Retrospective Voting and Natural Disasters that Cause No Damage: Accounting for the Selective Reporting of Weather Damage

Justin Gallagher*

August 11, 2021

Abstract

The decision on how to handle missing data is critical for the reliability of a model's results. This paper makes three contributions. First, we show that a popular weather damage database suffers from a nonrandom missing data problem. Second, we follow the applied statistics literature and demonstrate an imputation procedure that relies on an instrument to estimate missing values under a nonrandom selection mechanism. Third, we apply the instrument-based imputation procedure to account for missing damage and reevaluate a seminal study on retrospective voting following a natural disaster. The original findings are mostly reversed once accounting for the missing data.

Keywords: Retrospective Voting; Natural Disasters; Missing data; Imputation

JEL Classification: H84; Q54; C18; P16

*Contact information: Montana State University, P.O. Box 172920, Bozeman, MT 59717 (email: justin.gallagher1@montana.edu). The author is grateful for helpful comments from Mark Anderson, David Clingingsmith, Jacques-Emmanuel Galimard, Nick Hagerty, Dan Kaffine, Jason Lindo, Roderick J. A. Little, Neil Malhotra, Fernando Rios-Avila, Brock Smith, Carly Urban, and seminar participants at the Association of Environmental and Resource Economists annual meeting. Isaac Birdwell, Katiana Carestia, Simin Gao, Andrew Hoover, Tristan Jones, and especially Greta Linse and Kyle Musser provided outstanding research assistance. A special thanks to Christian Cassell, Lead Meteorologist NWS Great Falls, for his assistance in locating historical NWS documents. The author thanks the Initiative for Regulation and Applied Economic Analysis for research support.

1 Introduction

Economists frequently estimate models using large databases built by combining information from a number of separate sources. Utilizing a variety of data sources can increase the likelihood that a key independent variable in the econometric model suffers from missing data.[1] The decision on how to handle observations with missing data is critical for the reliability of a model's results.

Missing data can broadly be categorized into one of three cases (e.g. Little and Rubin [2020]). If the data are Missing Completely at Random then a researcher can proceed with a complete case analysis of observations that have no missing data. The complete case sample is a random subsample and can provide consistent estimates for parameters in the full sample.

If the data are Missing at Random then the missing observations are random conditional on the other covariates. A model estimated on the complete case sample can provide consistent parameter estimates for the population of the selected subsample. However, these model estimates will generally be inconsistent for the population in the full sample (e.g. Rubin [1976]). The typical Missing at Random approach for a researcher interested in testing hypotheses about the entire sample population is to impute the missing information, run a battery of diagnostic checks, and then to estimate the model of interest on the full sample (e.g. Abayomi et al. [2008]). A critical step is for a researcher to specify the selection equation used to impute the missing values.[2] It is not possible to know for certain whether the selection equation is specified correctly, unless the researcher has access to external information regarding the true selection mechanism (e.g. Raghunathan [2016], p7). Economists are often skeptical of imputation as an approach to correct for missing data due to concern over misspecification of the selection equation (e.g. Hirsch and Schumacher [2004]; Bollinger and Hirsch [2006]; DiNardo et al. [2006]).

Missing Not at Random is the third missing data case. The observations with complete data are a selected subsample and the selection mechanism is unknown. Parametric and semi-parametric models of the selection process are common approaches to address missing data for a *dependent variable* when a researcher is not comfortable assuming a Missing at Random mechanism (e.g. Heckman [1976]; Vella [1998] provides a review). The popularity of these models follows from the fact that they can provide consistent estimates for the entire sample, while only having complete data on a selected subsample.

[1] For example, Abrevaya and Donald [2017] survey published papers from four top economics journals from 2006-8. The authors find that almost 40% of the papers estimate models with data missingness.

[2] A pattern mixing model is an equivalent way to characterize the missing data (Carpenter and Kenwood [2013], p17).

