Supplemental Appendix for:

Disasters and Elections: Estimating the Net Effect of Damage and Relief in Historical Perspective

November 8, 2016

Contents

A. Data and Variable Construction

Our raw data come from four primary sources. For county-level Presidential election results, we used data from ICPSR's Historical Election Returns collection.[1] For our measure of the black population in each county, we used the 1920 Census. To estimate the level of Protestantism in each county, we used the Census of Religious Bodies from 1926, conducted by the US Census Bureau and provided via ICPSR.[2] Finally, we gathered data on the extent and severity of flooding, as well as the relief efforts, from a Red Cross report written in 1929.[3]

The 1926 Census of Religious Bodies reports the number of members of various congregations at the county level, but—because it was not a full population census—provides no denominator in the form of county population. We interpolate county population—using the 1920 and 1930 Censuses—to the year 1926, to calculate the percentage of each county's population that adheres to a Protestant faith. We define Protestantism broadly to include Churches of Christ, who were more closely tied to other Protestant faiths at this time than they are today, as well as Anabaptists and Unitarians. The latter category in the 1926 Census data presumably refers to Biblical Unitarians, since the Universalist Unitarian movement did not exist in 1926.

To measure flood severity, we employ the Red Cross report on the number of individuals affected by the flood. This measure is reported as a raw count at the county level, which we convert to a percentage using total county population. Again, we interpolate county population using the 1920 and 1930 Censuses, this time to the year 1927. We also gathered data on land acreage flooded, but focused our analysis on population impacted because it is a better indicator of actual suffering by voters in the affected areas.

Finally, to measure relief efforts, we use the Red Cross' data on families provided aid.

[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] U.S. Dept. of Commerce, Bureau of the Census. Censuses of Religious Bodies, 1906-1936. ICPSR ed. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [producer and distributor], 1980. http://doi.org/10.3886/ICPSR00008.v1.

[3] "The Mississippi Valley Flood Disaster of 1927: Official Report of Relief Operations of The American National Red Cross." The American National Red Cross, 1929.

Because the Red Cross reported on families rather than individuals, we use 1920 Census data on average family size at the county level, as well as total population, to convert this measure to the percentage of individuals receiving aid in a county.[4]

[4]Specifically, Aid in county i in our data is measured as:
$Aid_i = \frac{A_i \times FS_i}{Pop_i} \times 100$,
where A_i is the number of families receiving aid in county i, FS_i is average family size in county i and Pop_i is total population in county i.

B. Framework for Statistical Inference

The best method for conducting statistical inference in synthetic control models is an open question. Abadie, Diamond and Hainmueller (2010) recommend placebo-based inference in research designs with a single treated case. In these situations, the unit treatment effect can be compared to placebo-treatment effects for all available untreated units using the same specification, allowing researchers to calculate exact p-values.

Because our research question diverges from the case study approach described by Abadie, Diamond and Hainmueller (2010), we conduct inference in two ways. The first is a version of Abadie, Diamond and Hainmueller's recommended approach, applied to cases with multiple treated units, as in Acemoglu et al. (2016). The second applies Fisher's randomization inference (1935) in the manner proposed by Rosenbaum (2002) and implemented in the synthetic control context previously by Heersink and Peterson (2016). We label these alternatives "ADH p-values" (Abadie, Diamond and Hainmueller) and "Fisher p-values," respectively.

To compute ADH p-values, we first estimate placebo models for each unit in our control groups. Consider our primary specification, which uses a restricted donor pool (Donor Pool #1) and limits the sample to southern counties: our treatment group consists of all flooded southern counties; our overall control group consists of southern counties that were not flooded and were not contiguous with flooded counties.[5] We are left with 95 treated and 618 control counties. We estimate a placebo model for each of 618 control counties, using the other 617 control counties as potential matches; we follow the same specification used for treated counties by restricting the donor pool to the closest 100 matches on the basis of pre-treatment vote share (see description in Section F). After computing unit treatment effects for all 618 control counties, we take random draws of 95 control counties and calculate the average placebo treatment effect, repeating this process for 10,000 draws.[6] We calculate a p-value by comparing our treatment estimate (the

[5]We also eliminate counties that are "second-order contiguous," i.e. those counties that are contiguous to contiguous-to-flooded counties. Finally, we also eliminate counties without a full time-series of Presidential election results and data on black and Protestant populations.

[6]Note that we do not include treated counties in our simulations at all—repeated draws of 95 counties

SATT) to the distribution of average placebo treatment effects across 10,000 iterations. We repeat this process for each of our specifications to generate p-values reported in Section F.

In contrast to ADH p-values, Fisher p-values are not calculated with reference to placebo effects among the control group. The ideas behind randomization inference date to Fisher's early work on experimental design (1935). Using randomization inference, researchers are able to calculate exact p-values without making parametric assumptions, based strictly on the physical act of randomization. The basic process involves randomly re-assigning units to treatment and control groups and re-calculating the treatment effect many times. Under the sharp null hypothesis—no treatment effect for any unit—we can calculate an empirical null distribution of potential treatment effects that would occur under alternative versions of randomization.[7] By comparing our estimate to this distribution, we can calculate an exact p-value.

Importantly, Fisher's method was developed with reference to "physical randomization," and has seen its greatest use in experimental research. However, as Rosenbaum (2002) shows, Fisher's method can be applied to observational studies if treatment assignment is unconfounded. Because our use of the synthetic control approach accounts for confounding from both observed and unobserved covariates, a comparison of treated units *to their synthetic control units* meets Rosenbaum's criterion.

Again, consider our primary specification, which restricts the sample to southern counties. We have 95 pairs of units, each consisting of one treated county and one synthetic control county. We randomly assign treatment within pairs and calculate our test statistic—the average difference between treated and synthetic control units. We repeat this process 10,000 times to generate the empirical null distribution, calculating a "Fisher p-value" by comparing our treatment effect estimate to the null distribution.

We note possible objections to our approach: neither approach to calculating p-values

are limited to control counties. The likely effect is to understate the p-value we calculate, a good reason to focus on both the Fisher p-values we present and the substantive effect sizes in question.

[7]The sharp null hypothesis described here is restrictive and, ultimately, a lower bar than the more typical null hypothesis of "no average treatment effect."

produce typical p-values strongly supported by theory. We hope that researchers will continue to develop methods for statistical inference that can be utilized alongside the synthetic control method. Our preference is to focus on substantive effect sizes; however, we provide estimates of uncertainty in the form of p-values for the sake of transparency. We do not report p-values in the primary manuscript—including them only in the Supplemental Appendix—to focus attention on substantive effect sizes and to increase brevity.

C. Disaster Severity and Aid Delivery

In this section of the Supplemental Appendix, we report the results of several linear models predicting the distribution of relief aid from the Red Cross. Figure 2 in the primary manuscript shows the strong correlation between flood severity (percent of population affected by the flood) and relief efforts (percent of population receiving aid from the Red Cross). These models assess whether obvious political motivations shaped the relief effort.

We do not find any evidence that the racial composition of counties influenced the extent of relief that they received, nor do we find that Republican vote share in 1924 influenced the distribution of relief in 1927. In the south-only sample, 1924 vote share is marginally significant at the 10 percent level. However, the direction of effect is the opposite we would expect to see if the Republican Coolidge administration were distributing relief to its supporters—a higher Republican vote share in 1924 is correlated with lower levels of relief. More importantly, any political considerations are overpowered by the correlation between flood severity and the relief effort. Bivariate models of the correlation have high R^2 values, which are not improved by the inclusion of additional political and demographic variables.

Table A.1: Models Predicting Receipt of Relief

	Full Sample		South-Only Sample	
	(1)	(2)	(3)	(4)
Flood Severity	0.691***	0.676***	0.699***	0.696***
	(0.02)	(0.03)	(0.03)	(0.03)
Repub. Vote Share, 1924		-0.085		-0.147
		(0.05)		(0.08)
Pct. Black		-0.005		-0.023
		(0.03)		(0.04)
N	133	133	98	98
R^2	0.865	0.868	0.865	0.870

Linear models of relief efforts. Relief is the percentage of county population receiving aid from the Red Cross. Sample is limited to flooded counties.
Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

D. Alternative Operationalizations of Flood Severity

In this section, we consider a wide range of operational definitions for flood severity. First, we model flooding as a binary treatment variable, to estimate the net effect of the flood on vote share, averaging across all flooded counties. Our estimates suggest that, in the mean flooded county, Hoover lost 10.8 percentage points compared to the counterfactual outcome (Table A.2, Model 2). The treatment effect appears to have spilled over into contiguous counties, which witnessed a 7.4 percentage point drop in Republican vote share, compared to expectations (Table A.2, Model 3).

It is important to note that effect estimates reported in the primary manuscript are presented in this section alongside many estimates that were excluded. Specifically, Figure 3 in the primary manuscript includes treatment effect estimates from a model with a binary treatment variable (Table A.2), a model with subjective cutpoints (Table A.3, Column 1), a model splitting treatment severity into three quantiles (Table A.3, Column 3) and a model splitting treatment severity into five quantiles (Table A.4, Column 3). In the manuscript's discussion and conclusion, we focus on the effect estimate from a binary model (Table A.2, Model 2) because it is a straightforward quantity, easily interpretable, restricted to the most relevant within-region (southern) comparison, and is also among the more conservative estimates across our many specifications.

To investigate treatment effect heterogeneity, we estimate a series of models in which the treatment variable is divided into a number of categories. As discussed in the primary manuscript, we use both subjective and objective cutpoints. These models all take the following basic form:

$$v^i_{1928} - v^i_{1924} = \beta_0 + \alpha\gamma_i + \theta\delta_i + \mu,$$

where γ is a vector of covariates and δ is a vector of treatment dummies. In this difference-in-differences model, we focus on the value of θ—in some specifications we incorporate a single binary treatment dummy; in other specifications we separate treatment by severity according to quantiles and subjective judgments. Subjective cutpoints divide the sample into four categories of flood severity: 0.1 to 5 percent, 5 to 20 percent, 20 to 50 percent and 50+ percent. Objective cutpoints divide the data into various numbers of quantiles: 3, 4, 5, and 8. These results are reported in Tables A.3, A.4, and A.5.

We also exploit the continuous nature of our treatment variable (percent of population affected by the flood) to study treatment effect heterogeneity in greater depth. In Table A.5, Columns 3 and 4, we estimate models that incorporate a binary treatment variable (flood = 1) and a continuous variable capturing flood severity. In this model, the coefficient on the binary treatment variable represents the effect of moving from 0 percent to 0.1 percent flood severity; the coefficient on the continuous treatment variable represents the effect of increasing flood severity *within* the treated category.

The results in a sample restricted to the south indicate that lightly-flooded counties exhibit large negative effects and that the treatment effect diminishes as flooding becomes more severe (a positive coefficient estimate for the continuous measure of flood severity). However, in the full nationwide sample, this effect reverses: we still observe a negative point estimate for the flood/treatment dummy; however, as flood severity increases, the effect on Republican vote share becomes larger and more negative. To illustrate the expected treatment effect at varying levels of treatment intensity, we plot the predicted effects when flood severity varies across the range observed in our data (from 0.1 percent of the population affected to over 100 percent of the population affected[8]). As Figure A.1 shows, the negative effect of flooding diminishes as severity increases in the southern sample, but when we widen the sample, we find that increasing severity increases the flood's negative effect. Overall, the evidence for heterogeneous treatment effects is mixed across different specifications.

[8]Flood severity in our data goes beyond 100 percent because the Red Cross estimated the number of individuals impacted by the flood in each county, and we convert this to a percentage using census population figures, interpolated to 1927.

