Supplemental Appendix for:

Truman Defeats Dewey: The Effect of Campaign Visits on Election Outcomes

Boris Heersink* Brenton D. Peterson†

June 8, 2017

* Ph.D Candidate, Department of Politics, University of Virginia.

† Ph.D Candidate, Department of Politics, University of Virginia, and Research Affiliate, Strathmore University.

Contents

A. Data and Variable Construction

The data used in our analysis consists of a combination of readily-available public data on electoral results, economic growth, demographic characteristics, and data concerning candidates' campaign visits gathered from archival and other sources. Detailed records of candidate campaign visits in both the 1944 and 1948 campaigns were collected from: the Thomas E. Dewey Papers collection at the University of Rochester, a list of campaign appearances and a separate list of Truman's public addresses from the Truman Library & Museum, a list of public addresses from the Roosevelt Library, descriptions of the 1944 campaign from David M. Jordan's (2011) history of the 1944 Election, and systematic searches of historical newspaper accounts using ProQuest.

In addition to data on campaign visits, our outcome variable is county-level Presidential vote share for the Democratic and Republican parties, provided by ICPSR's Historical Election Returns collection.[1] Our primary models focus on Democratic and Republican vote share; in an alternative model—the results of which are available upon request—we also utilize two-party vote share, based on the same underlying data.

To test the robustness of our results, we also estimate models using a set of predictor variables, as described in Section B of this Supplemental Appendix. Data for predictor variables are all publicly-available. Congressional vote share was provided by the same ICPSR collection as our Presidential election results. The black percentage of each county's population in 1930 was gathered from US Census records, using the University of Virginia's historical census browser. Income data was used to calculate both per capita income by state and year-on-year changes in per capita income by state. Income data was not disaggregated to the county level during this period. As a result, we rely on state-level income data from Barro & Sala-i-Martin (1992), helpfully archived by Brian C. Jenkins.[2]

[1] Inter-university Consortium for Political and Social Research. United States Historical Election Returns, 1824-1968 [Computer File]. ICPSR00001-v3. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 1999-04-26. http://doi.org/10.3886/ICPSR00001.v3.

[2] Available on the web at http://www.briancjenkins.com/data/usconvergence/

B. Alternative Model Specifications

The results reported in the primary manuscript are based on a series of, arguably arbitrary, modeling decisions. Our primary models minimize MSPE over a seven-election pre-treatment time period (1920-1944). Our choice balances two considerations. On one hand, matching counties over a longer time-series increases our confidence in the quality of matches between each treated unit and its synthetic control unit, subsequently increasing our confidence in the results. On the other hand, matching counties over a longer time period requires additional years of data, which are not available for all counties. This applies to both donor and treated counties. By expanding the length of the pre-treatment period to eight elections, we are unable to estimate models for two additional counties visited by Truman. Expanding the length to ten or twelve elections forces us to remove an additional four and 23 treated units, respectively.

In addition to the pre-treatment period, we also made decisions regarding the donor pool, the set of untreated counties from which each synthetic control unit is constructed. Again, this decision balances a number of considerations. A larger donor pool, with more untreated counties, may produce a closer fit between treated and synthetic control units over the pre-treatment period. By definition, a larger donor pool will never result in a *worse* fit. At the same time, a larger donor pool requires additional computing time. And including all untreated counties may provide marginal benefits, if many of those counties are poor matches for the treated unit in question. Worse, combining untreated counties with extreme values to match a treated county with moderate values can result in interpolation biases. For this reason, we selected a moderate approach, in which we limit the donor pool for each treated county to 800 untreated counties. We select a different donor pool for each treated county. The donor pool consists of the 800 untreated counties that match treated unit i most closely in terms of pre-treatment vote share.[3]

Finally, we made decisions about the predictor variables that were used in the synthetic control models. Because of the time period that we study, the set of available predictor

[3]More precisely, we select untreated counties for inclusion on the basis of MSPE over the period 1920 to 1944.

variables is limited. In our primary models, we do not incorporate any predictor variables into the model, other than a full set of pre-treatment presidential election outcomes. This produces very close matches between treated counties and their synthetic control units, on average, as illustrated in this section.

To check the robustness of our results to modeling decisions such as these, we estimated a series of models employing alternative approaches. First, we varied the number of pre-treatment elections included in the models from seven to eight. Second, we varied the size of our donor pools, from 800 untreated counties down to 300. Third, we incorporated a set of predictor variables. Specifically, we use annual state income per capita from 1930 to 1946, change in state income per capita in the year prior to each presidential election from 1932 to 1944, congressional vote share from 1926 to 1946 and percent black in 1930.[4] These choice to include these predictors is consistent with models of recent presidential elections (e.g., Abramowitz 1988; Nadeau and Lewis-Beck 2001).

Table A.1 reports results from our primary models and a series of alternative models. The top panel focuses on counties visited by either Truman or Dewey, but not both. The first set of models, using seven years of pre-treatment elections and 800 counties in the donor pool, constitute the primary models reported in the paper, with estimated effects of 3.06 percentage points for Truman and -0.20 percentage points for Dewey. The second row reduces the size of the donor pool to 300 counties, while the third row increases the donor pool back to 800 counties but lengthens the pre-treatment period to eight elections. Finally, the last row of the top panel incorporates the set of predictor variables described above, while keeping other aspects of our primary models fixed.

Across the four models, we observe broadly consistent findings. In each case, Truman's visits produce a strong positive effect on Democratic vote share.[5] At its smallest, Truman's impact on county vote share is 2.14 percentage points, with a p-value of less than .01. Our results for Dewey hold similarly—Dewey's impact is never positive and always ranges between -0.20 percentage points and -0.94 percentage points. None of the models

[4]The sources for and construction of these variables are provided in Section A of this appendix.

[5]As described in Section D of this appendix, we report two-sided p-values derived from randomization inference.

Years & Donor Pool	Truman Effect	Dewey Effect
Primary Effects - Exclusive Visits		
7 years; 800 Donors	3.06 ($p < 0.01$)	-0.20 (0.76)
7 years; 300 Donors	2.14 ($p < 0.01$)	-0.94 (0.19)
8 years; 800 Donors	2.74 ($p < 0.01$)	-0.59 (0.35)
7 years; 800 Donors; Predictors	2.72 ($p < 0.01$)	-0.30 (0.64)
Overlapping Visits		
7 years; 800 Donor Counties	2.19 ($p = 0.01$)	-0.75 (0.21)
7 years; 300 Donor Counties	0.97 ($p = .19$)	-0.55 (0.34)
8 years; 800 Donor Counties	2.34 ($p < 0.01$)	-1.15 (0.07)
7 years; 800 Donor; Predictors	1.74 ($p = 0.04$)	-0.62 (0.32)
Adjacent to Exclusive Visits		
7 years; 800 Donor Counties	2.64 ($p < 0.01$)	-0.15 (0.60)

Table A.1: Average effects from alternative synthetic control models. Models in each panel vary along three dimensions: the length of the pre-treatment period over which MSPE is minimized; the number of untreated counties that comprise the donor pool; and the inclusion or exclusion of additional predictor variables. The top panel reports "primary effects," from counties visited by either Truman or D2016ewey alone. The middle panel reports results from counties which both candidates visited. The bottom panel reports results from counties adjacent to a county that was visited exclusively by either candidate.

produce statistically significant effects for Dewey, in either direction.

In the primary manuscript, we also analyze the impact of overlapping visits by the two candidates, i.e. counties which both candidates visited at least once. These results are of particular interest, because they shed light on the relative quality of the two candidates' appearances. If Truman's campaign visits were of high quality relative to those of Dewey, we would expect to see an increase in Democratic vote share, similar to the increase we observe in counties where Truman was the only candidate who visited. The results of counties with "overlapping visits" need to be interpreted with caution, however, as they mix two distinct treatments, the effects of which are difficult to separate.

Table A.1 includes the results of alternative models of counties with overlapping visits,

listed in the middle panel. The top row describes our primary results. The second through fourth rows report results from models with a decreased donor pool, a longer pre-treatment time series, and a set of predictor variables, respectively.

Relative to counties which he visited exclusively, Truman performed slightly worse in these "overlapping" counties. Nonetheless, he consistently gained points in these counties. Although the results are less robust—in one set of models, the effects are not statistically significant—the overall results show that Truman gained votes in counties where both candidates visited. In contrast, our aggregate estimates of Dewey's performance are—like those from counties which he visited exclusively—consistently negative.

The final panel of Table A.1 reports results of models focused on counties adjacent to a visit from either candidate, results which also appeared in the manuscript.

B.1. Models of Primary Effects

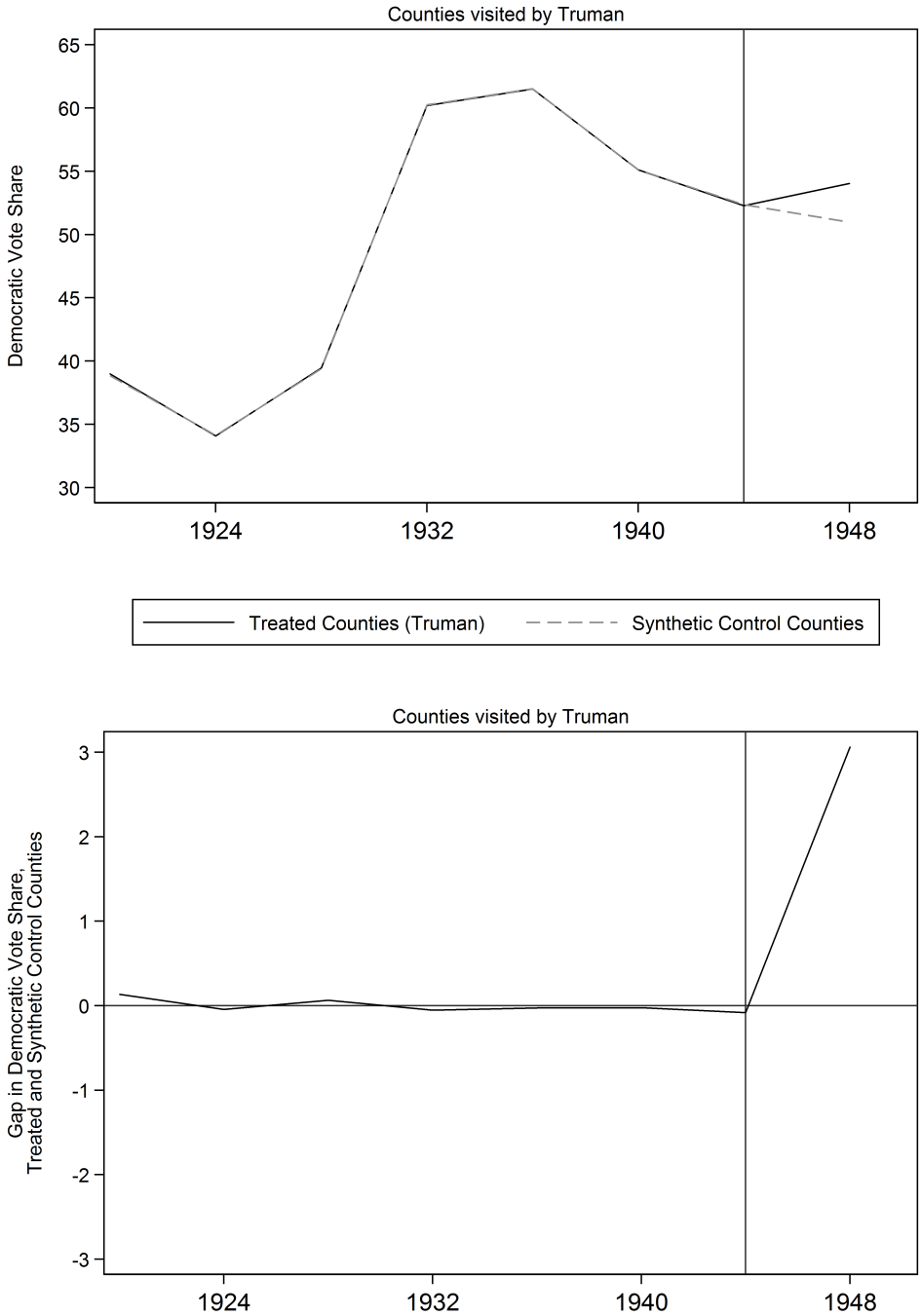

Figure A.1: Top panel: treated and synthetic control units' Democratic vote share, 1920-1948 (Treated n = 149). Bottom panel: the difference between treated and synthetic control units' Democratic vote share, 1920-1948. Treated units were visited exclusively by Truman. The donor pool consists of the closest 800 matches to county i in terms of pre-treatment (1920-1944) Democratic vote share.

