-order autocorrelation ($DW = 1.45$, $p < 0.01$). In addition to models with the standard assumption of independent errors, we thus fit models with an autocorrelation structure. Finally, in the absence of prior research on the functional form linking web search volume and electoral behavior, we allow for non-linear effects using a flexible spline term (Ruppert, Wand, and Carroll 2003). We specify models in which the measure of web searches in state s on day t ($srch_{s,t}$) enters as a linear predictor (equation (3)), or is modeled using the flexible spline term (equation (4)):

$$\eta_{s,t} = \alpha_0 + \alpha_w w_t + \alpha_{md} md_{s,t} + \alpha_{pd} pd_{s,t} + \alpha_{od} od_{s,t} + \alpha_{ed} ed_{s,t} + \alpha_{srch} srch_{s,t} \quad (3)$$

$$\eta_{s,t} = \alpha_0 + \alpha_w w_t + \alpha_{md} md_{s,t} + \alpha_{pd} pd_{s,t} + \alpha_{od} od_{s,t} + \alpha_{ed} ed_{s,t} + f(srch_{s,t}; \beta),$$

$$\text{where } f(srch_{s,t}; \beta) = \beta_1 srch_{s,t} + \sum_{j=1}^J \beta_{j+1} \left(|srch_{s,t} - \widetilde{srch}_j|^3 - |\widetilde{srch}_j|^3 \right) = \mathbf{z}_{s,t}' \beta. \quad (4)$$

We model $f(srch; \beta)$ with modified low-rank thin plate splines, which have been shown to exhibit better Markov chain Monte Carlo properties than other spline formulations (Crainiceanu et al. 2005). In equation (4), J is the number of knots and \widetilde{srch}_j , $j = 1, \dots, J$ are the knot locations (Ruppert, Wand, and Carroll 2003). We opt to use $J = 15$ knots at equally spaced quantiles of the Google search volume observed in all 50 states over the study period (and obtain similar results with more or fewer knots).

For models with independent errors, we simply estimate $\log(\mu_{s,t}) = \eta_{s,t}$. To model autocorrelation, we assume that $\log(\mu_{s,1}) = \eta_{s,1}$, and that

$$\log(\mu_{s,t}) = \eta_{s,t} + \rho \{ \log(\lambda_{s,t-1}) - \eta_{s,t-1} \}, \quad (5)$$

for $t = 2, \dots, 67$, $s = 1, \dots, 50$. The latter term in equation (5) denotes the latent residual from the previous day, and $\rho \in [-1, 1]$ measures the correlation between adjacent latent residuals (Hay and Pettitt 2001).

To complete the Bayesian model specification, we use vague priors for the parameters in the likelihood. For the parameters indicating weekends and the various deadlines (α 's), we use vague $\mathcal{N}(0, 10^5)$ priors, where $\mathcal{N}(\nu, \sigma^2)$ denotes a Gaussian distribution with mean ν and variance σ^2 . For the relationship between registration numbers and Google search volume (β 's), we specify the low-rank thin plate spline prior detailed in Crainiceanu et al. (2005). We model the remaining parameters with vague priors as follows: $\sigma_\beta \sim \mathcal{U}(0.01, 100)$, $\kappa \sim \mathcal{G}(0.01, 100)$ and $\rho \sim \mathcal{U}(-0.99, 0.99)$, where $\mathcal{U}(l, u)$ denotes a uniform distribution on $[l, u]$. To estimate the posteriors of all the parameters and the predictive posteriors for the unobserved $Y_{s,t}$'s, we use a Gibbs sampler built by JAGS (Plummer, 2003). We assessed convergence empirically using potential scale reduction factors (i.e., the Gelman and Rubin [1992] diagnostic, known as “R-hat”), and visually with trace plots (Supplementary Fig. S6). We found all the parameters to have an R-hat value of less than 1.1, which is indicative of convergence, and the trace plots show good mixing. The code for these models and the convergence assessment is included with the replication data (Street et al. 2015).

6 Results

Table 1 presents results from four models, with the linear and spline functional forms, and with independent and autoregressive errors. The fit of the models is good, explaining around 75% of the deviance. We prefer the spline model with autoregressive errors, due to the evidence of autocorrelation (Supplementary Fig. S4 shows the spline estimate of the relationship between search and registration numbers). Notably, all four models yield similar estimates and fit (on the DIC and pD, see Spiegelhalter et al. 2002). Table 1 shows coefficients for the deadline indicators, which are multiplicative because the models were fit using the log link. As expected, the coefficients show that search activity on deadlines predicted considerably more registrations than on other days. For example, a given level of search activity on Election Day, in an EDR state, was associated with around 11 times as many registrations as the same level of activity on a non-deadline weekday. To aid interpretation of the models, Table 1 presents predicted numbers of registrants at varying levels of search activity, from the first to the 99th percentile. For instance, we estimate that a non-deadline weekday at the 90th percentile of search activity saw around 10,000 new registrations in the relevant state.