This paper makes three main contributions. First, we show that the Spatial Hazard Events and Losses Database for the United States (SHELDUS) suffers from a severe and poorly understood missing data problem. SHELDUS is a popular data source for researchers examining questions related to natural disasters. Studies using SHELDUS are published in top general interest economics (e.g. Barrot and Sauvanat [2016]), finance (e.g. Bernile et al. [2017]), and political science (e.g. Gasper and Reeves [2011]) journals.[3] The widespread use of SHELDUS is likely due to the paucity of alternative weather damage databases. We are not aware of another public database that combines the spatial detail (county), observation frequency (monthly), and long time horizon (60 years). We show that the missing SHELDUS observations are not Missing Completely at Random. Failure to account for the missing weather damage data will likely result in biased model estimates.

Second, we follow the recent applied statistics literature and show how a selection model, combined with an instrument, can be used as part of a data imputation procedure to estimate missing values for an *independent variable* under an unknown and potentially nonrandom selection mechanism (e.g. Galimard et al. [2016]; Ogundimu and Collins [2019]; Gomes et al. [2020]). Model estimates will be inconsistent when an independent variable has missing information and selection depends on the dependent variable (either directly or indirectly via factors omitted from the analysis). Imputation using this approach addresses the concern that perfect knowledge of a fully specified selection equation may be unrealistic. Instead, the model relies on an exclusion restriction (or instrument), whereby a variable is included in the selection equation, but excluded from the model of interest. Valid imputation, and ultimately valid inference, depends on correct specification of the exclusion restriction, rather than the entire selection equation.

The imputation procedure offers a flexible way to account for the SHELDUS missing data problem. We leverage the organizational structure of the National Weather Service field office reporting zones, and historical changes to this structure, as the basis for the instrument. Our application shows how researchers studying *any topic* can continue to use the SHELDUS database, while avoiding bias due to selective non-reporting.

To our knowledge, we are both the first economics study to apply this imputation method, and the first to apply this method to a panel data model with unit fixed effects. We follow the above literature and use a probit model to estimate the selection equation. The challenge in estimating a probit selection model with unit fixed effects is bias due to the incidental parameters problem. Fernandez-Val and Vella [2011] provide a formula to correct for this bias. We demonstrate the imputation approach via a simulation exercise.

[3]These studies are published in the *Quarterly Journal of Economics*, *The Journal of Finance*, and the *American Journal of Political Science*, respectively.

Our third contribution is our reanalysis of Gasper and Reeves [2011], a seminal study on voting behavior that relies on the weather damage information from SHELDUS. Models of voting behavior often assume that the electorate is retrospective (e.g. Key [1966]; Wittman [1989]; Persson et al. [1997]). A large empirical literature in both economics (e.g. Ferrez and Finan [2008]; Ferrez and Finan [2011]) and political science (e.g. Conover et al. [1986]; Healy and Lenz [2014]) examines how voters evaluate political performance and react to different types of information when considering an incumbent politician (Anderson [2007] and Healy and Malhotra [2013] provide a reviews). Whether voters hold incumbents responsible for random events outside of their control, or only for the political response to these events, is a key topic in the literature.

The early empirical literature on retrospective voting focuses on how the electorate responds to economic conditions when voting for incumbent politicians or political parties (e.g. Fair [1978]). Gasper and Reeves [2011] (hereafter GR) are among the first to use weather damage and the political response to the damage as a quasi-experiment to examine retrospective voting. Random weather damage allows for a causal interpretation for both the exogenous damage and the subsequent political actions on the reelection vote share. The appeal of using extreme weather events as a quasi-experiment has spawned a sub-literature on retrospective voting.[4]

GR examine US gubernatorial and presidential elections from 1970-2006. The authors find evidence of a "responsive" electorate, whereby greater weather damage in the months before an election leads to larger reductions in the county vote share for incumbents. The authors also find evidence of an "attentive" electorate. Voters appear attentive to the actions of politicians, even when these actions do not change the outcome. The authors conclude that, overall, the negative vote share impact of a natural disaster "is dwarfed by the response of attentive electorates to the actions of their officials" (p1).

