

On Data ‘Janitor Work’ in Political Science: The Case of Thermostatic Support for Democracy*

Yue Hu Yuehong ‘Cassandra’ Tai Frederick Solt

Abstract

Data ‘janitor work’, the task of getting data into a format appropriate for analysis, has grown increasingly important as political science research has come to depend on data drawn from hundreds and thousands of sources. One tempting solution is to simply enter the data by hand, but this approach raises serious risks of data-entry error, a difficult-to-catch problem with the potential to fatally undermine our conclusions. Underscoring these points, we identify data-entry errors in a prominent recent article, the 2020 study by Claassen that examines the effect of changes in democracy on public support for democracy. We then show that when these errors are corrected, the work’s models provide no support for its conclusion that publics react thermostatically to changes in democracy. Researchers should refrain from hand-entering data as much as possible, and we offer additional suggestions for avoiding errors.

A growing quantity of political science research can be characterized as data science: it employs large quantities of data, often drawn from a large number of different sources. For such projects, data wrangling, the task of getting these data into the format required to perform analyses, is notoriously the bulk of the work (see, e.g., Lohr 2014). Such data ‘janitor work’ is often viewed as tiresome, as something to be delegated to research assistants, to someone—indeed anyone—else (see Torres 2017). Data wrangling is, however, critically important to scientific inquiry, and errors in the process can undermine our conclusions.

We focus here on one particularly insidious problem that can affect the ‘janitor work’ of any researcher: data-entry errors. Faced with the task of getting data into the correct format, even some very sophisticated researchers will conclude that the most straightforward means to that end is to simply copy the needed data into a spreadsheet manually. This technique may be straightforward, but it is very much prone to error. Barchard and Pace (2011) found that ‘research assistants’ assigned in an experiment to carefully enter data manually, even

*Corresponding author: yuehong-tai@uiowa.edu. Current version: June 10, 2022.

those instructed to double-check their entries against the original, had error rates approaching 1% in just a single roughly half-hour session. Rates likely go up as the tedious task goes on.

Like ‘janitor work’ itself, data-entry errors have thus far gained little attention in political science. In this piece, we illustrate the pernicious threat this problem poses by carefully scrutinizing a prominent recent work that examines how changes in democracy affect democratic support among the public (Claassen 2020a). We document the data-entry errors that slipped past both the author and the journal’s strict replication policy and how these errors affect the paper’s results and conclusions. Claassen (2020a, 51) concludes that when “elected leaders start dismantling democratic institutions and rights, public mood is likely to swing rapidly toward democracy again, providing something of an obstacle to democratic backsliding.” We show that, after the data-entry errors are corrected, there is no empirical evidence that public support responds thermostatically to changes in democracy in this way. On this basis, we conclude with four practical suggestions to help political scientists to reduce data-entry errors and their impact.

Data-Entry Errors and Democratic Support

With democracy under increasing threat in countries around the world, how the public reacts is a crucial question. According to a now-classic literature, it is experience with democratic governance that generates robust public support for democracy (see, e.g., Lipset 1959). Claassen (2020a) argues instead that democratic support behaves thermostatically: that increases in democracy yield an authoritarian backlash in the public, while democratic backsliding prompts the public to rally to democracy’s cause.

The evidence it offers in support of this latter argument takes advantage of recent advances in modeling public opinion as a latent variable to measure democratic support. This approach provided estimates of the paper’s dependent variable for over one hundred countries for up to nearly three decades, constituting a much larger evidentiary base than any previous study. These latent-variable estimates, in turn, were based on thousands of nationally aggregated responses to dozens of different questions from cross-national survey projects (Claassen 2020a, 40). Two pieces of data were collected for each distinct survey item in each country and year it was asked: the number of respondents to give a democracy-