We used the models to calculate the posterior predictive distributions of registrations in the period from the deadline to Election Day in each state. The total prediction from each model is reported toward the bottom of Table 1. Summing across states that did not allow EDR in 2012, our models suggest that around 3.5 million people would have registered in the post-deadline period, if this had been possible (these results are reported separately by state in Supplementary Table S3). This would have added 2% points to the total number registered nationwide. High turnout among late registrants, and full turnout among those who register on Election Day, implies that 80% or more of these people would have voted, producing a 3% point increase in turnout (see Supplementary Section S.4 for details on turnout among late registrants). In order to test whether our results were driven by the specifications described above, we also fit an array of different models, including a Poisson-normal mixture, and linear models with random slopes or intercepts by state. The estimates from these models were consistently between 3 and 4.5 million total additional registrants across the country (see Supplementary Section S.5 for more on alternative specifications).

7 Evaluating our Critical Assumption

We now report evidence on the plausibility of our assumption that web searches for “register to vote” (and similar terms) provide an equally valid measure of voter registration potential in the pre- and post-deadline periods. We begin with the possibility that this assumption is violated because different kinds of people entered such a query before versus after the deadline. Among the people who entered our five queries, the official web site with information on how to register was the most-clicked link in over 90% of the days in our sample period, and in many states was the most-clicked link on every single day. Even on Election Day, very few of the people who searched for “voter registration” appear to have had other intentions, such as checking their registration status or finding their polling place (see Supplementary Section S.2). Nonetheless, we saw some differences, after the deadline passed. In 19 states, the mean daily click-through rate to the relevant web sites was significantly lower in the post-deadline period. We found no significant difference between pre- and post-deadline click-through rates in 15 states, and found that click-through rates were actually higher, after registration closed, in a further 12 states (we used the conventional $p < 0.05$ threshold; Supplementary Fig. S5 shows trends in click-through rates, before and after the deadline).⁶ In the median state, the click-through rate after the deadline was one-third of a standard deviation lower than before the deadline. While the differences are not dramatic, they suggest that our critical assumption may not hold, and in the next section of the paper we assess the sensitivity of our

⁶We ran this test in 46 states. Data were missing for New Mexico and the District of Columbia. North Dakota does not require voters to register, and in Maine in-person registration does not close, precluding the pre-/post-deadline comparison.

Table 1 Coefficient estimates, predictions, and fit statistics from four models

| <i>Model</i> | <i>Linear</i> | <i>Linear AR(1)</i> | <i>Spline</i> | <i>Spline AR(1)</i> |
|----------------------------|--|-------------------------|-------------------------|-------------------------|
| Indicator variable | Multiplicative coefficients (95% CI) | | | |
| Weekend | 0.28 (0.24, 0.32) | 0.24 (0.20, 0.27) | 0.28 (0.24, 0.32) | 0.24 (0.21, 0.27) |
| Walk-in deadline | 1.92 (0.92, 3.73) | 1.82 (0.92, 3.42) | 2.04 (1.02, 3.83) | 1.99 (1.02, 3.66) |
| Mail-in deadline | 1.78 (0.76, 3.61) | 1.84 (0.84, 3.60) | 1.52 (0.70, 2.97) | 1.63 (0.76, 3.11) |
| Online deadline | 1.82 (0.41, 5.79) | 2.26 (0.52, 7.31) | 2.80 (0.65, 8.91) | 3.38 (0.78, 10.86) |
| Election Day | 13.43 (5.78, 30.11) | 13.33 (5.62, 29.71) | 11.09 (4.83, 24.70) | 11.41 (4.87, 25.17) |
| % of search volume | Predicted number of registrants (95% CI) | | | |
| 1% | 184 (161, 210) | 244 (204, 296) | 115 (96, 139) | 154 (122, 197) |
| 10% | 287 (256, 323) | 373 (318, 443) | 257 (229, 289) | 329 (280, 390) |
| 25% | 660 (603, 725) | 822 (724, 945) | 848 (739, 976) | 1004 (848, 1201) |
| 50% | 2001 (1849, 2168) | 2357 (2115, 2641) | 2709 (2335, 3233) | 2906 (2472, 3490) |
| 75% | 4794 (4390, 5250) | 5407 (4829, 6108) | 4951 (4330, 5656) | 5405 (4600, 6317) |
| 90% | 9917 (8921, 11,076) | 10,786 (9439, 12,458) | 8415 (7066, 9798) | 9635 (8031, 11,405) |
| 99% | 38,271 (32,852, 44,848) | 38,921 (32,090, 47,470) | 31,683 (25,192, 40,703) | 31,195 (23,982, 41,613) |
| Total prediction, millions | 3.89 (3.40, 4.51) | 3.67 (3.11, 4.35) | 3.66 (3.24, 4.17) | 3.49 (3.03, 4.05) |
| DIC (pD) | 7888 (769) | 7914 (780) | 7898 (777) | 7923 (790) |

findings to this possibility. At the aggregate level, we found no evidence that post-deadline searches were more common in states with less competitive elections (see Supplementary Section S.6).

