The Republicans Should Pray for Rain:
Weather, Turnout, and Voting in U.S. Presidential Elections

Supplemental Appendix

Brad T. Gomez, University of Georgia
Thomas G. Hansford, University of California, Merced
George A. Krause, University of Pittsburgh

In our paper, we reference a number of methodological issues and supplementary analyses. This document provides a further discussion of these issues and presents the results of the additional analyses.

Accounting for the Panel Aspect of Our Data

Because we utilize panel data in our analyses, several statistical issues must be addressed. In the text of our paper, we succinctly address our decision to employ a random effects model allowing for county-specific heterogeneity while including fixed effects for election-specific heterogeneity. Here, we provide additional details regarding these choices.

Our panel is heavily cross-section dominant, i.e., the number of counties included in the analysis is much larger than the number of included elections ($T = 14$; $N = 3{,}115$ (max); Total Observations = 43,340). Some econometricians note that because differences between timewise fixed effects and timewise random effects becomes larger when a panel is cross-section dominant, a Hausman test should be employed to determine the appropriate method for handling unobserved temporal heterogeneity in panel models (Hsiao 2003, 51). We, however, cannot obtain a suitable means to jointly estimate both cross-sectional and timewise random effects in a manner that does not leave the model unidentified. Hence, we handle the temporal heterogeneity with fixed effects. This appears to have been a valid choice, since that the election year (time) dummies are statistically significant and possess large joint F-statistic.

The standard approaches for TSCS research designs where $T \geq N$ by a slight or moderate amount (i.e., GLS-ARMA, Beck and Katz panel corrected standard errors) are not applicable for our substantive data problem (e.g., see Beck and Katz 1995, 644; Stimson 1985, 928-929 for political science treatments).

Baltagi (1999, 309) also notes that accounting for cross-sectional fixed effects (CSFEs) is not a sound practice for our panel design because of problems arising from collinearity—even when one has a more balanced panel design consisting of $N = 50$ and $T = 10$, for example.

Because the number of cross-sectional units exceeds time units by a factor of just over 220, a county-level fixed effects modeling strategy is inappropriate on both econometric and substantive grounds. In the former case, modeling cross-sectional fixed effects (CSFEs) in these voter turnout and partisan vote share equations is problematic due to collinearity (see Baltagi 1999: 309) and the standard rank condition assumption pertaining to the CSFEs will not be met (Assumption FE.2: Wooldridge 2003: 269).[1] Substantively, our model contains several county-level control variables that are viewed as critical determinants of cross-sectional variance in turnout and vote share models. The inclusion of county-level fixed effects would unnecessarily diminish these explanatory factors. Furthermore, we refrain from employing fixed effect variance decomposition estimation methods (see Plumper and Troeger 2004) in our study, since not only do most of the county-level independent variables vary considerably through time, but more importantly, we possess what resembles a traditional panel design (a very large number of cross-sections [N] relative to time points [T]) as opposed to a pooled time series cross-section design where $N > T$ by a factor ranging between 1 and 20.

Alternatively, we choose to model cross-sectional heterogeneity via theoretical variables. This omits spurious relationships that might exist between weather and turnout due to differences across counties in socioeconomic status or institutional factors related to the cost of voting. We treat timewise heterogeneity via the modeling of fixed effects. The substance underlying this decision is simple: we

[1] Hsiao (2003: 35) echoes this sentiment by contending that "if the explanatory variables contain some time-invariant variables, z_i, their coefficients cannot be estimated by CV (covariance estimation), because the covariance transformation eliminates z_i from the covariance transformed equation." Thus, modeling the cross-sectional heterogeneity as either random or deterministic process risks "throwing out the baby with the bathwater" by treating important substantive cross-sectional differences as nuisance.

treat each presidential election is a unique event. We believe that in a particular presidential election, voters across counties have a common view of the issues that define the electoral contest, as well as the relative performance of the incumbent party residing in the White House.[2] On an econometric level, the use of time dummies is appropriate in shorter panels, since proper stochastic modeling of the dependent variable is difficult when T is small (Arellano 2003: 60-64). Therefore, it is preferable to allow for time-varying intercepts when one has a cross-sectional dominant panel ($N >> T$) (Wooldridge 2003: 170).