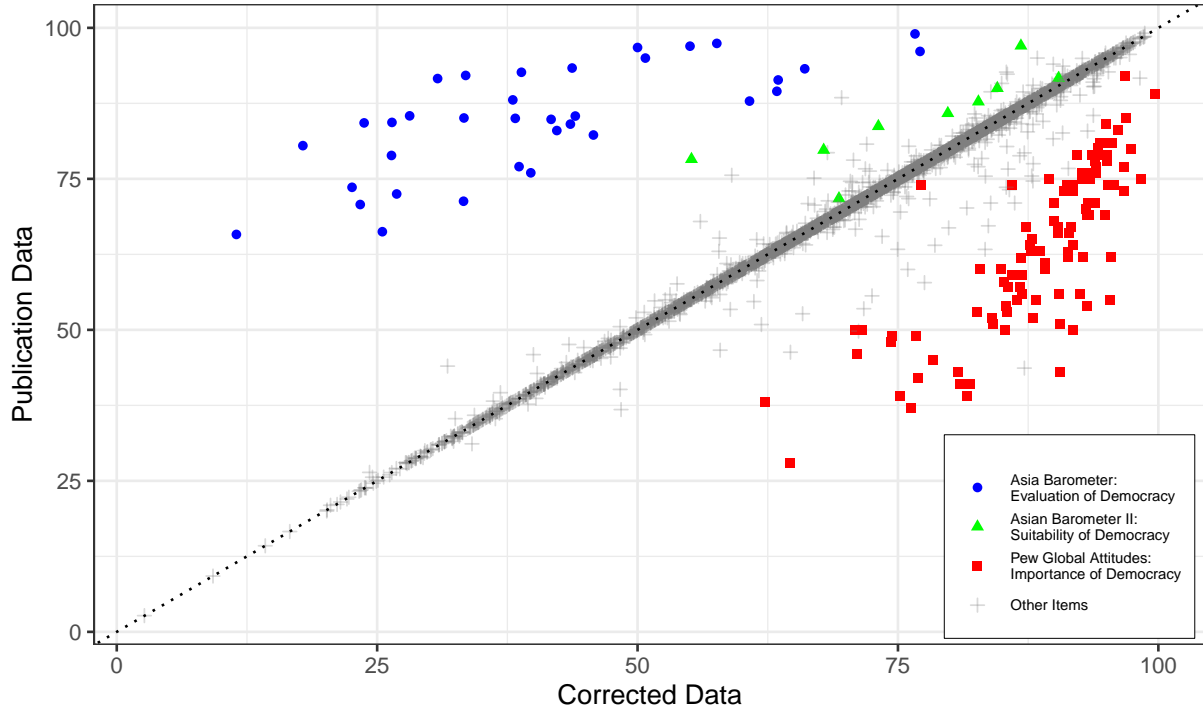
supporting response—defined, for ordinal responses, as those above the median value of the scale (Claassen 2020a, Appendix 1.3)—and the total number of respondents to whom the question was posed. Each of these 7,538 pieces of source data is recorded in a spreadsheet.¹

We re-collected all of the source data for the publication from the original surveys. We identified the variables of the survey items used by the article within each survey dataset, and then we used an automated process to collect the needed data from the survey datasets while avoiding data-entry errors (see Solt, Hu, and Tai 2018). In Figure 1, we compare the percentage of respondents to give a democracy-supporting response in the publication spreadsheet with the percentage we found using our automated process of wrangling these same data. When points fall along the plot’s dotted line, it indicates that the publication’s source data and our own automated workflow reported the same percentages. Points above this diagonal represent observations for which the publication data overestimated the actual percentage of respondents who offered a democracy-supporting response, while points below this line are observations where the publication data underestimated this percentage.

For 85% of the country-year-item observations, the difference between these percentages was negligible—less than half a percent—yielding points approximately along the plot’s dotted line. But for the remaining observations, the difference was often substantial as a result of data-entry errors in the publication data. For example, the Asia Barometer asked respondents in 35 country-years to indicate whether they thought “a democratic political system” would be very good, fairly good, or bad for their country. According the study’s coding rules (see Claassen 2020a, Appendix 1.3), only answers above the median of the response categories should be considered as democracy supporting, yet in this case the lukewarm intermediate category was coded as supporting democracy as well.² This led to overestimations of the percentage of democracy-supporting responses ranging from 19 to 63 percentage points and averaging 44 points.

¹The article’s replication materials include only the latent variable estimates without the original survey aggregates that serve as their source data (see Claassen 2020c). Fortunately, however, the spreadsheet recording these original source data is included in the replication materials for a companion piece that employed the identical estimates (see Claassen 2020b).