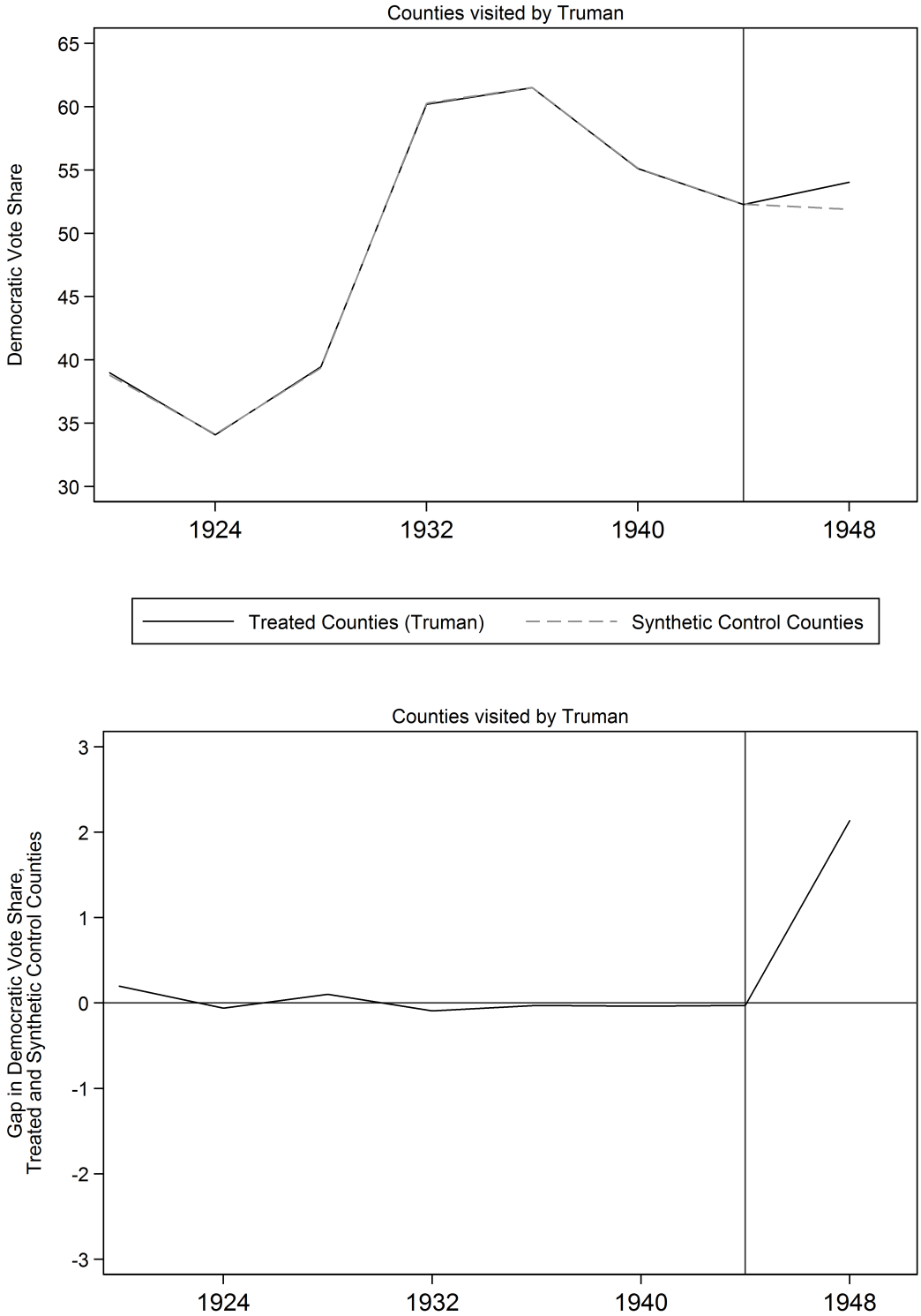

Figure A.2: Top panel: treated and synthetic control units' Democratic vote share, 1920-1948 (Treated n = 149). Bottom panel: the difference between treated and synthetic control units' Democratic vote share, 1920-1948. Treated units were visited exclusively by Truman. The donor pool consists of the closest 300 matches to county i in terms of pre-treatment (1920-1944) Democratic vote share.

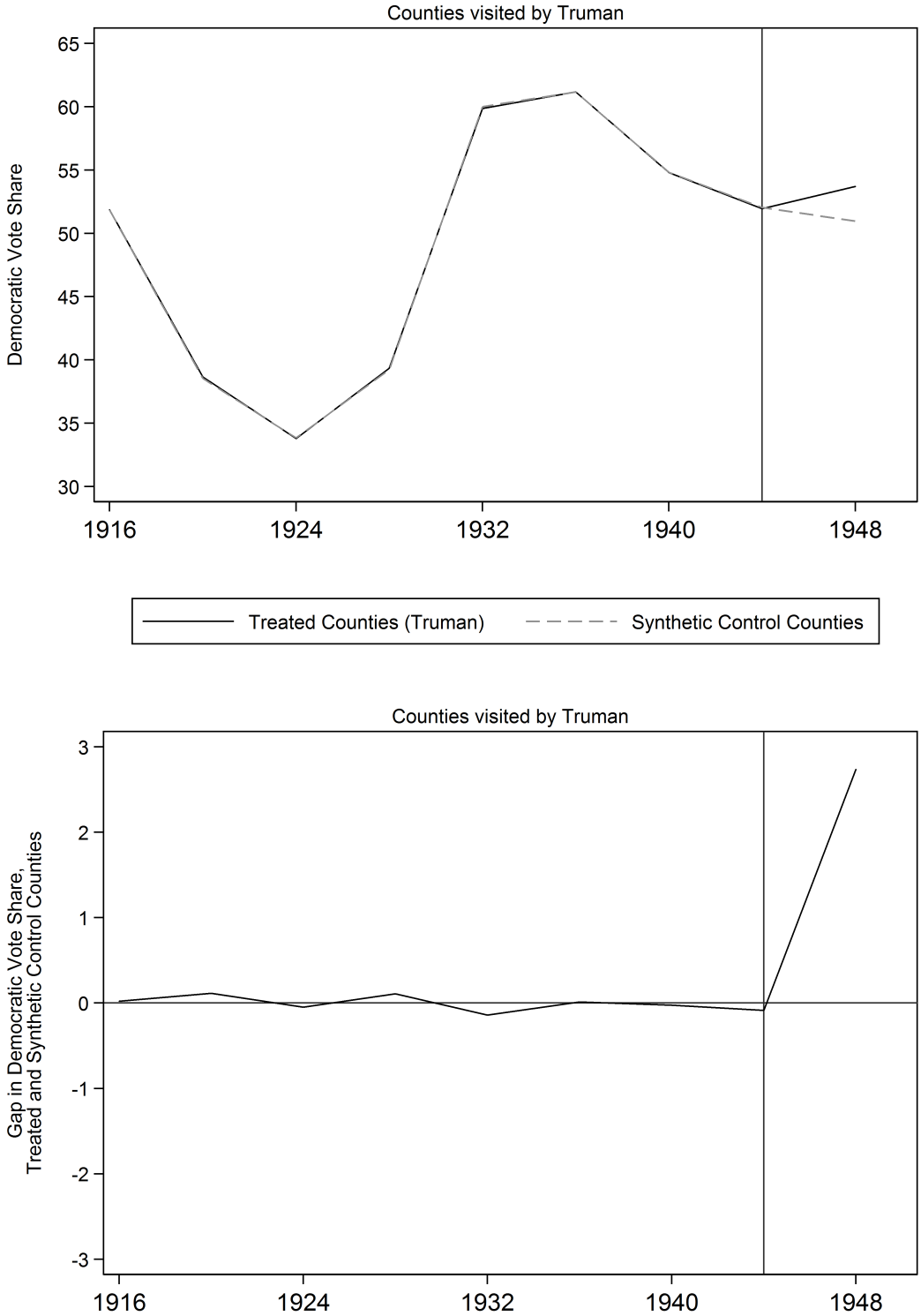

Figure A.3: Top panel: treated and synthetic control units' Democratic vote share, 1916-1948 (Treated n = 147). Bottom panel: the difference between treated and synthetic control units' Democratic vote share, 1916-1948. Treated units were visited exclusively by Truman. The donor pool consists of the closest 800 matches to county i in terms of pre-treatment (1916-1944) Democratic vote share.

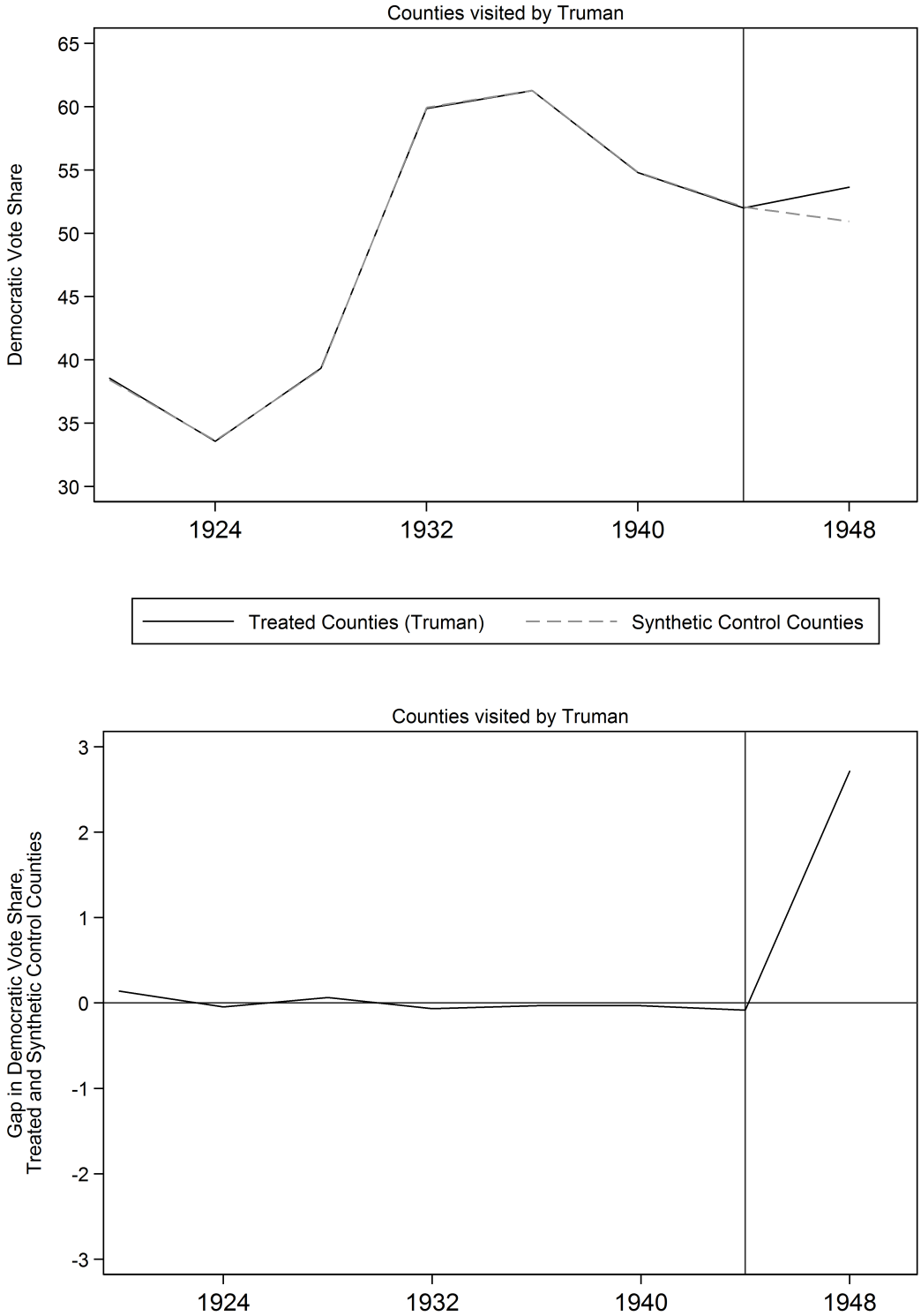

Figure A.4: Top panel: treated and synthetic control units' Democratic vote share, 1920-1948 (Treated n = 144). Bottom panel: the difference between treated and synthetic control units' Democratic vote share, 1920-1948. Treated units were visited exclusively by Truman. The donor pool consists of the closest 800 matches to county i in terms of pre-treatment (1920-1944) Democratic vote share. Synthetic control models included state per capita income, change in state per capita income, percent black and Congressional vote share as predictor variables.

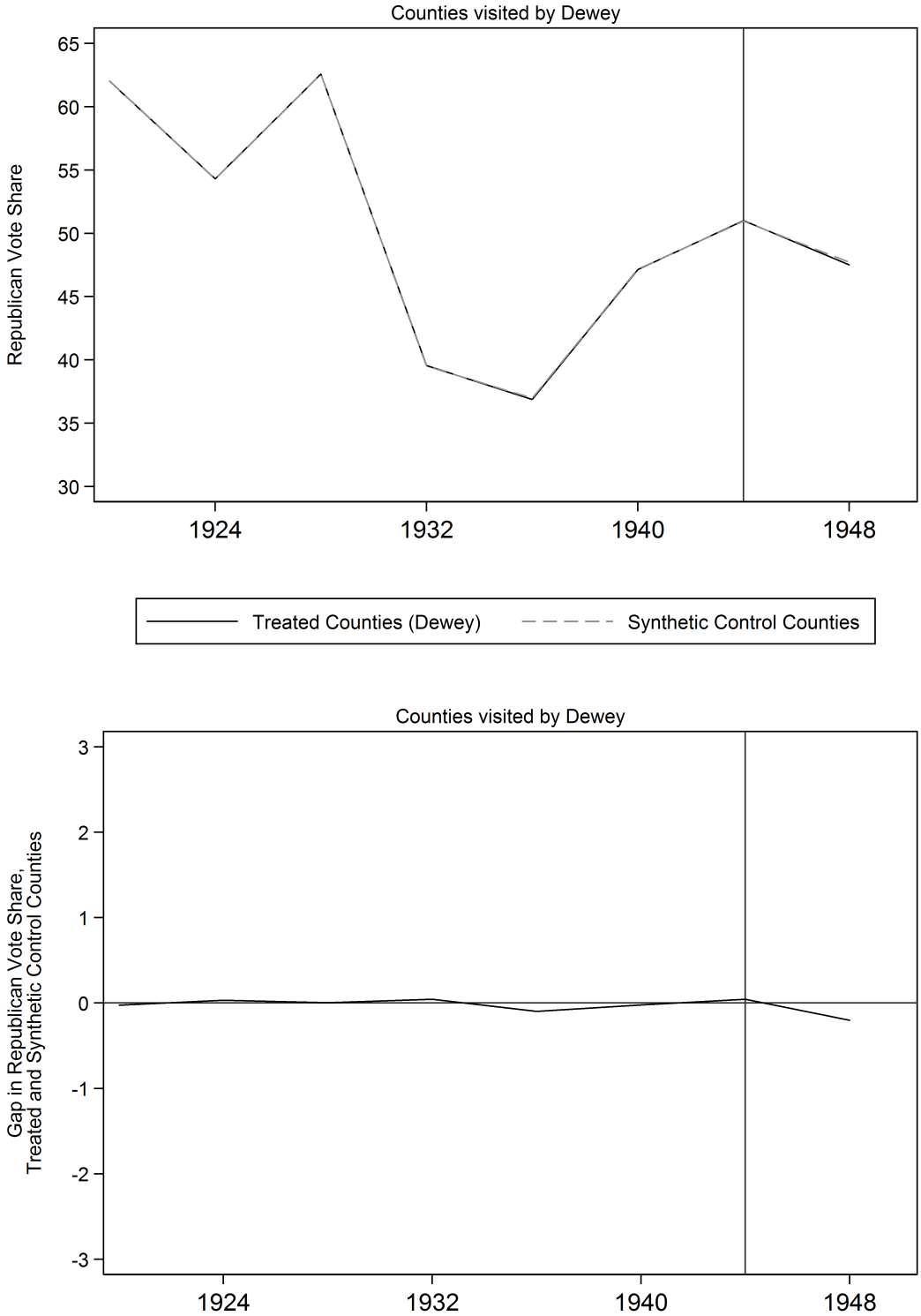

Figure A.5: Top panel: treated and synthetic control units' Republican vote share, 1920-1948 (Treated n = 62). Bottom panel: the difference between treated and synthetic control units' Republican vote share, 1920-1948. Treated units were visited exclusively by Dewey. The donor pool consists of the closest 800 matches to county i in terms of pre-treatment (1920-1944) Republican vote share.