Table A.2: Flooding and Republican Vote Share

	South-Only Sample			Full Sample		
	(1)	(2)	(3)	(4)	(5)	(6)
Treat (binary)	-18.323***	-10.814***	-12.342***	-1.680	-5.486***	-5.301***
	(1.61)	(1.66)	(1.67)	(1.61)	(1.56)	(1.56)
Contiguous (1st-order)	-10.373***		-7.367***	3.412*		0.790
	(1.71)		(1.66)	(1.56)		(1.46)
Contiguous (2nd-order)	-5.918**		-4.184	6.068***		2.618
	(2.28)		(2.16)	(1.83)		(1.70)
Pct. Black		-0.260***	-0.240***		0.154***	0.152***
		(0.02)	(0.02)		(0.02)	(0.02)
Pct. Protestant		0.164***	0.138***		0.332***	0.331***
		(0.04)	(0.04)		(0.02)	(0.02)
N	996	980	980	3077	3049	3049

Linear models of Republican vote share using a binary treatment variable for flooded counties and binary measures of contiguity to flooded counties (first-order contiguous counties are adjacent to flooded counties; second-order contiguous counties are adjacent to counties that are first-order contiguous).
Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A.3: Flooding and Republican Vote Share

	Subjective Cutpoints		Three Quantiles	
	South	Full Sample	South	Full Sample
	(1)	(2)	(3)	(4)
Low Severity	-13.089**	-7.006		
	(4.19)	(3.59)		
Med-Low Severity	-10.223***	-0.736		
	(2.46)	(2.27)		
Med-High Severity	-15.453***	-10.433***		
	(2.89)	(3.04)		
High Severity	-4.517	-10.025**		
	(3.21)	(3.66)		
Pct. Black	-0.266***	0.164***	-0.266***	0.161***
	(0.02)	(0.02)	(0.02)	(0.02)
Pct. Protestant	0.167***	0.324***	0.169***	0.326***
	(0.04)	(0.02)	(0.04)	(0.02)
Quantile_3 1			-9.992***	-3.467
			(2.57)	(2.56)
Quantile_3 2			-14.882***	-3.475
			(2.63)	(2.53)
Quantile_3 3			-7.219**	-10.090***
			(2.73)	(2.68)
N	980	3049	980	3049

Linear models of Republican vote share. Subjective cutpoints (0.1 - 5 percent, 5 - 20 percent, 20 - 50 percent and 50+ percent) are reported in Models 1 & 2. Results based on dividing the treatment variable into three quantiles are reported in Models 3 & 4. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.4: Flooding and Republican Vote Share

	Four Quantiles		Five Quantiles	
	South	Full Sample	South	Full Sample
	(1)	(2)	(3)	(4)
Quantile_4 1	-9.642**	-3.998		
	(2.95)	(2.94)		
Quantile_4 2	-12.338***	-1.898		
	(2.97)	(2.91)		
Quantile_4 3	-15.789***	-7.151*		
	(3.00)	(2.96)		
Quantile_4 4	-4.629	-9.611**		
	(3.15)	(3.08)		
Quantile_5 1			-8.263*	-5.494
			(3.28)	(3.24)
Quantile_5 2			-13.164***	2.281
			(3.33)	(3.31)
Quantile_5 3			-14.238***	-4.500
			(3.27)	(3.25)
Quantile_5 4			-14.023***	-11.420***
			(3.41)	(3.36)
Quantile_5 5			-3.419	-9.296**
			(3.43)	(3.39)
Pct. Black	-0.268***	0.160***	-0.271***	0.163***
	(0.02)	(0.02)	(0.02)	(0.02)
Pct. Protestant	0.168***	0.327***	0.171***	0.324***
	(0.04)	(0.02)	(0.04)	(0.02)
N	980	3049	980	3049

Linear models of Republican vote share. Treatment (flood severity) is separated into four quantiles (Models 1 & 2) or five quantiles (Models 3 & 4).
Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.5: Flooding and Republican Vote Share

	Eight Quantiles		Continuous Treatment	
	South	Full Sample	South	Full Sample
	(1)	(2)	(3)	(4)
Quantile_8 1	-13.001**	-8.047*		
	(4.19)	(4.08)		
Quantile_8 2	-6.418	0.247		
	(4.03)	(4.20)		
Quantile_8 3	-13.578**	1.972		
	(4.16)	(4.08)		
Quantile_8 4	-11.062**	-5.851		
	(4.16)	(4.09)		
Quantile_8 5	-19.784***	-3.310		
	(4.22)	(4.20)		
Quantile_8 6	-11.934**	-10.875**		
	(4.08)	(4.13)		
Quantile_8 7	-7.852	-11.667**		
	(4.24)	(4.28)		
Quantile_8 8	-1.142	-7.932		
	(4.35)	(4.21)		
Treat (Binary)			-13.520***	-3.148
			(2.18)	(2.02)
Flood Severity (Continuous)			0.090	-0.093
			(0.05)	(0.05)
Pct. Black	-0.271***	0.163***	-0.269***	0.162***
	(0.02)	(0.02)	(0.02)	(0.02)
Pct. Protestant	0.171***	0.325***	0.172***	0.326***
	(0.04)	(0.02)	(0.04)	(0.02)
N	980	3049	980	3049

Linear models of Republican vote share. Treatment (flood severity) is separated into four quantiles (Models 1 & 2) or five quantiles (Models 3 & 4).
Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

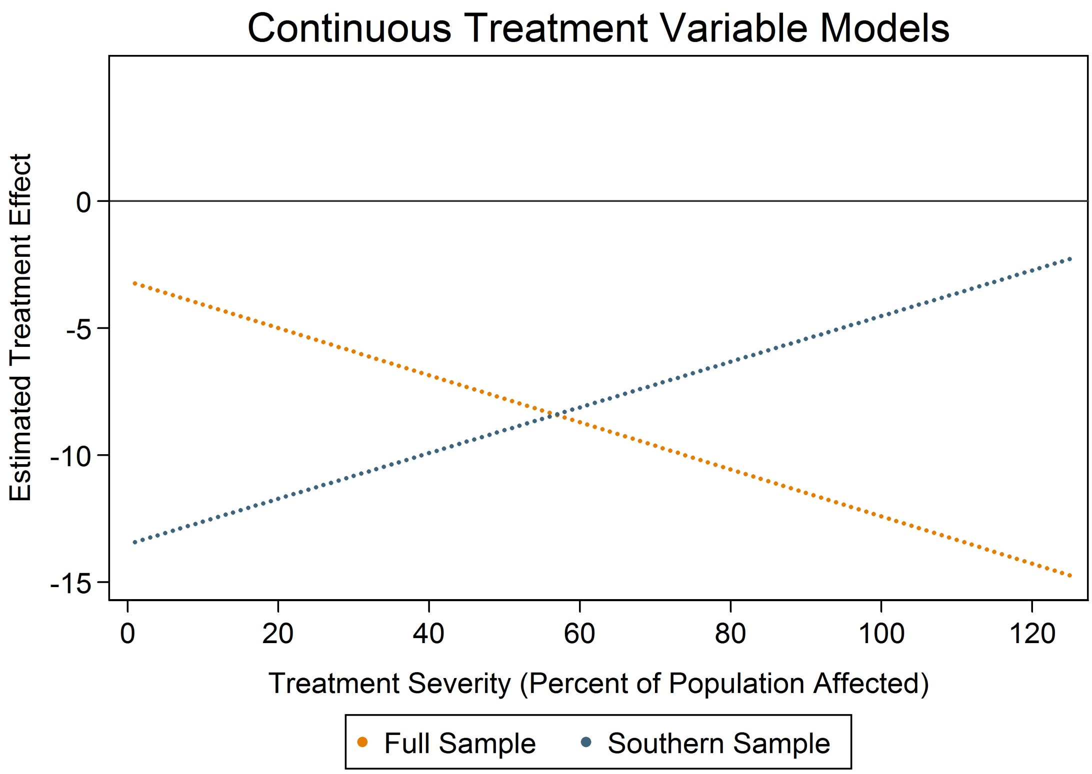

Figure A.1: Heterogeneous treatment effects at varying levels of flood severity. At higher levels of flood severity, the estimated treatment effect is statistically indistinguishable from zero.

E. Alternative Model Specifications

In the primary manuscript, we focus our attention on the *total* effect of flooding in the US south in 1927, estimating the effect of flood damage, without accounting for differentiation in terms of relief efforts. In contrast, Gasper and Reeves (2011) and other scholars in this literature have often modeled incumbent vote share as a function of both disaster severity (monetary damages, for instance) and relief efforts (such as disaster declarations). Our motivations for this departure are two-fold: first, flood severity and the extent of relief aid are closely correlated in our case. This makes it very difficult to parse the two effects accurately, as multicollinearity decreases the stability of our estimates.[9] Second, our interest is not in whether increased levels of aid marginally increase support for Hoover in 1928—rather, we are interested in the overall or total effect of flooding, including the disaster response that followed it. It may be the case that aid increases vote share. But if it only does so at the margins, and voters still—in total—punish the incumbent for the flood's overall impact, it suggests that blind retrospection dominates attentive retrospection in this case.

To ensure full transparency, we report a series of models that mimic those of Gasper and Reeves in this section. In these models, we incorporate separate variables that capture flood severity and aid efforts (both measured as a percentage of county population). The results for a southern-only sample are reported in Table A.6, column 1. Although the coefficient estimates are in the expected direction—more severe flooding reduces Republican vote share, while more generous aid increases it—neither coefficient is statistically significant. Moreover, these results are arguably driven by the modeling choice: if we incorporate a treatment dummy as well, the coefficient signs flip, though they remain insignificant. The choice to add a treatment dummy is defensible, because a linear model without it assumes that the difference between 0% flooding and 1% flooding is the same as that between 1% flooding and 2% flooding. In contrast, the model we report in column 2 (and in other specifications elsewhere in this Appendix) allows the effect of moving from

[9] Gasper and Reeves, to take one example, have a significantly larger n; moreover, their blunter measure of aid (disaster declarations) is also less correlated with disaster severity than our's.

no flooding to very light flooding to differ from that of increasing flood intensity, among flooded counties. Because we estimate such large differences between non-flooded and lightly-flooded counties—as shown in Table A.6 as well as throughout the paper—it is important that we incorporate such a treatment dummy.

Table A.6, columns 3 and 4, and Table A.7 reinforce this point. In the former, we incorporate an interaction effect between flood severity and relief aid. In the latter, we repeat these analyses in the full sample. Across these many specifications, the distinct effects of flood severity and relief efforts vary wildly. We argue that this is a result of their close correlation and our inability to parse the distinct effects accurately, problems that threaten other research in this literature.

Finally, we apply a more flexible specification in Table A.8, using polynomials to allow the effect of flood severity to vary more dramatically than in other models. We use linear, quadratic and cubic terms to capture the impact of flood severity; in some models we also include flexible polynomial terms for the size of each county's black and Protestant populations. Because models with higher-order polynomials make interpretation of overall effects difficult, we plot predicted values across the range of flood severity in Figure A.2.

Table A.6: Flooding, Relief Aid and Republican Vote Share — Southern Sample

	(1)	(2)	(3)	(4)
Flood Severity (Continuous)	-0.289*	0.053	-0.233*	0.032
	(0.11)	(0.13)	(0.11)	(0.13)
Relief Aid (Continuous)	0.291	0.053	-0.427	-0.114
	(0.17)	(0.17)	(0.24)	(0.25)
Pct. Black	-0.286***	-0.269***	-0.280***	-0.269***
	(0.02)	(0.02)	(0.02)	(0.02)
Pct. Protestant	0.211***	0.173***	0.193***	0.172***
	(0.04)	(0.04)	(0.04)	(0.04)
Treat (Binary)		-13.350***		-11.920***
		(2.24)		(2.74)
Severity x Aid			0.007***	0.002
			(0.00)	(0.00)
N	980	980	980	980

Linear models of Republican vote share. Treatment is, alternately, measured as a continuous variable or separated into both a continuous variable (among flooded counties) and a dummy variable (distinguishing flooded from non-flooded counties).
Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.7: Flooding, Relief Aid and Republican Vote Share — Full Sample

	(1)	(2)	(3)	(4)
Flood Severity (Continuous)	-0.185	-0.082	-0.162	-0.091
	(0.12)	(0.14)	(0.12)	(0.14)
Relief Aid (Continuous)	0.066	-0.015	-0.167	-0.107
	(0.17)	(0.18)	(0.26)	(0.27)
Pct. Black	0.159***	0.162***	0.161***	0.162***
	(0.02)	(0.02)	(0.02)	(0.02)
Pct. Protestant	0.328***	0.326***	0.327***	0.326***
	(0.02)	(0.02)	(0.02)	(0.02)
Treat (Binary)		-3.198		-2.617
		(2.11)		(2.45)
Severity x Aid			0.002	0.001
			(0.00)	(0.00)
N	3049	3049	3049	3049

Linear models of Republican vote share in the full, nationwide sample. Treatment is, alternately, measured as a continuous variable or separated into both a continuous variable (among flooded counties) and a dummy variable (distinguishing flooded from non-flooded counties).
Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.8: Flooding and Republican Vote Share –
Polynomial Regression Models

