SUPPORTING INFORMATION

STREET, MURRAY, BLITZER AND PATEL, 2015,

"ESTIMATING VOTER REGISTRATION DEADLINE EFFECTS WITH WEB SEARCH DATA"

S.1 The number of registered voters

The numbers of registered voters in each US state and in the District of Columbia, at the time of the 2012 general election, are included with the replication data. These data were collected from Secretary of State websites in January 2013. Numbers were not available for Mississippi, and North Dakota does not require voters to register. The total was 185,408,692. Our source for the Voting-Eligible Population (VEP), which subtracts foreign residents and those barred from voting by criminal records from the Voting-Age Population (VAP), is the U.S. Elections Project run by Dr. Michael McDonald. See `http://elections.gmu.edu/index.html`. The VEP for the 2012 general election (excluding Mississippi and North Dakota) was 214,744,339. Thus 86.3% of the VEP was estimated to have been registered in November 2012.

Note that, because voter files contain some people who have moved or are even deceased, the number of registered voters typically exceeds the number registered and able to vote in their current state of residence (Ansolabehere and Hersh 2012). Thus, these data provide a conservative estimate of the number unable to vote in 2012 because they were not registered. An alternative source of information on voter registration is the November supplement to the Current Population Survey. Individuals from a representative sample of the US population were asked whether they voted in the recent election, and whether they were registered. In November 2012, 71% of the US citizen population said they were registered, 15% said they were not, and 13% said they did not know or refused to answer the question. By this measure, we could infer that only slightly more than two thirds of US citizens were registered. However, focusing on those who answered the question yields results more similar to the figures from

Secretary of State websites. Among those who answered the question, 82% said they were registered. In short, there is reason to think that *at least* one in seven Americans was unable to vote in the 2012 presidential election because they were not registered.

S.2 Data on web search timing

Standardized web search totals for each of the 67 days and 49 states (plus the District of Columbia, but without North Dakota) in our study are included with the replication materials. We do not provide raw search data, or the absolute number of searches on a given day, because this information is commercially sensitive. As Lazer, et al. (2014: 1205) note, "It is impossible for Google to make its full arsenal of data available to outsiders, nor would it be ethically acceptable, given privacy issues." Indeed, the non-Google authors on this paper were only given access to the aggregated data that we are making available.

In order to facilitate replications using alternative search terms or time periods, we also designed this study to rely on data similar to those that are publicly available from the Google Trends (GT) website (`http://google.com/trends`). The longest period for which daily data are available from GT is three calendar months, which is one reason why we opted to study search volume from September to November 2012. We also used only five queries because this is the maximum allowed at the GT site. However, while much of our analysis in this paper is replicable with data from the GT site, we opted to use internal Google logs. This is because the GT site does not allow direct comparison of search volume for multiple search terms across states, and because daily data are not available from GT for less populous states (e.g. Rhode Island or Montana).

One advantage of using internal Google data is that we were able to test the plausibility of our assumption that people searching the web for "register to vote" actually wished to register. We observed which links were commonly selected by people who entered the relevant search terms. For example, in California, the top links were sites run by the state Secretary

of State, including (http://www.sos.ca.gov/elections/elections_vr.htm) and (http://registertovote.ca.gov/). The only two states in which the top link was not run by a state office such as the Secretary of State or the DMV were New Hampshire and New Mexico, where the highest-rated link was run by a non-profit group called Project Vote Smart, which provides similar information on voter registration requirements. In each case, the next most popular links were run by state offices.

One possibility that would endanger our estimates is if people who searched for "register to vote" were in fact seeking information on *where* to vote, or wished to check their registration status. These intentions were probably more relevant, closer to the election, implying that we might over-estimate the interest in registering that falls at the very end of the campaign period. If the true intention was to find information on polling place locations, this would be a problem for our approach, since, presumably, many of the people seeking this information were already registered. The intention to check one's voter status poses less of a problem since it is also driven by a desire to vote, and presumably a large share of those who did not know their registration status were not, in fact, registered.

We used click-through data to test for these alternative intentions. Secretary of State websites in many states provide separate pages with a) information on how to register, b) online registration (where available), c) registration status checks, and d) polling place locators (in some states, c and d are combined). In fact, only a small share of those who used Google to search for "voter registration," or the other queries that we used as predictors, clicked on the links to registration checks or polling place locators. Even on Election Day, the polling place locator page was not among the ten most commonly chosen links in any of the states in our sample. The registration status link was rarely among the ten most popular choices, and in the few states where it was, was selected by 10-20% of users. There is thus no evidence that large numbers of people who searched for "register to vote" in fact intended to search for "check voter registration" or "polling place." Of course, these alternative queries were *also* common, especially on Election Day, but the queries that we used as predictors in

this paper appear to provide clean measures of interest in registering.

