Supplementary Materials for "Randomization Inference with Rainfall Data: Using Historical Weather Patterns for Variance Estimation"

Alicia Dailey Cooperman

Contents

1 Appendix: Tables and Figures

Replication materials available at Cooperman (2017).

List of Tables

List of Figures

Table 1: Replication of Main Findings in Gomez et al. (2007) with Additional Rainfall Data Sources and Measurement – Covariates Shown

	Dependent variable:				
	Voter Turnout				
	(1)	(2)	(3)	(4)	(5)
Rain (inches) - Gomez et al. (2007)	-0.833^{***}	-0.833^{***}			
	(0.107)	(0.107)			
Rain (inches) - Author			-1.052^{***}		
			(0.109)		
Rain Index - Author				-0.143^{***}	
				(0.029)	
Rain SPI - Author					-0.208^{***}
					(0.029)
Snow	-0.153^{*}	-0.156^{*}	-0.146	-0.080	-0.038
	(0.091)	(0.091)	(0.091)	(0.093)	(0.093)
% High School Graduates	0.537^{***}	0.541^{***}	0.536^{***}	0.546^{***}	0.548^{***}
	(0.044)	(0.044)	(0.044)	(0.044)	(0.044)
Median HH Income	0.235^{**}	0.228^{**}	0.240^{***}	0.230^{**}	0.230^{**}
	(0.092)	(0.092)	(0.092)	(0.092)	(0.092)
% African American	-0.029^{***}	-0.029^{***}	-0.029^{***}	-0.029^{***}	-0.028^{***}
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Farms per Capita	21.405^{***}	21.357^{***}	21.322^{***}	21.385^{***}	21.460^{***}
	(0.896)	(0.896)	(0.897)	(0.894)	(0.893)
Closed Days Pre-Election	-0.031^{***}	-0.031^{***}	-0.032^{***}	-0.032^{***}	-0.032^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Motor Voter	0.036	0.034	0.056	0.015	0.035
	(0.110)	(0.110)	(0.110)	(0.110)	(0.110)
Property Requirement Vote Law	-3.096^{***}	-3.100^{***}	-3.097^{***}	-3.110^{***}	-3.080^{***}
	(0.316)	(0.316)	(0.316)	(0.316)	(0.316)
Literacy Vote Law	-0.170	-0.167	-0.168	-0.161	-0.182^{*}
	(0.104)	(0.104)	(0.104)	(0.104)	(0.104)
Poll Tax Vote Law	-6.092^{***}	-6.094^{***}	-6.095^{***}	-6.071^{***}	-6.040^{***}
	(0.136)	(0.136)	(0.136)	(0.136)	(0.136)
Gubernatorial Elec.	-0.083	-0.086	-0.073	-0.099	-0.086
	(0.065)	(0.065)	(0.065)	(0.065)	(0.065)
Senatorial Elec.	0.016	0.021	0.022	0.025	0.026
	(0.051)	(0.051)	(0.051)	(0.051)	(0.051)
Turnout (lag)	0.758^{***}	0.758^{***}	0.758^{***}	0.759^{***}	0.759^{***}
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Intercept	13.204^{***}	13.215^{***}	13.247^{***}	13.061^{***}	13.081^{***}
	(0.234)	(0.234)	(0.234)	(0.233)	(0.233)
Sample	*Original*	*Limited*	*Limited*	*Limited*	*Limited*
Observations	43,340	43,305	43,305	43,305	43,305
Log Likelihood	-131,340	-131,212	-131,196	-131,231	-131,218
Akaike Inf. Crit.	262,741	262,485	262,453	262,523	262,497
Bayesian Inf. Crit.	263,001	262,745	262,713	262,783	262,757
Note:					*p<0.1; **p<0.05; ***p<0.01

Note: Table replicates the findings from Gomez et al. (2007) using a maximum-likelihood linear cross-sectional random effects model (CSRE) with random effects for each county, fixed effects for each year, and conventional standard errors (Gomez et al., 2015). Columns 1 and 2 use data from Gomez et al. (2007) replication package of rainfall in inches. Column 1 reflects the full dataset from Gomez et al. (2007), and Columns 2-5 reflect a dataset with 35 fewer observations due to 7 counties not being present when calculating rainfall using county lines from 2000. Columns 3-5 use rainfall data processed by the author in inches and as indices relative to the county's historical mean. Column 4 uses a rainfall index calculated using a z-score, and Column 5 uses a Standardized Precipitation Index based on the empirical cumulative distribution function.