The states in our sample were chosen based on the availability of registration data. To test our ability to predict beyond this sample, we successively held out each state for which we have data, and calculated new predictions. We then tested whether the observed number of registered voters in the held-out state, during the period when registration was open, was within the PI. Observed registrations were within the 90% PI in 14 of the 16 states in our sample. The exceptions were Delaware and New York, where the observed numbers were at the 0.2 and 0.7 percentile of predictions, respectively. This is higher than the expected number of extreme results, but may (partly) reflect peculiarities in the registration records.⁷ Figure 2 illustrates the ability of a model, fit on data that excluded Arkansas, to predict daily registration numbers in that state. The black line shows observed registrations, the dashed gray line shows search volume, and the dotted gray line with points shows predictions. To the nearest thousand, the predicted number of pre-deadline registrations was 59,000 (90% PI 35,000–96,000), and the observed number was 62,000. Overall, this cross-validation exercise suggests that our approach is moderately robust to the inclusion of some states, but not others, in the sample. Of course, this does not guarantee that our post-deadline, counterfactual predictions were equally accurate.

Finally, we conducted a general test of our ability to predict post-deadline behavior based solely on the pre-deadline relationship between search and registration timing. To do so, we used historical data from a state that recently changed its voter registration rules. Iowa allowed EDR in 2008 and 2012, but not in 2004. We fit a model to web search and registration data in Iowa in 2004, and used the coefficients to predict from search to registration in Iowa in 2008 and 2012.⁸ The models were similar to those used for our counterfactual predictions, except that for the purpose of predicting the timing of registration we moved the deadline indicator to Election Day in 2008 and 2012 (we still used only one deadline indicator in the predictions for each year).

Search data from Iowa in each year were normalized against all other states in the same time period, and standardized using the procedure described above. Transforming the data in this way is necessary to control for temporal trends in web search volume, and for trends in the composition of internet users, though it does not rule out the possibility that web search, or the habits of search engine users, changed in unusual ways in Iowa. We compared the resulting predictions against the observed registration totals. The predictions were reasonably accurate. To the nearest thousand, we predicted 148,000 registrations from September to November 2008, and observed 103,000 (the observed value fell at the 3rd percentile of predictions, that is, slightly outside the 90% PI). The fact that EDR was new to Iowa in 2008 may help explain the high search volume and our over-prediction. In the same period in 2012, we predicted 100,000 registrations and observed 128,000 (the observed value was around the 84th percentile of predictions). Figures 3 and 4 display our results for the final weeks of the campaigns. In each figure, the black line (and left axis) shows observed registrations, while the dashed gray line (and right axis) shows search volume. The dotted gray line with points shows predictions, and the shaded area shows 90% PIs. Of all our cross-validation exercises, the Iowa data provide the best evidence on our approach of modeling post-deadline behavior based on pre-deadline data, since we are able to compare the predictions to actual outcomes. One limitation of the analysis, of course, is that we only have data from one state that recently changed its registration deadline (data from Montana, which also recently introduced EDR, were not available), and we can't rule out the possibility that Iowa is atypical.

⁷In Delaware, registration was closed in early September, and numbers jumped when it re-opened. Delaware was atypical in that the state saw no spike in registrations on the day of the deadline. In New York, about 70,000 voters were recorded as registering by mail in the week after the deadline, but also as having voted on November 6, 2012, suggesting that they did in fact register by the deadline. New York was the only state in our sample to show such large numbers of people as having registered in the week after the deadline. Officials in New York and other states assured us that the dates in voter registration files refer to the date of submission rather than the date of processing. Nonetheless, it is possible that errors occurred. Excluding New York and Delaware made little difference to our overall results.

⁸Registration totals were provided by Iowa Secretary of State officials; see Supplementary Section S.7.