S.3 Data on registration timing

Although a few states, including Maryland, Florida and Colorado, report monthly registration numbers on Secretary of State websites, in no case are records of the daily number of voter registrations available. We thus turned to voter files. These are not available for research purposes in all states, however. We were also constrained by the cost of obtaining voter files from some states. For example, Alabama and Arizona both charge around $30,000, and Wisconsin charges $12,500. The internal diversity of the states in our sample, the range of state sizes and the presence of both Democratic- and Republican- leaning states, all provide reasons for expecting that our results should generalize to the rest of the US. The share of registered voters in the VEP is also very similar in our sample and the rest of the US (86.6% in our sample; 86.3% overall).

In most states, small numbers of people continue to register even after the deadline has passed. In many cases, these people register when completing driver or vehicle registration forms. People who register after the deadline are not eligible to vote in the coming election. Voter files from several states record thousands of registrations on Election Day, even though EDR is not allowed in these states. According to election officials in these states, this is due to people who were allowed to update their registration data and cast provisional ballots (many had been registered at another address in the same state). Rather than counting them as new registrants, we treat the data for this day as missing in these states. This does not affect our predictions, since our models are based on days when registration was open. It is also worth noting that some of the changes in voter registration files reflect updates, e.g. from people who moved within the state, rather than new registrations. We do not see this as a problem for our approach, since such updates are necessary to vote (in some cases, updates are possible on Election Day).

In addition to the sixteen states in our sample, we also obtained the voter file for the District of Columbia, but found that search volume was extremely high, relative to population size, presumably because of web searches conducted by people who commute into the city but reside elsewhere, or due to campaign activities in the capital city. We thus discarded the DC data, although including registration data from DC in our model made little difference to the overall predictions.

Voter registration officials in several states assured us that files contain the date of registration, rather than the date on which the application was processed (which might be several days later, for applications submitted by mail). Since, as is evident in Figure 1, the timing of registration and search activity was closely related, we opted not to include lagged search volume in the model (though the autoregressive terms in the models capture correlation in residuals over time within states). We opted not to standardize the registration data, since the count of registered voters is the most readily interpretable outcome in this context.

S.4 Turnout among late registrants

Late registrants vote at high rates, suggesting that many people who register shortly before an election intend to vote in that election (Cain and McCue 1985; Gimpel, Dyck and Shaw 2007). In our data, for instance, the turnout rate among people who registered in Ohio in September and early October 2012 was 78%, compared to 68% for those who registered in the first eight months of the year. Turnout among those registered on Election Day is effectively 100%.

S.5 Alternative models

One alternative is a Poisson-normal mixture, which assumes $Y_{s,t} \sim \mathcal{P}(\lambda_{s,t})$, and, $\log(\lambda_{s,t}) \sim \mathcal{N}(\log(\mu_{s,t}), \sigma_\mu^2)$. This results in a model with constant overdispersion, whereas in the

Poisson-gamma mixture the overdispersion increases with the mean. Based on the DIC, the Poisson-gamma mixture provides a better fit. We also estimated models with random intercepts by state, and with random slopes by state, while assuming a linear effect for Google search volume. These yielded similar results. Using the same model described in the text (the spline with autoregressive errors), but focusing only on the eleven non-EDR states for which we were able to obtain both registration and search data, we estimate that allowing registration through Election Day would increase the number registered by 1.27 million (90% PI 1.01 million, 1.64 million).

S.6 State-level variation in search timing, competitiveness, and political information

For additional robustness tests, we studied state-level correlates of our search data that could cast doubt on our core assumption. Because the data are aggregated, rather than individual-level, there is a risk of spurious results due to problems related to the ecological fallacy. Nonetheless, these tests could hint at problems with our approach. We tested state-level variation in the competitiveness of the presidential race, and in levels of political information, against the share of search activity that fell after the deadline. To calculate the share of post-deadline search activity, we normalized the data by the number of days before and after the deadline. This is necessary in order to make meaningful comparisons between states with 30-day deadlines (e.g. Tennessee) and states where registration was closed for a shorter period (e.g. Nebraska, where registration was closed for 11 days).