Table 2: Fixed Effects Model Using Rainfall (inches) with Clustered Standard Errors and RI P-Value

	Clustered Standard Errors							RI
	None	CWA-Year	CWA	State-Year	State	Region-Year	Year	US-level
Coefficient	-1.181	-1.181	-1.181	-1.181	-1.181	-1.181	-1.181	-1.181
Standard Error	0.11	0.316	0.344	0.416	0.477	0.491	0.598	-
P-Value	<.001	<.001	0.001	0.005	0.013	0.016	0.048	0.051
No. Clusters	NA	1610	115	671	48	56	14	-

Note: All columns use rainfall (inches) data processed by the author. Column 1 replicates the findings from Gomez et al. (2007) using a two-way fixed effects model instead of the CSRE model, as used by Gomez and Hansford (2010) with similar data (Gomez et al., 2015). All covariates included but omitted from table. Column 2 uses clustered errors at the County Warning Area - Year level. Columns 3-7 use a two-way fixed effects model with standard errors clustered at the unit described. Column 8 provides the two-tailed p-value estimated with randomization inference from sampling distributions of 1000 estimated ATEs under the sharp null hypothesis of no effect, also using the two-way fixed effects model with US-level assignment. Potential randomizations are drawn from county-level rainfall on election or would-be election days from 1940-2012. US assignment assumes that treatment assignment is clustered within the entire United States in a given year, and this specification draws data for all counties within the same year when creating the potential randomizations for the sampling distribution.

Table 3: Fixed Effects Model Using Rainfall SPI with Clustered Standard Errors and RI P-Value

	Clustered Standard Errors							RI
	None	CWA-Year	CWA	State-Year	State	Region-Year	Year	US-level
Coefficient	-0.232	-0.232	-0.232	-0.232	-0.232	-0.232	-0.232	-0.232
Standard Error	0.029	0.091	0.107	0.131	0.163	0.173	0.217	-
P-Value	<.001	0.011	0.03	0.078	0.155	0.181	0.287	0.244
No. Clusters	NA	1610	115	671	48	56	14	-

Note: All columns use rainfall data processed by the author and transformed to a Standardized Precipitation Index. Column 1 replicates the findings from Gomez et al. (2007) using a two-way fixed effects model instead of the CSRE model, as used by Gomez and Hansford (2010) with similar data (Gomez et al., 2015). All covariates included but omitted from table. Column 2 uses clustered errors at the County Warning Area - Year level. Columns 3-7 use a two-way fixed effects model with standard errors clustered at the unit described. Column 8 provides the two-tailed p-value estimated with randomization inference from sampling distributions of 1000 estimated ATEs under the sharp null hypothesis of no effect, also using the two-way fixed effects model with US-level assignment. Potential randomizations are drawn from county-level rainfall on election or would-be election days from 1940-2012. US assignment assumes that treatment assignment is clustered within the entire United States in a given year, and this specification draws data for all counties within the same year when creating the potential randomizations for the sampling distribution.

Table 4: Randomization Inference Estimates of ATE Variance Under Different Assumptions, Including Estimates Using Pre-2001 Weather Patterns

		P-Values by Assumed Unit of Weather Assignment					
Variable	Estimated ATE	County	County Warning Area	State	Weather Region	US	US (Pre-2001)
Rainfall (in)	-1.052	<.001	<.001	0.014	0.069	0.084	0.105
Rainfall Index	-0.143	<.001	0.072	0.237	0.340	0.405	0.380
Rainfall SPI	-0.208	<.001	0.012	0.102	0.267	0.314	0.296

Note: Table draws estimated ATEs from Columns 3-5 in Table 2 in the original paper. Two-tailed p-values are estimated with randomization inference from sampling distributions of 1000 estimated ATEs under the sharp null hypothesis of no effect using five different assumptions of the clustered randomization procedure. Potential randomizations are drawn from county-level rainfall on each election day or potential election day from 1940-2012. The first row uses rainfall (in), the second row uses a z-score index, and the third row uses a Standardized Precipitation Index. County assignment assumes that each county is independent and supposes simple random assignment at the county-level. County warning area (CWA) and region are National Weather Service classifications. For CWA, state, and region specifications, I assume that treatment assignment is clustered within that unit in a given year but each unit is independent of other units. US assignment assumes that treatment assignment is clustered within the entire United States in a given year. The last column also uses US assignment but limits the potential randomizations to be from 1940-2000.