Critically, GR assume that all of the counties with missing SHELDUS weather damage observations incurred no damage in those months. This assumption is not valid and in sharp contrast to best practices on how to handle missing data. Further, there is no reason to expect that using inaccurate damage information will lead to conservative estimates that are biased towards zero (e.g. Loken and Gelman [2017]). We allow for the non-random selection of the reported weather damage in SHELDUS. In our reanalysis, the negative effect on vote share from weather damage is much larger than in GR. We find little evidence in favor of an attentive electorate. On balance, the negative vote response of the electorate to weather

[4] Retrospective voting studies following natural disasters include: Bechtel and Hainmueller [2011]; Cole et al. [2012]; Chen [2013]; Fair et al. [2017]; Heersink et al. [2017]; Nyhan [2017]; Heersink et al. [2020]; Rodriguez-Valadez and Marinez-Alvarez [2021].

damage either cancels out or outweighs the positive, attentive influence attributed to the post-disaster actions of a politician. Overall, the findings of GR are mostly reversed once accounting for the missing data.

McDonough and Millimet [2017] also use an instrumental variables imputation approach to address missing data for an independent variable. McDonough and Millimet [2017] focus on a setting where an endogenous independent variable has missing data. Our setting is different in that we do not assume that the variable with missing data is endogenous in the model of interest. Instead, the selection of whether information is missing for an independent variable is allowed to be endogenous.

The rest of the paper proceeds as follows. In Section 2, we describe the missing data problem in SHELDUS. We also introduce an instrument that relies on the geographic boundaries of the counties included in the National Weather Service field office reporting zones, and which strongly predicts if weather damage is reported. Section 3 reviews our imputation approach. In this section, we also describe the results of a simulation exercise. Section 4 applies the imputation approach to a reanalysis of Gasper and Reeves [2011]. Section 5 concludes.

2 SHELDUS Weather Damage

SHELDUS is a loss and hazard database currently maintained by the Center for Emergency Management and Homeland Security at Arizona State University. The database website lists more than one hundred academic publications that use SHELDUS weather damage information[5] The primary source for SHELDUS is the National Centers for Environmental Information (formerly National Climatic Data Center) monthly Storm Data publications. A critical feature of the Storm Data publications is that the weather damage information is voluntarily reported by regional National Weather Service (NWS) offices. One consequence of the self-reporting is that the monthly Storm Data publications, and by extension SHELDUS, are susceptible to unreported (i.e. "missing") data. In fact, the Storm Data publications include the following disclaimer: "due to difficulties inherent in the collection of this type of data, it is not all-inclusive" (Storm Data 1995, p2). Gallagher [2014] was among the first to point out the severity of the missing data problem.

Table 1 shows SHELDUS statistics for a sample of counties and months from 1972-2004. The sample matches the panel months used in Gasper and Reeves [2011], and includes observations from May-October for all US counties for the nine presidential election years during this period. The table illustrates the magnitude of the missing data problem. Three-

[5]The SHELDUS website provides a list.

quarters of the SHELDUS database in our sample contain missing observations. The only thing we know for sure is that no regional NWS office reported damage information to the National Climatic Data Center during these months for these counties. The second panel in Table 1 considers "yearly" observations for each county, where each year is comprised of the six sample months. GR sum the six months each year in their county by year panel. Fewer than 1% of the counties have non-missing weather damage reported in SHELDUS for each month during the year. Approximately, one-third of the counties contain only missing weather damage during the year.

The table also provides initial evidence for why it is incorrect to assume that counties with missing weather damage incurred zero dollars in damage. Zero dollars in damage is reported in SHELDUS. Twelve percent of all non-missing observations in the sample report zero dollars in damage. Moreover, 20% of the counties that have a mix of missing and non-missing observations for the year report at least one month with zero dollars in damage. Overall, more than two-thirds of the counties in the sample report at least one month with zero dollars in damage.