We compared the (normalized) share of search activity after the deadline, by state, to the margin of victory for Obama in the 2012 presidential race. We excluded three states where in-person registration did not close: Maine, Montana and North Dakota. We also excluded the District of Columbia from the analysis, because it was an outlier with an 84 percentage point margin of victory for Obama. There was no significant relation: $n = 47$, Pearson's correlation 0.06, p=0.68. We obtained similar results when DC was included. Comparing

battleground states to those with much less competitive races, we nonetheless found similar numbers of people seeking information when it was already too late to register.

The very fact that the web data were generated implies that the searchers were not fully informed about voting procedures. But it is possible that late searchers knew even less. Such people may be less likely to participate, in which case it would be wrong to treat their web search behavior as equally strong evidence of an intent to register. To test for this possibility, we built an index of political knowledge using the pre-election ANES time series data from 2012. We used seven knowledge items: the religion of Obama and Romney, the Presidential term limit, an item on the size of the deficit, the length of the US Senate term, a definition of medicare, and a comparison of government spending on a number of programs. The index is the sum of correct answers to these questions (Cronbach's α=0.61). Interviews were conducted from September through November, the same period in which our web search data were generated. We found no significant relationship between mean political knowledge per state and the share of post-deadline search activity (p=0.41, n=5914).

S.7 Iowa registration data

Daily registration numbers were calculated by Iowa officials at our request using the voter file from early 2013. This risks undercounting of registration data in previous years, since some people have moved within the state and registered at a new address since 2004 or 2008, and only the most recent date of registration is reported in the voter file. Election officials in Iowa also cautioned us about the possibility of double-counting people who were new registrants in the state, and those who were newly registered in Iowa counties. Apparently the Iowa voter file keeps separate records of these events, but with some overlap. To avoid double-counting, we used only the data on new registrants in each county, which was closer to the total number of voters actually registered on Election Day in 2008 and 2012, according to Secretary of State figures (we also obtained similar results with models of the lower number

of people recorded as registering for the first time in any county in the state). Because of these issues we did not include the 2012 Iowa data in the multi-state model. Given the high cost of obtaining the Iowa voter file, this was the only form in which these data were available. Another state that recently allowed EDR is Montana (since 2006), but high costs prevented us from obtaining the voter file or the daily registration totals.

Table S1: Estimates of the effects of extending voter registration deadlines in presidential elections since the 1960s. We cite the percentage point increase in turnout predicted to result from moving from the typical deadline (three to four weeks before the election) to either no registration or EDR.

Year, author(s)	**Method**	**Estimate**
1967, Kelley, Ayres and Bowen	Cross-sectional multivariate regression; turnout in 104 large cities in 1960	30-day to no deadline: 4.8%
1978, Rosenstone and Wolfinger	Cross-sectional multivariate probit; individuals living in states with varying deadlines in 1970	30-day to no deadline: 6.1%
1992, Teixera	Cross-sectional multivariate probit; individuals in states with varying deadlines in 1970, '80, '84	30-day to no deadline: 4.8%
1994, Fenster	Difference in means tests; turnout in states with and without EDR in 1970s	Nationwide EDR: 4.7%
1995, Mitchell and Wlezien	Pooled cross-sectional multivariate logit; individuals living in different states in 1972-82	Nationwide EDR: 7.6%
1995, Rhine	Pooled cross-sectional regression; state turnout from 1972-92, with state and year fixed effects	30-day to no deadline: 3%
1996, Rhine	Multivariate probit; individuals in states with varying deadlines in 1992	Nationwide EDR: 14%
2001, Brians and Grofman	Pooled multivariate logit; individual turnout from 1972-96, with state and year fixed effects	Nationwide EDR: 7%
2001, Knack	Linear regression; state turnout in 1970s and 1990s with controls for lagged turnout by state	Nationwide EDR: 3%
2005, Fitzgerald	Pooled cross-sectional regression; state turnout from 1972-02, with state and year fixed effects	Nationwide EDR: 1.4%
2006, Ansolabehere & Konisky	Panel methods; turnout in counties in NY and OH from 1954 to 2000, including some counties forced by state laws to introduce registration	Early registration reduces turnout by 3-5%
2011, Knee and Green	Pooled regression; state turnout from 1980 to 2006, with state and year fixed effects	30-day to no deadline: 3%
2013, Neiheisel and Burden	Panel methods; turnout in municipalities in WI forced by state law to introduce EDR in 1976	EDR: 3%
2013, Keele and Minozzi	Regression discontinuity; municipalities with populations above and below cut-off for introducing EDR due to state legislation	EDR: -1.5% (WI) 0.7% (MN), not significant
2014, Burden et al.	Cross-sectional logit, turnout in 2004 and 2008, with individual and state controls; also panel methods for county turnout in 2004 and 2008	EDR: 3-4%

Table S2: Spearman's correlation between web search volume and daily registration totals, in the period when registration was open.