Table 5: Correlations of Rainfall (in) Between Counties in Each State

State	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Num. Counties
ALABAMA	0.215	0.693	0.789	0.765	0.873	0.995	67
ARIZONA	0.113	0.480	0.738	0.681	0.876	0.991	15
ARKANSAS	0.107	0.659	0.805	0.757	0.895	0.995	75
CALIFORNIA	-0.079	0.145	0.578	0.486	0.768	0.995	58
COLORADO	-0.102	0.205	0.390	0.431	0.654	0.992	63
CONNECTICUT	0.706	0.855	0.903	0.891	0.951	0.975	8
DELAWARE	0.707	0.799	0.891	0.845	0.915	0.939	3
FLORIDA	-0.010	0.370	0.549	0.550	0.733	0.995	67
GEORGIA	0.328	0.663	0.768	0.752	0.856	0.994	159
IDAHO	-0.013	0.283	0.574	0.529	0.733	0.988	44
ILLINOIS	0.023	0.401	0.621	0.602	0.803	0.996	102
INDIANA	0.184	0.548	0.730	0.703	0.874	0.997	92
IOWA	0.078	0.474	0.705	0.663	0.872	0.997	99
KANSAS	-0.025	0.433	0.651	0.609	0.808	0.997	105
KENTUCKY	-0.037	0.458	0.648	0.622	0.814	0.995	120
LOUISIANA	0.039	0.384	0.646	0.602	0.836	0.995	64
MAINE	0.260	0.597	0.756	0.732	0.907	0.994	16
MARYLAND	0.189	0.670	0.821	0.773	0.921	0.992	24
MASSACHUSETTS	0.183	0.480	0.745	0.682	0.917	0.988	14
MICHIGAN	-0.030	0.315	0.607	0.564	0.814	0.995	83
MINNESOTA	0.336	0.739	0.825	0.812	0.908	0.998	87
MISSISSIPPI	0.137	0.501	0.672	0.647	0.802	0.994	82
MISSOURI	0.040	0.356	0.575	0.567	0.779	0.995	115
MONTANA	-0.086	0.224	0.387	0.421	0.601	0.990	56
NEBRASKA	0.081	0.427	0.626	0.612	0.798	0.992	93
NEVADA	-0.033	0.164	0.399	0.423	0.631	0.996	17
NEW HAMPSHIRE	0.619	0.851	0.911	0.879	0.953	0.984	10
NEW JERSEY	0.407	0.808	0.888	0.858	0.950	0.995	21
NEW MEXICO	0.009	0.463	0.679	0.630	0.824	0.989	33
NEW YORK	0.251	0.580	0.701	0.694	0.831	0.999	62
NORTH CAROLINA	0.130	0.522	0.649	0.656	0.794	0.997	100
NORTH DAKOTA	0.087	0.485	0.616	0.632	0.809	0.993	53
OHIO	0.022	0.464	0.672	0.637	0.833	0.995	88
OKLAHOMA	0.205	0.661	0.824	0.769	0.911	0.996	77
OREGON	0.194	0.523	0.697	0.673	0.855	0.989	36
PENNSYLVANIA	0.218	0.666	0.770	0.753	0.859	0.995	67
RHODE ISLAND	0.956	0.970	0.982	0.978	0.988	0.993	5
SOUTH CAROLINA	0.278	0.595	0.697	0.714	0.851	0.989	46
SOUTH DAKOTA	0.207	0.669	0.822	0.776	0.910	0.996	66
TENNESSEE	0.115	0.674	0.803	0.773	0.901	0.997	95
TEXAS	-0.103	0.148	0.338	0.389	0.613	0.997	254
UTAH	0.229	0.540	0.648	0.638	0.754	0.968	29
VERMONT	0.592	0.801	0.886	0.860	0.939	0.982	14
VIRGINIA	0.094	0.474	0.662	0.648	0.853	1.000	135
WASHINGTON	0.061	0.575	0.711	0.682	0.814	0.992	39
WEST VIRGINIA	0.560	0.833	0.914	0.886	0.962	0.998	55
WISCONSIN	0.278	0.587	0.754	0.732	0.883	0.996	72
WYOMING	0.047	0.445	0.665	0.606	0.790	0.965	23