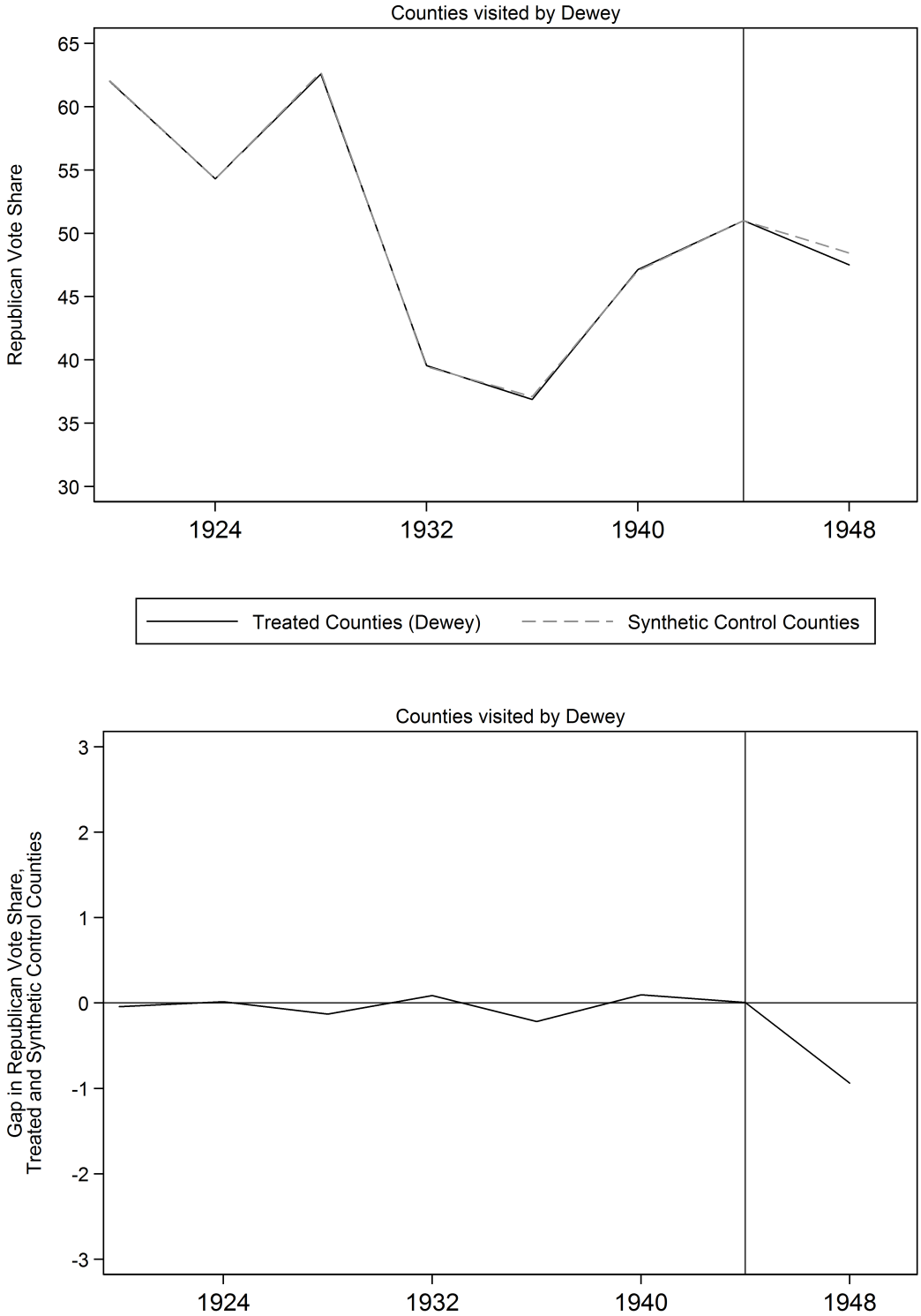

Figure A.6: Top panel: treated and synthetic control units' Republican vote share, 1920-1948 (Treated n = 62). Bottom panel: the difference between treated and synthetic control units' Republican vote share, 1920-1948. Treated units were visited exclusively by Dewey. The donor pool consists of the closest 300 matches to county i in terms of pre-treatment (1920-1944) Republican vote share.

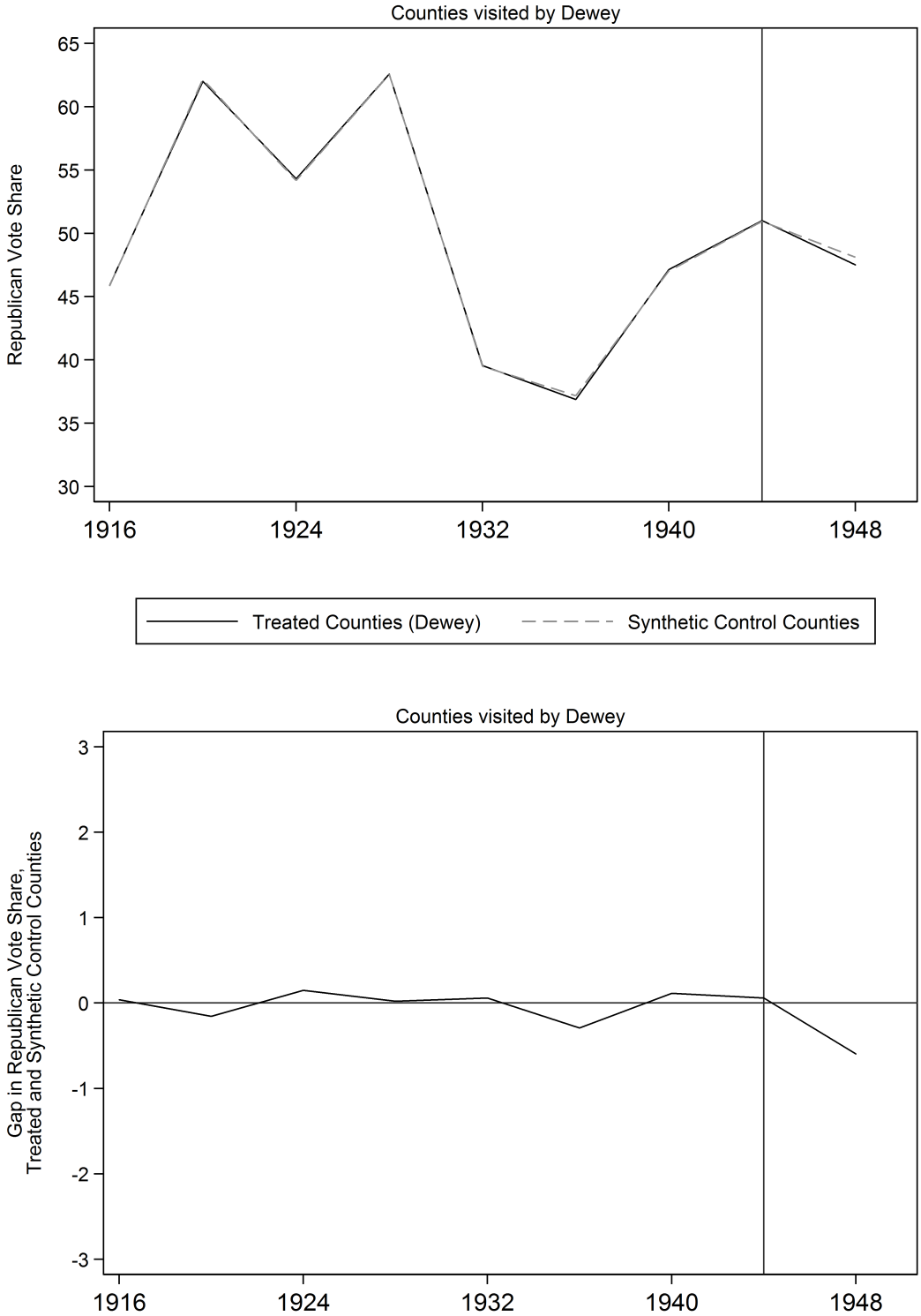

Figure A.7: Top panel: treated and synthetic control units' Republican vote share, 1916-1948 (Treated n = 62). Bottom panel: the difference between treated and synthetic control units' Republican vote share, 1916-1948. Treated units were visited exclusively by Dewey. The donor pool consists of the closest 800 matches to county i in terms of pre-treatment (1916-1944) Republican vote share.

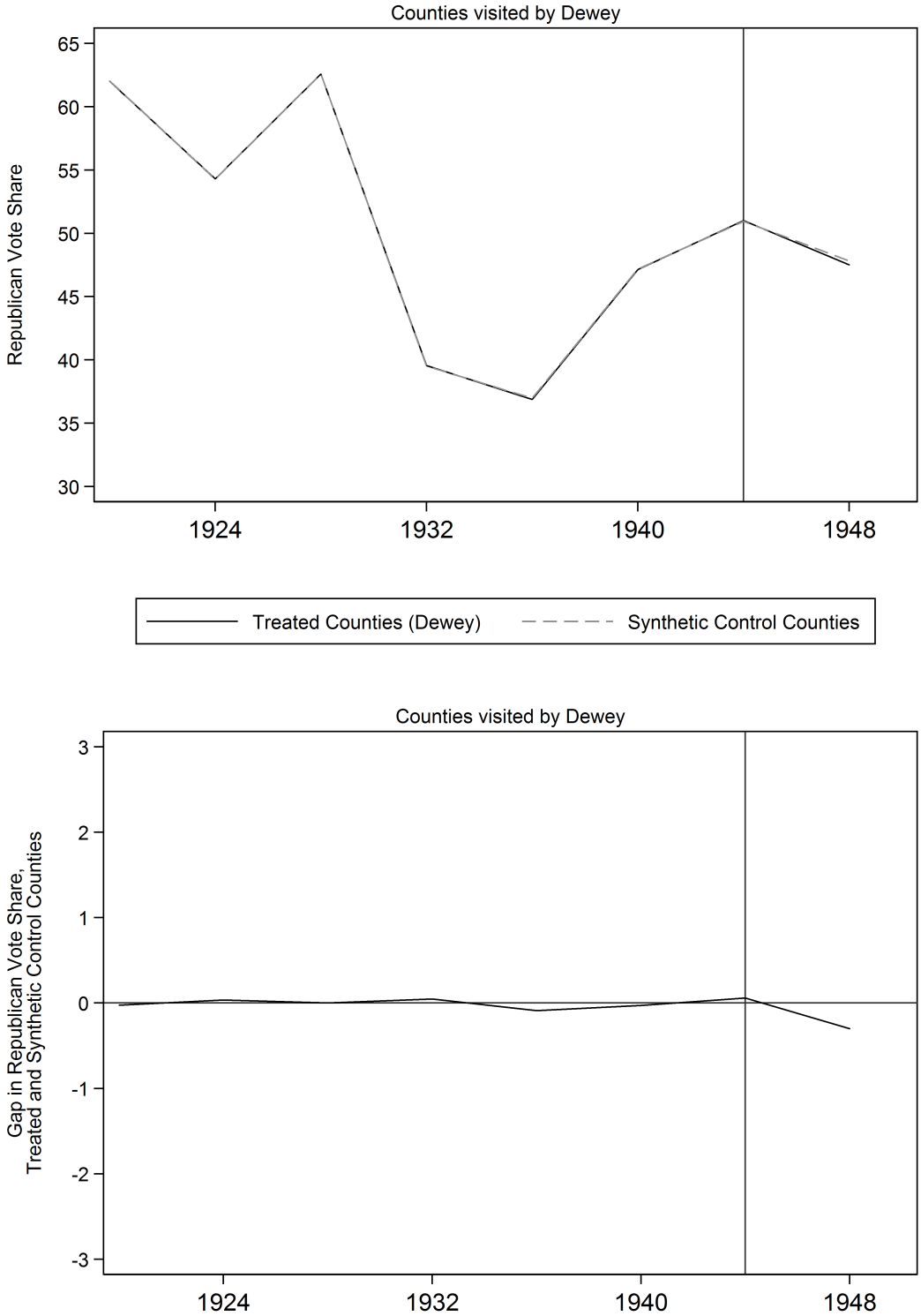

Figure A.8: Top panel: treated and synthetic control units' Republican vote share, 1920-1948 (Treated n = 62). Bottom panel: the difference between treated and synthetic control units' Republican vote share, 1920-1948. Treated units were visited exclusively by Truman. The donor pool consists of the closest 800 matches to county i in terms of pre-treatment (1920-1944) Republican vote share. Synthetic control models included state per capita income, change in state per capita income, percent black and Congressional vote share as predictor variables.