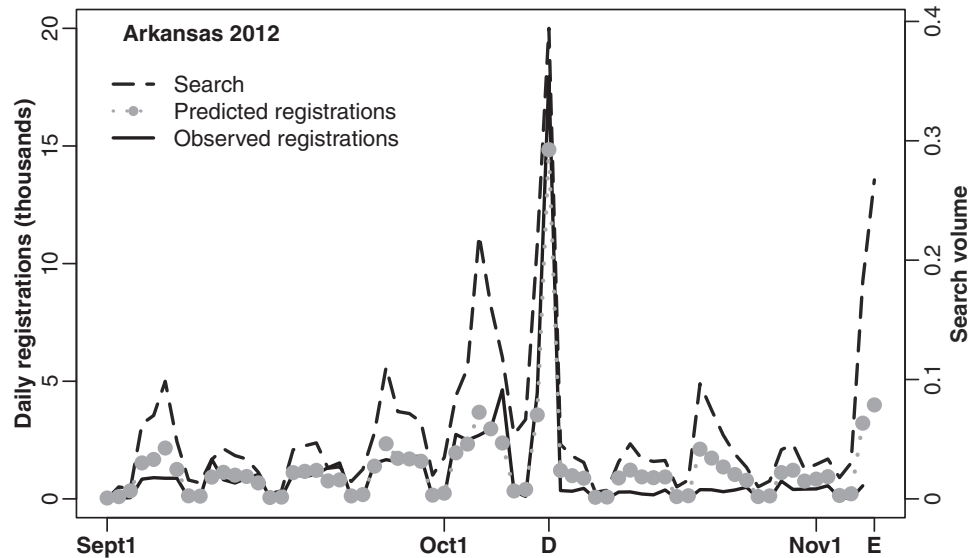


Fig. 2 Predicting voter registration from search activity in Arkansas, using a model fit on data from other states. The black line and left axis show daily registrations, in thousands, the dashed gray line and right axis show search volume, and the dotted gray line with points shows predicted registrations. D shows the mail and in-person registration deadline, and E marks the date of the election.

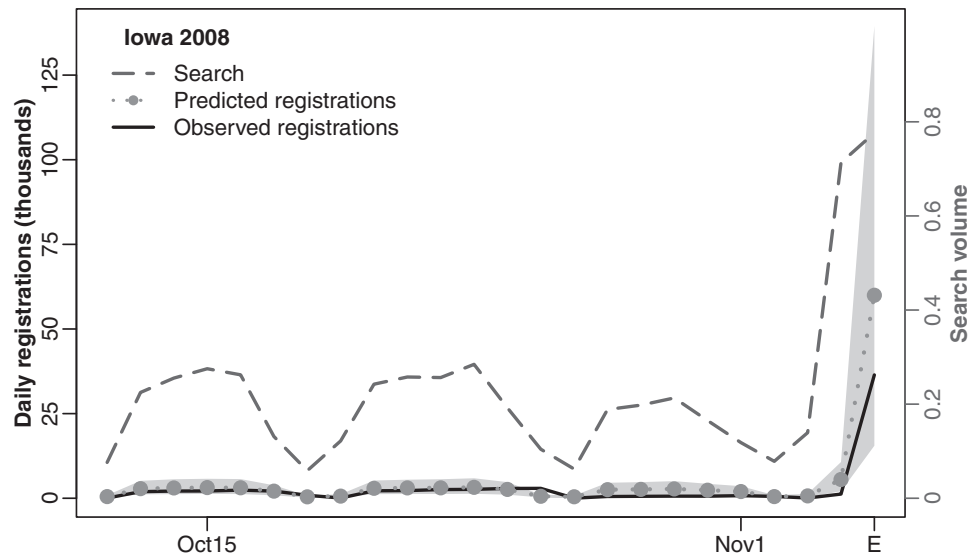


Fig. 3 Predicting voter registration in Iowa in 2008, using the relationship between search and registration estimated in 2004. Shaded gray areas show 90% PIs.

8 Sensitivity Analysis

We now assess how violations of our critical assumption would affect our results. We report how different the relationship between web search and voter registration would have to be, in the post-deadline period, in order to produce substantively different outcomes. To do this, we allow for a post-deadline main effect, and calculate which values of this effect would be needed to yield predictions ranging from 1 million to 6 million additional registrants. We assume the same functional form as in our core results and again use the log link, so that the model can be summarized as $E[Y | \text{search volume} = \text{srch, after deadline}] = \exp(f(\text{srch}))$. However, we now allow for the