Nevertheless, to demonstrate the robustness of our findings, we also estimated our turnout and vote share models using county-level fixed effects. The results of these alternative analyses are contained in Tables A1 and A2. In both of the turnout models and both of the vote share models, the inferences drawn with the fixed effects approach are very similar to those drawn from the random effects approach (results presented in the paper). Two differences are worth noting. First, in the first turnout model (Table A1, Model 1) the coefficient estimate for *Election Day Snow* is statistically significant when the fixed effects approach is employed. Second, in both Models 1 and 2 in Table A1, the coefficient estimates for each of our precipitation variables are larger in magnitude than those found in Table 1 of the main text. Thus, the results presented in the paper are conservative by comparison.

Alternative Model Estimation: Arellano-Bond GMM

A potential problem with the estimated models reported in the article is the possibility of nonzero covariance between the lagged turnout or vote share variables and the stochastic residual disturbances. If this is the case, then our coefficient estimates may yield a Nickell (1981) bias. Normally, this bias is only problematic when the nonzero covariance assumption of panel models is violated, the coefficient on the lagged dependent variable approaches unity (1.0), and when T is of modest size (in most applications, $T \leq$

[2] This is not to suggest that all counties view the voter turnout decision or retrospective voting calculus similarly in *absolute terms*, but rather that they do so in *relative terms* insofar that each county's level of voter turnout and GOP partisan electoral support should vary either above or below its mean historical level in response to the unique circumstances surrounding each presidential election.

10). Although we do not think Nickell bias poses a serious problem for our random effects models given that the coefficients on our lagged dependent variables do not exceed .758 (Baseline Voter Turnout model) and $T = 14$, we address this issue here, nonetheless.

Specifically, we employ Arellano and Bond's (1991) GMM estimation strategy, whereby, we employ alternative lagged level(s) of the lagged dependent variable in question (i.e., y_{t-2}; y_{t-2}, y_{t-3}; y_{t-2}, y_{t-3}, y_{t-4}) to eliminate any such coefficient biases that might plague our statistical results reported in the paper. This model is re-estimated in first difference form (see Baltagi (2005, 135-142) for the technical details of this approach). Because the statistical findings are substantively similar across alternative instrument lag structures, for brevity, we only report the results from the simple instrument involving a single lag (i.e., y_{t-2}). The results of these models are reported in Tables A3 and A4.

On a substantive level, the turnout model results from the Arellano-Bond GMM estimation approach comport very well with what we obtained from our random effects MLE estimation procedure. Although the 2nd order serial correlation and Sargan tests are both significant at conventional levels, we are not concerned in this instance about an underspecified set of instruments since these results are consistent with those containing additional lags as instruments (which also happen to reject the null for these diagnostic hypotheses tests). Moreover, we are limited in the number of instruments (lags) that we can adopt given that we have a limited T ($T = 14$). The results for the vote share models are somewhat different than those presented in the paper. In the additive model, the inference drawn about (*Election Day Rain – Normal Rain*) stays the same, but the effect of *Snow* is not significant when the Arellano-Bond approach is used. The Arellano-Bond results also do not support the "two-effects model," unlike the results presented in the manuscript. These differences in results are noted in footnote 25.

Using Just Observed Precipitation in the Vote Share Models

In our paper, we include (*Election Day Rain – Normal Rain*) and (*Election Day Snow – Normal Snow*) in our vote share models. The substantive results change very little if we instead simply include *Election Day Rain* and *Election Day Snow*. The results for this specification are presented in Table A5. The only distinction between these results and those presented in the paper is that in the Two Effects

Model *Election Day Rain* is negative and significant while in the results in the paper the estimate for (*Election Day Rain – Normal Rain*) is insignificant. Given the presence of the interaction term, this difference only has implications for counties where Republican vote share in the previous three elections approaches zero. Otherwise, the same pattern in the data emerges. The more Republican a county is, the more rain helps Republican vote share.