Table 2 uses administrative data from the Federal Emergency Management Agency (FEMA) to confirm that missing weather damage information in SHELDUS should not be interpreted as zero dollars of damage. The FEMA damage information includes grants to cities to repair public infrastructure caused by a natural disaster and grants to individuals with verified losses due to a disaster. The disaster grants are authorized by a Presidential Disaster Declaration. The Presidential Disaster Declaration system is a formalized process to request and receive federal assistance following large natural disasters. Disaster declarations occur at the county-level. A governor of a US state that experiences a natural disaster requests a Presidential Disaster Declaration in a written letter to FEMA. The letter must contain a list of proposed counties and preliminary damage estimates. FEMA then makes an official recommendation to the US president, who decides whether or not to grant the request. There is no damage threshold for a Presidential Disaster Declaration. However, the aim is to assist with "acts of God" that are of "such severity and magnitude that effective response is beyond the capacities of the state and the affected local governments" (Daniels and Trebilcock [2006]).

We compare the county-specific FEMA disaster information to SHELDUS for counties with a Presidential Disaster Declaration. We received the FEMA damage information via a Freedom of Information Act Request and though the FEMA website. The appendix provides further details. FEMA reports disaster damage and SHELDUS reports no information for 40% of the county-months in the FEMA-SHELDUS overlap sample. If we allow for the possibility that the disaster damage was reported in either the month of the declaration or the previous month (possible due to a lengthy delay in declaring a disaster), then there is no information in SHELDUS for 25% of the county-months in the overlap sample. The FEMA administrative data cover a subset of the Presidential Disaster Declaration sample (1990-2004) and is heavily skewed towards more recent disasters. The missing SHELDUS weather damage information during this time period is all the more striking as we might expect recent SHELDUS reporting to be more reliable due to modernized computer systems and improved NWS communication within the organization.[6]

2.1 SHELDUS Damage is Not Missing Completely at Random

Missing damage information in SHELDUS does not imply no damage. Still, the purpose of Storm Data is to promulgate information related to weather events. The publication typically includes several sections and follows a similar format each month. For example, the first section in the May 1992 issue provides an overview of the national weather during the month (Storm Data 1992). The second section highlights salient storm events. The third section, "Storm Data and Unusual Weather Phenomena," lists county-level damage, the number of persons killed, and the number of persons injured, as reported by the regional NWS offices to the National Climatic Data Center.[7] The NWS offices may be more likely to report damage information to the National Climatic Data Center when there is a larger storm that causes damage to one or more counties in their reporting region. Table 1 suggests that this is indeed the case, as 88% of non-missing damage observations report positive damage.

Table 3 shows estimation results from a linear probability model that investigates the likelihood that a county is missing damage information in SHELDUS for a particular month. We consider the same 1972-2004 sample of counties as in the Table 1 and estimate three model specifications. Each specification includes four weather event variables and four demographic variables. FEMA is the source of the weather information and the demographic data are from the US Census.[8] The first column estimates a model that includes year fixed effects to control for common yearly factors that may impact reporting, and month fixed effects to control for seasonality. The second column adds county fixed effects to control for county characteristics that are constant over the panel (e.g. whether a county typically receives a lot of rain, boarders the ocean, or is located in a high tornado risk region). The third column

[6] The latter portion of our sample is after NWS "modernization" which NRC [2012] summarize as: "[Leading to] greater integration of science into weather service activities and improved outreach and coordination with uses of weather information [...] The modernized NWS was achieved through the development and deployment of new observational and computational systems and redefining the NWS field office structure to best utilize the investment in the new technologies" (p.vii).

[7] Electronic copies of Storm Data can be accessed via the National Centers for Environmental Information.

[8] The appendix contains a detailed discussion of the data sources and panel construction.

adds NWS forecast zone fixed effects. We argue below that these NWS forecast region indicators satisfy a selection model exclusion restriction and can be interpreted as exogenous predictors for whether damage information is missing for a county during a particular month. Standard errors in the model are clustered at the state by year level to allow for spatial correlation in disaster declarations and damage.

Weather damage in SHELDUS is not missing completely at random. First, damage information is less likely to be missing when there is a natural disaster. A county damage observation is approximately 20 percentage points or 27% less likely to be missing during the month of a Presidential Disaster Declaration (probability value less than 0.01 in all specifications), and 10 percentage point or 13% less likely to be missing the month before a Disaster Declaration (probability value less than 0.05 in all specifications). The latter correlation is explained by the fact that there is sometimes a delay in declaring a disaster. Second, damage is less likely to be missing in counties with larger populations and in counties where residents have higher incomes. Third, the estimated coefficients on the year fixed effects (not shown) imply that the weather damage is more likely to be missing in earlier panel years. Fourth, weather damage is less likely to be missing in the summer months (May-August) then during the fall months (September-October). Finally, as described in the next section, the NWS forecast zone to which a county belongs strongly predicts the likelihood that weather damage is missing.