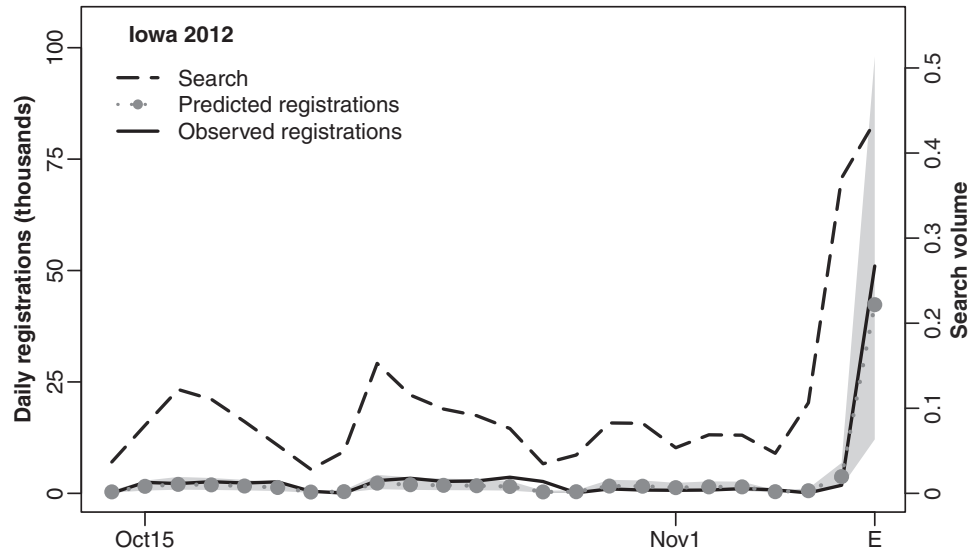


Fig. 4 Predicting voter registration in Iowa in 2012, using the relationship between search and registration estimated in 2004. Shaded gray areas show 90% PIs.

Table 2 Sensitivity of the predicted number of additional registrants to hypothetical post-deadline effects

| <i>Expected number of additional registrants (million)</i> | $\exp(\alpha_{post})$ |
|--|-----------------------|
| 1 | 0.286 |
| 2 | 0.571 |
| 3 | 0.857 |
| 3.5 | 1 |
| 4 | 1.143 |
| 5 | 1.429 |
| 6 | 1.714 |

alternative that $E[Y | \text{search volume} = srch, \text{after deadline}] = \exp\{\alpha_{post} + f(srch)\}$, where α_{post} denotes the post-deadline main effect. We cannot estimate α_{post} because we do not observe unrestricted registration activity after the deadline, but we can calculate how a range of values of α_{post} affect our predictions. Table 2 shows the results; we take the exponent of α_{post} in order to report values on a linear scale.

As Table 2 shows, in order for our estimate to fall from 3.5 million to 1 million, post-deadline search activity would have to be indicative of around 70% fewer registrations, compared with searches for the same terms in the pre-deadline period. If the relationship were only around half as strong, we would still predict an additional 2 million late registrants. In contrast, if searches on and immediately before Election Day indicated higher registration potential, for example, 40% higher, our prediction would rise to 5 million people, enough to add 4% to the electorate (equal to Obama's margin of victory over Romney in the popular vote). These calculations do not account for all possibilities. The relationship between search and registration might change in more complex ways, modifying $f(srch)$. But this simple exercise conveys some implications and limitations of our approach.

9 Discussion and Conclusions

Web search data provide insights into citizens' interests and intentions. In this article we measured web queries about voter registration, as Election Day approached, in order to estimate pre- and

post-deadline interest in registering. In 2012, millions of Americans searched online for information on registering to vote, after the deadline in their state had already passed. Our results suggest that extending registration deadlines to Election Day would have allowed 3–4 million more Americans to register to vote.

Given the limits of any single study, our results should be interpreted in light of other research. In this spirit, it is striking that our estimate of a 3% increase in turnout is similar to findings from over-time comparisons in districts that introduced EDR (e.g., Ansolabehere and Konisky 2006; Knee and Green 2011; Neiheisel and Burden 2012). Without randomization to lend credence to the claim that some units can be used to estimate counterfactual outcomes for others, all identification strategies are fragile. But our approach of modeling the counterfactual relies on different assumptions than prior research on the effects of voter registration rules. The fact that we obtain similar results with different methods should add to our confidence that extending registration to Election Day would allow significantly more people to vote—albeit not as many as some advocates of easier registration hope (e.g., Piven and Cloward 2000).

We have gone beyond prior research by showing that much of the late interest in registering is concentrated at the very end of the campaign period. Across the country, 26% of the post-deadline search activity occurred in the final 2 days of the 2012 campaign. This may help explain the limited impact of the National Voter Registration Act (NVRA) of 1993, which has puzzled scholars (Highton 2004; Berinsky 2005). The NVRA was expected to increase turnout by mandating motor-voter, public agency, and mail registration, and by regulating how states purge voter files. However, turnout in subsequent presidential elections actually fell. Extending deadlines to Election Day is among the few steps that would allow last-minute interest to feed through to electoral participation.

Predictions based on web search data have recently come in for some criticism. Lazer et al. (2014) raise two concerns. The first is “big data hubris [...] the often implicit assumption that big data are a substitute for, rather than a supplement to, traditional data collection and analysis.” Our article draws on large sources of data: Google search logs and voter registration files. However, compiling the data did not require heavy computation, and the resulting data set is small (50 states over 67 days). Our Bayesian models can be fit on ordinary laptops. Nor do we face the common problem with big data that the number of predictors greatly exceeds the number of outcome measures ($p \gg n$), which can lead to over-fitting. We avoided this problem by selecting a small number of relevant queries for construct validity. Hence, our research may not even qualify as “big data.” More importantly, we are clear that our approach should complement rather than substitute for other methods.