Why Not Ecological Inference?

We employ standard panel econometric methods instead of ecological inference (EI) techniques in our empirical investigation for clear substantive and methodological reasons. Substantively, our aim is to make aggregate county-level assessments regarding the effect of weather on voter turnout and partisan vote shares, as opposed to individual-level inferences. This choice is grounded in the fact that we treat weather in a given county as being fixed across all individuals who comprise the electorate. EI applications in political science, on the other hand, attempt to gain empirical purchase on individual characteristics (e.g., race, split-ticket voting) relating to individual-level electoral behavior (e.g., King 1997; Burden and Kimball 1998). This is not our aim. On a methodological level, using panel methods allows us to deal with the unobserved heterogeneity, via random effects estimation, that plagues our cross-sectional dominant data design. An EI approach does not afford us the opportunity to model such heterogeneity in a straightforward manner.

Details Regarding Changes to Electoral College Outcomes under Hypothetical Weather Scenarios

In our paper, we consider two hypothetical weather scenarios and then use our models to predict the winner of each of the states in presidential elections from 1948 to 2000. In the first scenario, the election day in question is perfectly dry. In the second, each county experiences the heaviest rainfall/snowfall that is observed for the county on election day during the 1948-2000 time period. We then compare the results of these simulations with actual Electoral College outcomes to determine what effect very dry or very wet weather would have had. Table A6 provides the specific changes that our models predict would have happened under the “dry” and “wet” scenarios.

TABLE A1. Fixed Effects Model of County-Level Voter Turnout in U.S. Presidential Elections, 1948-2000.

Independent Variable	Model 1 Coefficient Estimate (Robust Standard Error)	Model 2 Coefficient Estimate (Robust Standard Error)
Election Day Rain	-.977*	---
	(.137)	
Election Day Snow	-.553*	---
	(.129)	
Election Day Rain - Normal Rain	---	-1.112*
		(.138)
Election Day Snow - Normal Snow	---	-.626*
		(.129)
% High School Graduates	.163	.167
	(.111)	(.111)
Income	.501*	.492*
	(.190)	(.190)
% African American	-.218*	-.219*
	(.019)	(.019)
Rural	16.054*	16.181*
	(2.645)	(2.649)
Registration Closing Date	-.036*	-.036*
	(.002)	(.002)
Motor Voter	-.402*	-.397*
	(.100)	(.100)
Property Requirement	-4.090*	-4.081*
	(.335)	(.336)
Literacy Test	-1.509*	-1.504*
	(.155)	(.155)
Poll Tax	-8.187*	-8.196*
	(.216)	(.216)
Gubernatorial Election	-1.126*	-1.127*
	(.111)	(.111)
Senate Election	.061	.061
	(.051)	(.051)
Turnout_{t-1}	.602*	.602*
	(.006)	(.006)
Constant	24.838*	24.718*
	(.554)	(.553)
σ_μ	3.391	3.394
ρ	.334	.334
Number of Observations	43,340	43,340
R-squared	.845	.845
F Test	2,232*	2,232*

* $p \leq .05$ (two-tailed test). Model also includes fixed effects for election. Estimates are available from authors.

TABLE A2. Fixed Effects Model of County-Level Republican Candidate Vote Share in U.S. Presidential Elections, 1948-2000.

	Conventional Model	*Two Effects Model*
Independent Variable	Coefficient Estimate (Robust Standard Error)	Coefficient Estimate (Robust Standard Error)
(Election Day Rain - Normal Rain)	2.865* (.164)	.984 (.748)
(Election Day Snow - Normal Snow)	.304* (.113)	.982 (.734)
(Election Day Rain - Normal Rain) × Previous Republican Vote Share	---	.043* (.015)
(Election Day Snow - Normal Snow) × Previous Republican Vote Share	---	-.013 (.014)
Previous Republican Vote Share	.577* (.006)	.579* (.006)
Constant	16.899* (.244)	16.877* (.244)
σ_μ	3.890	1.567* (.075)
ρ	.170	.031* (.003)
Number of Observations	43,294	43,294
R-squared	.665	.665
F test	4,289*	3,825*

* $p \leq .05$ (two-tailed test). Model also includes fixed effects for election; coefficient estimates can be obtained from the authors.