Note: Table calculates correlations between all counties in each state from 1940-2012. For each state, I calculated the correlation of rainfall in inches between county i and county j (including where i=j, creating an n x n matrix), and then I calculated summary statistics on all cells above the diagonal. Rainfall inches calculated by author.

Table 6: Correlations of Rainfall (SPI) Between Counties in Each State

State	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Num. Counties
ALABAMA	0.577	0.768	0.833	0.830	0.896	0.993	67
ARIZONA	0.328	0.616	0.708	0.693	0.792	0.953	15
ARKANSAS	0.457	0.739	0.822	0.809	0.888	0.991	75
CALIFORNIA	0.026	0.508	0.731	0.659	0.846	0.984	58
COLORADO	0.107	0.443	0.583	0.591	0.742	0.988	63
CONNECTICUT	0.866	0.935	0.950	0.945	0.967	0.985	8
DELAWARE	0.832	0.886	0.940	0.909	0.948	0.956	3
FLORIDA	0.057	0.436	0.595	0.586	0.745	0.983	67
GEORGIA	0.197	0.666	0.781	0.759	0.869	0.996	159
IDAHO	0.258	0.575	0.686	0.679	0.803	0.977	44
ILLINOIS	0.174	0.596	0.768	0.723	0.868	0.994	102
INDIANA	0.523	0.761	0.840	0.827	0.905	0.993	92
IOWA	0.239	0.696	0.795	0.772	0.880	0.991	99
KANSAS	0.231	0.584	0.700	0.692	0.816	0.993	105
KENTUCKY	0.219	0.669	0.795	0.761	0.878	0.995	120
LOUISIANA	0.292	0.674	0.785	0.760	0.871	0.992	64
MAINE	0.541	0.779	0.845	0.839	0.931	0.981	16
MARYLAND	0.478	0.781	0.869	0.842	0.930	0.986	24
MASSACHUSETTS	0.489	0.761	0.859	0.823	0.913	0.987	14
MICHIGAN	0.173	0.516	0.707	0.674	0.849	0.991	83
MINNESOTA	0.316	0.619	0.744	0.728	0.852	0.990	87
MISSISSIPPI	0.534	0.736	0.813	0.806	0.884	0.994	82
MISSOURI	0.201	0.573	0.684	0.690	0.814	0.991	115
MONTANA	0.082	0.494	0.614	0.614	0.747	0.974	56
NEBRASKA	0.308	0.608	0.717	0.708	0.813	0.985	93
NEVADA	0.063	0.591	0.703	0.671	0.783	0.976	17
NEW HAMPSHIRE	0.659	0.843	0.910	0.884	0.949	0.975	10
NEW JERSEY	0.663	0.882	0.925	0.910	0.952	0.991	21
NEW MEXICO	0.276	0.598	0.706	0.693	0.801	0.968	33
NEW YORK	0.372	0.650	0.742	0.741	0.841	0.992	62
NORTH CAROLINA	0.423	0.665	0.765	0.756	0.853	0.991	100
NORTH DAKOTA	0.312	0.619	0.723	0.721	0.835	0.986	53
OHIO	0.387	0.725	0.805	0.793	0.877	0.990	88
OKLAHOMA	0.425	0.684	0.775	0.768	0.858	0.987	77
OREGON	0.419	0.662	0.760	0.751	0.845	0.982	36
PENNSYLVANIA	0.539	0.775	0.839	0.830	0.893	0.994	67
RHODE ISLAND	0.956	0.972	0.976	0.976	0.983	0.992	5
SOUTH CAROLINA	0.558	0.747	0.822	0.816	0.892	0.987	46
SOUTH DAKOTA	0.460	0.732	0.823	0.810	0.896	0.992	66
TENNESSEE	0.216	0.729	0.845	0.804	0.911	0.995	95
TEXAS	-0.104	0.355	0.524	0.526	0.697	0.988	254
UTAH	0.368	0.586	0.714	0.694	0.802	0.965	29
VERMONT	0.664	0.802	0.891	0.869	0.935	0.982	14
VIRGINIA	0.397	0.702	0.805	0.788	0.878	0.998	135
WASHINGTON	0.378	0.694	0.774	0.760	0.836	0.987	39
WEST VIRGINIA	0.704	0.829	0.875	0.874	0.921	0.990	55
WISCONSIN	0.449	0.695	0.816	0.789	0.895	0.994	72
WYOMING	0.143	0.563	0.704	0.674	0.802	0.948	23