State	Correlation between search and registration
AK	0.87
AR	0.82
CA	0.84
DE	0.58
FL	0.71
ID	0.75
ME	0.66
MI	0.75
NC	0.68
NJ	0.68
NV	0.72
NY	0.72
OH	0.63
RI	0.71
WA	0.35
WY	0.65

Table S3: Predicted additional registrations based on post-deadline searches, by state.

State	Extra days for registering, if deadline moved to Election	Prediction, to nearest thousand (90% prediction interval)
AL	11	37,000 (16,000 to 67,000)
AZ	28	103,000 (63,000 to 154,000)
AR	28	28,000 (17,000 to 43,000)
CA	15	289,000 (146,000 to 490,000)
CO	28	131,000 (82,000 to 192,000)
CT	5	19,000 (6,000 to 40,000)
DE	24	12,000 (7,000 to 20,000)
FL	28	179,000 (111,000 to 267,000)
GA	28	178,000 (109,000 to 269,000)
HI	29	24,000 (14,000 to 36,000)
IL	28	212,000 (129,000 to 320,000)
IN	28	100,000 (60,000 to 153,000)
KS	21	35,000 (20,000 to 54,000)
KY	28	51,000 (31,000 to 76,000)
LA	28	65,000 (40,000 to 100,000)
MD	21	84,000 (48,000 to 132,000)
MA	20	87,000 (48,000 to 141,000)
MI	28	101,000 (59,000 to 160,000)
MS	31	23,000 (14,000 to 35,000)
MO	27	88,000 (54,000 to 134,000)
NE	11	16,000 (7,000 to 30,000)
NV	21	40,000 (23,000 to 60,000)
NJ	21	111,000 (63,000 to 176,000)
NM	28	23,000 (14,000 to 35,000)
NY	25	320,000 (182,000 to 506,000)
NC	3	46,000 (10,000 to 108,000)
OH	28	127,000 (77,000 to 190,000)
OK	25	37,000 (22,000 to 57,000)
OR	21	54,000 (32,000 to 82,000)
PA	28	180,000 (107,000 to 281,000)
SC	30	94,000 (58,000 to 142,000)
SD	15	9,000 (4,000 to 17,000)
TN	28	76,000 (48,000 to 111,000)
TX	28	317,000 (193,000 to 477,000)
UT	15	53,000 (28,000 to 87,000)
VA	22	111,000 (64,000 to 174,000)
VT	6	3,000 (1,000 to 7,000)
WA	8	32,000 (13,000 to 61,000)
WV	21	10,000 (5,000 to 17,000)

Figure S1: Searches for "voter registration" and related terms from September 1 to November 10, 2012, in 20 states. Vertical axes are in standardized units that preserve scale across states.