2.2 NWS Forecast Zones Predict Missing SHELDUS Damage

The National Weather Service is responsible for forecasting weather in the US. The NWS relies on a decentralized organizational structure that includes regional NWS offices and forecast zones. There were 52 Weather Service Forecast Offices during the first period (1972 - July 1984) of the sample (e.g. NWS [1978]). Each office was responsible for forecasting weather and reporting on weather conditions in a specific "area of responsibility." These areas were often a single state, but sometimes included multiple states, or only a portion of a state.

Two major NWS structural reorganizations impacted the reporting of weather damage from 1972-2004. The first reorganization occurred in August 1984 (NWS [1985]). The 52 Weather Service Forecast Offices remained in place, but local forecasts and severe weather information were now primarily the responsibility of a network of approximately 200 Weather Service Offices (WSOs). Each WSO covered a single warning area that consisted of a group of counties. The median number of counties in each warning area was 11. A WSO was responsible for "issuing special and severe weather statements," warning counties of "impending severe weather conditions which may cause the loss of life or property," and "issuing local statements to keep the public informed of the local hurricane effects" (NWS [1985], p27).

The second reorganization took effect in 1997. Referred to as "modernization and restructuring," the 1997 reorganization eliminated the Weather Service Forecast Offices and dramatically reduced the number of WSOs (National Research Council [2012]). Each remaining WSO now covered (on average) a larger county warning area. The median number of counties in each WSO post-modernization is 23.

The decentralized and changing structure of the NWS offices from 1972-2004 is the basis of our selection model exclusion restriction. We create indicator variables for the NWS forecast areas. The indicators are identified in the model from the counties belonging to each NWS forecast area in the cross section, and from the changes over time to the NWS office forecast area geographical boundaries. Figure 1 provides an illustrative example. Each of the three panels covers a different time period. The figure shows the county boundaries for Indiana and Ohio and for counties in adjacent states that share a NWS forecast area as a county in Indiana or Ohio. The forecast areas for Indiana and Ohio were essentially statewide from 1972 - July 1984, with the exception of the northwestern-most county in Indiana which was part of the Illinois forecast area. The number of forecast areas covering Indiana and Ohio expanded to 15 from August 1984 - 1996, seven of which included counties from multiple states. The number of forecast areas was reduced to nine in 1997 following NWS modernization.

Table 3 column 3 includes the NWS forecast office indicator variables as explanatory variables in the missing weather damage linear probability model. The NWS indicators strongly predict whether damage information is missing in SHELDUS. The model specification that only includes the NWS indicators (not shown) has an R-squared statistic of 0.093. This simple model can explain about 50% more of the variation in the missing weather data than can the specification that includes the eight weather event and demographic variables, along with year and month fixed effects (Table 3 column 1). Adding the NWS indicators to a specification that already includes county fixed effects still explains substantially more variation in the missing weather data (Table 3 column 3). The F-statistic from a hypothesis test that each NWS indicator has an estimated coefficient equal to zero is 7.53.[9]

The second requirement is for the NWS indicators to satisfy the exclusion restriction. The NWS indicators are based on geography. The inclusion of county fixed effects in the model is potentially critical for the validity of the exclusion restriction as there is spatial

[9] A traditional F-test assumes that all observations are independent. The default in the statistical software Stata is to use the number of clusters less the number of regressors as the denominator degrees of freedom. When we assume that all observations are independent the F-statistic is 22.17.

correlation in weather damage. For example, some counties have higher historical tornado risk. The county fixed effects control for any constant geographic correlation in weather damage at a finer geographic unit than the NWS indicators. The assumption is that there is no correlation between weather damage and the NWS indicators beyond the geographic correlation captured by the county fixed effects.