	(1)	(2)	(3)	(4)
Flood Severity	-0.266**	-0.536**	-0.691***	-0.391
	(0.10)	(0.20)	(0.20)	(0.33)
Flood Severity, Sq.	0.002	0.009	0.014*	0.008
	(0.00)	(0.01)	(0.01)	(0.01)
Flood Severity, Cubed		-0.000	-0.000*	-0.000
		(0.00)	(0.00)	(0.00)
Pct. Black	0.161***	0.872***	1.784***	1.794***
	(0.02)	(0.05)	(0.11)	(0.11)
Pct. Black, Sq.		-0.012***	-0.049***	-0.050***
		(0.00)	(0.00)	(0.00)
Pct. Black, Cubed			0.000***	0.000***
			(0.00)	(0.00)
Pct. Protestant	0.327***	0.474***	0.322*	0.323*
	(0.02)	(0.07)	(0.14)	(0.14)
Pct. Prot, Sq.		-0.004***	-0.001	-0.001
		(0.00)	(0.00)	(0.00)
Pct. Prot., Cubed			-0.000	-0.000
			(0.00)	(0.00)
Treat (Binary)				-3.522
				(3.04)
N	3049	3049	3049	3049

Polynomial models of Republican vote share in the full, nationwide sample. Flood severity is captured by linear, quadratic and cubic terms. In some specifications, covariates are also allowed to enter as polynomial terms.
Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

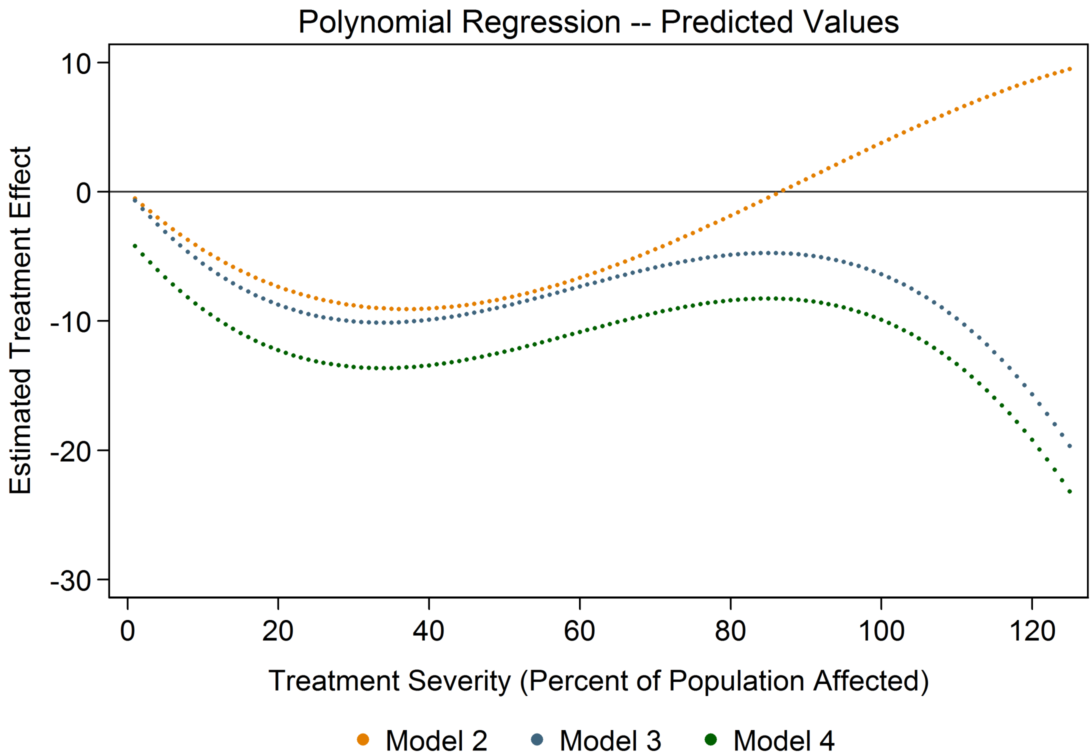

Figure A.2: Estimated effects of flooding on Republican vote share at varying levels of flood severity, based on polynomial regression models (Table A.2).

F. Synthetic Control Models

As discussed in the primary manuscript, we check the robustness of our empirical results using the synthetic control method. In the manuscript, we report the results of one iteration of this strategy, in which we restrict the sample to southern counties and construct the donor pool for county i as the 100 most closely-matched non-treated, non-contiguous southern counties in terms of pre-treatment vote share (1896-1924). To ensure that our synthetic control results are not sensitive to specific modeling choices, we expanded our analysis in five ways. First, we expanded the analysis to a full nationwide sample. Second, we employed an alternative metric for trimming the donor pool, by selecting the counties most closely matched to treated county i in terms of both percent black and percent Protestant; third, we used a larger donor pool (500 control counties) rather than the 100 reported in the manuscript. We repeated this expansion to 500 control counties for models trimmed on the basis of pre-treatment vote share *and* for models trimmed on the basis of pct. black and pct. Protestant. Fourth, we estimated an extremely conservative model, in which the donor pool was restricted to counties that bordered flooded counties. In the absence of treatment, these counties have the most prima facie validity as a control group. This model is extremely conservative, however, because we expect treatment spillovers to influence Republican vote share in adjacent counties; by restricting the donor pool to these adjacent counties, we expect these models to be biased toward a null effect. Fifth, we estimated a slightly less conservative model, in which we restrict the donor pool to counties in flooded states (i.e. states that experienced flooding in at least one county), but do not include contiguous counties in the donor pool.

In the interest of brevity, we report point estimates from each set of model runs in Table A.9. This table also includes p-values derived using the process described in Section B. We also report the results of each run as a set of two time-series graphs. In each run, the first graph shows Republican two-party vote share in the treated and synthetic control units over time; the second graph shows the difference between the treatment group and synthetic control group over time. In the most extreme case, Donor Pool #1 using southern cases only (Figure A.3), we estimate a nearly 20 percentage point decline in support for Herbert Hoover. In the most conservative case, Donor Pool #5 using both southern and northern cases (Figure A.13), we estimate a decline of 4.09 percentage points, a substantively large effect.[10] We summarize treatment effect estimates and p-values across model specifications in Table A.9.

Donor Pool Descriptions

-Donor Pool 1: The donor pool for each treated county i is composed of the 100 counties most closely matched to county i in terms of pre-treatment vote share from 1896-1924 (mean of squared differences).

-Donor Pool 2: The 100 counties most closely matched to county i in terms of percent black and percent Protestant.

-Donor Pool 3: The 500 counties most closely matched to county i in terms of pre-treatment vote share.

-Donor Pool 4: The 500 counties most closely matched to county i in terms of percent black and percent Protestant.

-Donor Pool 5: All contiguous-to-flooding counties, i.e. counties which touched a county that was flooded.

-Donor Pool 6: All non-flooded, non-contiguous counties in flooded states, i.e. counties in states that were flooded *except* those counties that were, themselves, flooded or which were contiguous to a flooded county.

Finally, to study the heterogeneity in treatment effects in our synthetic control results, we plot each treated units' treatment effect against flood severity. As Figure A.15 shows, our treatment effect estimates are generally stable across severity levels; contrary to the linear models reported earlier, there does not appear to be systematically diminishing or increasing treatment effects at higher levels of flooding.

[10] This case is arguably too conservative, as we expect cross-county spillover effects. As we show in Table A.2, contiguous counties also experienced a decline in support for Hoover. Spillovers are theoretically plausible based on local economic impacts or familial ties that cross county borders. Therefore, our 4.09 point estimate should be considered a lower bound on the likely treatment effect.

Table A.9: Treatment Effect Estimates Across Samples and Specifications

	Treatment Effect	ADH P-Value	Fisher P-Value
Southern Sample			
Donor Pool #1	-19.77	< 0.0001	< 0.0001
Donor Pool #2	-12.02	< 0.0001	< 0.0001
Donor Pool #3	-16.34	< 0.0001	< 0.0001
Donor Pool #4	-15.95	< 0.0001	< 0.0001
Donor Pool #5	-5.29	< 0.0001	< 0.0001
Donor Pool #6	-9.91	< 0.0001	< 0.0001
Full Sample			
Donor Pool #1	-15.46	< 0.0001	< 0.0001
Donor Pool #2	-9.88	< 0.0001	< 0.0001
Donor Pool #3	-12.12	< 0.0001	< 0.0001
Donor Pool #4	-11.72	< 0.0001	< 0.0001
Donor Pool #5	-4.09	< 0.0001	< 0.0001
Donor Pool #6	-8.21	< 0.0001	< 0.0001

Treatment effects, estimated using the synthetic control method, across samples and donor pool specifications. Southern sample n = 95; full sample n = 130.

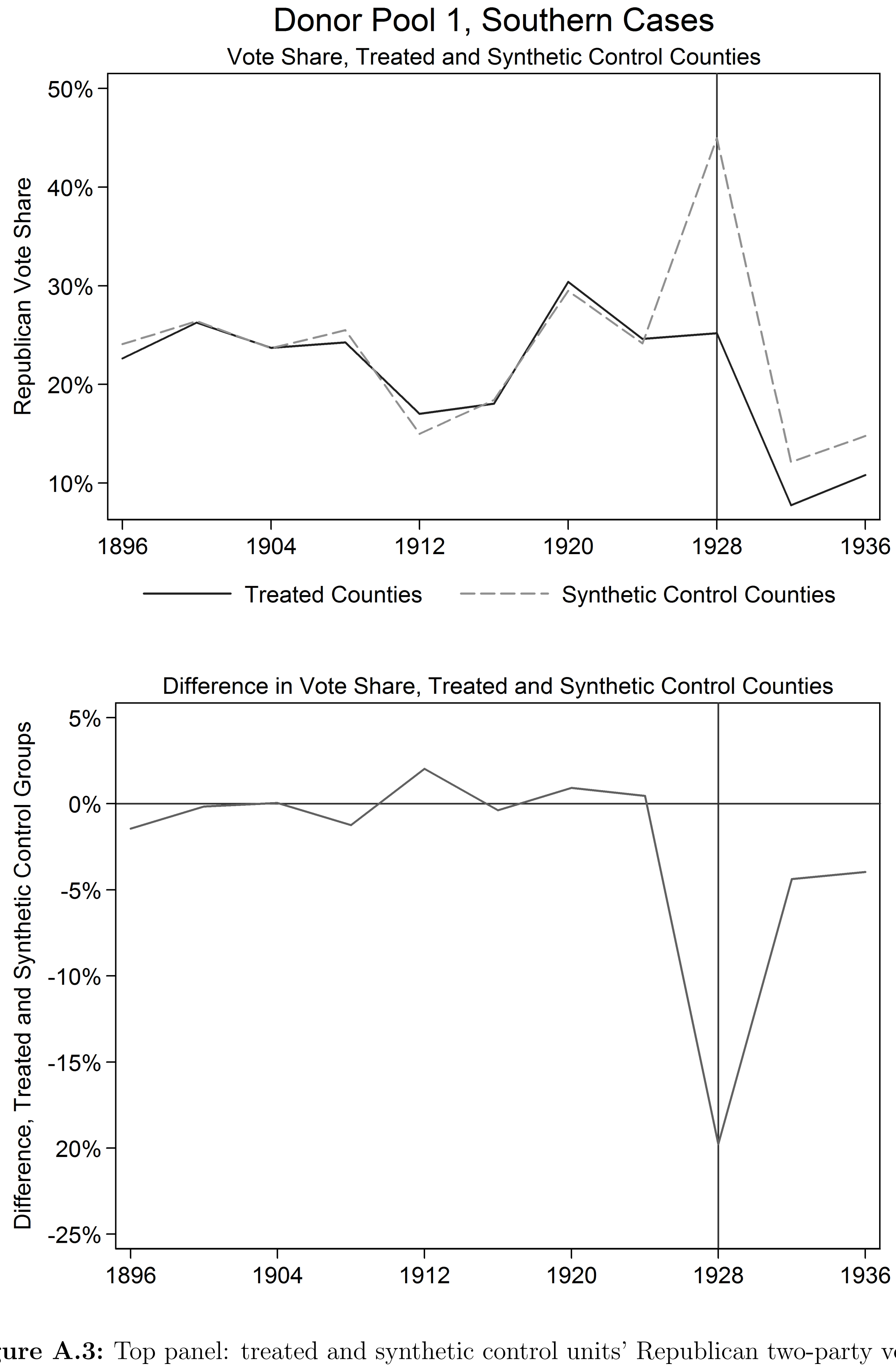

Figure A.3: Top panel: treated and synthetic control units' Republican two-party vote share, 1896-1936. Bottom panel: the difference between treated and synthetic control units' Republican vote share, 1896-1936. The sample is restricted to southern counties, and the donor pool consists of the closest 100 matches to county i in terms of pre-treatment vote share.