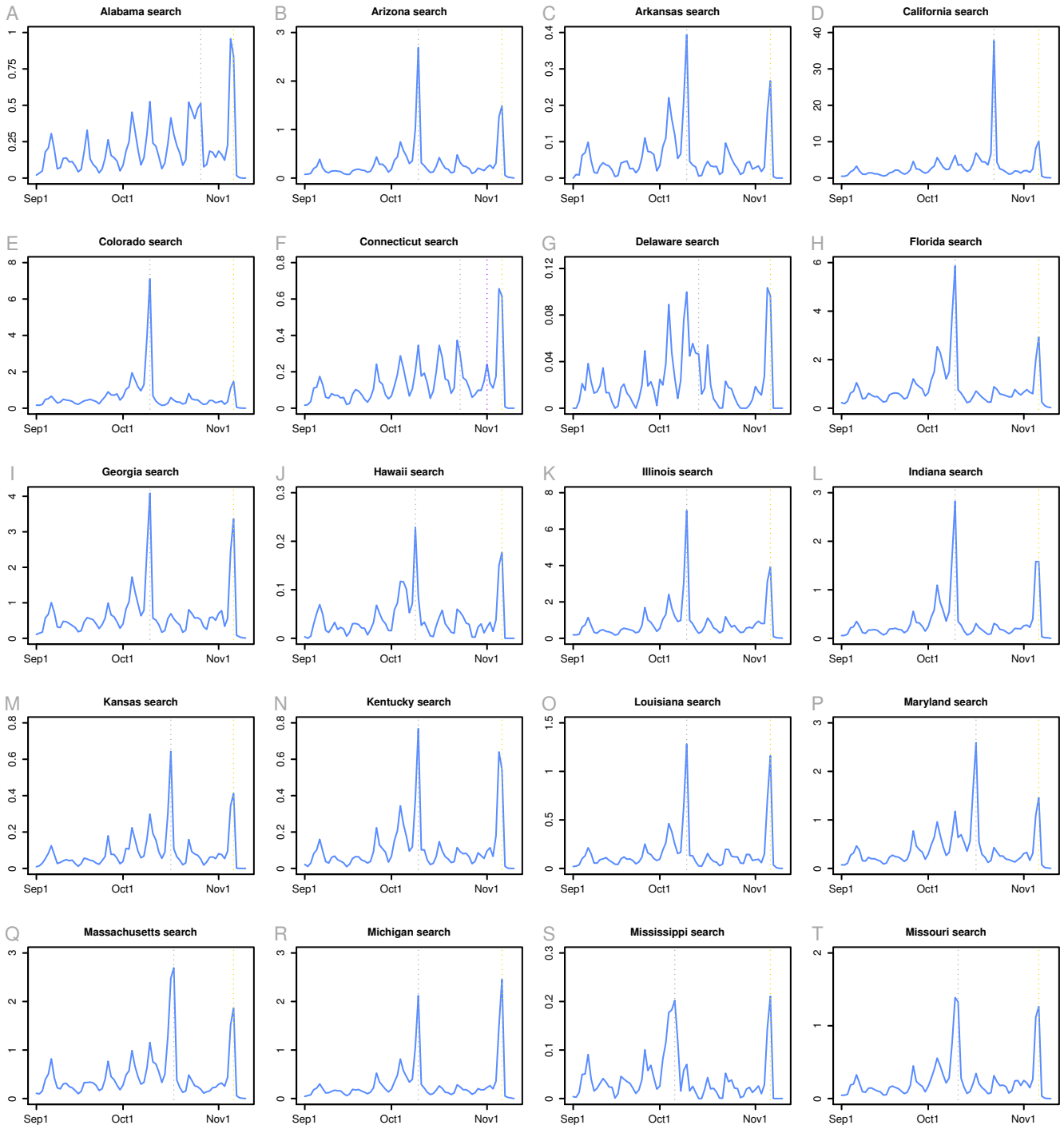

Figure S2: Searches for "voter registration" and related terms from September 1 to November 10, 2012, in a further 19 states. Vertical axes are in standardized units that preserve scale across states.