The exclusion restriction is similar to Barnighausen et al. [2011] who use interviewer identity as a predictor for whether an individual will agree to a HIV test as part of a health survey. Interviewers are randomly assigned geographically and some interviewers are more persuasive at convincing individuals to complete the test. Interviewer characteristics such as gender and charisma correlate with the likelihood that an interviewee agrees to a HIV test, while having no direct impact on the test outcome. However, since some geographic regions have higher underlying levels of HIV infection in the population it is important to control for geographic fixed effects.

3 Imputation using a Selection Model and an Exclusion Restriction

It is not possible to know for certain whether the selection equation is specified correctly, unless the researcher has access to external information regarding the true selection mechanism (e.g. Raghunathan [2016], p7). Unknown factors that are omitted from the selection model could invalidate the Missing at Random (MAR) assumption. The statistics literature emphasizes the critical role of imputation diagnostic checks (e.g. Abayomi et al. [2008]; Nguyen et al. [2017]). For example, if the MAR assumption is valid, then the distributions of the non-missing observations and the imputed (missing) observations for a variable should be similar, conditional on the other variables included in the selection equation. Still, diagnostic checks can only provide evidence that a MAR selection mechanism is reasonable given the observed data and can not rule out that the data are Missing Not at Random (MNAR).

Data imputation using a selection model has two attractive features. First, the Heckman Selection Model, and the parametric and non-parametric selection model literature more broadly, is already an established approach to account for non-random selection when there is missing data for a dependent variable (e.g. Heckman [1976]; Vella [1998]).[10] The recent innovation in the applied statistics literature is to use the model as part of an imputation procedure to address missing data for an *independent* variable (e.g. Galimard et al. [2016];

[10] We model the selection process following Heckman [1976], and for the ease of exposition refer to this selection model as the Heckman Selection Model throughout the rest of the paper.

Ogundimu and Collins [2019]; Gomes et al. [2020]).[11] This allows the model to be applied to a wider range of settings involving missing data.

Second, provided the Heckman Selection Model assumptions are satisfied, then imputation is valid under both MAR and MNAR selection mechanisms. Correct specification of a MAR selection equation is no longer the central challenge. Instead, the credibility of the imputation procedure hinges on the validity of the exclusion restriction (or instrument). In other words, use of a Heckman Selection Model is a move towards a more "design-based" imputation approach that does not attempt to specify the "true" selection mechanism, but rather focuses on credible modeling of a single causal relationship (Angrist and Pischke [2017]).

Researchers have highlighted the difficulty in finding an instrument for the probability of missingness that satisfies a selection model exclusion restriction (e.g. Bushway et al. [2007]; DiNardo et al. [2006]; Lee [2009]; Kline and Santos [2013]). This difficulty is one motivation for a literature that introduces approaches to bound how the missing values impact a model's estimates (e.g. Horowitz and Manski [2000]; Lee [2009]; Kline and Santos [2013]). For example, Lee [2009] develops a method that does not require an exclusion restriction and provides worst-case scenario bounds based on trimming the outcome distribution. However, this bounding method is not directly applicable when the missing data are for an independent variable in a model.

3.1 Selection Model

Equation 1 is our model of interest and matches the GR model that we reanalyze in Section 4. Y_{cm} is the dependent variable for county c during month m. X^*_{cm} is a key independent variable that is only partially observed in the panel. We refer to this variable as having missing observations. In our reanalysis of GR, Y_{cm} is the two-party (Democrat and Republican) Presidential vote share and X^*_{cm} is SHELDUS weather damage.

$$Y_{cm} = \theta X^*_{cm} + \boldsymbol{W_{cm}}\boldsymbol{\beta} + \boldsymbol{\alpha_c} + \boldsymbol{\gamma_t} + \eta_{cm} \tag{1}$$

$\boldsymbol{W_{cm}}$ are a vector of covariates. θ and (potentially) $\boldsymbol{\beta}$ are parameters of interest in the model. $\boldsymbol{\alpha_c}$ are county fixed effects that control for county characteristics such as location and topography. $\boldsymbol{\gamma_t}$ are year fixed effects that control for changes in county voting preferences over time. η_{cm} is an error term.