Lazer et al. (2014) also argue that predictions based on web search data are unstable, because engineers frequently update search engine algorithms. While valid in some contexts, this critique is not relevant here, since our models and predictions apply only to a short time period, in which no major changes to the search algorithm were made.⁹ More generally, while Lazer et al. are correct to observe that much existing research uses web search data to predict future outcomes, this is not the only way in which scholars can use these data. We have illustrated a new application, modeling counterfactual outcomes. We now discuss some promising paths for future research.

An obvious extension of our work would be to use web search data to study other aspects of election administration. The effects of electoral rules have acquired new relevance in the wake of the Supreme Court decision in *Shelby County v. Holder* (2013), which invalidated a key section of the 1965 Voting Rights Act, allowing many more jurisdictions to amend electoral rules without federal oversight. As Kousser and Mullin (2007) observe, the United States is cross-cut by boundaries for elections at the local, state, and federal levels. Besides variation in registration deadlines, U.S. elections differ in myriad ways, such as the presence or absence of voter ID laws (Bentele and O'Brien 2013), the availability of mail ballots, sample ballots, and other information (Wolfinger, Highton, and Mullin 2005; Kousser and Mullin 2007), or the accessibility of polling places (Brady

⁹For the 2014 general election, Google has started displaying boxes with links to information on local registration rules, in response to queries such as “register to vote.” These may change click-through behavior, meaning that a replication of our approach for the 2014 election would require slightly different methods.

and McNulty 2011). The onus is on the voter to find out about the details, and searching online is now the obvious way to do so.

Jurisdictional or district boundaries can be used to study the effects of different electoral rules, under certain conditions (Keele and Titiunik 2015). But this is only possible if data are available across the border. One great advantage of web search data is granularity. We expect that search data will become available at the local level and, given the rise of mobile devices—over half of web search volume now comes from such devices—that it will be possible to study search behavior by specific location and time. Evidence on information-seeking behavior could be compared with registration or voting patterns, and potentially to demographic information (e.g., from voter files or the census). This could reveal the effects of election laws, or the effects of the varying implementation of those laws on certain sectors of the electorate (Atkeson et al. 2010). It may be possible to build on our methods in this paper by using additional data on search engine users—such as information on other searches generated by the same person—to find out which kind of people exhibit the most similar behavior on either side of administrative deadlines. This information could be used to refine models of counterfactual outcomes, using the best set of synthetic controls. Concerns over privacy will need careful attention, and models for collaboration between scholars and industry may need to be improved in order to make the data available (King 2011), but we foresee many opportunities for micro-level research on information and elections.

Future research could also seek to explain information-seeking behavior, taking web search data as the outcome variable. Indeed, while political scientists have produced rich literatures on (typically low) levels of political knowledge among the public (e.g., Lippmann 1922; Converse 1964), and on how citizens process political information (e.g., Sniderman, Brody, and Tetlock 1991; Lupia and McCubbins 1998), much less is known about the ways in which members of the public go about acquiring the limited political information that they *do* obtain. As Lau and Redlawsk (2006, 3) argue in the context of voting, “Most of our existing models of the vote choice are relatively static, based in a very real sense on cross-sectional survey data, taking what little (typically) voters know about the candidates at the time of the survey as a given with almost no thought to how they went about obtaining that information in the first place.” The same point could be made for our understanding of political information more broadly.

Research has been limited, in part, by a lack of tools for measuring how people go about collecting information. Scholars have made progress by studying temporal dynamics in information-seeking, or how emotions motivate inquiry. In so doing, they have relied on surveys or laboratory experiments (Marcus, Neuman, and MacKuen 2000; Valentino, Hutchings, and Williams 2004; Lau and Redlawsk 2006). These research methods have advantages: surveys and the laboratory provide controlled environments and allow the collection of data on individual attributes. But these are not the natural contexts in which members of the public seek political information. Data on web search activity provide a new source of leverage, at a time when the ubiquity of the internet is reducing the cost of acquiring information. Web search data are available from commercial search engines in large volumes, for a range of geographical units and time scales. These new measures will allow scientists to address old questions in new ways, and also promise to open new areas of study.

References

- Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association* 105(490):493–505.
- Angrist, Joshua, and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.
- Ansolabehere, Stephen, and David M. Konisky. 2006. The introduction of voter registration and its effect on turnout. *Political Analysis* 14(1):83–100.
- Ansolabehere, Stephen, and Eitan Hersh. 2012. Validation: What big data reveal about survey misreporting and the real electorate. *Political Analysis* 20(4):437–59.
- Atkeson, Lonna R., Lisa A. Bryant, Thad E. Hall, Kyle Saunders, and Michael Alvarez. 2010. A new barrier to participation: Heterogeneous application of voter identification policies. *Electoral Studies* 29(1):66–73.