²Although this may be interpreted as an exercise of researcher judgment as to what constitutes a democracy-supporting response rather than a data-entry error, examination of similar answers to similar questions shows that similarly lukewarm responses at and below the median response category (e.g., in the Arab Barometer, that democracy was “somewhat appropriate” for the country) were coded as not supportive.



Notes: Each point represents the percentage of respondents in a country–year to give a democracy–supporting response to a particular survey item. Publication data is as reported in Claassen (2020b); the corrected data was collected directly from the original surveys. The Asia Barometer’s item on the evaluation of democracy accounts for most overreports, and the Pew Global Attitudes item on the importance of democracy accounts for most substantial underreports. In both cases, as well as the overreports of the suitability of democracy item in the second wave of the Asian Barometer, the issues can be easily explained by errors in transcribing the data. Deviations in other items result from inconsistent treatment of missing data and/or survey weights, reflecting in part differences in codebook reporting practices across surveys.

Figure 1: Comparing Democracy-Supporting Responses in the Publication Data and the Corrected Data

Similarly, the four waves of the Asian Barometer included the following item: “Here is a similar scale of 1 to 10 measuring the extent to which people think democracy is suitable for our country. If 1 means that democracy is completely unsuitable for [name of country] today and 10 means that it is completely suitable, where would you place our country today?” In accordance with the coding rules of the study, responses of 6 through 10 are considered democracy supporting, and that is how the first, third, and fourth waves of the survey are coded. For the second wave, however, 5 was erroneously also included among the democracy supporting-response. This data-entry error resulted in overestimates of as much as 23 percentage points in 9 country-years.

A third example comes from the Pew Global Attitudes surveys’ four-point item asking about the importance of living in a country with regular and fair contested elections: the

question wording is “How important is it to you to live in a country where honest elections are held regularly with a choice of at least two political parties? Is it very important, somewhat important, not too important or not important at all?” In this case, rather than including respondents who gave both responses above the median—“very important” and “somewhat important”—only those respondents who answered “very important” were entered as supporting democracy. This error caused substantial underreporting of the extent of democratic support in 91 country-years.

While these issues involve mistakes in recording the numerator of the percentage, the number of respondents who provided a democracy-supporting answer, entering the denominator, the total number of respondents asked a question, was also problematic on occasion. For example, when the Americas Barometer surveyed Canada in 2010, it included an item asking whether, when “democracy doesn’t work,” Canadians “need a strong leader who doesn’t have to be elected through voting.” It posed this question to only half of its sample. Those who were not asked the question, however, were included in the total number of respondents as if they had refused to answer. According to the study’s coding rules, refusing to answer is equivalent to answering in a fashion not supporting democracy, that is, in this case, agreeing that Canada needed a strong leader who need not bother with elections (see Claassen 2020a, Appendix 1.3). This rule may be a reasonable coding choice, but including in this category those who were never asked the question at all is clearly a data-entry error.