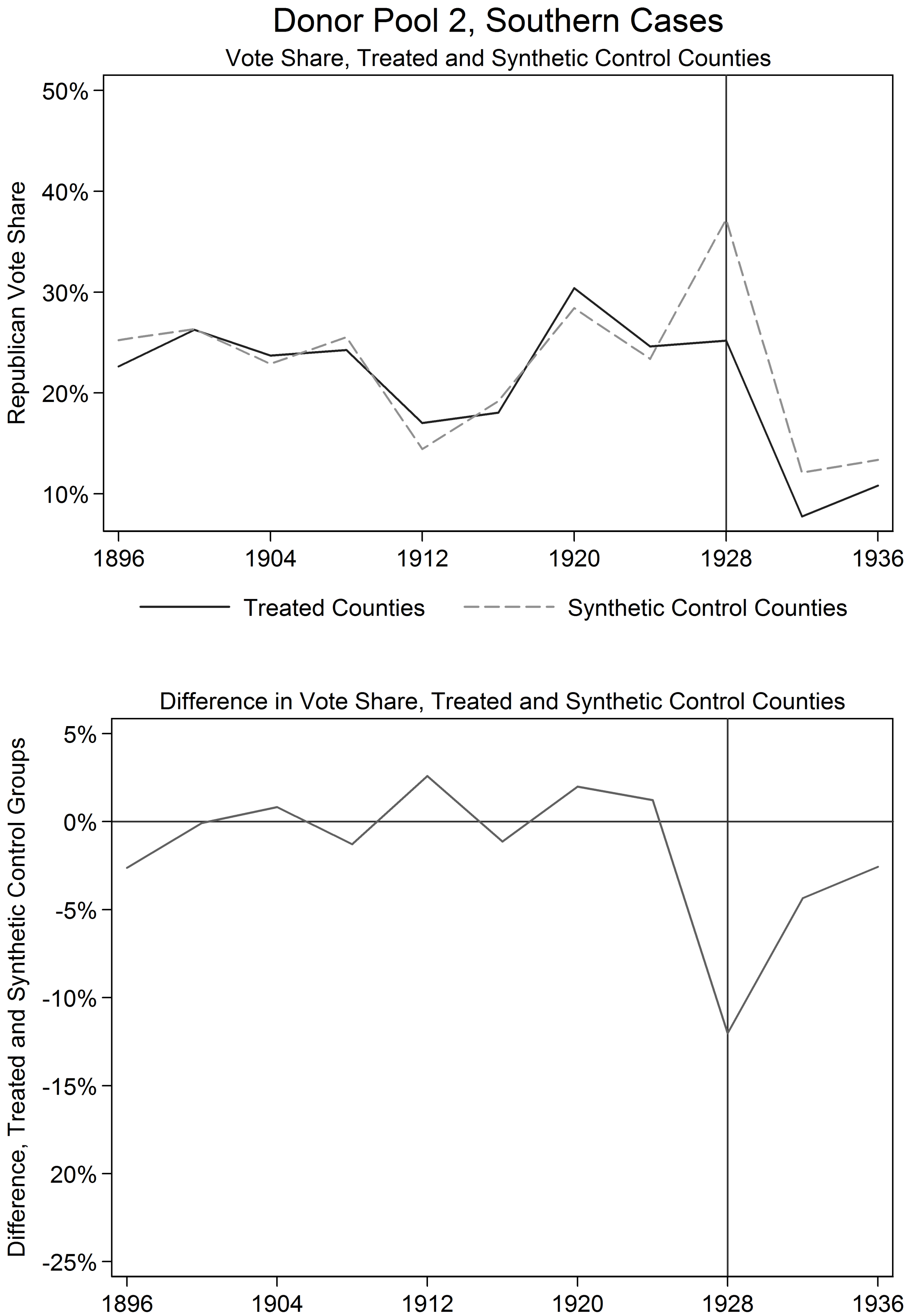

Figure A.4: Top panel: treated and synthetic control units' Republican two-party vote share, 1896-1936. Bottom panel: the difference between treated and synthetic control units' Republican vote share, 1896-1936. The sample is restricted to southern counties, and the donor pool consists of the closest 100 matches to county i in terms of black and Protestant populations.

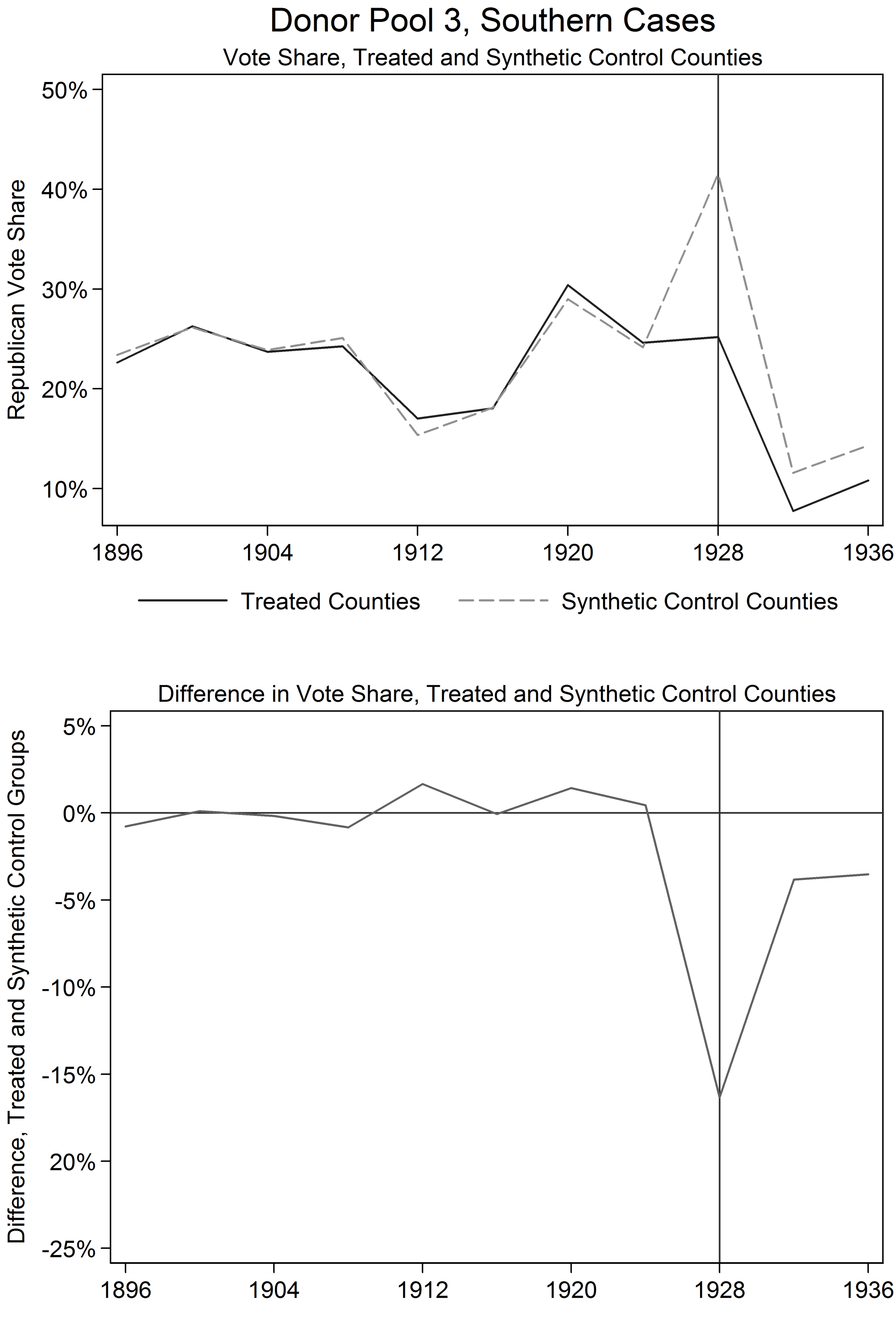

Figure A.5: Top panel: treated and synthetic control units' Republican two-party vote share, 1896-1936. Bottom panel: the difference between treated and synthetic control units' Republican vote share, 1896-1936. The sample is restricted to southern counties, and the donor pool consists of the closest 500 matches to county i in terms of pre-treatment vote share.

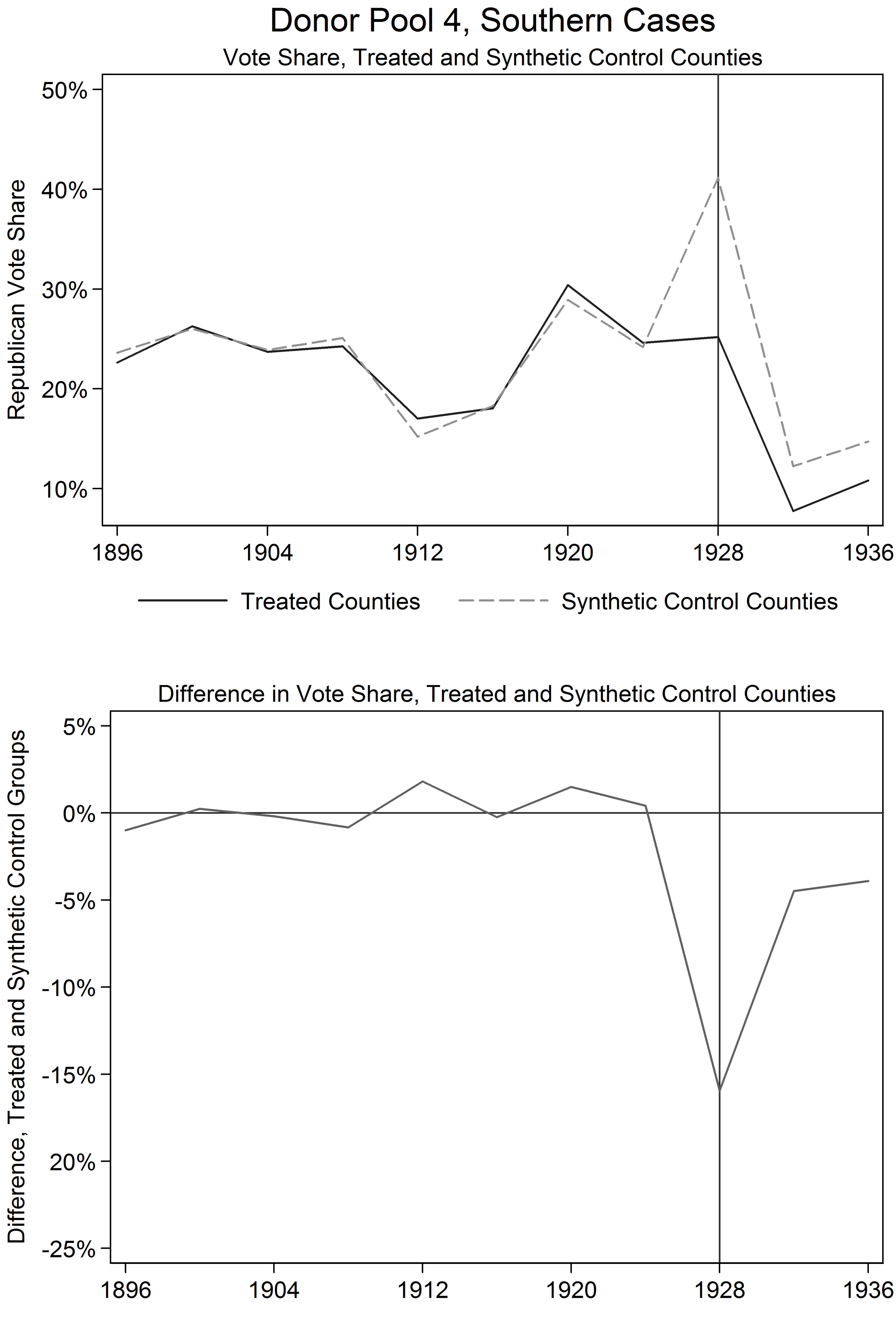

Figure A.6: Top panel: treated and synthetic control units' Republican two-party vote share, 1896-1936. Bottom panel: the difference between treated and synthetic control units' Republican vote share, 1896-1936. The sample is restricted to southern counties, and the donor pool consists of the closest 500 matches to county i in terms of black and Protestant populations.