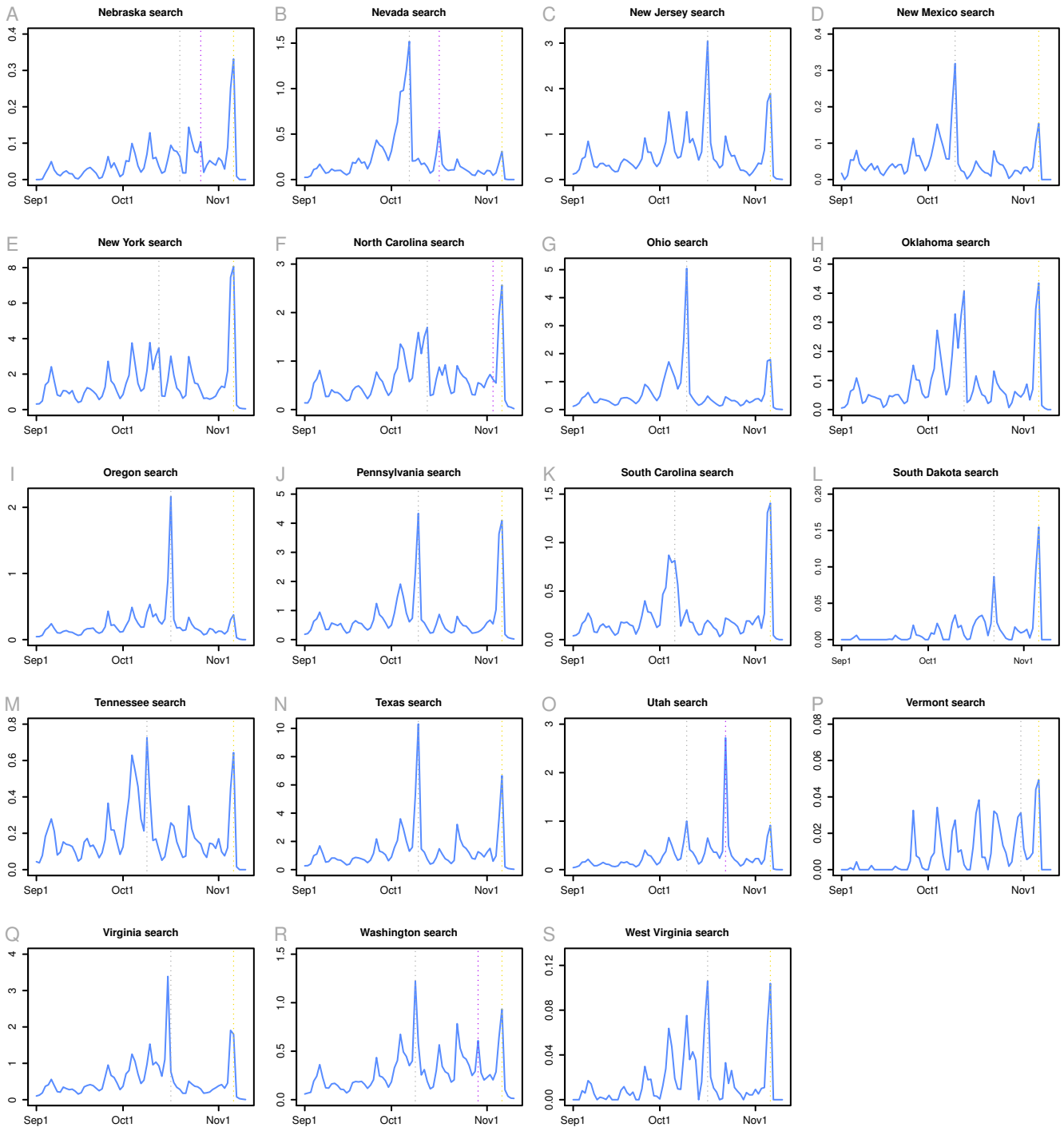

Figure S3: Searches for "voter registration" and related terms from September 1 to November 10, 2012, in states that allowed Election-Day Registration (in some cases only for the Presidential ballot), or had no registration requirement (North Dakota) in 2012.