[11] Little and Rubin [2020] discuss the Heckman Selection Model as a method to address MNAR missing data for the dependent variable, but write that the approach "could also be used to model a predictor variable with missing values" (p362).

Equations 2 and 3 follow the standard selection model setup (Heckman [1976]). Equation 2 is a model for X_{cm}. Equation 3 is a selection model. X^*_{cm} is observed and equal to X_{cm} when $S_{cm} \geq 0$, and missing otherwise (if $S_{cm} < 0$).

$$X_{cm} = \boldsymbol{W_{cm}}\boldsymbol{\beta^x} + \boldsymbol{\alpha_c} + \boldsymbol{\gamma_t} + \epsilon_{cm} \tag{2}$$

$$S_{cm} = \omega Y_{cm} + \boldsymbol{W_{cm}}\boldsymbol{\beta^s} + \boldsymbol{Z_{cm}}\boldsymbol{\delta} + \boldsymbol{\alpha_c} + \boldsymbol{\gamma_t} + \upsilon_{cm} \tag{3}$$

$\boldsymbol{W_{cm}}$, $\boldsymbol{\alpha_c}$, and $\boldsymbol{\gamma_t}$ are included in both equations.[12] Y_{cm} is included in the selection equation. If the coefficient on Y_{cm} is zero (and the error terms η_{cm} and υ_{cm} are uncorrelated), then a complete case analysis using Equation 1 will provide consistent model estimates for the full sample (e.g. Little and Zang [2011]). We return to this point when interpreting the results in Section 4.

As we describe in the next section, Equation 2 is the basis for a MAR imputation procedure. Imputation will lead to consistent parameter estimates provided the right hand side variables account for all of the sources of selection (modeled in Equation 3) that also affect the value of Y_{cm}. A critical factor when imputing the missing X^*_{cm} observations is to allow for the same underlying variance-covariance structure from the economic model (Equation 1) in the imputation model (e.g. Raghunathan [2016], p101). When we impute using Equation 2 we also include Y_{cm} as a covariate on the right hand side.

The NWS forecast zone indicators, $\boldsymbol{Z_{cm}}$, are excluded from Equation 2 and allow for identification in the selection model without relying on the functional form assumptions. The exclusion restriction is valid if the NWS indicators are predictive of weather damage information being voluntarily reported to the national office for a county, and uncorrelated with the actual level of damage (after controlling for fixed county characteristics).

Heckman [1979] shows that potential bias from estimating Equation 2 on the selected sample can be reformulated as an omitted variable problem. The sample selection problem can be solved using a 2-step estimation procedure. First, estimate the selection equation on the full sample using a probit model and calculate the estimated Inverse Mills Ratio.[13] Second, estimate a version of Equation 2 that also includes the estimated Inverse Mills Ratio. Until recently, the incidental parameter problem precluded the inclusion of unit fixed effects

[12] For simplicity we use the same notation for the county and year fixed effects in Equations 1 - 4. The estimated coefficients (which are not of primary interest) can vary between models.

[13] We make the standard selection model assumptions that $(\epsilon_{cm}, \upsilon_{cm})$ is independent of $(\boldsymbol{W_{cm}}, \boldsymbol{Z_{cm}})$ with mean zero, $\upsilon_{cm} \sim$ Normal(0,1), and linearity of the population regression of ϵ_{cm} on υ_{cm}. Of note, we do not need to assume that the error term in Equation 2 is distributed normally (e.g. Wooldridge [2002], p562). The Inverse Mills Ratio is defined as the ratio of the pdf of the normal distribution to the cdf of the normal distribution, evaluated using the estimated coefficients from Equation 3.

when running a Heckman Selection Model (e.g. Lancaster [2000]).[14] We follow Fernandez-Val and Vella [2011] who provide a formula to correct for this bias.[15]

$$X_{cm} = \boldsymbol{W_{cm}\beta} + \psi IMR_{cm} + \boldsymbol{\alpha_c} + \boldsymbol{\gamma_t} + \epsilon_{cm} \tag{4}$$

Equation 4 is the same as Equation 2, except that it includes the Inverse Mills Ratio (IMR_{cm}) constructed from estimating the selection equation and applying the Fernandez-Val and Vella [2011] bias correction. As we discuss below, imputation using Equation 4 (again including Y_{cm} on the right hand side) avoids concerns over misspecification of the exact selection equation and can improve the credibility of study results that use SHELDUS damage information. Finally, we also estimate Equation 1 using a lagged dependent variable model. We replace the county fixed effects with a twice lagged dependent variable when we impute and estimate the lagged dependent variable model.