- Bentele, Keith G., and Erin E. O'Brien. 2013. Jim Crow 2.0? Why states consider and adopt restrictive voter access policies. *Perspectives on Politics* 11(4):1088–1116.
- Berinsky, Adam J. 2005. The perverse consequences of electoral reform in the United States. *American Politics Research* 33(4):471–91.
- Brady, Henry E., and John E. McNulty. 2011. Turning out to vote: The costs of finding and getting to the polling place. *American Political Science Review* 105(1):115–34.
- Brians, Craig L., and Bernard Grofman. 2001. Election day registration's effect on US voter turnout. *Social Science Quarterly* 82(1):170–83.
- Brodersen, Kay H., Fabian Gallusser, Jim Koehler, Nicolas Remy, and Steven L. Scott. Forthcoming. Inferring causal impact using Bayesian structural time-series models. *Annals of Applied Statistics*.
- Burden, Barry C., David T. Canon, Kenneth R. Mayer, and Donald P. Moynihan. 2014. Election laws, mobilization, and turnout: The unanticipated consequences of election reform. *American Journal of Political Science* 58(1):95–109.
- Cain, Bruce E., and Ken McCue. 1985. The efficacy of registration drives. *Journal of Politics* 47(4):1221–1230.
- Carlin, Bradley P., and Thomas A. Louis. 2009. *Bayesian methods for data analysis*. Boca Raton, FL: CRC Press.
- Choi, Hyunyoung, and Hal Varian. 2012. Predicting the present with Google trends. *Economic Record* 88(1):2–9.
- Converse, Philip E. 1964. The nature of belief systems in mass publics. In *Ideology and discontent*, ed. David E. Apter, 206–61. New York: Free Press of Glencoe.
- Crainiceanu, Ciprian, David Ruppert, Gerda Claeskens, and Matthew P. Wand. 2005. Exact likelihood ratio tests for penalised splines. *Biometrika* 92(1):91–103.
- Fitzgerald, Mary. 2005. Greater convenience but not greater turnout: The impact of alternative voting methods on electoral participation in the United States. *American Politics Research* 33(6):842–67.
- Gallego, Aina, and Daniel Oberski. 2012. Personality and political participation: The mediation hypothesis. *Political Behavior* 34(3):425–51.
- Gelman, Andrew, and Donald B. Rubin. 1992. Inference from iterative simulation using multiple sequences. *Statistical Science* 7(4):457–511.
- Gerber, Alan S., Gregory A. Huber, David Doherty, Conor M. Dowling, Connor Raso, and Shang E. Ha. 2011. Personality traits and participation in political processes. *Journal of Politics* 73(03):692–706.
- Gimpel, James G., Joshua J. Dyck, and Daron R. Shaw. 2007. Election-year stimuli and the timing of voter registration. *Party Politics* 13(3):351–74.
- Ginsberg, Jeremy, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski, and Larry Brilliant. 2008. Detecting influenza epidemics using search engine query data. *Nature* 457(7232):1012–1014.
- Goel, Sharad, Jake M. Hofman, Sébastien Lahaie, David M. Pennock, and Duncan J. Watts. 2010. Predicting consumer behavior with web search. *Proceedings of the National Academy of Sciences* 107(41):17486–17490.
- Green, Donald P., and Alan S. Gerber. 2008. *Get out the vote: How to increase voter turnout*. Washington, DC: Brookings Institution Press.
- Hanmer, Michael J. 2009. *Discount voting: Voter registration reforms and their effects*. New York: Cambridge University Press.
- Hay, John L., and Anthony N. Pettitt. 2001. Bayesian analysis of a time series of counts with covariates: an application to the control of an infectious disease. *Biostatistics* 2(4):433–44.
- Herron, Michael C., and Daniel A. Smith. 2012. Souls to the polls: Early voting in Florida in the shadow of House Bill 1355. *Election Law Journal* 11(3):331–47.
- . 2013. The effects of House Bill 1355 on voter registration in Florida. *State Politics & Policy Quarterly* 13(2):279–305.
- Highton, Benjamin. 2004. Voter registration and turnout in the United States. *Perspectives on Politics* 2(3):507–15.
- Jackman, Robert W. 1987. Political institutions and voter turnout in the industrial democracies. *American Political Science Review* 81(2):405–23.
- Keele, Luke, and Rocio Titiunik. 2015. Geographic boundaries as regression discontinuities. *Political Analysis* 23(1):127–155.
- Keele, Luke, and William Minozzi. 2013. How much is Minnesota like Wisconsin? Assumptions and counterfactuals in causal inference with observational data. *Political Analysis* 21(2):193–216.
- Key, Valdimer O. 1949. *Southern politics in state and nation*. Knoxville: University of Tennessee Press.
- Keyssar, Alexander. 2009. *The right to vote: The contested history of democracy in the United States (Rev. Ed.)*. New York: Basic Books.
- King, Gary. 2011. Ensuring the data-rich future of the social sciences. *Science* 331(6018):719–21.
- Knack, Stephen. 2001. Election-day registration: The second wave. *American Politics Research* 29(1):65–78.
- Knee, Matthew R., and Donald P. Green. 2011. The effects of registration laws on voter turnout: An updated assessment. In *Facing the challenge of democracy: Explorations in the analysis of public opinion and political participation*, eds. M. Sniderman Paul and Benjamin Highton, 312–28. Princeton, NJ: Princeton University Press.
- Kousser, J. Morgan. 1974. *The shaping of southern politics: Suffrage restriction and the establishment of the one-party South, 1880–1910*. New Haven, CT: Yale University Press.
- . 1999. *Colorblind injustice: Minority voting rights and the undoing of the second reconstruction*. Chapel Hill: University of North Carolina Press.
- Kousser, Thad, and Megan Mullin. 2007. Does voting by mail increase participation? Using matching to analyze a natural experiment. *Political Analysis* 15(4):428–45.