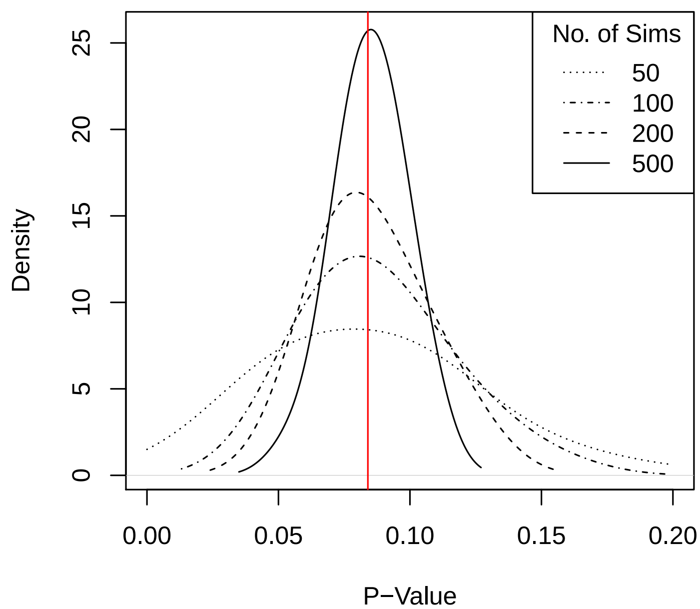
Note: Table calculates correlations between all counties in each state from 1940-2012. For each state, I calculated the correlation of the rainfall standardized precipitation index (SPI) between county i and county j (including where i=j, creating an n x n matrix), and then I calculated summary statistics on all cells above the diagonal. Rainfall SPI calculated by author.

Table 7: Summary Statistics for P-Values Using RI by Number of Simulations

	Mean	SD	Min	Max
No. Sims = 50	0.084	0.040	0.000	0.200
No. Sims = 100	0.088	0.026	0.050	0.160
No. Sims = 200	0.086	0.019	0.055	0.125
No. Sims = 500	0.085	0.012	0.054	0.108

Note: Table shows summary statistics for p-values based on estimated ATEs for the Rainfall (inches) and US-Year specifications using different numbers of simulations. Each row shows summary statistics for 50 p-values. Two-tailed p-values were calculated by comparing the observed ATE to the sampling distribution of estimated ATEs from simulations of alternate randomizations from national weather patterns. To calculate the p-values in the last row, I performed the RI procedure using 500 simulations of alternate randomizations. I repeated this 50 times, and the solid line represents the 50 p-values estimated using 500 simulations each. I did the same for 200, 100, and 50 simulations. The results are sensitive to the random sample of weather patterns that are drawn and are less precise when calculating the RI p-value with fewer simulations of alternate randomizations from national weather patterns.

Figure 1: Distribution of P-Values Using RI by Number of Simulations



Note: Figure shows density plots for p-values based on estimated ATEs for the Rainfall (inches) and US-Year specifications using different numbers of simulations. Each density plot represents 50 p-values. The vertical line represents the RI p-value of 0.084 calculated by comparing the observed ATE to the sampling distribution of estimated ATEs from 1000 simulations of alternate randomizations from national weather patterns. To calculate the p-values in the solid line, I performed the RI procedure using 500 simulations of alternate randomizations. I repeated this 50 times, and the solid line represents the 50 p-values estimated using 500 simulations each. I did the same for 200, 100, and 50 simulations. The results are sensitive to the random sample of weather patterns that are drawn and are less precise when calculating the RI p-value with fewer simulations of alternate randomizations from national weather patterns.