B.2. Models of Mixed Effects from Overlapping Visits

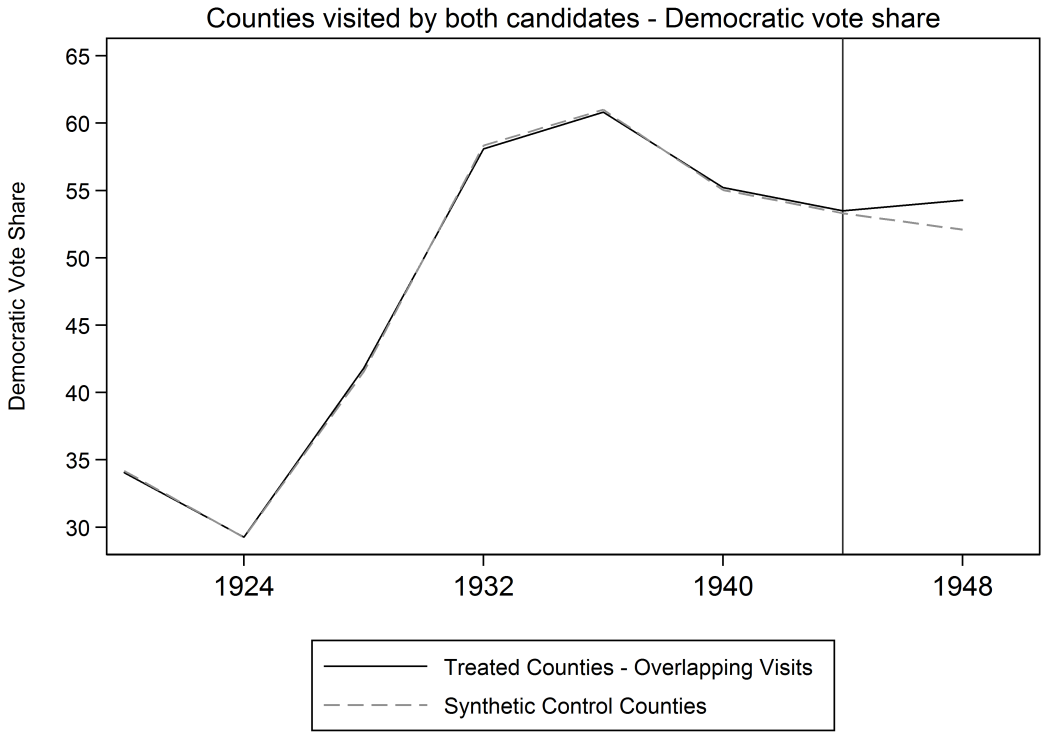

Figure A.9: The difference in Democratic vote share between double-visited counties and their synthetic control units, 1920-1948 (Treated n = 55). Treated units were visited by both Truman and Dewey during the 1948 campaign. The outcome variable for Truman's effect is Democratic vote share. The donor pool consists of the closest 800 matches to county i in terms of pre-treatment (1920-1944) Democratic vote share.

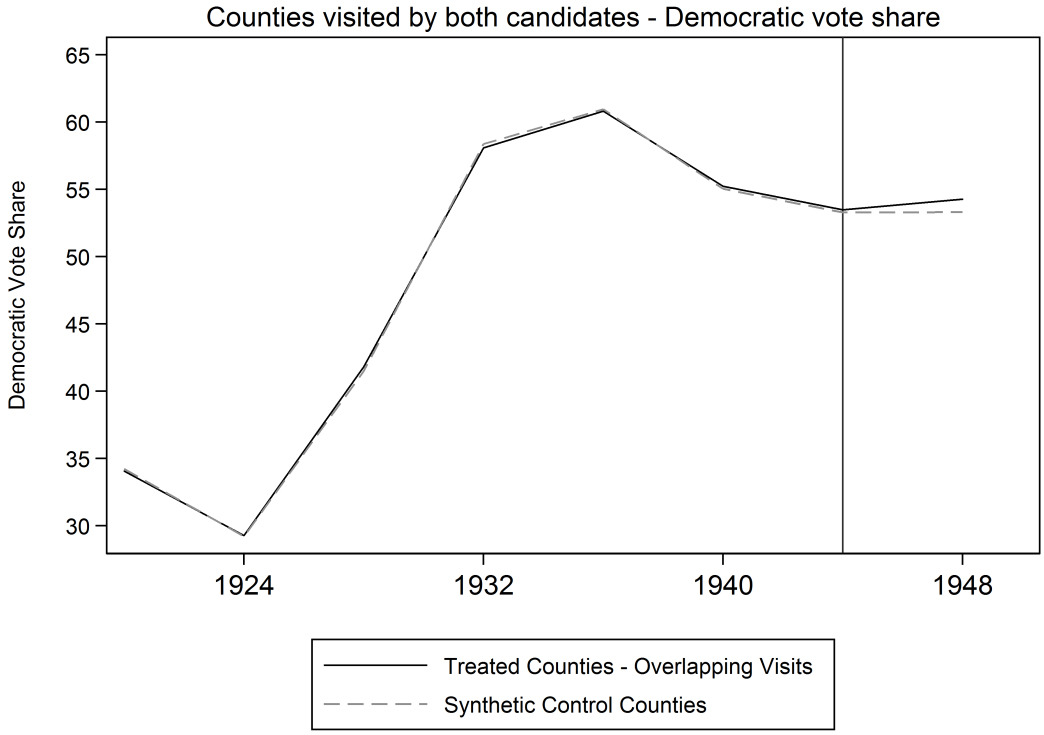

Figure A.10: The difference in Democratic vote share between double-visited counties and their synthetic control units, 1920-1948 (Treated n = 55). Treated units were visited by both Truman and Dewey during the 1948 campaign. The outcome variable for Truman's effect is Democratic vote share. The donor pool consists of the closest 300 matches to county i in terms of pre-treatment (1920-1944) Democratic vote share.

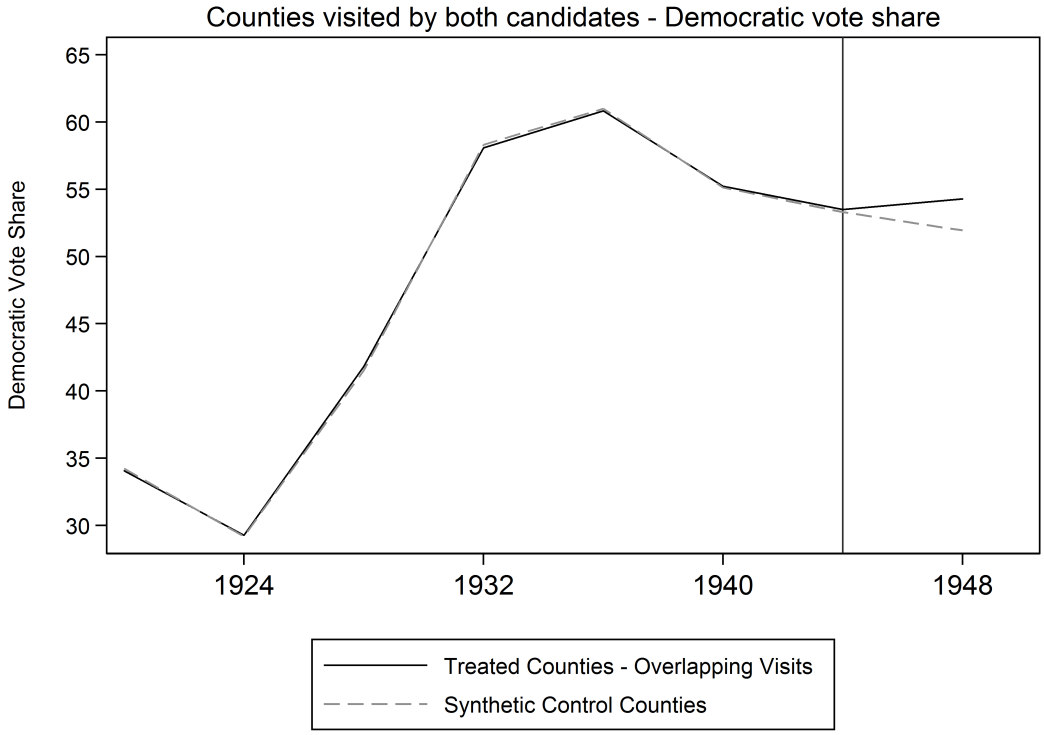

Figure A.11: The difference in Democratic vote share between double-visited counties and their synthetic control units, 1916-1948 (Treated n = 55). Treated units were visited by both Truman and Dewey during the 1948 campaign. The outcome variable for Truman's effect is Democratic vote share. The donor pool consists of the closest 800 matches to county i in terms of pre-treatment (1916-1944) Democratic vote share.

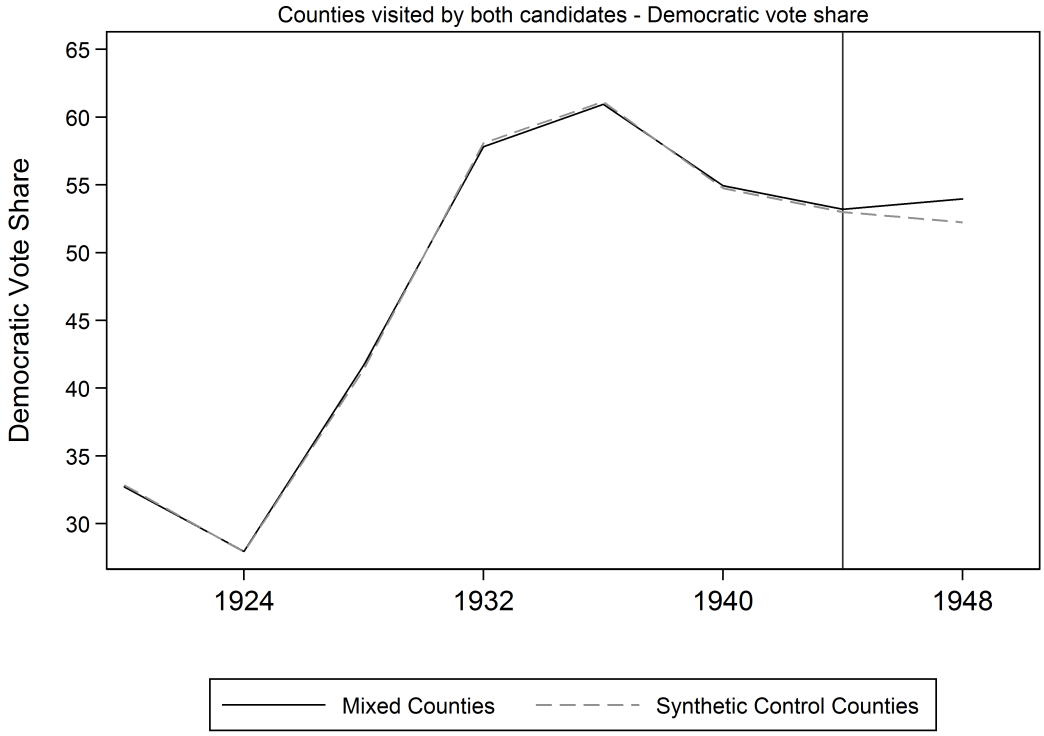

Figure A.12: The difference in Democratic vote share between double-visited counties and their synthetic control units, 1920-1948 (Treated n = 55). Treated units were visited by both Truman and Dewey during the 1948 campaign. The outcome variable for Truman's effect is Democratic vote share. The donor pool consists of the closest 800 matches to county i in terms of pre-treatment (1920-1944) Democratic vote share. Synthetic control models included state per capita income, change in state per capita income, percent black and Congressional vote share as predictor variables.

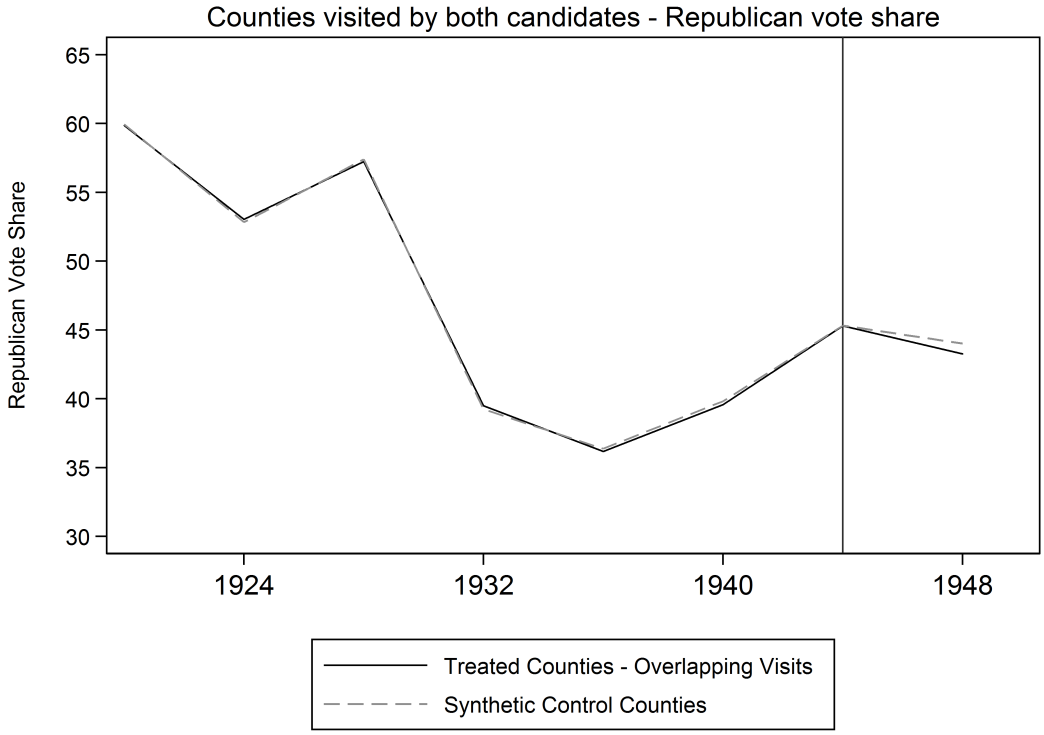

Figure A.13: The difference in Republican vote share between double-visited counties and their synthetic control units, 1920-1948 (Treated n = 55). Treated units were visited by both Truman and Dewey during the 1948 campaign. The outcome variable for Dewey's effect is Republican vote share. The donor pool consists of the closest 800 matches to county i in terms of pre-treatment (1920-1944) Republican vote share.