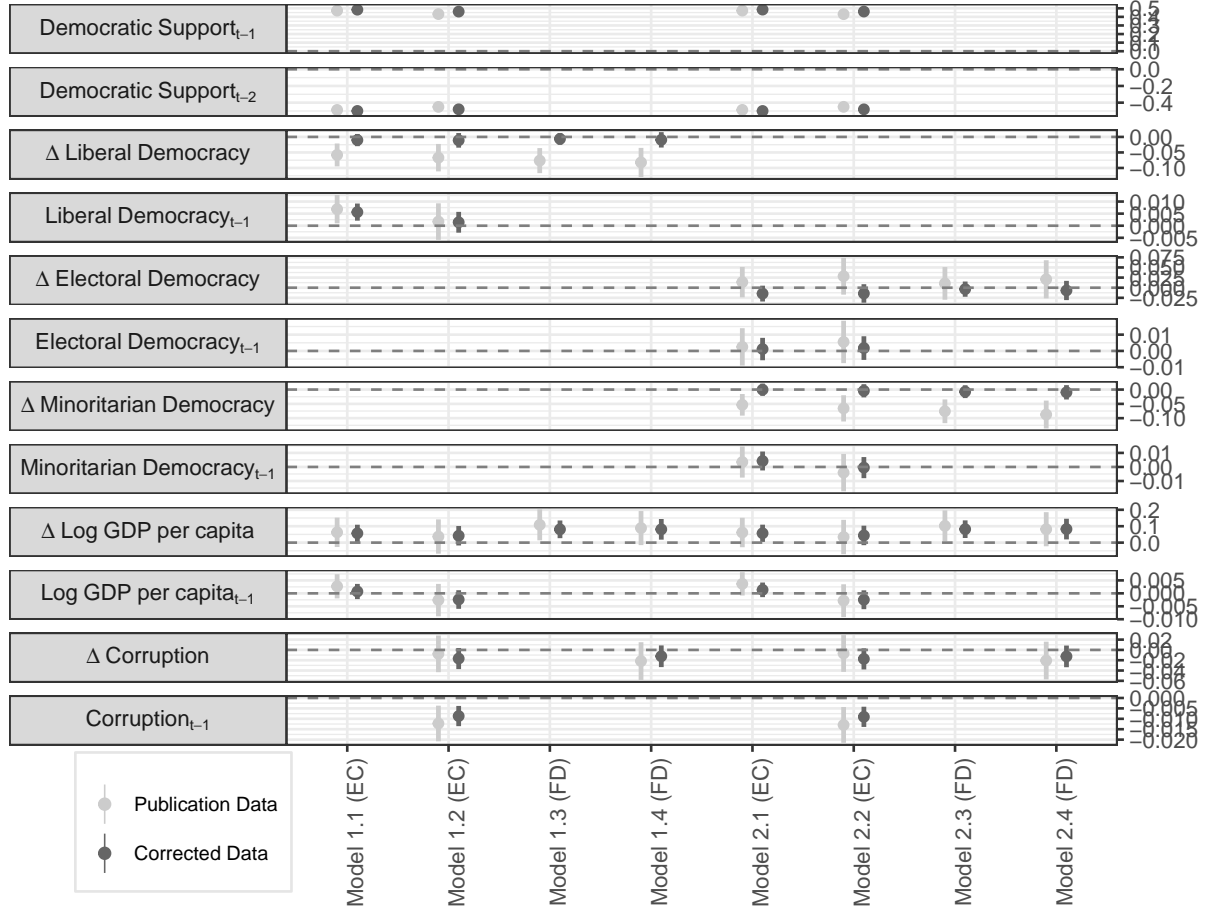
Another source of data-entry errors here involves survey weights. Weighting raw survey results to maximize the extent to which they are representative of the target population is important. Relying on topline reported in codebooks rather than the survey data itself evidently caused some mistakes in correctly entering the needed information here, as codebooks do not always take survey weights into account. These errors shifted the percentage of democracy-supporting responses in both directions, typically by relatively small amounts.

Consequences for Inference

Data-entry errors of this sort can yield erroneous conclusions. After generating the latent variable of democratic support with the corrections to the errors we describe above, we replicated each of the models presented in Claassen (2020a) exactly using both the original

and the new version of the latent variable. The results using the corrected data reveal only limited support for the classic argument that democracy generates its own demand, at least in the short run, and none at all for a thermostatic relationship.

Dependent Variable: Change in Public Democratic Support



Notes: Replications of Claassen (2020), Table 1, 47, and Table 2, 49. Models denoted 'EC' are error-correction models; those marked 'FD' are first-difference models.

Figure 2: The Effect of Democracy on Change in Public Support

Figure 2 presents our results in a “small multiple” plot (Solt and Hu 2015) for a clear comparison of the coefficients of each variable in the article’s models. In the plot, the dots represent point estimates and the whiskers show the associated 95% confidence intervals. Models 1.1 through 1.4, which replicate those presented in Table 1 of Claassen (2020a, 47), examine the effects of overall liberal democracy using error-correction models and first-difference models. As Claassen (2020a, 46) notes, the thermostatic theory predicts that the

estimated coefficient of the change in liberal democracy will be negative, while the classic theory suggests that lagged levels of liberal democracy will be positive. When using the original publication data with their data-entry errors, we replicate the results of the article exactly: the coefficients estimated for the change in liberal democracy are large, negative, and statistically significant across all four models, just as the thermostatic theory predicts. The positive and statistically significant result for the lagged level of liberal democracy found in Model 1.1—supporting the classic theory—disappears when corruption is taken into account in Model 1.2, suggesting that “this effect is not particularly robust” (Claassen 2020a, 47).