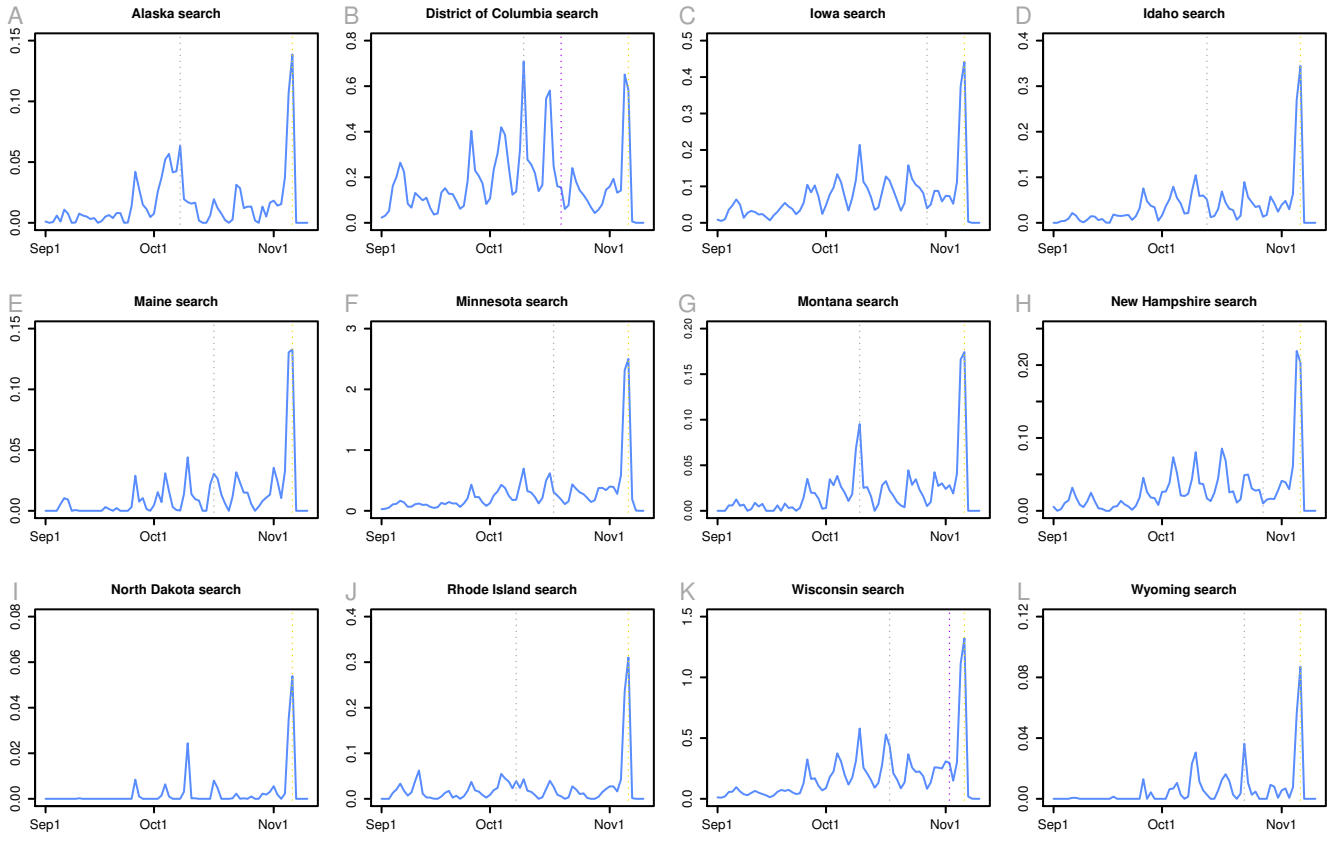

Figure S4: Spline fit of relationship between search volume and registration numbers. Dashed lines show 90% credible intervals.

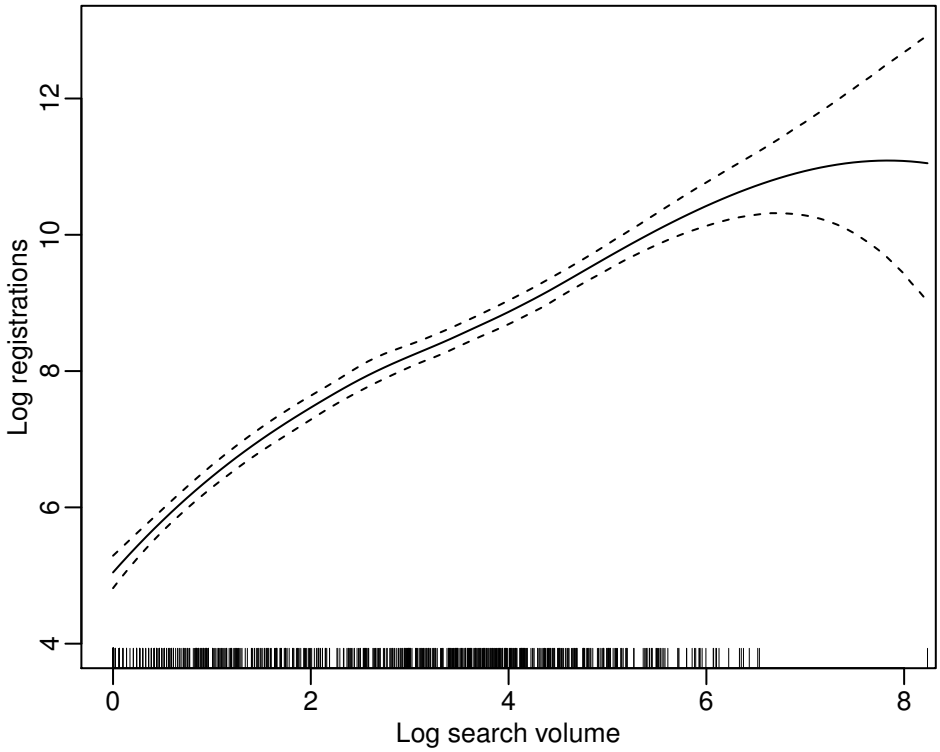

Figure S5: Trends in the share of users in each state who searched for "voter registration" and then clicked the link to the official (usually Secretary of State) website with relevant information, from Sept. 1 to Nov. 6, 2012. States are grouped by mail registration deadlines, shown by the grey vertical line in each panel. Zero on the vertical axis is the state mean click-through rate across all days. We standardized the click-through proportion by state so that the variance equaled one, and set vertical axes at +/- four times the variance. We do not report actual click- through proportions so as to avoid revealing proprietary information.