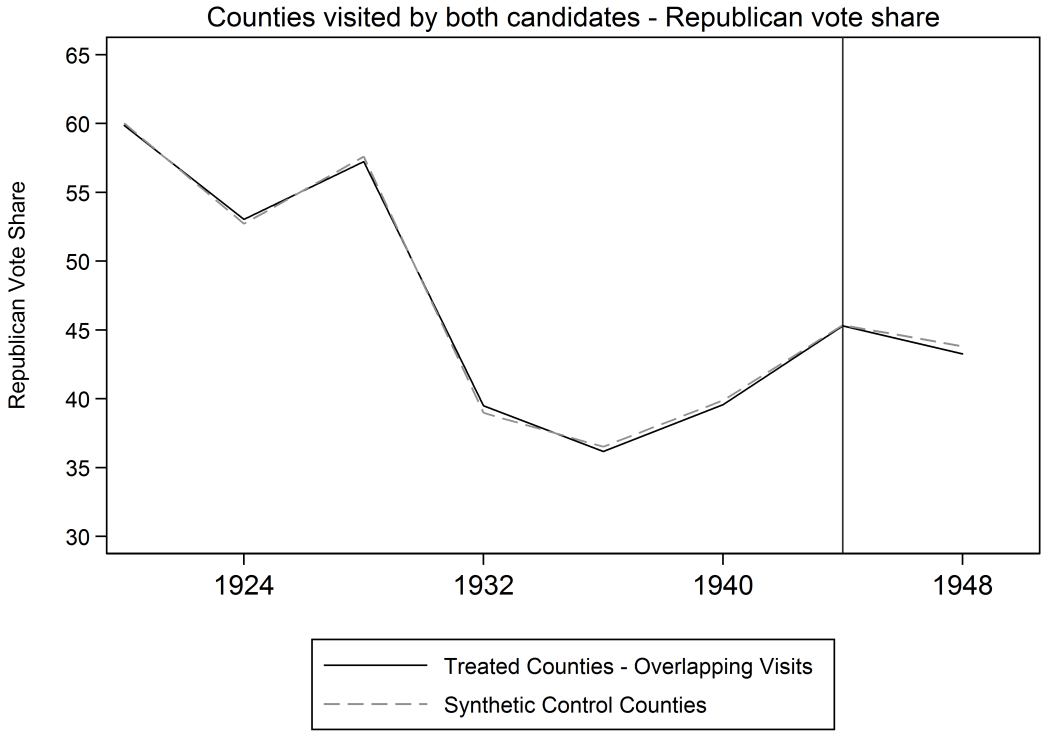

Figure A.14: The difference in Republican vote share between double-visited counties and their synthetic control units, 1920-1948 (Treated n = 55). Treated units were visited by both Truman and Dewey during the 1948 campaign. The outcome variable for Dewey's effect is Republican vote share. The donor pool consists of the closest 300 matches to county i in terms of pre-treatment (1920-1944) Republican vote share.

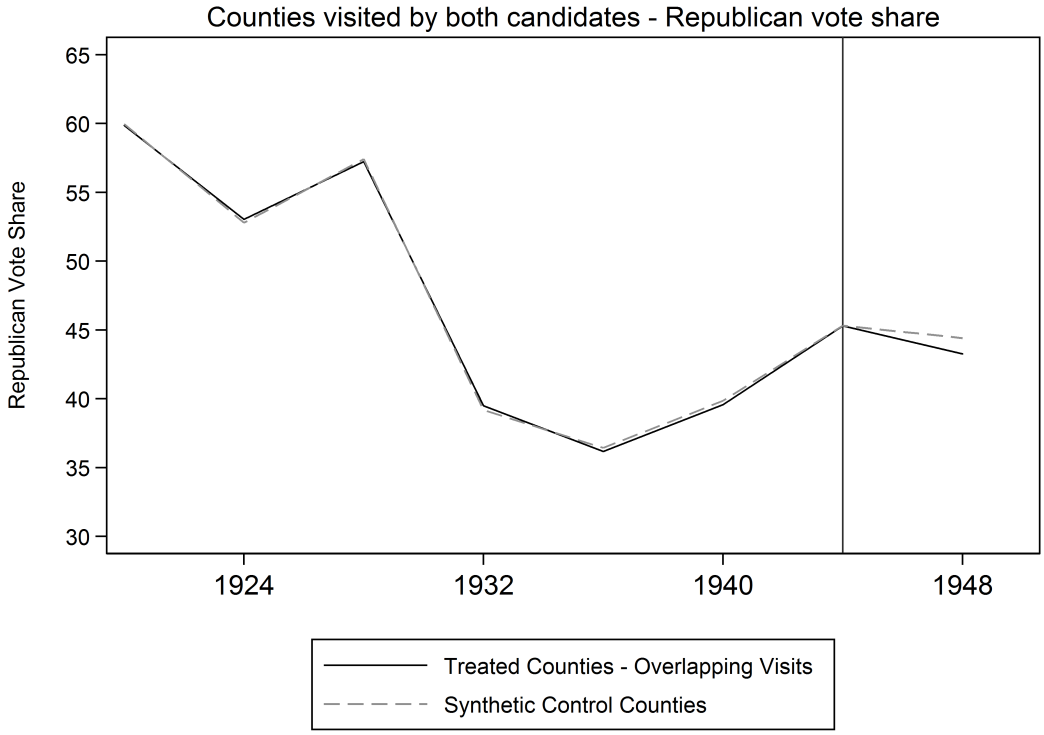

Figure A.15: The difference in Republican vote share between double-visited counties and their synthetic control units, 1916-1948 (Treated n = 55). Treated units were visited by both Truman and Dewey during the 1948 campaign. The outcome variable for Dewey's effect is Republican vote share. The donor pool consists of the closest 800 matches to county i in terms of pre-treatment (1916-1944) Republican vote share.

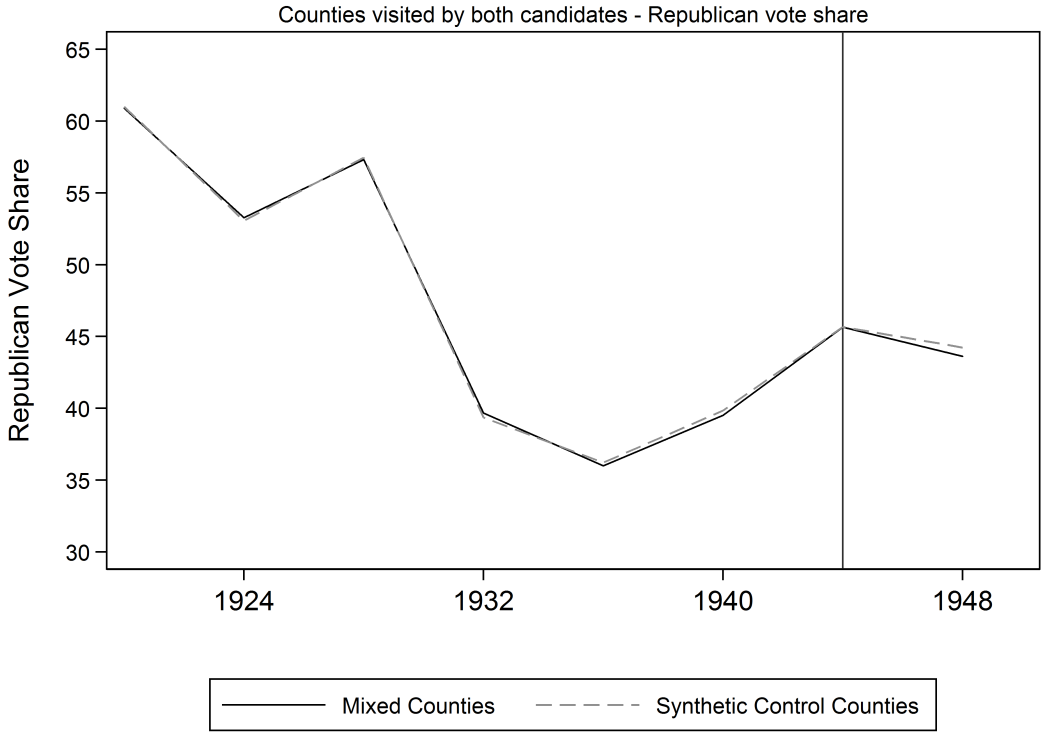

Figure A.16: The difference in Republican vote share between double-visited counties and their synthetic control units, 1920-1948 (Treated n = 55). Treated units were visited by both Truman and Dewey during the 1948 campaign. The outcome variable for Truman's effect is Republican vote share. The donor pool consists of the closest 800 matches to county i in terms of pre-treatment (1920-1944) Republican vote share. Synthetic control models included state per capita income, change in state per capita income, percent black and Congressional vote share as predictor variables.

B.3. Models of Adjacent County Effects

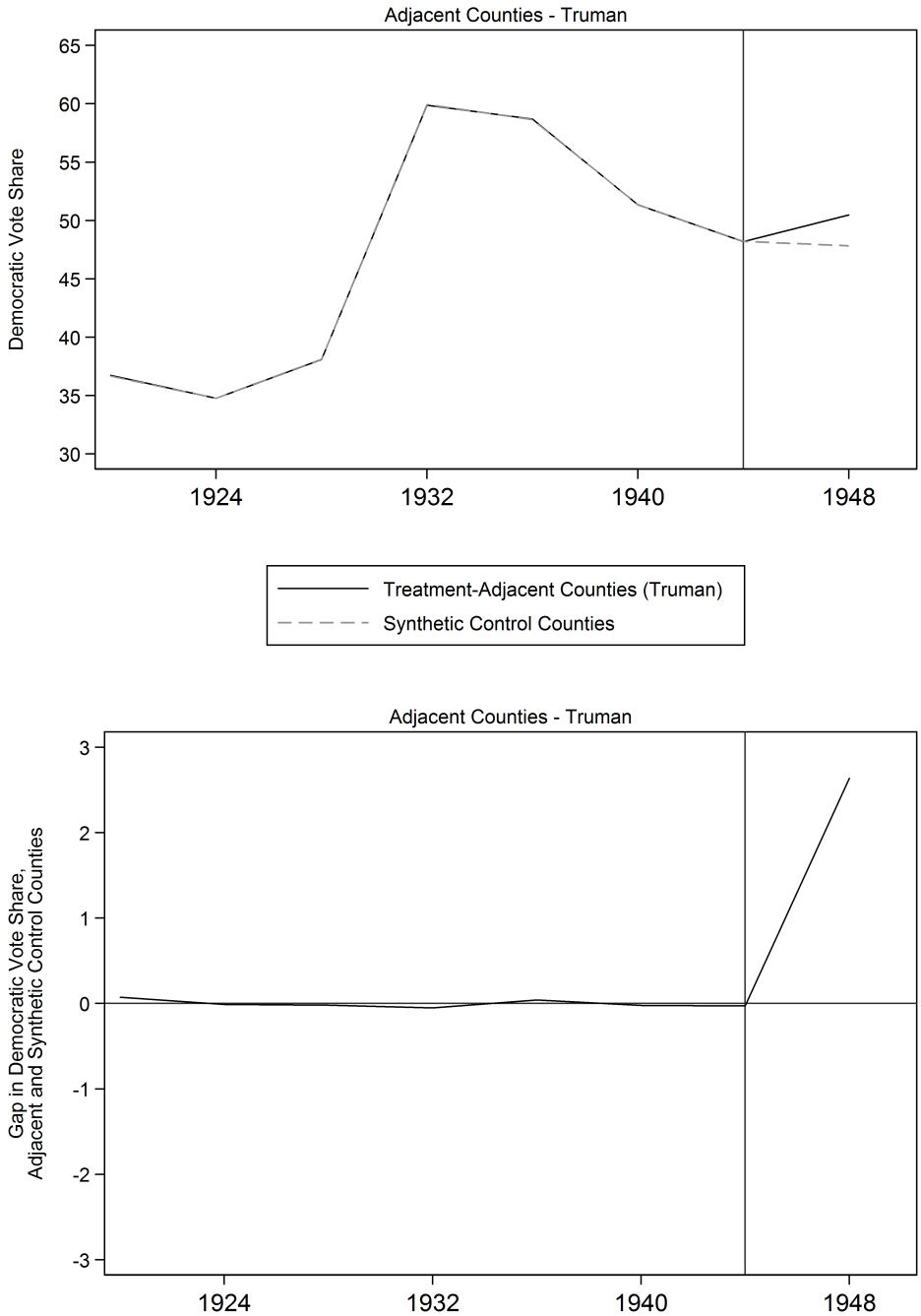

Figure A.17: Top panel: adjacent counties and their synthetic control units' Democratic vote share, 1920-1948 (Treated n = 357). Bottom panel: the difference between adjacent and synthetic control units' Democratic vote share, 1920-1948. Adjacent counties are those that are adjacent to a county visited exclusively by Truman. The donor pool consists of the closest 800 untreated matches to county i in terms of pre-treatment (1920-1944) Democratic vote share.