When the data-entry errors are corrected, however, the results for these models suggest a very different set of conclusions. The standard errors shrink across the board—indicating that the models are better estimated in the corrected data—but so do the magnitudes of the coefficients. The positive and statistically significant result for the lagged level of liberal democracy remains in Model 1.1. The estimate is only slightly smaller than in the publication data, and as with the publication data, it disappears when corruption is added in Model 1.2: the evidence, such as it is, for the classic theory, operationalized as a short-run process, remains substantively unchanged. On the other hand, the estimates for the change in liberal democracy that provided support for the thermostatic theory are much smaller—very nearly exactly zero—and fail to reach statistical significance in any of these four models. Models 2.1 through 2.4, which break liberal democracy into its electoral democracy and minoritarian democracy components, similarly undermine claims for the thermostatic theory. The strong and statistically significant negative coefficients for the change in minoritarian democracy on public democratic support that are found using the publication dataset evaporate when the data-entry errors are corrected. There is no support for the thermostatic theory.

This is not, we contend, a particularly surprising finding. As much as those who favor democracy might wish it were so, and as well as the thermostatic theory performs with regard to many other topics in public opinion, it is not a particularly likely candidate for explaining trends in democratic support—the mechanism required for it to operate is not present. In its original formulation, the theory requires citizens to possess a level of knowledge of politics that a long line of public opinion research shows is unrealistic, and as recently re-elaborated it requires the issue in question to be debated by political parties so as to provide cues to the

broader public as to what is going on (Atkinson et al. 2021, 5–6). But few parties actually engaged in eroding democracy put their actions in such terms: instead they claim to be defending democracy, or saving democracy, or putting forth a different model of democracy that better suits the nation’s needs. And to the extent they succeed, their opponents are increasingly unable to make their case to the public at all. Absent its mechanism, the thermostat cannot operate on the public’s democratic support.

Discussion

The analysis above reveals that data-entry errors are an especially pernicious threat to the credibility of our results. The threat is a subtle one that is not easily detected. To discern it requires close scrutiny of every manual entry; merely examining the data and their distribution will uncover few errors (Barchard and Pace 2011, 1837–38). Although failure to find support for a research hypothesis may prompt us to undertake a such a close review, an analysis that yields statistical significance is unlikely to trigger what will likely be, as in the above example, a time-consuming and difficult effort (see Gelman and Loken 2014, 464).

This leads us to recommend researchers take the following steps to reduce the possibility of data-entry errors. First, **automate data entry**: researchers should minimize reliance on manual data entry and maximize the extent to which data wrangling is performed computationally. Automating ‘janitor work’ will sometimes require considerable programming effort, but often software is available that makes the task straightforward, such as the `readtext` R package (Benoit, Obeng, et al. 2016) for formatting the contents of text files for text analysis or the `DCP0tools` R package (Solt, Hu, and Tai 2018) that we employed above. In addition to minimizing data-entry errors, writing computer code that starts from the raw source material and works forward has the added benefit of making research much more reproducible (see, e.g., Benoit, Conway, et al. 2016) and hence more credible (see, e.g., Wuttke 2019).

Second, **use the double-entry method**: when manual data entry *cannot* be avoided, each entry should be made twice. Double entry is labor intensive, but experiments have shown that it reduces error rates by thirty-fold even when done immediately after the initial collection and by the same person (Barchard and Pace 2011, 1837). Given that data-entry errors can completely undermine the validity of our conclusions, as in the example above,

double entry is worth the extra effort.

Third, **embrace teamwork**: for any project involving entering data by hand, splitting the task up among team members will reduce the risk of errors going undetected. When double entries are performed by different people, discrepancies will be noted, discussed, and resolved correctly; having two sets of eyes on complex materials like survey codebooks also increases the chances that nuances of the presentation like survey weights will be uncovered. Further, by dividing the load, teamwork also lessens the probability of errors due to fatigue arising in the first place.