2 Appendix: Data Processing

2.1 Rainfall Data Source

Rainfall data is from the NOAA NCDC GHCN-Daily dataset. It was previously known as NCDC DSI 3200, which is the Summary of the Day database cited by Gomez et al. (2007). Detailed information about the dataset is available at `www.ncdc.noaa.gov/oa/climate/ghcn-daily` and the data is available in a .csv file for each year at `ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/daily/by_year/`. To calculate rainfall values, I used weather station data from GHCN-Daily version 3.21.

While the dataset is portrayed as having 20,000 weather stations across the U.S., and was pitched this way by the company EarthInfo, Inc. that disseminated the information at the time of the original publication, the number of active stations at any given time is closer to 8,000. By 2014, there have technically been over 51,000 stations ever used in the U.S., but only 9,417 stations were used during the year 1996 and only 7,637 stations were used on November 5, 1996 election day. The weather station data points were formatted using R to create a shapefile with projection NAD 1983, which was the projection of the data as of Version 2.2 of the NOAA Weather and Climate Toolkit (`www.ncdc.noaa.gov/wct/newfeatures.php`).

2.2 County Boundaries Source

The county boundaries were downloaded as a shapefile (`co99_d00_shp.zip`) from the U.S. census website and are the bounds from 2000 at `https://www.census.gov/geo/maps-data/data/cbf/cbf_counties.html`. Documentation describes them as being based on the NAD 1983 projection, which used the "Define Projection" in ArcMap 10.2.2.

2.3 Kriging Interpolation

I used python code for ArcMap to interpolate county values in the following way. I projected the weather station points to the USA Contiguous Equidistant Conic projection with units of meters, which preserves distances and is appropriate for the Kriging process. After selecting points from each day, I ran the Kriging procedure using Universal linear drift with cell size of 4,000 meters squared, as done in the original Gomez et al. (2007) paper.

The default and preferred specification is to create a cell value based on the 12 nearest data points, since not all counties have weather stations in them. The Kriging process incorporates the distance and weighs the closest stations more heavily. The more neighboring points are used, the more the data is smoothed across the country and the more spatial correlation that exists. There is a trade-off between using more points for more information while not oversmoothing the data and losing the variation in actual rainfall.

The correlation coefficient between the Gomez et al. (2007) rainfall measure and mine is 0.95 with Kriging estimates using 12 nearest stations and 0.90 with Kriging estimates using 100 nearest stations. However, the difference between my calculations and data from Gomez et al. (2007) can vary by year. For example, the values that I calculated using 12 nearest neighbors were only correlated at $\rho = 0.6$ with data from Gomez et al. (2007) for the election day in 1996. When I used 100 nearest neighbors, this increased to $\rho = 0.79$. This paper uses the Kriging values from 12 nearest neighbors.

2.4 Zonal Statistics by County

The raster values from the Kriging analysis were Projected to USA Contiguous Albers Equal Area Conic, to preserve area in order to take spatial averages across county areas. The raster was Resampled to cell sizes of 1000 meters squared, and Zonal Statistics as Table was used on the resampled, projected kriging values to export a .dbf file with an average rainfall value for each county-date. Since the GHCN data is in tenths of a millimeter, this was converted to inches by dividing by 254. The Kriging interpolation process did not account for the fact that rainfall amounts are censored at 0, and some counties with no rainfall received Kriging-based estimates of negative rainfall. These were recoded to equal 0. This only led to four county-year observations having a different above-average treatment status (z-score greater than 0), among 226,884 county-year rainfall observations (3108 counties from 1940-2012).

The python code for spatial interpolation and zonal statistics is available on request.