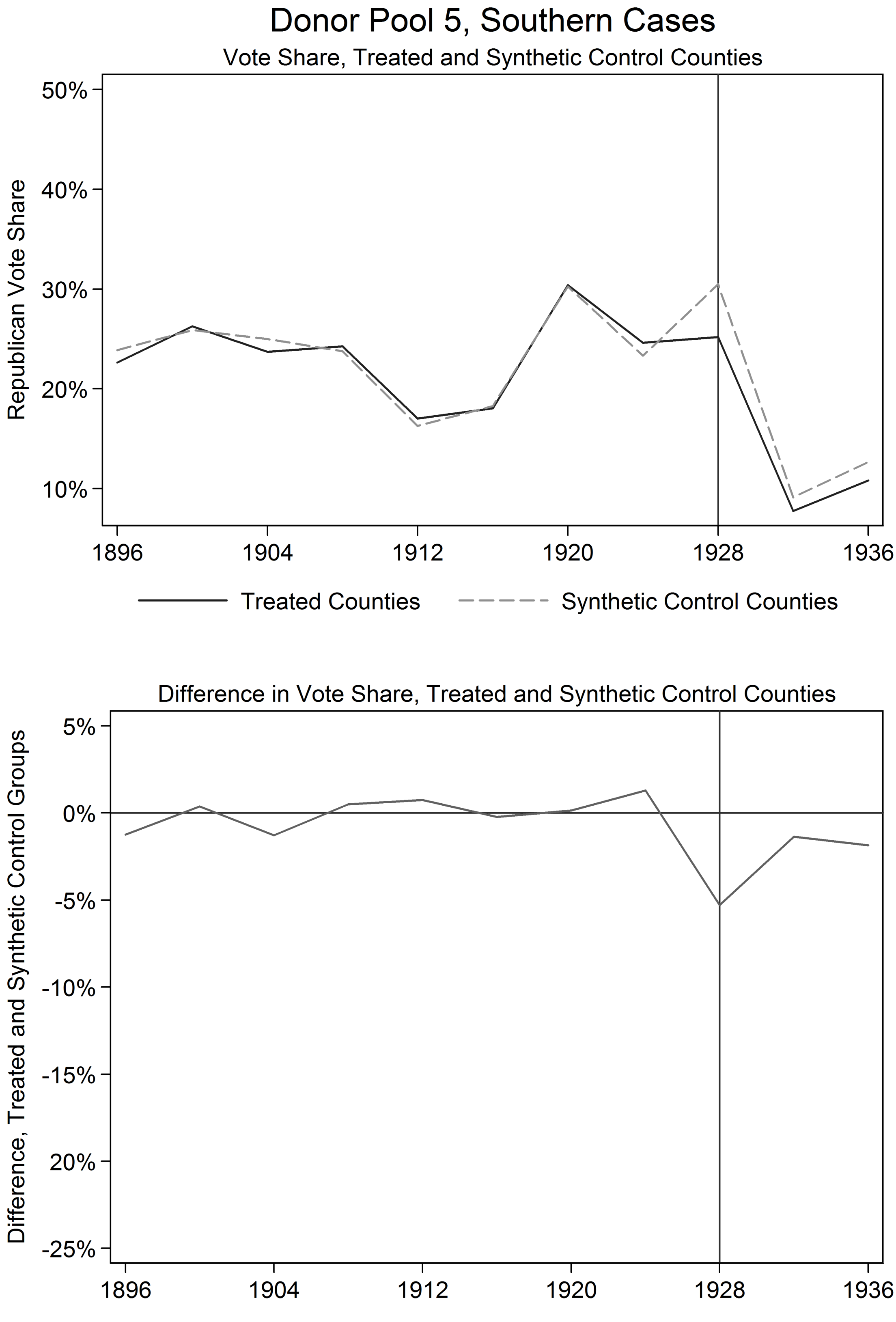

Figure A.7: Top panel: treated and synthetic control units' Republican two-party vote share, 1896-1936. Bottom panel: the difference between treated and synthetic control units' Republican vote share, 1896-1936. The sample is restricted to southern counties, and the donor pool consists of flood-adjacent counties (i.e. counties contiguous to a flooded county).

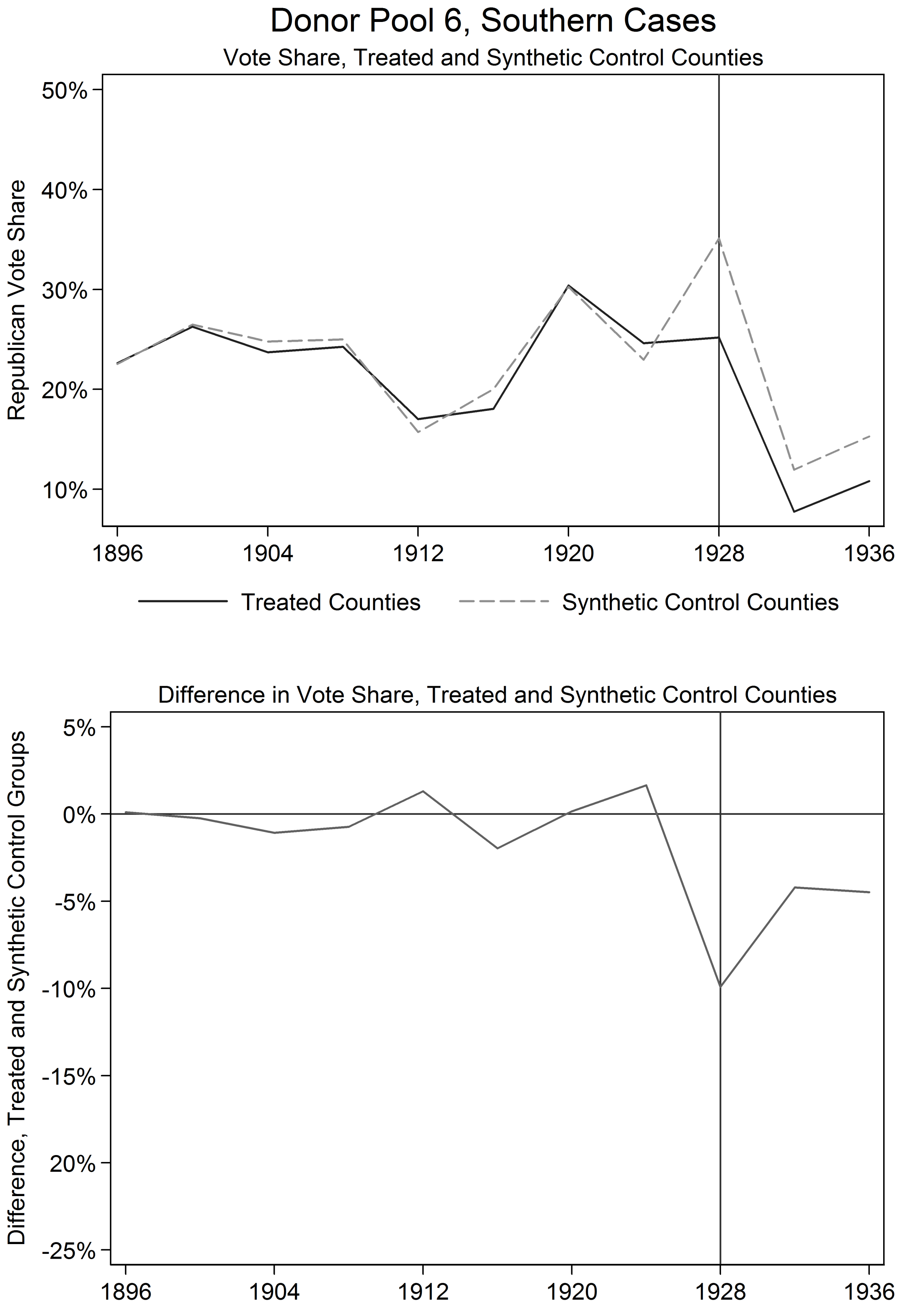

Figure A.8: Top panel: treated and synthetic control units' Republican two-party vote share, 1896-1936. Bottom panel: the difference between treated and synthetic control units' Republican vote share, 1896-1936. The sample is restricted to southern counties, and the donor pool consists of non-flooded, non-contiguous counties in states that experienced flooding.

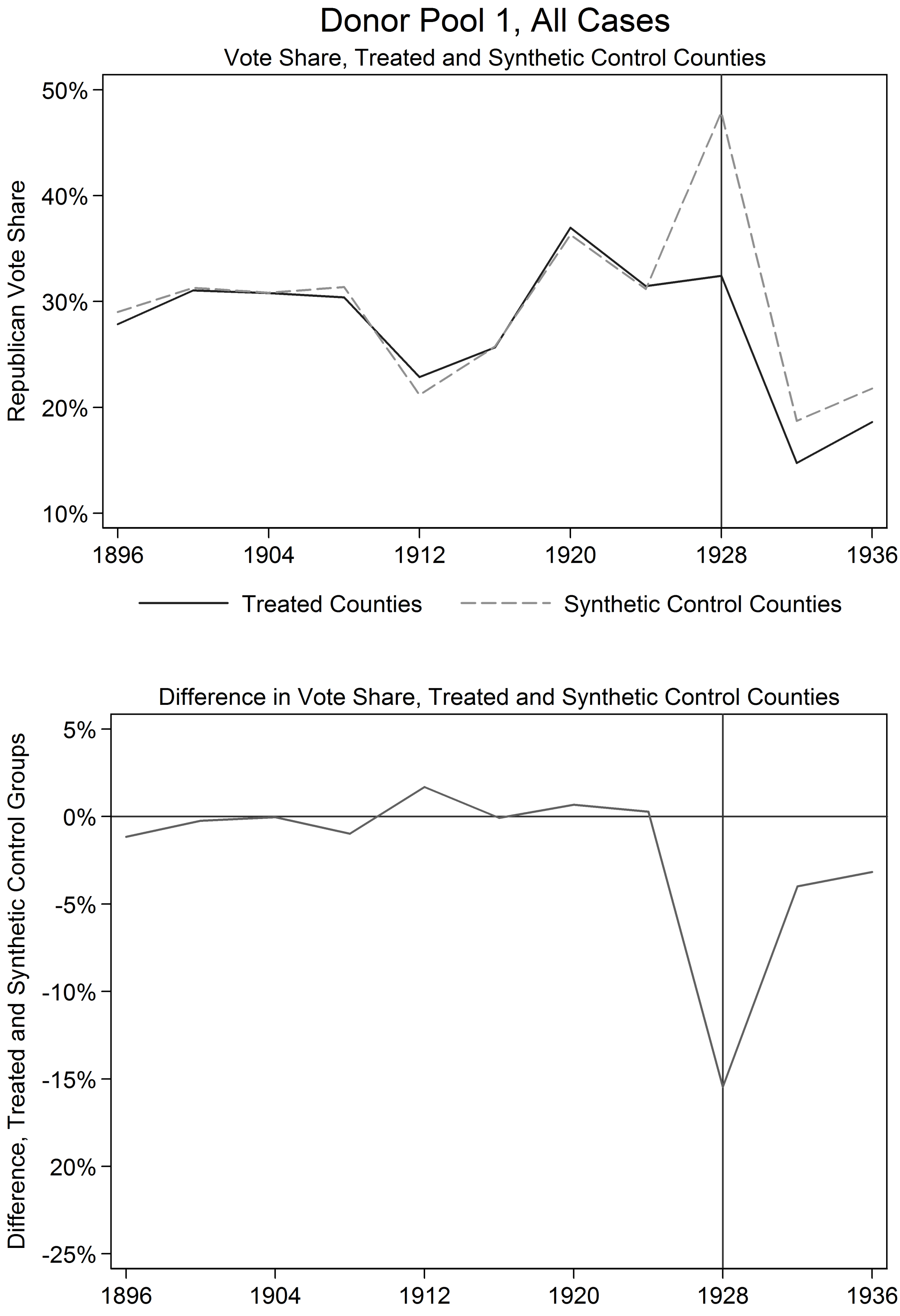

Figure A.9: Top panel: treated and synthetic control units' Republican two-party vote share, 1896-1936. Bottom panel: the difference between treated and synthetic control units' Republican vote share, 1896-1936. The sample consists of all US counties, and the donor pool consists of the closest 100 matches to county i in terms of pre-treatment vote share.