Fourth, **be aware of the threat of data-entry error**: this final recommendation is especially for manuscript reviewers. If data-entry errors are invisible to the authors themselves, they are doubly so to reviewers (though if editors provided reviewers with replication materials at the time of the review it may help them to better assess the work’s credibility). But the case described above nevertheless suggests a valuable heuristic: when a work’s conclusions suggest that a difficult problem will be easily solved—that democratic erosion will reflexively trigger a backlash and a renewed public support for democracy, in the present instance—it warrants especially careful scrutiny.

Data-entry errors are inevitable, and even following these recommendations is unlikely to eliminate them entirely. With careful attention, however, the threat of data-entry errors to our ‘janitor work’, our research, and our understanding of the world can be minimized.

References

- Atkinson, Mary Layton, K. Elizabeth Coggins, James A. Stimson, and Frank R. Baumgartner. 2021. *The Dynamics of Public Opinion*. Cambridge: Cambridge University Press.
- Barchard, Kimberly A., and Larry A. Pace. 2011. “Preventing Human Error: The Impact of Data Entry Methods on Data Accuracy and Statistical Results.” *Computers in Human Behavior* 27 (5): 1834–39.
- Benoit, Kenneth, Drew Conway, Benjamin E. Lauderdale, Michael Laver, and Slava Mikhaylov. 2016. “Crowd-Sourced Text Analysis: Reproducible and Agile Production of Political Data.” *American Political Science Review* 110 (2): 278–95.
- Benoit, Kenneth, Adam Obeng, Paul Nulty, Aki Matsuo, Kohei Watanabe, and Stefan Müller.

2016. “readtext: Import and Handling for Plain and Formatted Text Files.” Available at the Comprehensive R Archive Network (CRAN).
- Claassen, Christopher. 2020a. “In the Mood for Democracy? Democratic Support as Thermostatic Opinion.” *American Political Science Review* 114 (1): 36–53.
- . 2020b. “Replication Data for: Does Public Support Help Democracy Survive?” <https://doi.org/10.7910/DVN/HWLW0J>, American Journal of Political Science Data-verse.
- . 2020c. “Replication Data for: In the Mood for Democracy? Democratic Support as Thermostatic Opinion.” <https://doi.org/10.7910/DVN/FECIO3>, American Political Science Review Dataverse.
- Gelman, Andrew, and Eric Loken. 2014. “The Statistical Crisis in Science.” *American Scientist* 102 (6): 460–65.
- Lipset, Seymour Martin. 1959. “Some Social Requisites of Democracy: Economic Development and Political Legitimacy.” *American Political Science Review* 53: 69–105.
- Lohr, Steve. 2014. “For Data Scientists, ‘Janitor Work’ Is Hurdle to Insights.” *New York Times*, B4.
- Solt, Frederick, and Yue Hu. 2015. “dotwhisker: Dot-and-Whisker Plots of Regression Results.” Available at the Comprehensive R Archive Network (CRAN). <http://CRAN.R-project.org/package=dotwhisker>.
- Solt, Frederick, Yue Hu, and Yuehong ‘Cassandra’ Tai. 2018. “DCPOtools: Tools for Dynamic Comparative Public Opinion.” <https://github.com/fsolt/DCPOtools>.
- Torres, Rachel. 2017. “Me: Shouldn’t There Be Someone in a Basement That We Just Pay to Do All This Awful Data Cleaning? Advisor: That’s Who You Are.” Twitter. <https://twitter.com/torrespolisci/status/886993701855268865>.
- Wuttke, Alexander. 2019. “Why Too Many Political Science Findings Cannot Be Trusted and What We Can Do about It: A Review of Meta-Scientific Research and a Call for Academic Reform.” *Politische Vierteljahresschrift* 60 (1): 1–19.