2.5 Measurement Error and Spatial Correlation

The clustered random assignment of rainfall described in the paper is further complicated by spatial correlation in measurement error of county rainfall. Relying on one weather station in a large county may not reflect the political story, especially if the weather station is in a relatively drier part of the county while the more populated areas experience heavier rainfall. In addition, many counties lack weather stations and their rainfall amount on election day is likely to be interpolated from weather stations in neighboring counties, including those hundreds of miles away. Weather patterns are already correlated across counties, and the interpolation required in data analysis only exacerbates this problem.

My sample is smaller than the dataset from Gomez et al. (2007) by 35 observations since I used county lines from 2000 to estimate county mean rainfall levels, and 7 counties had been redistricted and no longer existed. I therefore did not calculate historical rainfall amounts for them and needed to exclude their political data from a few presidential elections. The land area for these seven counties was redistricted to be part of other neighboring counties, and the mean rainfall for counties in that area may be slightly incorrect in some years due to changing political boundaries.

I use the same weather station source as Gomez et al. (2007), but I have slightly different estimates for rainfall amounts on election days. The correlation coefficient between their rainfall measure and mine is 0.95 with Kriging estimates using 12 nearest stations and

0.90 with Kriging estimates using 100 nearest stations. To calculate rainfall values, I used weather station data from GHCN-Daily version 3.21, while Gomez et al. (2007) likely used version 2.0, which was placed online on December 21, 2006. Version 2.1 was released on November 23, 2009. Based on the status updates at `ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/daily/status.txt`, it seems that errors in the database have been corrected over time and weather stations have been added. I expect that the differences in rainfall estimates are due to retroactive updates to the data when additional weather station data were introduced. Other differences may stem from the spatial processing procedures used to calculate county mean rainfall from weather station data, especially in places with fewer weather stations and therefore more spatial interpolation to calculate county rainfall estimates.

2.6 Standardized Indices

The first index, referred to as "rainfall index," is a z-score for each county's rainfall values during 1940-2012:

$$\frac{x_{it} - \bar{x}_i}{s_i} \tag{1}$$

where x_{it} is the rainfall value on election or would-be election day for county i in year t, $\bar{x}_i$ is the mean rainfall value on election or would-be election days in county i during 1940-2012, and s_i is the standard deviation of rainfall values on election or would-be election days in county i during 1940-2012.

The second index, referred to as "rainfall SPI," uses an empirical cumulative distribution function (ECDF) and symmetrizing procedure to create a index that reduces the extremity and sensitivity to outliers. The R function is as follows and was drawn from `http://joewheatley.net/20th-century-droughts/`.

```
getSPI <- function(y){
  #empirical, symmetrized Standard Precipitation Index
  fit.cdf <- ecdf(y)
```

```
  fit.rcdf <- ecdf(-y)
  cdfs <- fit.cdf(y)
  rcdfs <- 1 -fit.rcdf(-y)
  #invert normal
  spi.t <- qnorm(cdfs)
  spi.tp <- na.omit(spi.t[ spi.t != Inf ]) #drop Inf
  #reversed
  rspi.t <- qnorm(rcdfs)
  rspi.tp <- na.omit(rspi.t[ rspi.t != -Inf ]) #drop Inf
  #symmetrise
  spi.sym <- (spi.t+rspi.t)/2
  spi.sym[which(spi.t == Inf)] <- rspi.t[which(spi.t==Inf)]
  spi.sym[which(rspi.t == -Inf)] <- spi.t[which(rspi.t==-Inf)]
  spi <- spi.sym - mean(spi.sym)
  spi <- spi/sd(spi)
  return(spi)
}
```

References

Cooperman, A. D. (2017). Replication data for: Randomization inference with rainfall data: Using historical weather patterns for variance estimation. *Harvard Dataverse* *http://dx.doi.org/10.7910/DVN/RJF61A*.

Gomez, B. T. and T. G. Hansford (2010). Estimating the electoral effects of voter turnout. *American Political Science Review 104*(02), 268–288.

Gomez, B. T., T. G. Hansford, and G. A. Krause (2007). The republicans should pray for rain: Weather, turnout, and voting in us presidential elections. *Journal of Politics 69*(3), 649–663.

Gomez, B. T., T. G. Hansford, and G. A. Krause (Accessed May 18, 2015). Replication data for: The republicans should pray for rain. *http://myweb.fsu.edu/bgomez/research.html*.