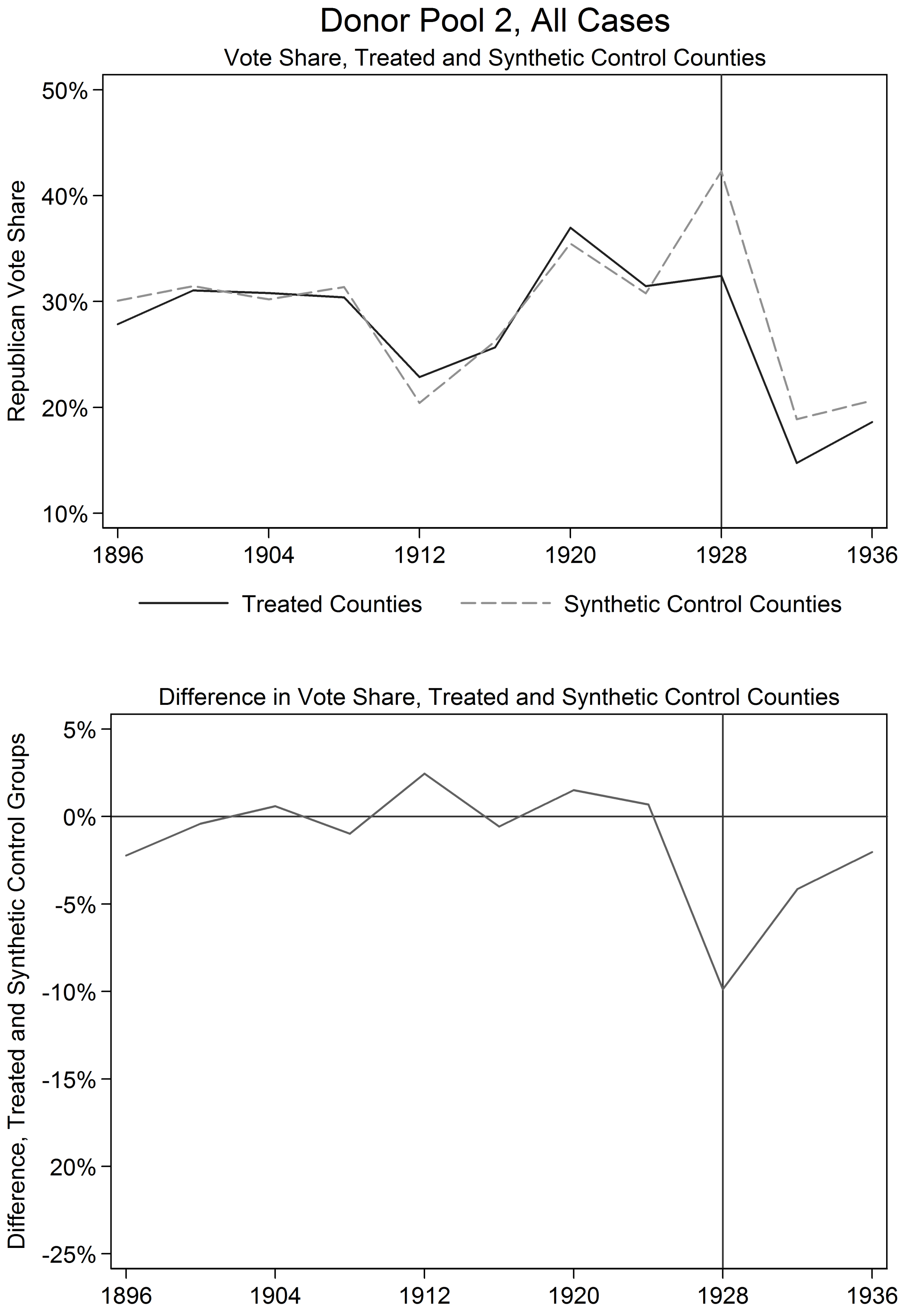

Figure A.10: Top panel: treated and synthetic control units' Republican two-party vote share, 1896-1936. Bottom panel: the difference between treated and synthetic control units' Republican vote share, 1896-1936. The sample consists of all US counties, and the donor pool consists of the closest 100 matches to county i in terms of black and Protestant populations.

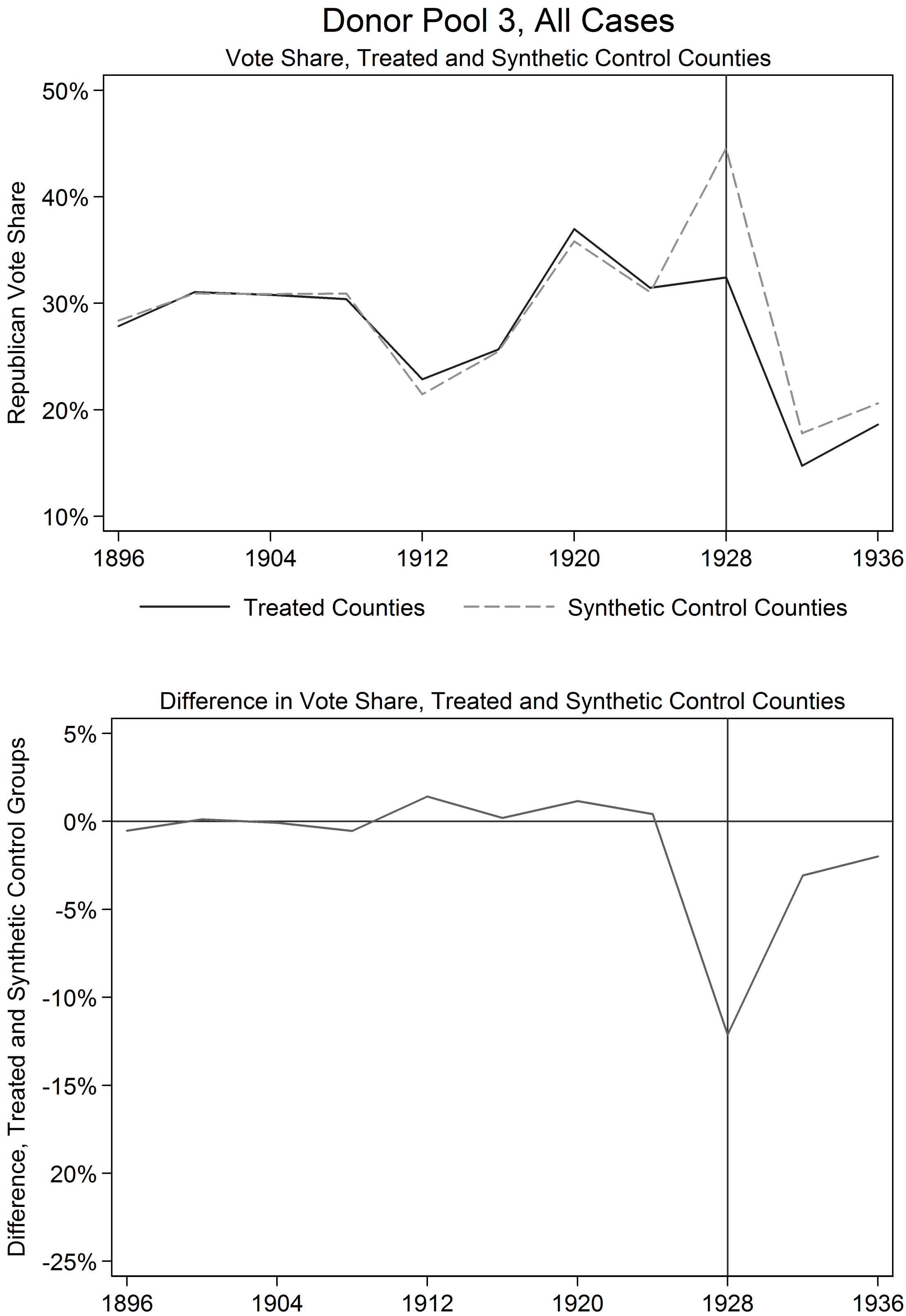

Figure A.11: Top panel: treated and synthetic control units' Republican two-party vote share, 1896-1936. Bottom panel: the difference between treated and synthetic control units' Republican vote share, 1896-1936. The sample consists of all US counties, and the donor pool consists of the closest 500 matches to county i in terms of pre-treatment vote share.

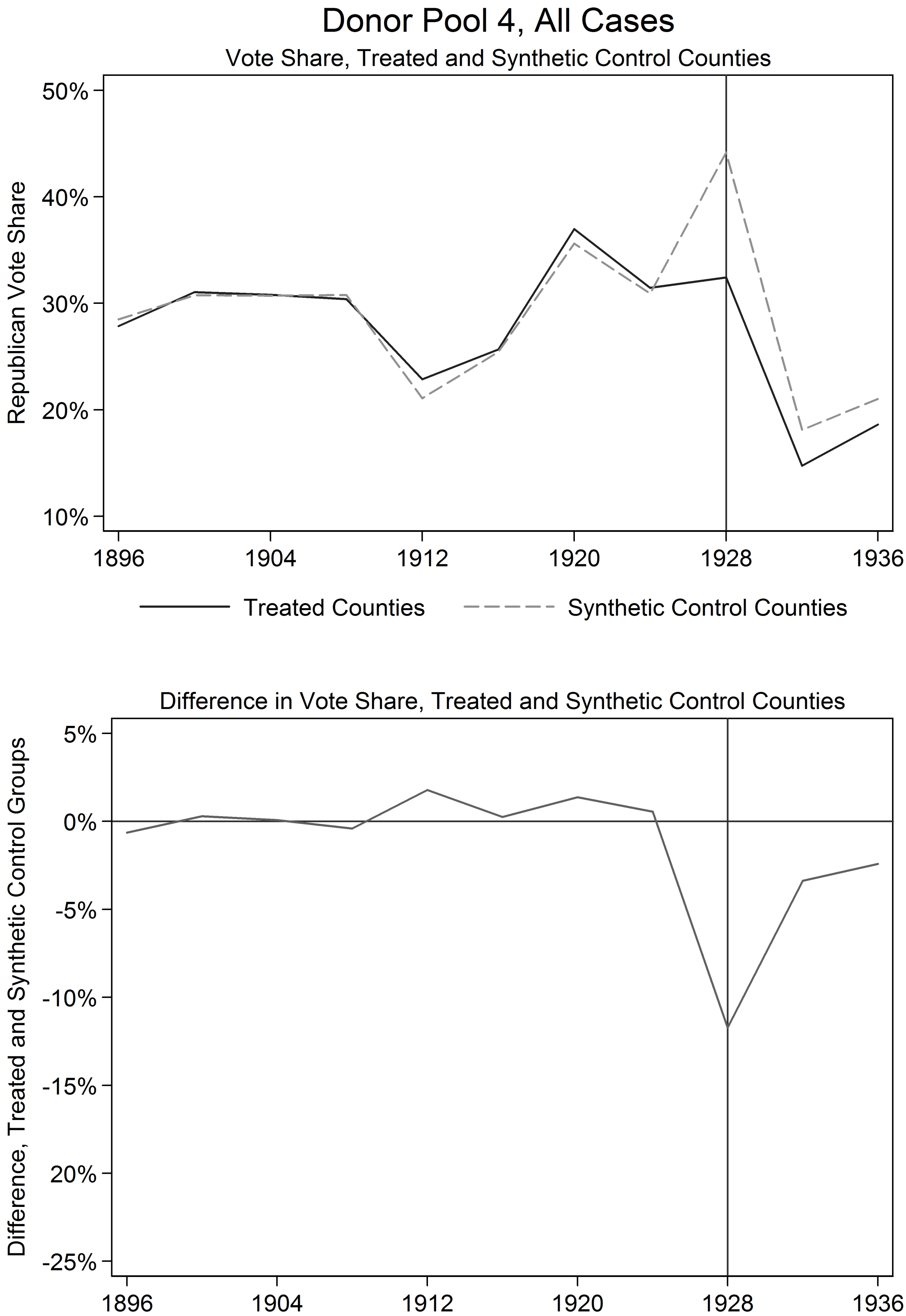

Figure A.12: Top panel: treated and synthetic control units' Republican two-party vote share, 1896-1936. Bottom panel: the difference between treated and synthetic control units' Republican vote share, 1896-1936. The sample consists of all US counties, and the donor pool consists of the closest 500 matches to county i in terms of black and Protestant populations.