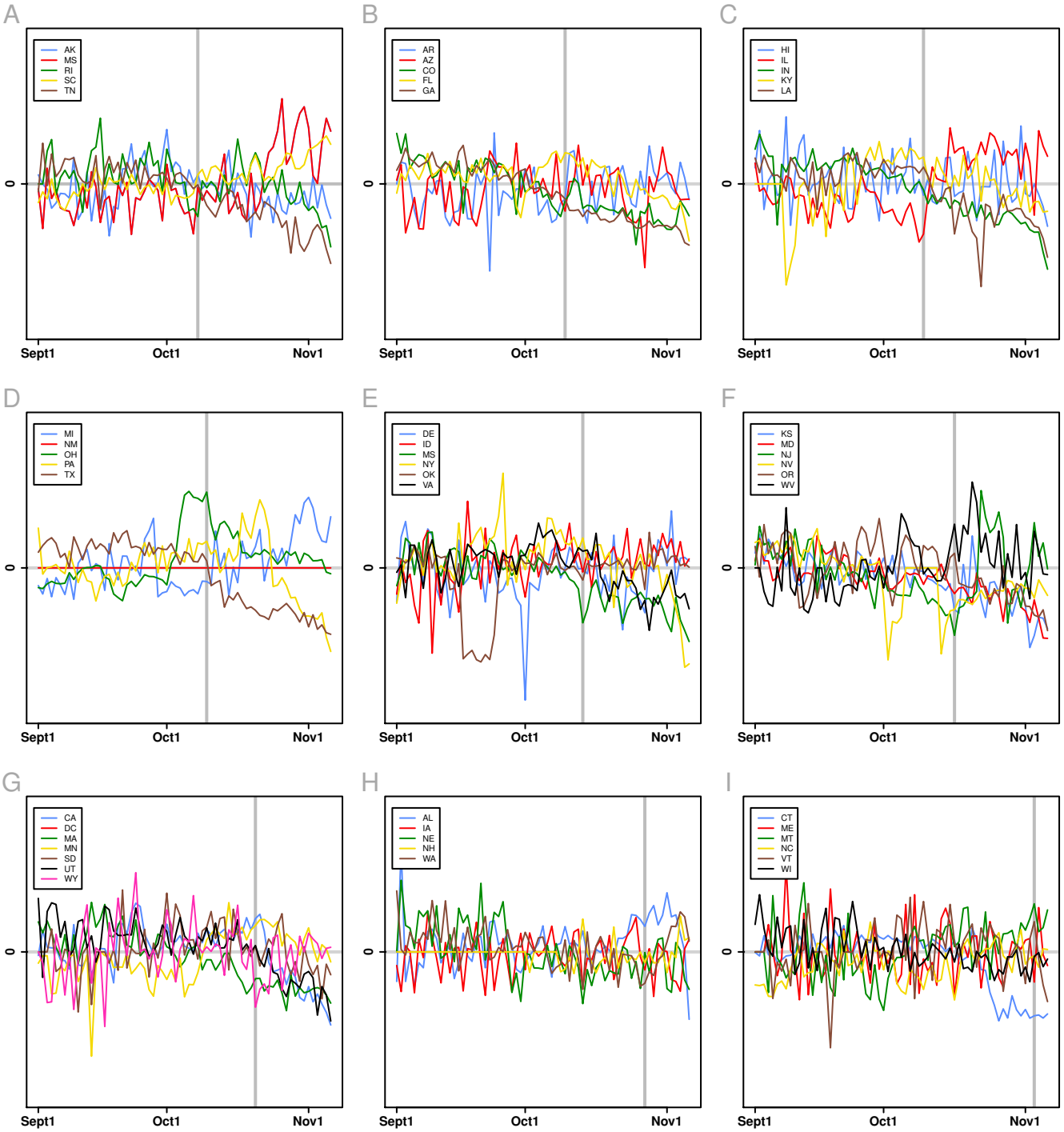

Figure S6: MCMC trace plots for selected parameters in the spline/AR1 model. Two MCMC chains were used for posterior estimation, with black depicting one and red depicting the other.