TABLE A3. Arellano-Bond Model of County-Level Voter Turnout in U.S. Presidential Elections, 1948-2000.

Independent Variable	Coefficient Estimate	Standard Error
Δ Rain	-1.792*	.127
Δ Snow	-.896*	.116
Δ % High School Graduates	-.568*	.135
Δ Income	.395	.246
Δ % African American	-.364*	.022
Δ Rural	.338	3.611
Δ Registration Closing Date	-.129*	.003
Δ Motor Voter	-.969*	.162
Δ Property Requirement	-4.775*	.571
Δ Literacy Test	1.439*	.249
Δ Poll Tax	-8.241*	.219
Δ Gubernatorial Election	-.260	.166
Δ Senate Election	.087	.051
Δ Turnout_{t-1}	.575*	.006
Constant	10.669*	.130
Number of Observations	40,211	
Wald Test (chi-square, 26 d.f.)	78,190*	
Sargan Test (chi-square, 90 d.f.)	5684*	
1st Order Arellano-Bond Autocorrelation Test (z stat.)	-82.96*	
2nd Order Arellano-Bond Autocorrelation Test (z stat.)	4.95*	

* $p \leq .05$ (two-tailed test) Model also includes fixed effects for each election (these estimates can be obtained from authors).

TABLE A4. Arellano-Bond Model of County-Level Republican Candidate Vote Share in U.S. Presidential Elections, 1948-2000.

Independent Variable	*Conventional Model* Coefficient Estimate (Standard Error)	*Two Effects Model* Coefficient Estimate (Standard Error)
Δ(Election Day Rain - Normal Rain)	4.207* (.242)	4.870* (.844)
Δ(Election Day Snow - Normal Snow)	.130 (.202)	2.812* (1.033)
Δ(Election Day Rain - Normal Rain) × Previous Republican Vote Share	---	-.016 (.019)
Δ(Election Day Snow - Normal Snow) × Previous Republican Vote Share	---	-.050* (.019)
Republican Vote Share$_{t-1}$	-.098* (.006)	-.099* (.006)
Republican Vote Share$_{t-2}$	.246* (.005)	.245* (.005)
Republican Vote Share$_{t-3}$	.216* (.005)	.216* (.005)
Constant	-1.437* (.044)	-1.437* (.015)
Number of Observations	33,983	33,983
Wald Test (chi-square, 15 d.f.)	39,765*	39,794*
Sargan Test (chi-square, 85 d.f.)	15,730*	15,730*
1st Order Arellano-Bond Autocorrelation Test (z stat.)	-124.51*	-124.29*
2nd Order Arellano-Bond Autocorrelation Test (z stat.)	11.16*	39.30*

* $p \leq .05$ (two-tailed test). Model also includes fixed effects for each election (these estimates can be obtained from authors).

TABLE A5. Random Effects Model of County-Level Republican Candidate Vote Share in U.S. Presidential Elections, Including Only Observed Precipitation.

	Conventional Model	*Two Effects Model*
Independent Variable	Coefficient Estimate (Standard Error)	Coefficient Estimate (Standard Error)
Election Day Rain	1.328* (.190)	-1.306* (.595)
Election Day Snow	.393* (.159)	-.103 (.808)
Election Day Rain × Previous Republican Vote Share	---	.061* (.013)
Election Day Snow × Previous Republican Vote Share	---	.009 (.015)
Previous Republican Vote Share	.734* (.004)	.579* (.006)
Constant	10.886* (.225)	16.877* (.244)
σ_μ	1.580* (.075)	1.585* (.076)
ρ	.032* (.003)	.032* (.003)
Number of Observations	43,294	43,294
Log-Likelihood	-155,730	-155,719
LR Test (chi-square, 16 and 18 d.f., respectively)	47,696*	47,735 *

* $p \leq .05$ (two-tailed test). Model also includes fixed effects for election; coefficient estimates can be obtained from the authors.