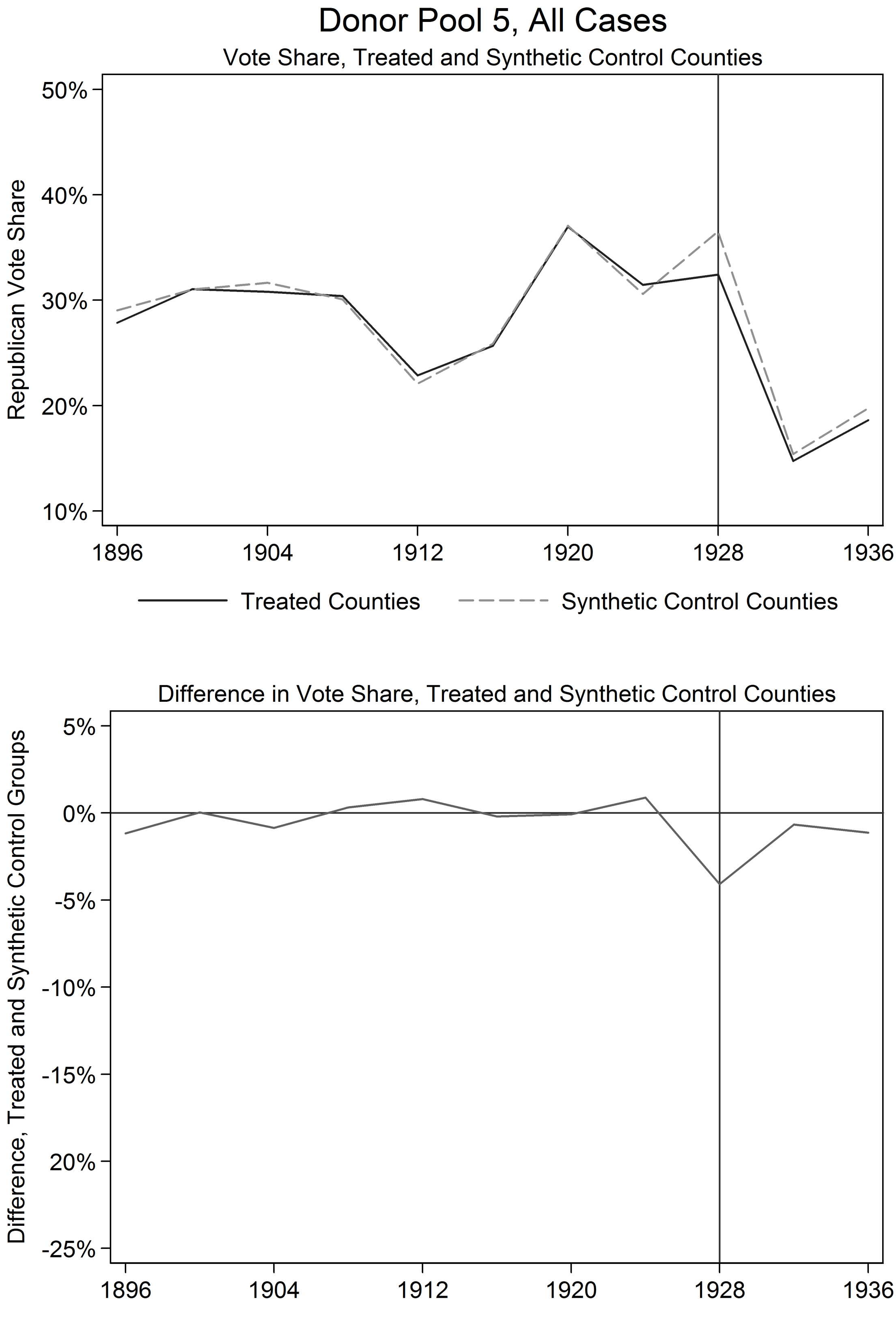

Figure A.13: Top panel: treated and synthetic control units' Republican two-party vote share, 1896-1936. Bottom panel: the difference between treated and synthetic control units' Republican vote share, 1896-1936. The sample consists of all US counties, and the donor pool consists of flood-adjacent counties (i.e. counties contiguous to a flooded county).

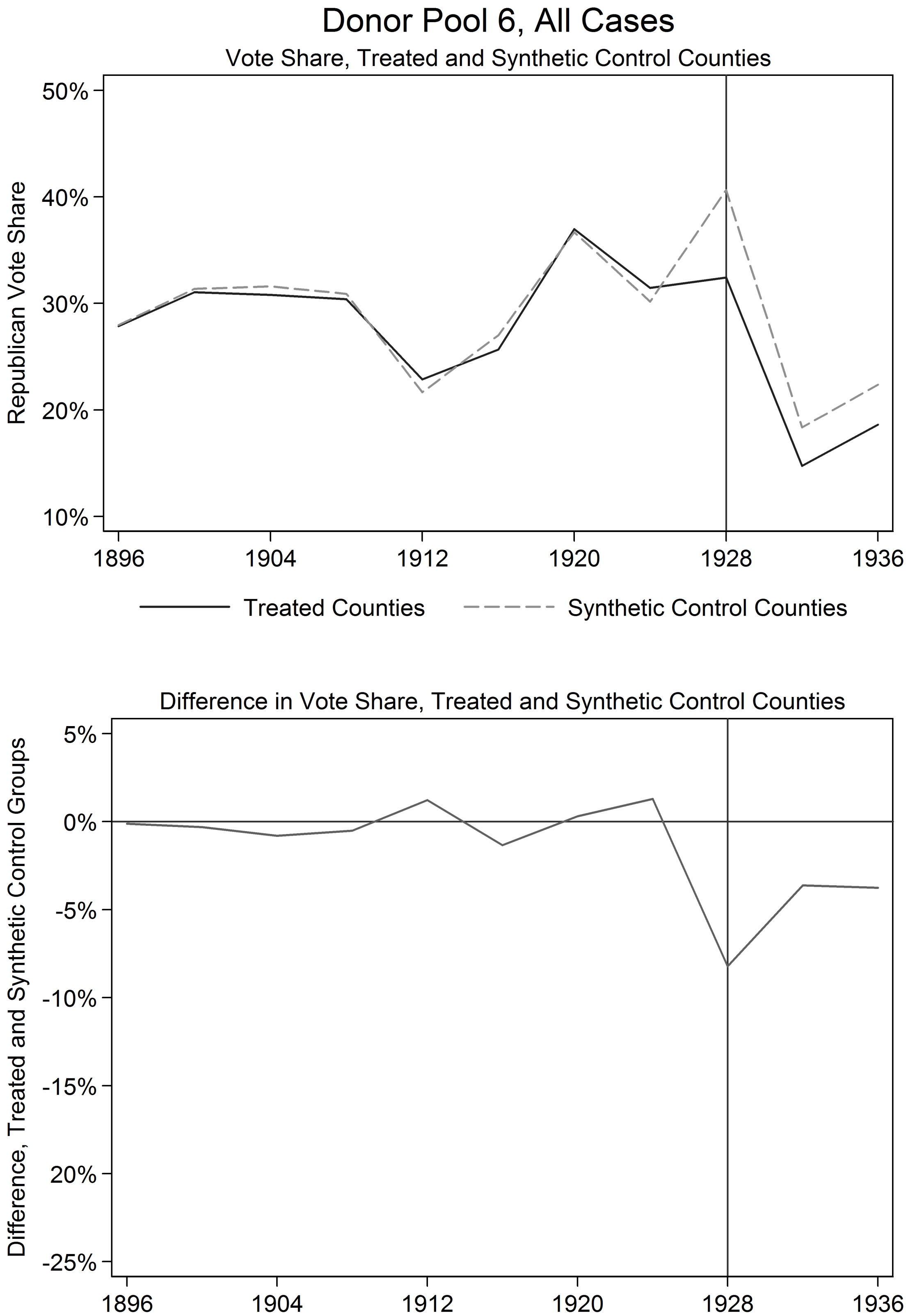

Figure A.14: Top panel: treated and synthetic control units' Republican two-party vote share, 1896-1936. Bottom panel: the difference between treated and synthetic control units' Republican vote share, 1896-1936. The sample consists of all US counties, and the donor pool consists of non-flooded, non-contiguous counties in states that experienced flooding.

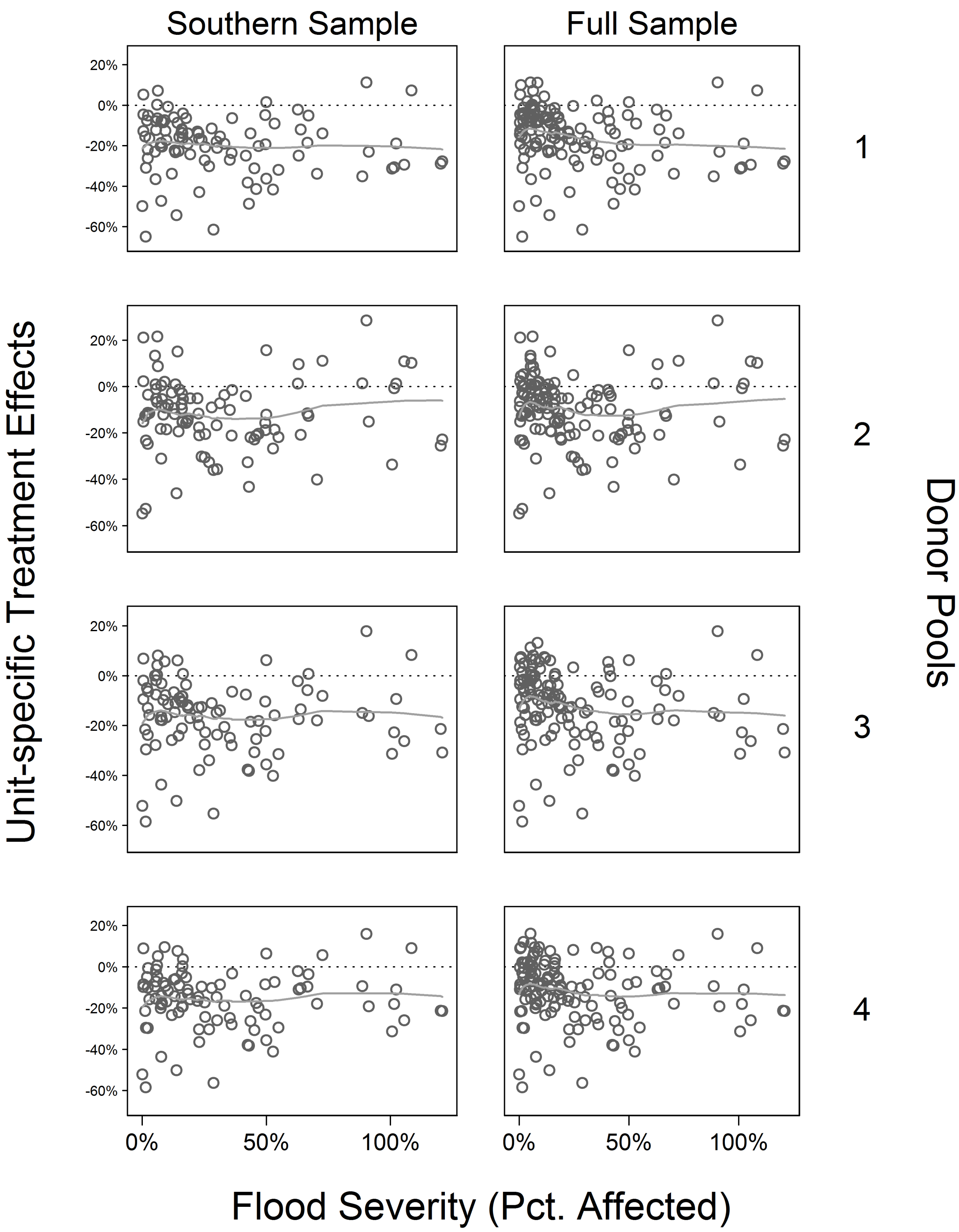

Figure A.15: Heterogeneous treatment effects, generated from synthetic control models under varied donor pool and sample selection criteria. Results based on a restricted sample appear in the left panels; results based on a full nationwide sample appear on the right. Results from different donor pools are arranged from top to bottom. In each panel, we plot unit-specific treatment effects against flood severity and illustrate the relationship between them with a locally-weighted regression line. In most cases, there does not appear to be any specific relationship between flood severity and treatment effect magnitude.

G. Placebo Test - 1924 Election

One possible threat to inference in our study arises from time-varying heterogeneity. In the regression models that we report, this would manifest in treatment and control groups that do not share parallel trends in Republican vote share under the counterfactual. This is a real possibility, given the nature of the area flooded: compared to southern control counties, flooded counties in the south have larger black populations and are less Protestant. As we note in the main manuscript, Hoover was alleged to be against segregation, and his opponent was the first Catholic nominated on a major-party ticket. Both of these characteristics of the 1928 election make the flooded counties—at least those in the south—less likely to support Hoover. A larger black population would motivate white voters to support Smith, out of fear that Hoover would support desegregation and undermine their positions of political and economic power. A less Protestant population, meanwhile, would mean that the anti-Smith backlash among Protestants would be less pronounced, resulting in less support for Hoover. Thus, even in the absence of a flood in 1927, we would expect Hoover to do particularly poorly in treated counties.