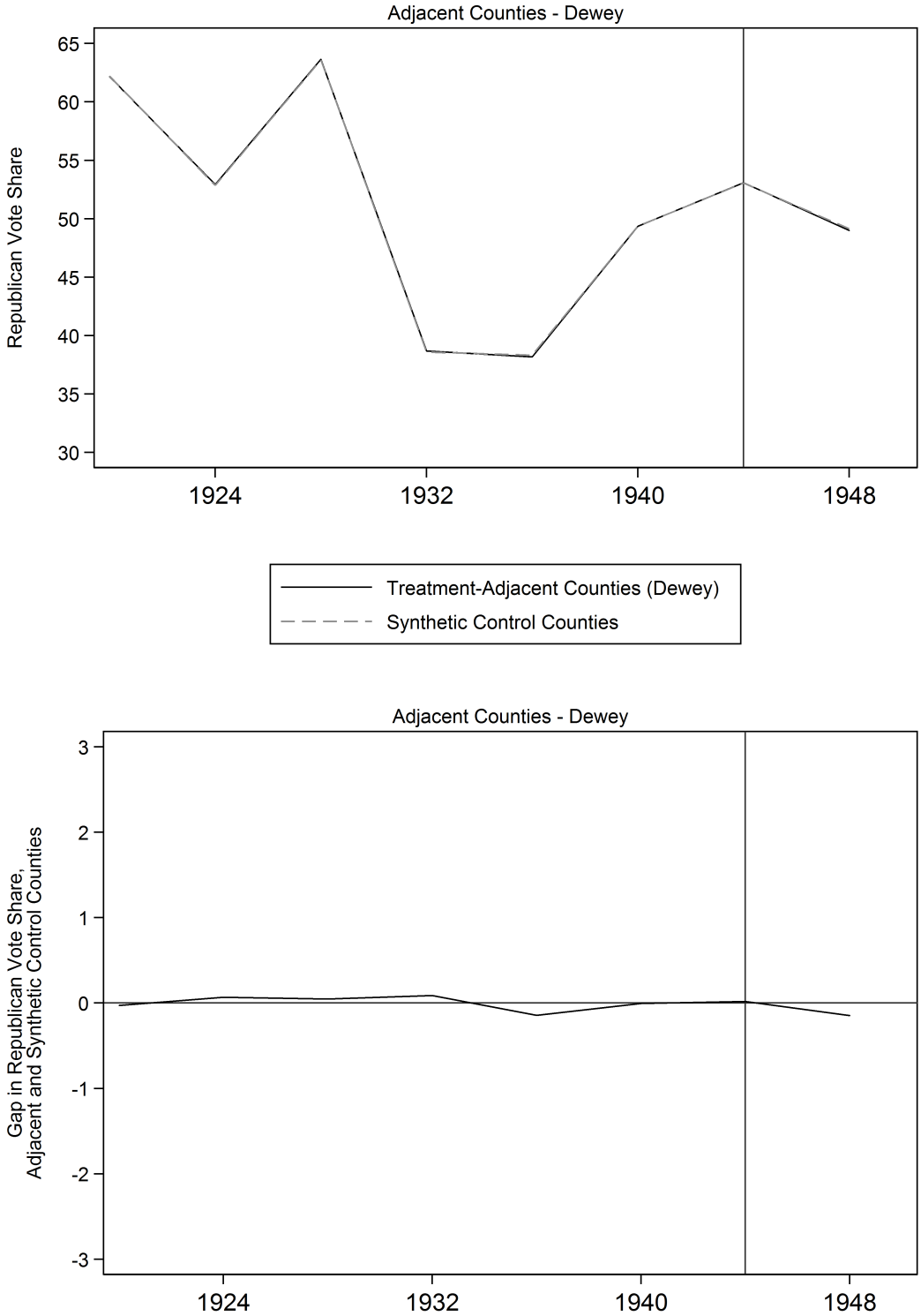

Figure A.18: Top panel: adjacent counties and their synthetic control units' Republican vote share, 1920-1948 (Treated n = 191). Bottom panel: the difference between adjacent and synthetic control units' Republican vote share, 1920-1948. Adjacent counties are those that are adjacent to a county visited exclusively by Dewey. The donor pool consists of the closest 800 untreated matches to county i in terms of pre-treatment (1920-1944) Republican vote share.

C. Additional Subgroup Effects

In the primary manuscript, we report a number of subgroup effects in Figure 7. Space limitations prevented us from reporting all relevant subgroup results, the remainder of which we include in this section.

The first set of results disaggregates our treated cases by the quality of matches with their synthetic control units. In Table A.2 we disaggregate treated effects by match quality, measured in terms of Mean Squared Prediction Error (MSPE). Our primary analysis uses all available treated cases for which we have the requisite data; the results of this analysis, discussed in detail in the primary manuscript and elsewhere, are reported for reference in the top row of Table A.2. In the second row, we limit the sample of treated counties to the 90 percent of cases with the lowest MSPE. We repeat this procedure for the top 70 percent of cases in terms of MSPE, the top 50 percent, and the top 30 percent. In the process, the number of treated cases declines, from 149 treated counties to just 44, in the case of counties visited exclusively by Truman. In general, our results are unchanged across these subsamples. We find no decline in the average effect size among Truman-visited counties when we restrict matches to those that are of the highest quality. Nor do we find that the null effect estimated in Dewey-visited counties changes appreciably when limiting the sample. Both findings imply that our results are not driven by any tenuous matches in our sample.

	Treatment Effect	
Percent of "Best Matches"	Truman Counties	Dewey Counties
100	3.06 (n = 149)	-.20 (n = 62)
90	3.38 (n = 134)	-.13 (n = 55)
70	3.34 (n = 104)	.44 (n = 43)
50	3.43 (n = 74)	-.01 (n = 31)
30	2.64 (n = 44)	-.13 (n = 18)

Table A.2: Estimated treatment effect of Truman and Dewey campaign visits, disaggregated by quality of pre-treatment match. Each row increasingly limits the sample of treated cases, with only the most closely matched cases—in terms of MSPE—remaining. At the most restrictive, the bottom row includes only the 30 percent of treated cases with the lowest MSPE over the pre-treatment period, 1920-1944.

A second set of results disaggregates treated cases according to their relative newspaper circulation. Subgroup analysis in the primary manuscript showed that treatment effects varied according to whether a treated county was the home of an active newspaper in 1948, with Truman experiencing larger electoral gains in treated counties that lacked a newspaper and Dewey experiencing larger drops in the same type of counties. However, not all newspapers are equal—a well-known and widely-read local newspaper will have a very different effect than an obscure outlet that few local residents are actually aware of or read.

To examine the moderating impact of newspaper circulation, we use data made available by Gentzkow, Shapiro, and Sinkinson on historical newspaper circulation.[6] Using this data, we estimate "newspaper penetration" for each treated county in our sample. Our measure is total circulation for all newspapers based in a county, over total county population. This measure is imprecise, because newspapers that are distributed across multiple counties often have large circulation numbers, which we attribute solely to the counties in which they are published. As a result, many counties—especially those that are the home of national newspapers like the New York Times—have penetration rates above 100 percent in our estimate. We cap this rate at 100 percent. While this obscures some degree of difference between counties at the upper end of newspaper penetration, our metric still provides a reasonable estimate of newspaper penetration among treated counties.

Figure A.19 plots county-specific treatment effects against county newspaper penetration rates. The top panel limits the sample to counties visited exclusively by Truman, while the bottom panel focuses on counties visited exclusively by Dewey. The subgroup results presented in the primary manuscript, which distinguished between counties with at least one newspaper and counties that lacked a newspaper altogether, are borne out here. The effect of a Truman visit is greatest in counties without a newspaper or with relatively low penetration rates. At the highest rates, 100 percent, the mean effect of

[6]We thank an anonymous reviewer for making us aware of this data and suggesting that we consider the importance of newspaper availability in shaping campaign visit effects.

a Truman visit declines to 1.26 percentage points. In contrast, Dewey's visits have the strongest negative effects in counties with low penetration rates. The mean effect of a Dewey visit only climbs above 0 percentage points at the highest level of penetration, as shown on the right-hand side of Figure A.19's bottom panel.[7]

[7]In additional analysis, reported in Section B1 of this Supplemental Appendix, we show that an alternative empirical strategy—matching treated counties to control counties partially on the basis of whether the county contains a newspaper—produces broadly similar results. Truman's largest effects were in counties without a newspaper, though he still gained over three percentage points, on average, in counties with one or more newspapers. Estimates of Dewey's impact were negative in both types of counties, but we find larger negative results in counties without a newspaper. Both findings are consistent with the subgroup analysis reported in Figure 7 of the primary manuscript and the analysis in this section.

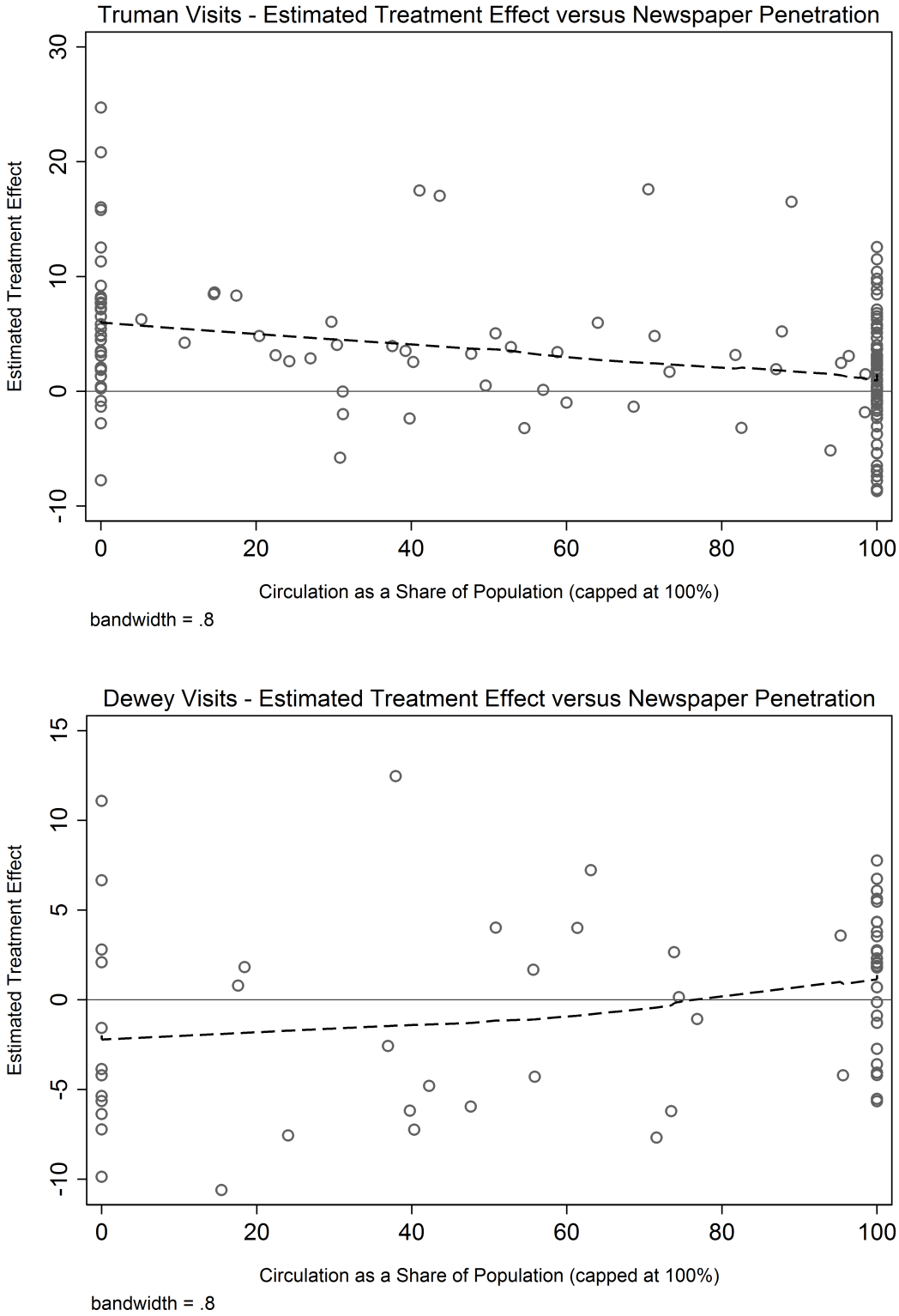

Figure A.19: Top panel: the treatment effect of Truman visits, as a function of county newspaper circulation (Treated n = 149). Bottom panel: the treatment effect of Dewey visits, as a function of newspaper circulation (Treated n = 62). Newspaper circulation is the total circulation of a newspaper based in county i, over total county population. Circulation rates are capped at 100 percent.

D. Placebo Test

The non-random nature of campaign visits raises the possibility that counties visited by Truman were becoming more Democratic over time. If this were the case, our estimate of the impact of Truman's campaign visits might overstate the true effect. For instance, if Truman tended to visit counties which had been safely Republican but were becoming increasingly competitive over time—in an attempt, perhaps, to "pick off" formerly safe Republican areas—our results could be biased upward by this trend.

Note that the synthetic control method specifically accounts for these types of heterogeneity between treated and control units, which motivated our research design. But placebo tests of this kind are standard in studies using synthetic control models. Focusing on the impact of Truman's visits, we study the 1944 election outcomes in counties that Truman visited in 1948. The central idea is that we should not observe an effect of Truman's visits on the 1944 election.

The typical approach is to estimate identical models in a pre-treatment period. Hence, we estimate models of the 1944 election that mirror our primary models in every way: they match counties over seven pre-treatment elections (1916-1940); each county is matched using a donor pool of 800 untreated counties, selected based on their similarity to the treated county over the pre-treatment period; and we use raw vote share, rather than two-party vote share. In general, we are skeptical of the value these tests add.[8] Nonetheless, we perform a placebo test using the Presidential election in 1944 to provide an additional robustness check of our results.

We estimate a treatment effect in 1944 for each county visited exclusively by Truman in 1948 (n = 150). Our expectation is that Truman's visits in 1948 could not impact

[8]By construction, the synthetic control units used in the placebo test differ from those of our primary models. Simply comparing treated units to their synthetic control units at time $t-1$ is uninformative, because the synthetic control units were chosen, in part, to minimize differences between treated and synthetic control units in $t-1$. However, using a different synthetic control unit—constructed to minimize differences over a slightly different time period—fundamentally changes the comparison in question. If we observe a substantial difference in Democratic vote share in 1944 it could reflect fundamental divergence between the treated and synthetic control units, suggesting that the effect found in time t is just the continuation of a pre-treatment trend. But the same pattern could be specific to the synthetic control units constructed in the placebo test, which differ from those used in the primary models. As a result, tests of this kind are only mildly informative.

the 1944 election, an expectation that is confirmed in Figure A.20. Our placebo estimate for Truman is a mere 0.32 percentage points. Using the same randomization approach to inference described in section D below, this estimate has a p-value of approximately 0.13. Relative to the substantive effect size estimated in 1948 (3.06) and its p-value ($p < 0.0001$), this test gives us additional confidence in our primary results. The effect we estimate in 1948 does not appear to be due to the continuation of a pre-treatment divergence among treated counties.[9]

[9] An equivalent test applied to Dewey's campaign stops yields a placebo estimate of -0.64 percentage points.