Online Supplementary Materials

List of Tables

A.1	The Effect of Democracy on Change in Public Support (Publication Data)	. .	A2
A.2	The Effect of Democracy on Change in Public Support (Corrected Data)	. .	A3

Table A.1: The Effect of Democracy on Change in Public Support (Publication Data)

	Model 1.1 (EC)	Model 1.2 (EC)	Model 2.1 (EC)	Model 2.2 (EC)	Model 1.3 (FD)	Model 1.4 (FD)	Model 2.3 (FD)	Model 2.4 (FD)
Democratic Mood (t-1)	0.473*** (0.026)	0.433*** (0.028)	0.473*** (0.025)	0.432*** (0.028)				
Democratic Mood (t-2)	-0.487*** (0.025)	-0.451*** (0.027)	-0.487*** (0.025)	-0.450*** (0.027)				
Liberal Democracy (Difference)	-0.058** (0.023)	-0.067** (0.031)			-0.076*** (0.028)	-0.082*** (0.034)		
Liberal Democracy (t-1)	0.007* (0.003)	0.002 (0.004)						
Electoral Democracy (Difference)			0.014 (0.031)	0.028 (0.039)			0.011 (0.033)	0.021 (0.040)
Electoral Democracy (t-1)			0.002 (0.006)	0.006 (0.006)				
Minoritarian Democracy (Difference)			-0.053** (0.022)	-0.066** (0.029)			-0.076*** (0.025)	-0.087*** (0.029)
Minoritarian Democracy (t-1)			0.003 (0.006)	-0.004 (0.006)				
Log GDP Per Capita (Difference)	0.063 (0.040)	0.037 (0.044)	0.062 (0.040)	0.034 (0.045)				
Log GDP per capita (Difference)					0.108* (0.052)	0.089+ (0.051)	0.102* (0.053)	0.082 (0.051)
Log GDP (t-1)	0.003 (0.002)	-0.003 (0.003)	0.004 (0.002)	-0.003 (0.003)				
Corruption (Difference)		-0.008 (0.016)		-0.007 (0.016)		-0.022 (0.017)		-0.021 (0.017)
Corruption (t-1)		-0.012** (0.004)		-0.013** (0.004)				
N observations	2300	1949	2300	1949	2435	2040	2435	2040
N countries	135	135	135	135	135	135	135	135

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table A.2: The Effect of Democracy on Change in Public Support (Corrected Data)

	Model 1.1 (EC)	Model 1.2 (EC)	Model 2.1 (EC)	Model 2.2 (EC)	Model 1.3 (FD)	Model 1.4 (FD)	Model 2.3 (FD)	Model 2.4 (FD)
Democratic Mood (t-1)	0.484*** (0.023)	0.462*** (0.025)	0.485*** (0.023)	0.462*** (0.025)				
Democratic Mood (t-2)	-0.500*** (0.023)	-0.480*** (0.025)	-0.500*** (0.023)	-0.480*** (0.025)				
Liberal Democracy (Difference)	-0.010 (0.010)	-0.011 (0.011)			-0.006 (0.009)	-0.009 (0.012)		
Liberal Democracy (t-1)	0.006** (0.002)	0.001 (0.002)						
Electoral Democracy (Difference)			-0.015 (0.013)	-0.014 (0.015)			-0.004 (0.009)	-0.007 (0.011)
Electoral Democracy (t-1)			0.001 (0.004)	0.002 (0.004)				
Minoritarian Democracy (Difference)			0.000 (0.012)	-0.004 (0.014)			-0.007 (0.012)	-0.010 (0.012)
Minoritarian Democracy (t-1)			0.004 (0.003)	0.000 (0.004)				
Log GDP Per Capita (Difference)	0.057* (0.023)	0.042 (0.021)	0.057* (0.023)	0.044 (0.021)				
Log GDP per capita (Difference)					0.081** (0.029)	0.081* (0.032)	0.082** (0.029)	0.083* (0.032)
Log GDP (t-1)	0.001 (0.001)	-0.002 (0.002)	0.001 (0.001)	-0.002 (0.002)				
Corruption (Difference)		-0.017 (0.010)		-0.017+ (0.010)		-0.012 (0.009)		-0.013 (0.009)
Corruption (t-1)		-0.009*** (0.003)		-0.009*** (0.003)				
N observations	2228	2012	2228	2012	2375	2101	2375	2101
N countries	135	134	135	134	135	135	135	135

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001