We control for both characteristics of the treated counties explicitly in our regression models, by including the percent black from the 1920 US Census, and the percent Protestant from the 1926 Census of Religious Bodies. In a difference-in-differences framework, our identifying assumption is that—conditional on observed covariates (percent black, percent Protestant)—trends in Republican vote share in the treatment and control groups are parellel under the counterfactual. Naturally, we cannot assess this assumption directly; researchers often check the plausibility of the parallel trends assumption by plotting each group's pre-treatment trends in the outcome variable. If pre-treatment trends in the treatment and control groups are roughly parallel, it strengthens our confidence in inferences drawn from standard difference-in-differences models. Using this metric, the parallel trends assumption does not appear to hold: as Figure A.16 shows, pre-treatment trends in the treatment and control groups do not mirror each other very closely.[11]

[11] Note that this approach ignores the "conditional on observed covariates" qualifier mentioned previously—thus, any claims based on plotting trends in this way should be considered tentative.

Figure A.16: Trends in Republican two-party vote share across flooded (treated) and non-flooded (control) counties in the south. Pre-treatment (1896-1924) trends do not match well, suggesting that the parallel trends assumption is violated.

Concern regarding the parallel trends assumption—or time-varying heterogeneity—motivated our use of the synthetic control method. Under assumptions briefly outlined in the manuscript—and described in much greater detail in Abadie, Diamond and Hainmueller (2010)—the synthetic control method identifies unit-specific treatment effects even in cases of time-varying heterogeneity between treatment and control units. We conceptualize this approach slightly differently, thinking of the synthetic control method as a type of data pre-processing analogous to matching (Ho et al. 2007), which makes the assumption of no time-varying heterogeneity more plausible. In essence, we trim the control group using the synthetic control algorithm to ensure that the remaining control units very closely match the treated group. The idea mimics studies that use cross-sectional matching methods prior to difference-in-differences estimation (see, e.g., Ladd and Lenz 2009) but improves upon them by using a much more robust approach to matching.[12]

However one views the synthetic control approach when applied to multiple treated units, we expect treatment effects from the 1927 flood to manifest only in the 1928 elections and—to the extent that there are residual effects—in elections that follow. If we observe non-zero treatment effects prior to the 1928 election when using the synthetic control method, it would raise concerns about our empirical strategy. For instance, if we observe a negative "treatment effect" in 1924, it suggests that our negative treatment effect estimates for 1928 could be a function of time-varying heterogeneity unaccounted for by the synthetic control approach.

To assess this possibility, we replicate our analysis of the 1928 election in the case of 1924. Specifically, we performed two sets of analysis, mirroring the analysis reported in the main manuscript. First, we estimated a linear model of the form

$$v^i_{1924} - v^i_{1920} = \beta_0 + \alpha\gamma_i + \theta\delta_i + \mu,$$

which is precisely analogous to the main model in the manuscript and to the description in

[12]For a similar interpretation of the synthetic control method, see Heersink and Peterson (2016). In practice we report the simple difference between treated and synthetic control units, because the differences in effect estimates are trivial.

Table A.10: Placebo Regression Models - 1924 Election

	(1)	(2)
Treat (binary)	0.555	-0.052
	(1.04)	(1.12)
Pct. Black		0.051***
		(0.02)
Pct. Protestant		0.032
		(0.03)
N	974	973

Linear models of vote share in flooded counties, in 1924. Placebo models use the 1927 Mississippi Flood to predict vote share at the county level in 1924.
Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Section D of this Supplemental Appendix. γ is our treatment variable—in this context, we focus on a binary treatment variable, rather than estimating a model allowing for heterogeneous treatment effects across levels of flood severity. Second, we replicate our synthetic control analysis, constructing synthetic control units for each flooded county over the period 1896-1920.[13]

Recall that our regression-based estimate of the flood's effect was -10.8 percentage points in a sample restricted to southern counties; among the full sample, the magnitude of our estimate declined to -5.5 percentage points, which was still significant at the 1% level. In contrast, equivalent regression models applied to the 1924 election—as described in the equation above—result in null results. In the southern sample, with control variables, our point estimate is -0.05 percentage points, with a p-value of approximately 0.96 and a confidence interval of [-2.26,2.15]. The full results of this model, and a sparse model without control variables, are reported in Table A.10.

Applying the synthetic control method, we estimate a placebo treatment effect on the 1924 election of 2.23 percentage points among treated counties in the south (n = 95) and 1.05 percentage points among all treated counties (n = 130). These estimates have p-values below 0.05 using either approach to statistical inference outlined in Section B,

[13] We replicate the analysis from the main manuscript, using Donor Pool #1, in which the set of donor counties is trimmed on the basis of pre-treatment vote share. For additional details, see the discussion in the previous section.

though the p-values are much larger than those of our main results from 1928.

These findings provide suggestive evidence that time-varying heterogeneity cannot account for our large negative estimates of the flood's impact on Hoover's performance in 1928. In the placebo regression, we estimate a fairly precise null effect of the flood on vote share in 1924. In the synthetic control models, our results in 1924 are statistically significant but the point estimates are of much smaller magnitude than our estimates for the 1928 election; importantly, they are also of the opposite sign. While reversion to the mean could explain a portion of our estimated effect in 1928, the large differences in magnitude from 1924 to 1928 mean that it cannot account for the full -19.8 percentage point effect that we estimate.

H. Turnout and Migration Effects as Alternative Mechanisms

The primary contribution of our paper is to show the large negative effect of the 1927 Mississippi Flood on Herbert Hoover's vote share, in affected counties, in the 1928 presidential election. We interpret substantial negative estimates of the flood's treatment effect as evidence for the blind retrospection hypothesis, as voters apparently punished the incumbent Republican party candidate for the misfortune they experienced from flooding.

Two alternative mechanisms that would account for a decline in Hoover's two-party vote share center on turnout in, and out-migration from, flooded counties. Turnout could account for a decline in support for Hoover if Hoover supporters were unable to reach polling stations or chose not to turnout. Note that this would require a differential impact of the flood on Hoover supporters' ability or willingness to turn out, relative to Smith supporters. Using data on county-level voter turnout, we explored the possibility that turnout in flooded counties was dampened in 1928—we show trends in turnout between treated and control counties over time in Figure A.17. In both the southern and the full sample, we find no evidence that turnout in 1928 fell below the trend line in flooded counties. In the full sample, for instance, voter turnout in non-flooded counties increased by 4.92 points from 1924 to 1928. Meanwhile, in flooded counties, voter turnout increased by 5.4 points over the same period. If turnout among Smith supporters in flooded counties increased, or turnout among Hoover supporters decreased, or both, a turnout effect could still explain our findings. However, we see little reason to expect Hoover supporters to be differentially impacted by the flood in terms of voter turnout; moreover, the differential effect would have to be dramatic to account for the magnitude of treatment effects that we estimate.

The second possible mechanism concerns out-migration from flooded counties. If Hoover supporters left flooded counties en masse between 1924 and 1928—either in response to the flood or for other reasons—it could explain Hoover's poor performance in

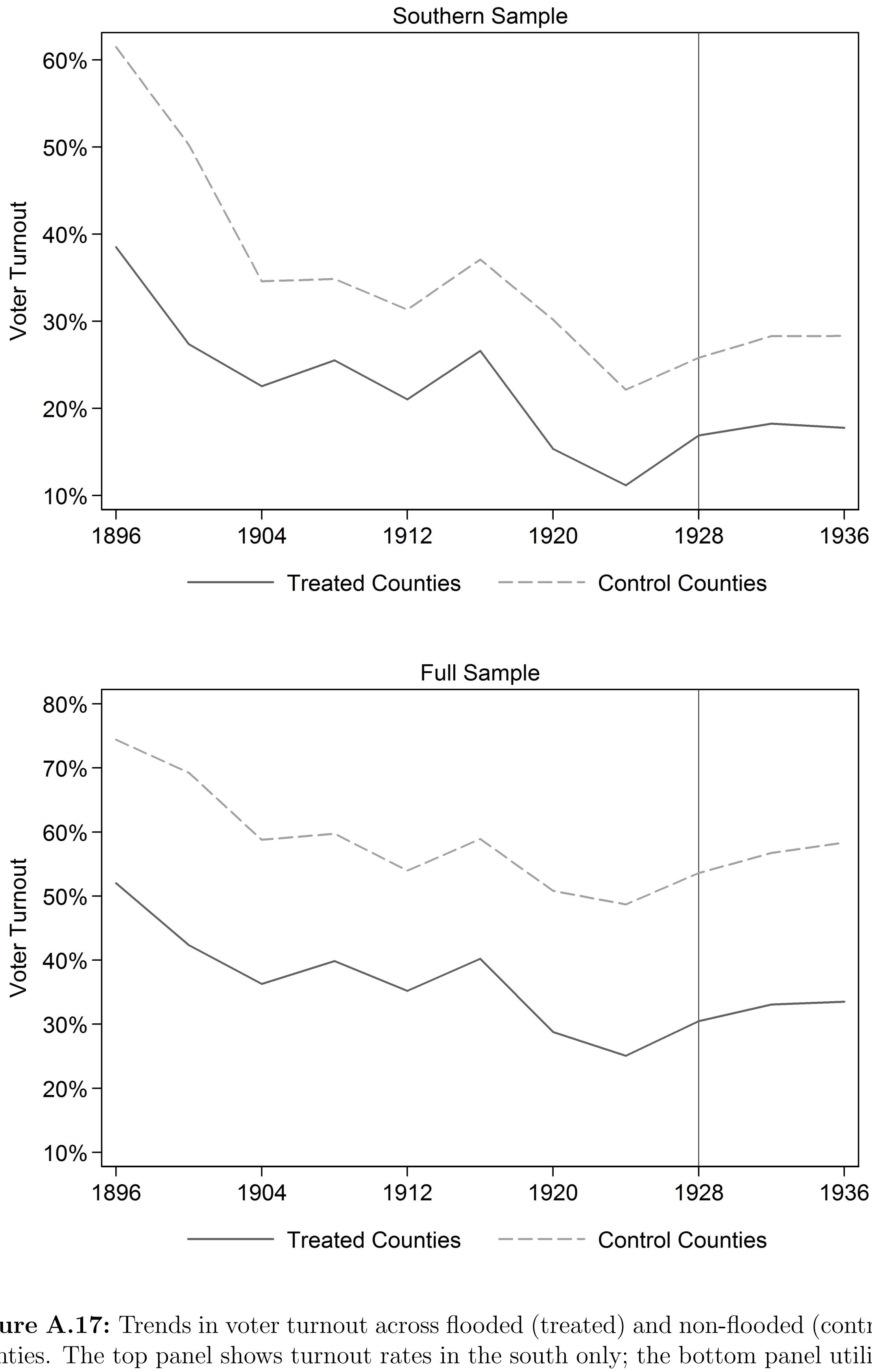

Figure A.17: Trends in voter turnout across flooded (treated) and non-flooded (control) counties. The top panel shows turnout rates in the south only; the bottom panel utilizes the full sample, including northern counties. Turnout rates do not appear to decrease in treated counties in 1928.

flooded counties. The period under consideration coincides with the Great Migration, which shifted the concentration of black residents from the rural south to the urban centers of the north. Based on population counts from the 1920 and 1930 censuses, however, we do not believe that an out-migration mechanism can explain our result. Among southern counties, the population of flooded areas actually grew faster, on average, than non-flooded areas over the period 1920-1930. Moreover, for out-migration to explain such a sharp drop in support for Hoover, Hoover supporters would have to be much more likely to leave flooded counties, and simultaneously no more likely to leave non-flooded counties, than Smith supporters. Theoretically, we find this possibility unlikely. Empirically, the magnitude of differential out-migration that would be necessary are implausible: to explain a 15 point decline in support for Hoover via out-migration alone, flooded counties would have to lose roughly two-thirds of their Republican voters, while keeping all of their Democratic voters.

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.

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.

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.

Ho, Daniel E., Kosuke Imai, Gary King and Elizabeth A. Stuart. 2007. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." *Political Analysis* 15:199–236.

Ladd, Jonathan McDonald and Gabriel S. Lenz. 2009. "Exploiting a Rare Communication Shift to Document the Persuasive Power of the News Media." *American Journal of Political Science* 53:394–410.

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