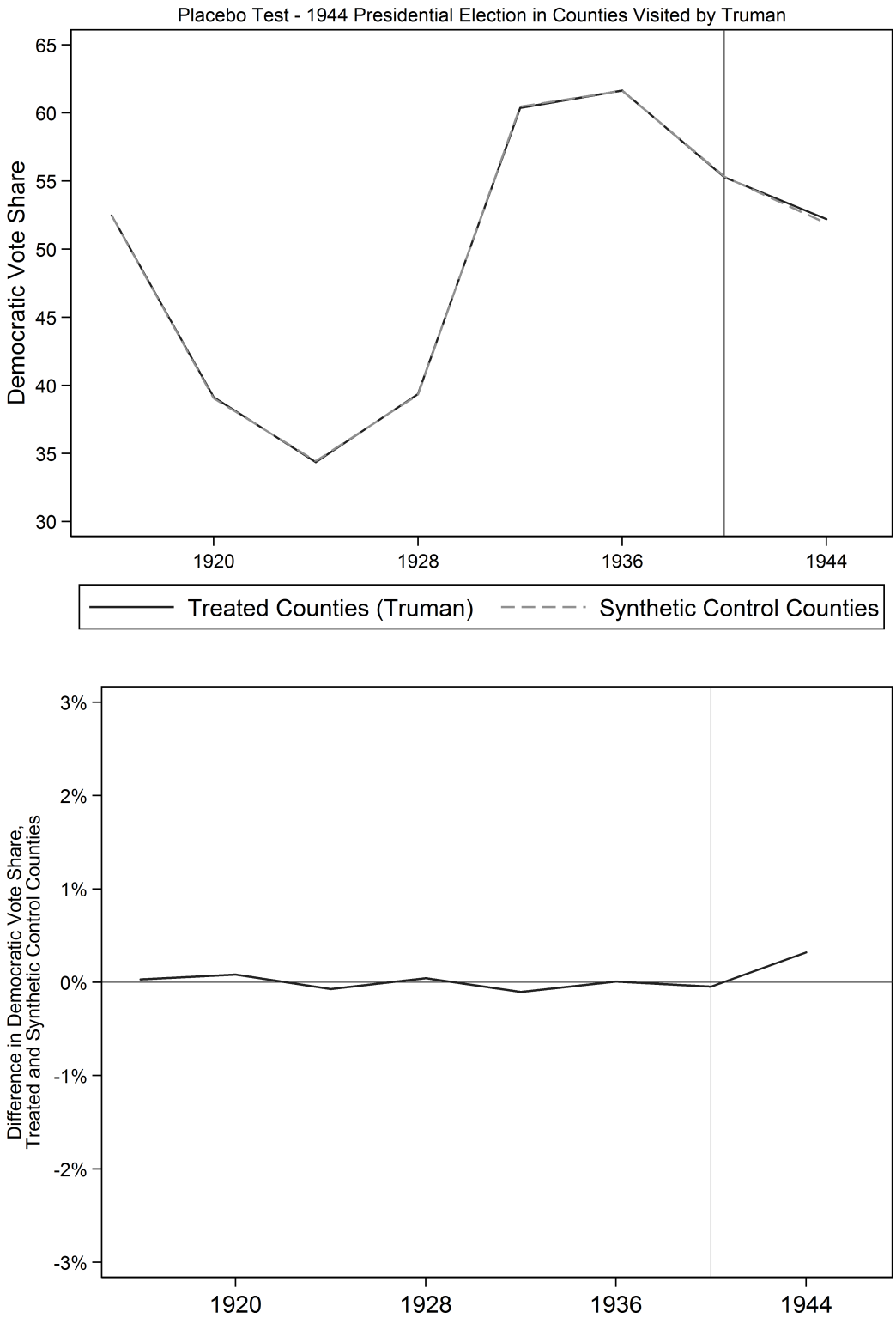

Figure A.20: A placebo test using the 1944 Presidential election, showing the impact of 1948 Truman campaign visits on 1944 Democratic vote share.

E. Statistical Inference

As we discussed briefly in the primary manuscript, no standard approach to statistical inference has emerged in the methodological literature concerning synthetic control models. Our preferred approach relies on a form of randomization inference, a framework for statistical inference first developed by R.A. Fisher (1935) and extended to non-experimental settings by Rosenbaum (2002). Inference in this framework produces an exact p-value (hereafter "Fisher p-value"), without relying on parametric assumptions. In this section, we describe the Fisherian approach to statistical inference and our application of it to the study of campaign visits; we report estimates of uncertainty surrounding our treatment effect estimates; and we briefly describe an alternative approach to inference recommended by Abadie, Diamond and Hainmueller (2010) and report a limited set of uncertainty estimates using this method.

To illustrate Fisherian randomization inference, consider a simple experimental setting, with five treated and five control units, in which the experimenter has randomized assignment to treatment and control. After completing the experiment and measuring the outcome, the researcher calculates the difference-in-means or other test statistic of interest (we focus on the difference-in-means, which we refer to as the SATT). Thanks to the virtues of randomization, potential confounders are balanced between treated and control groups. As Fisher (1935) proved, randomization also provides a basis for estimating the extremeness of our estimated treatment effect.

Assume a restrictive null hypothesis, that the treatment effect for each unit in our experiment is zero. If this null hypothesis were true, what would the distribution of estimates of SATT be? We can calculate this distribution by re-assigning units to treatment and control groups repeatedly and re-calculating the SATT under alternative assignments. This method takes advantage of the fact that, assuming the null hypothesis is true, we know the potential outcomes for each unit under both treatment and control.[10]

[10]Imagine a unit that, in our experiment, was assigned to treatment. It's observed outcome was a 45 percent Democratic vote share. Under the null hypothesis, this unit's outcome under an alternative assignment, in which it was part of the control group, would be 45 percent, because the treatment is assumed to have no effect on any unit.

By completing many iterations of this reassignment, we can construct the empirical null distribution—the distribution of estimates of the SATT that would obtain if the null hypothesis were true. We then compare our true estimate to this empirical null distribution to judge how "extreme" of a value we observed, and how likely it is to occur if the null hypothesis were true. The p-value for our estimate is the percentage of iterations with more extreme values than our true estimate.[11]

We use this approach to calculate Fisher p-values for all results reported in the primary manuscript and this appendix. To illustrate, consider the set of models estimating the impact of Truman's exclusive visits to 149 counties, reporting in the manuscript. The estimated SATT is 3.06 percentage points. We treat the 149 treated counties and their synthetic control units as a matched-pair design, in which each treated unit is matched to a single control unit.[12] To calculate the empirical null distribution, we reassign placebo-treatment *within* these matched pairs and calculate the placebo-SATT. We perform 5,000 iterations. The empirical null distribution for this particular set of models is shown in the top-left panel of Figure A.21, which plots the distribution of placebo effects in gray and marks the observed effect from our models in red. As this figure shows, our estimate was larger than all 5,000 iterations, implying a p-value equal to or less than .0002. The other panels in Figure A.21 illustrate results for Dewey's campaign visits, and different model specifications, as described in Section B.

Note that the p-values reported are also provided in Table A.1, in Section B. This section merely illustrates the full empirical null distribution used for conducting statistical inference. In addition to models of "exclusive visits," or counties visited by just one candidate, Figure A.22 provides results for inference related to counties visited by both candidates (i.e. "overlapping visits"). Finally, we also calculated p-values for models focused on spillover effects, estimating the impact of a visit by either Truman or Dewey

[11]The restrictive null hypothesis employed—that the treatment effect for all units is equal to zero—is an important drawback of the randomization inference approach. For excellent discussions of the approach, see Bowers and Panagopoulos (2011) and Keele, McConaughey, and White (2012).

[12]Importantly, we follow Rosenbaum (2002), who showed that randomization inference could be applied outside of experimental settings under specific assumptions of non-confounding, an assumption that is justified by the quality of matches that we are able to construct using the synthetic control method.

on vote share in adjacent counties, reported in Figure A.23.

Our approach, while relying on an established method of statistical inference, is not the only approach used in the applied synthetic control literature. Abadie, Diamond and Hainmueller (2010), in their seminal paper, suggest that researchers use a series of placebo tests to judge the uncertainty of their estimates. Specifically, they suggest an approach that requires three steps when applied to cases with multiple treated units. First, calculate a placebo synthetic control model for every untreated unit, using the same model specification as the primary models of interest. Second, take random draws of n placebo-treated units, where n is the number of treated units in the primary models of interest, and average their placebo treatment effects to calculate the SATT. Third, repeat this process for many permutations, to calculate a distribution of placebo treatment effects that would obtain under the null hypothesis, and compare the observed treatment effect to this distribution.

The Abadie, Diamond and Hainmueller approach, though originally described in the context a single treated unit, has been applied to cases with multiple treated units in research by Acemoglu et al. (2016) and Heersink and Peterson (2016), among others. Unfortunately, the properties of this approach are not fully proven, and there is lingering disagreement in the methodological literature about the proper approach to inference (see, e.g. Ando and Savje 2013). For this reason, we focus on Fisher's randomization inference. However, for the sake of completeness, we replicated our inferences regarding the impact of Truman's campaign visits (our primary models) using Abadie, Diamond and Hainmueller's approach. The full results and code are available upon request, but we performed 5,000 draws of 149 counties and calculated the average placebo treatment effect for each draw. Our observed treatment effect, a 3.06 point increase in Democratic vote share, was larger than all but 24 of the iterations, resulting in a two-sided p-value of .0096.

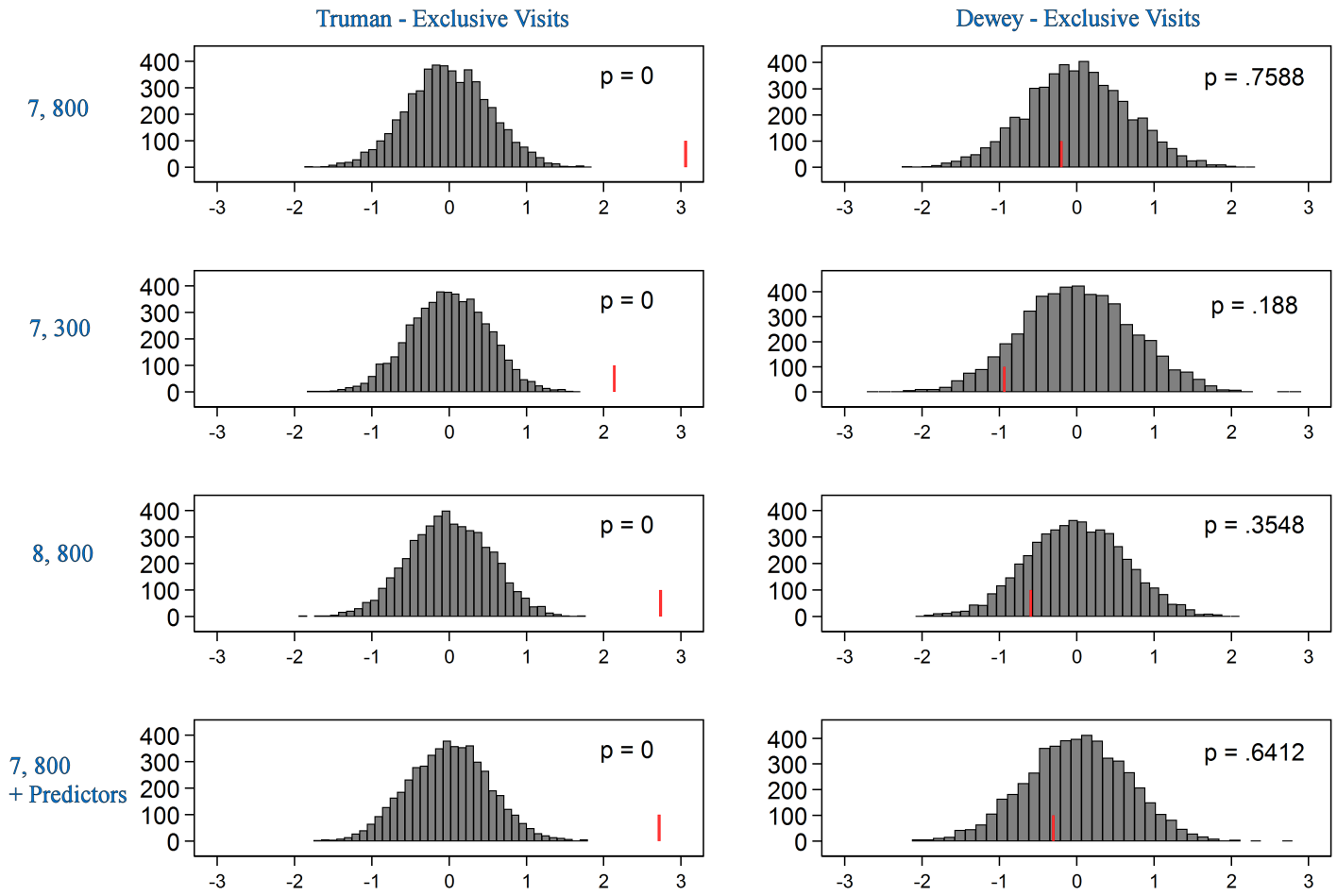

Figure A.21: Results of randomization inference for models of Truman and Dewey's "exclusive visits" (i.e. counties visited by one, but not both, candidates). Effects in counties visited by Truman are shown in the left column, and effects in counties visited by Dewey are shown in the right column. Alternative model specifications are shown across rows, with changing numbers of election years (7 or 8), sizes of donor pools (300 or 800 counties) and the inclusion of predictor variables.