TABLE A6. Changes to Electoral College Outcomes under Hypothetical Weather Scenarios.

Scenario	Election	State	Electoral College Votes	Historical Winner	Winner Under Hypothetical Scenario
Dry Election	1992	North Carolina	14	G. H. W. Bush	Bill Clinton
	2000	Florida	25	George W. Bush	Al Gore
Wet Election	1948	Illinois	28	Harry Truman	Thomas Dewey
	1948	Ohio	25	Harry Truman	Thomas Dewey
	1952	Kentucky	10	Adlai Stevenson	Dwight Eisenhower
	1956	Missouri	13	Adlai Stevenson	Dwight Eisenhower
	1960	Delaware	3	John Kennedy	Richard Nixon
	1960	Illinois	27	John Kennedy	Richard Nixon
	1960	Minnesota	11	John Kennedy	Richard Nixon
	1960	Missouri	13	John Kennedy	Richard Nixon
	1960	New Jersey	16	John Kennedy	Richard Nixon
	1960	New Mexico	4	John Kennedy	Richard Nixon
	1960	Pennsylvania	32	John Kennedy	Richard Nixon
	1968	Maryland	10	Hubert Humphrey	Richard Nixon
	1968	Texas	25	Hubert Humphrey	Richard Nixon
	1976	Mississippi	7	Jimmy Carter	Gerald Ford
		Ohio	25	Jimmy Carter	Gerald Ford
		Wisconsin	11	Jimmy Carter	Gerald Ford
	1984	Minnesota	10	Walter Mondale	Ronald Reagan
	1992	Georgia	13	Bill Clinton	G. H. W. Bush
	1996	Kentucky	8	Bill Clinton	Bob Dole
	2000	Wisconsin	11	Al Gore	George Bush

References

Angrist, Joshua D., Guido W. Imbens, and Alan Krueger. 1999. "Jackknife Instrumental Variables Estimation." *Journal of Applied Econometrics* 14 (January/February): 57-67.

Arellano, Manuel. 2003. *Panel Data Econometrics*. New York: Oxford University Press.

Baltagi, Badi, H. 1999. *Econometrics*. 2nd ed. New York: Springer-Verlag.

Baltagi, Badi H. 2005. *Econometric Analysis of Panel Data*. 3rd ed. New York: John Wiley & Sons.

Beck, Nathaniel, and Jonathan N. Katz. 1995. "What to Do (and Not to Do) with Time-Series Cross-Section Data." *American Political Science Review* 89(September): 634-647.

Burden, Barry C., and David Kimball. 1998. "A New Approach to the Study of Split-Ticket Voting" *American Political Science Review* 92(September): 533-544.

Hosking, J.R.M. 1980. "The Multivariate Portmanteau Statistic." *Journal of the American Statistical Association* 75(September): 602-608.

Hsiao, Cheng. 2003. *Analysis of Panel Data*. New York: Cambridge University Press.

King, Gary. 1997. *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data.* Princeton, NJ: Princeton University Press.

Murphy, Kevin M., and Robert H. Topel. 1985. "Estimation and Inference in Two-Step Econometric Models." *Journal of Business and Economic Statistics* 3(October): 370-379.

Nickell, Stephen. 1981. "Biases in Dynamic Models with Fixed Effects." *Econometrica* 49(November): 1417-1426.

Plümper, Thomas, and Vera E. Troeger. 2004. "Efficient Estimation of Rarely Changing Variables in Fixed Effects Models." University of Konstanz. Typescript.

Stimson, James A. 1985. "Regression in Space and Time: A Statistical Essay." *American Journal of Political Science* 29(November): 914-947.