- Lau, Richard R., and David P. Redlawsk. 2006. *How voters decide: Information processing in election campaigns*. New York: Cambridge University Press.
- Lazer, David, Ryan Kennedy, Gary King, and Alessandro Vespignani. 2014. The parable of Google flu: Traps in big data analysis. *Science* 343(6176):1203–1205.
- Lippmann, Walter. 1922. *Public opinion*. New York: Harcourt Brace.
- Lupia, Arthur, and Mathew D. McCubbins. 1998. *The democratic dilemma: Can citizens learn what they need to know?* New York: Cambridge University Press.
- Marcus, George E., W. Russell Neuman, and Michael MacKuen. 2000. *Affective intelligence and political judgment*. Chicago: University of Chicago Press.
- McDonald, Michael P. 2007. The true electorate a cross-validation of voter registration files and election survey demographics. *Public Opinion Quarterly* 71(4):588–602.
- Mondak, Jeffery J., Matthew V. Hibbing, Damarys Canache, Mitchell A. Seligson, and Mary R. Anderson. 2010. Personality and civic engagement: An integrative framework for the study of trait effects on political behavior. *American Political Science Review* 104(01):85–110.
- Nagler, Jonathan. 1991. The effect of registration laws and education on US voter turnout. *American Political Science Review* 85(4):1393–1405.
- Neiheisel, Jacob R., and Barry C. Burden. 2012. The impact of election day registration on voter turnout and election outcomes. *American Politics Research* 40(4):636–664.
- Piven, Frances F., and Richard A. Cloward. 2000. *Why Americans still don't vote: And why politicians want it that way*. Boston: Beacon Press.
- Plummer, Martyn. 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*, Vienna, Austria, March, 20–22.
- Powell, G. Bingham, Jr. 1986. American voter turnout in comparative perspective. *American Political Science Review* 80(1):17–43.
- Rosenstone, Steven J., and Raymond E. Wolfinger. 1978. The effect of registration laws on voter turnout. *American Political Science Review* 72(1):22–45.
- Rosenstone, Steven, and John M. Hansen. 1993. *Mobilization, participation and democracy in America*. New York: MacMillan Publishing.
- Ruppert, D., M. Wand, and R. Carroll. 2003. *Semiparametric regression*. New York: Cambridge University Press.
- Sniderman, Paul M., Richard A. Brody, and Philip E. Tetlock. 1991. *Reasoning and choice: Explorations in social psychology*. New York: Cambridge University Press.
- Spiegelhalter, Richard, Nicola G. Best, Bradley P. Carlin, and Angelika Van Der Linde. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B* 64(4):583–639.
- Street, Alex, Thomas A. Murray, John Blitzer, and Rajan S. Patel. 2015. Replication data for: Estimating voter registration deadline effects with web search data. <http://dx.doi.org/10.7910/DVN/28575>.
- Teixeira, Ruy A. 1992. *The disappearing American voter*. Washington, DC: Brookings Institution Press.
- Tokaji, Daniel P. 2008. Voter registration and election reform. *William & Mary Bill of Rights* 17(2):1–56.
- Valentino, Nicholas A., Vincent L. Hutchings, and Dmitri Williams. 2004. The impact of political advertising on knowledge, internet information seeking, and candidate preference. *Journal of Communication* 54(2):337–54.
- Varian, Hal R. 2014. Big data: New tricks for econometrics. *Journal of Economic Perspectives* 28(2):3–28.
- Verba, Sidney, Kay Lehman Schlozman, and Henry E. Brady. 1995. *Voice and equality: Civic voluntarism in American politics*. Cambridge, MA: Harvard University Press.
- Wolfinger, Raymond E., Benjamin Highton, and Megan Mullin. 2005. How postregistration laws affect the turnout of citizens registered to vote. *State Politics & Policy Quarterly* 5(1):1–23.
- Wolfinger, Raymond E., and Steven J. Rosenstone. 1980. *Who votes?* New Haven, CT: Yale University Press.
- Zeger, Scott L. 1988. A regression model for time series of counts. *Biometrika* 75(4):621–29.