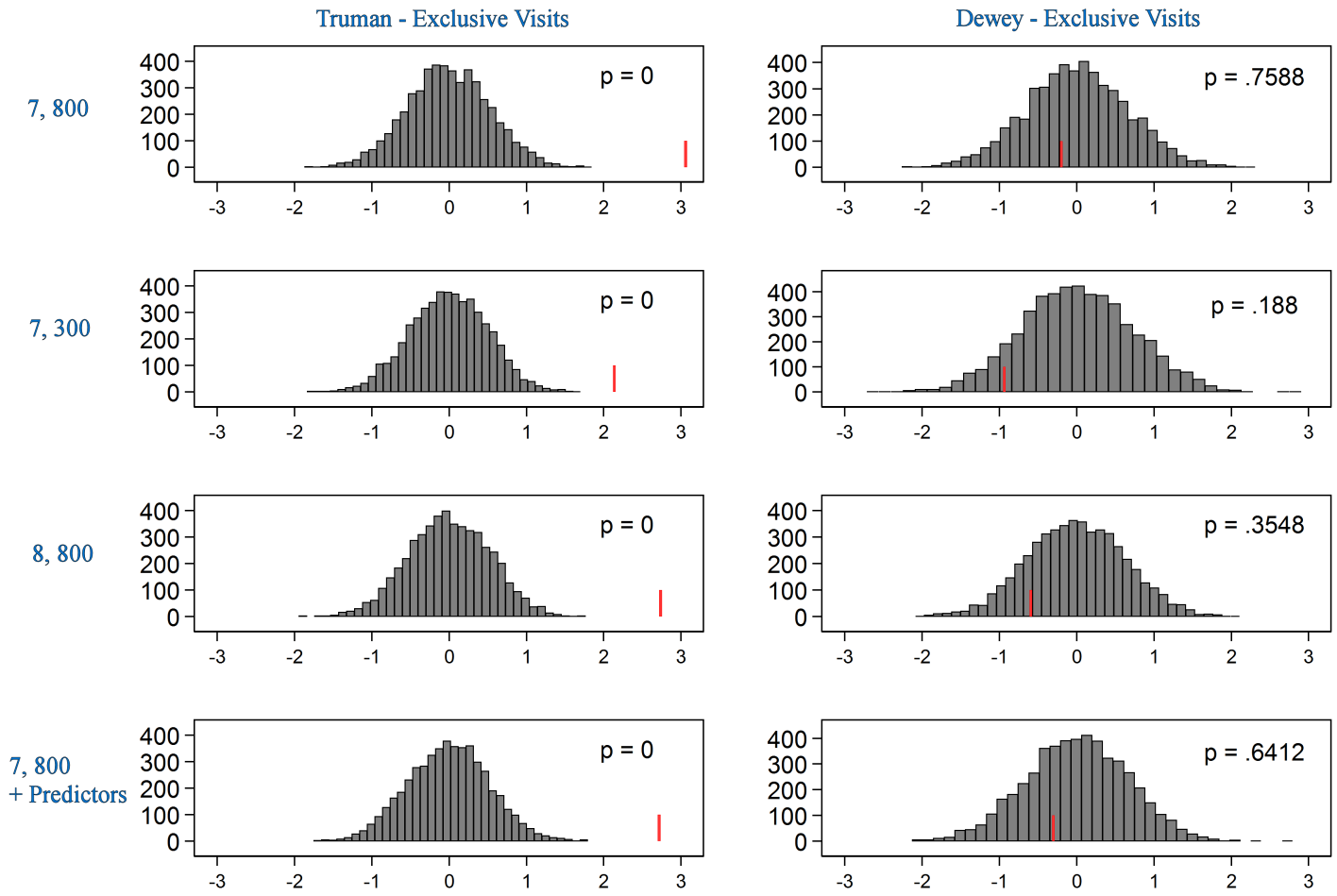

Figure A.22: Results of randomization inference for models of Truman and Dewey's "overlapping visits" (i.e. counties visited by both candidates). Effects in counties visited by Truman are shown in the left column, and effects in counties visited by Dewey are shown in the right column. Alternative model specifications are shown across rows, with changing numbers of election years (7 or 8), sizes of donor pools (300 or 800 counties) and the inclusion of predictor variables.

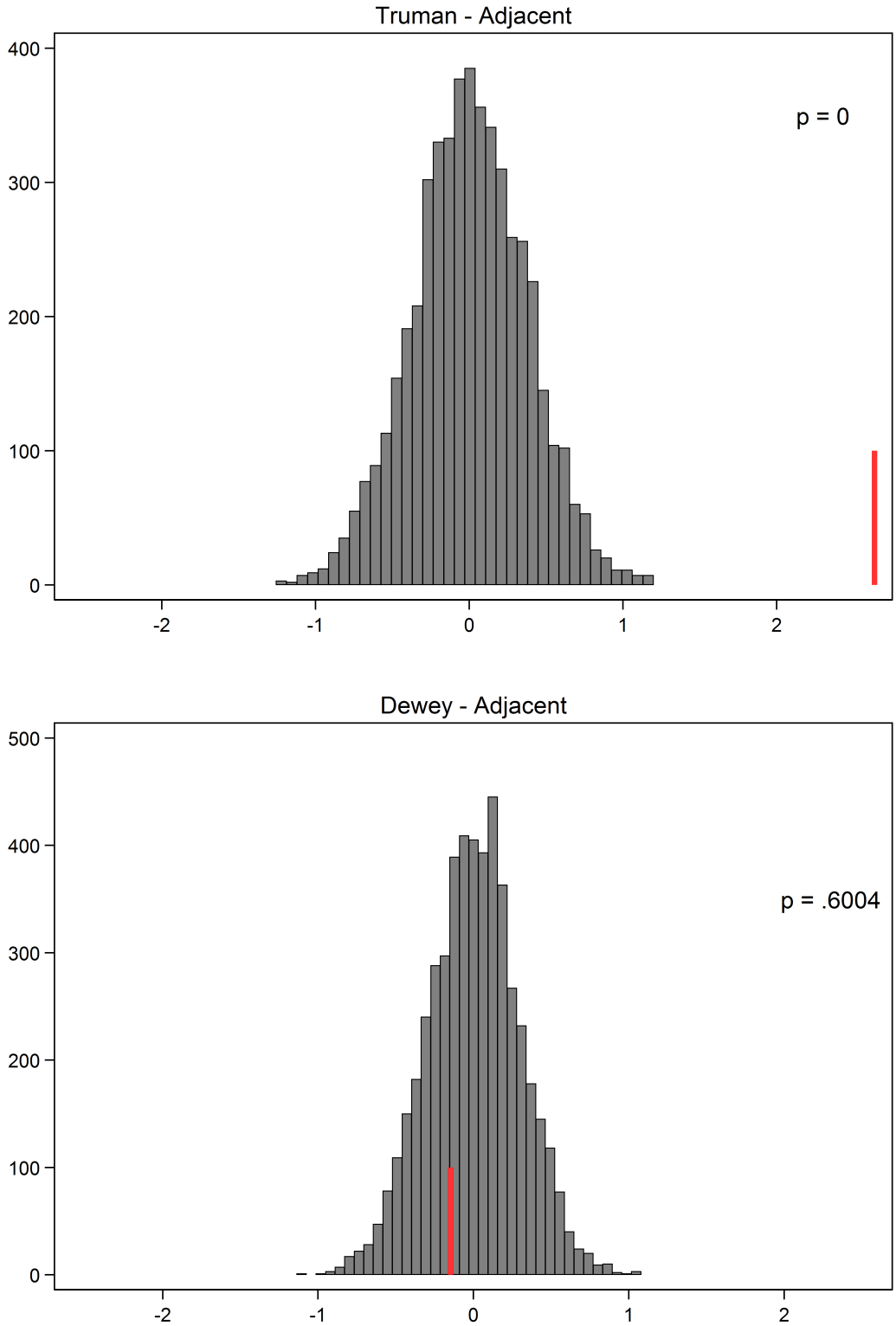

Figure A.23: Empirical null distributions for models of treatment effects in adjacent counties. The top panel is the distribution and treatment effect for counties adjacent to a Truman visit; the bottom panel provides the same information for counties adjacent to a Dewey visit.

F. Counterfactual Simulations

In the primary manuscript, we report the results of a simulation that attempts to shed light on how Truman's campaign scheduled might have shifted electoral outcomes in three critical, closely-contested states. Specifically, we investigated election outcomes in California, Illinois and Ohio, three states that Truman won with less than 50.5 percent of the two-party vote. The closest of these three states, Ohio, was won by just 7,171 votes, with Truman gaining 50.12 percent of the two-party vote.

Our simulation assessed whether Truman would have won these three states if he had eliminated his campaign stops in each completely. Our results suggested that Dewey likely would have won Ohio, if not for Truman's vigorous campaign calendar in the state; they also suggested that, in some scenarios, Dewey may also have won California and Illinois and—by extension—the presidency.

The simulations reported in the paper concerned two possible scenarios: one in which Truman eliminated all of his "exclusive" stops in a state, and one in which he eliminated *all* stops in a state, including those that overlapped with Dewey's visits. In the top panel of Table A.3, we investigate two less-drastic scenarios, in which Truman reduces his campaign visits in a state in half. Because the number of voters varies across counties, the outcomes of these simulations depend on which visits are eliminated. We approach this issue systematically by randomly selecting which half of visits to eliminate. We repeat this process 1000 times for each state, eliminating a randomly-selected half of all visits, calculating the total votes Truman would have won, and determining whether this was sufficient to hold onto the state. We report expected outcomes when Truman eliminates half of his exclusive visits to a state (row 2) and half of *all* visits to a state (row 4) as a percentage—the share of iterations in which the state would have flipped. Rows 1 and 3 repeat the results reported in Table 2 of the primary manuscript for reference. The outcomes follow the same pattern as those observed when eliminating all visits: we are relatively confident that Ohio would have been won by Dewey if Truman had eliminated either half or all of his visits in the state. We are less confident that California or Illinois

Visits Eliminated	California	Illinois	Ohio
	Counterfactual: Average Effect (SATT)		
All Exclusive Visits	No Change	No Change	Flip
1/2 Exclusive Visits	0% Flip	0% Flip	97.7% Flip
All Visits	Flip	Flip	Flip
1/2 Visits	73.7% Flip	49.4% Flip	99.4% Flip
	Counterfactual: County-Specific Effects		
All Exclusive Visits	No Change	No Change	Flip
1/2 Exclusive Visits	0% Flip	0% Flip	65.5% Flip
All Visits	No Change	No Change	No Change
1/2 Visits	0% Flip	0% Flip	33.2% Flip

Table A.3: Simulations of Truman's performance in three competitive states under varied counterfactuals. Democratic vote share in county i is assumed to decline by an amount equal to the average effect or SATT (top panel) or the county-specific treatment effect estimated for each county (bottom panel). Simulations of a reduction in Truman's visits by half were performed with 1000 iterations by randomly assigning half of Truman-treated counties in each state to the counterfactual "no visit" condition.

would have changed columns, but it is possible.

Thus far we have assumed that—in the absence of a visit by Truman—Democratic vote share in each county would have fallen by 3.06 percentage points. That is, we assumed that the average treatment effect applied to each county. Of course, not all counties were affected equally by a campaign visit. An alternative assumption is to use the county-specific treatment effects provided by the synthetic control method instead, allowing for heterogeneous treatment effects across counties. We consider this assumption less defensible because our confidence is highest in the aggregate treatment effect that we estimate, rather than county-specific estimates that are based on a single case. Nonetheless, we report the results of our counterfactual simulation using this alternative assumption in the bottom panel of Table A.3.[13]

[13]Both simulation approaches make a series of other, extremely conservative assumptions, as outlined in our paper, including assuming a null effect in adjacent counties, and assuming that votes lost by Truman through an eliminated visit are not redistributed to Dewey. These two assumptions both

The results of this simulation differ somewhat from those reported in the primary manuscript. In this simulation, Truman wins California and Illinois in nearly every case. Importantly, however, Ohio is still vulnerable. Even in cases where Truman reduces his campaign schedule in the state by half, he would have lost the state in 65.5 percent or 33.2 percent of iterations, depending on which visits were dropped.

understate the impact of Truman's visits, and work against the possibility of finding a state that flips in our counterfactual simulation.

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* 105:493–505.

Abramowitz, Alan I. 1988. "An Improved Model for Predicting Presidential Election Outcomes." *PS: Political Science & Politics* 21(4):843–847.

Acemoglu, Daron, Simon Johnson, Amir Kermani, James Kwak and Todd Mitton. 2016. "The Value of Connections in Turbulent Times: Evidence from the United States." *Journal of Financial Economics* 121(2):368–391.

Ando, Michihito and Fredrik Savje. 2013. "Hypothesis Testing with the Synthetic Control Method." Unpublished working paper. Available at: http://www.eea-esem.com/files/papers/eea-esem/2013/2549/scm.pdf.

Barro, Robert J. and Xavier Sala i Martin. 1992. "Convergence." *Journal of Political Economy* 100(2):223–251.

Bowers, Jake and Costas Panagopoulos. 2011. "Fisher's Randomization Mode of Statistical Inference, Then and Now." Unpublished working paper. Available at: www.jakebowers.org/PAPERS/BowPan-Fisher.pdf.

Fisher, Ronald A. 1935. *The design of experiments.* London: Oliver and Boyd.

Heersink, Boris and Brenton D. Peterson. 2016. "Measuring the Vice-Presidential Home State Advantage with Synthetic Controls." *American Politics Research* 44(4):734–763.

Jordan, David M. 2011. *FDR, Dewey, and the Election of 1944.* Bloomington, IN: Indiana University Press.

Keele, Luke, Corrine McConnaughy and Ismail White. 2012. "Strengthening the Experimenter's Toolbox: Statistical Estimation of Internal Validity." *American Journal of Political Science* 56:484–499.

Nadeau, Richard and Michael S. Lewis-Beck. 2001. "National Economic Voting in U.S. Presidential Elections." *Journal of Politics* 63(1):159–181.

Rosenbaum, Paul R. 2002. "Covariance Adjustment in Randomized Experiments and Observational Studies." *Statistical Science* 17:286–327.