3.2 Imputation that assumes a MAR model

Standard missing data imputation approaches typically assume that the missing data are MAR (Abayomi et al. [2008]). In our setting, there is a single variable, SHELDUS damage, with missing information. The observed SHELDUS distribution does not follow a normal distribution. There are no negative values, a discrete probability mass at zero, and a roughly normal distribution to the right of zero. We use Predictive Mean Matching (PMM) as part of our imputation procedure. PMM was developed, in part, to better handle situations where the missing data are not thought to follow a normal distribution. Our estimation results in Section 4 are similar if we use a regression-based imputation process that does not involve PMM.

Imputation of the missing data and estimation of the model of interest can be divided into five steps. First, estimate Equation 2 on the subsample with no missing data. Calculate the

[14] Lancaster [2000] summarizes a seminal paper on the incidental parameters problem by Neyman and Scott [1948], writing: "A sequence of independent random variables whose probability laws involve parameters of two types. The first type appears in the probability law of every random variable; the second type appears in the law of only a finite number, possibly one. Parameters of the first type are 'structural'; parameters of the second type are 'incidental', in Neyman and Scott's terminology. In so far as each observation provides information about the parameters of its probability law, information about the incidental parameters stops accumulating after a finite number have been taken" (p392). Lancaster [2000] further states that "the incidental parameter problem is typically seen to arise (only) with panel data models when allowance is made for agent specific intercepts in a regression model" (p395).

[15] Fell and Kaffine [2018] estimate a Heckman Selection Model with unit fixed effects and apply the Fernandez-Val and Vella [2011] bias correction to study the reduction in coal-fired electricity generation in the US since 2007. We thank the authors for sharing their Stata script, which we use as a template for the bias correction portion of our code. Fernandez-Val and Vella [2011] provide example code to implement the bias correction in MatLab on the website of one of the authors.

estimated parameter distributions using the coefficient point estimates and the estimated error variance. Second, create a new dataset by jointly drawing from the parameter distributions in step one, and then using Equation 2 to calculate an imputed value for each observation. When imputing each observation, an error term is randomly drawn using the estimated distribution from the first step. X_{cm} is calculated for each observation regardless of whether X^*_{cm} is missing. Third, we use PMM. For each observation with a missing X^*_{cm}, we select the five observations with a non-missing X^*_{cm} that have the closest imputed values to the observation with the missing X^*_{cm}. We then randomly select the *actual* X^*_{cm} from one of the five matches to use for the missing observation. Fourth, we estimate the model of interest (Equation 1). Fifth, we repeat steps 2-4 twenty times. We combine the twenty estimates for each parameter from step four using Rubin's rules (Rubin [1987]), which is essentially an averaging of the coefficients across each imputation.[16]

3.3 Imputation using a model that is robust to MNAR

The critical step in the MAR imputation process outlined above is setting the imputation model. Imputation will (likely) lead to inconsistent parameter estimates in Equation 1 if X^*_{cm} is MNAR. Estimating a Heckman Selection Model as part of an imputation procedure relaxes the strong assumption that the researcher has complete knowledge of the selection equation, and can lead to consistent estimation of Equation 1 even when X^*_{cm} is MNAR. Specifically, imputation using a Heckman Selection Model replaces Equation 2 in steps one and two (of the MAR imputation) with Equation 4.

Galimard et al. [2016] demonstrate that imputation based on a two-step Heckman Selection Model can provide consistent estimates for the model of interest when the missing variable is MNAR. We follow Galimard et al. [2016] and expand on their approach in three ways. First, Galimard et al. [2016] examine a setting where there is missing data for two variables in the model of interest. The dependent variable is MNAR, while the independent variable is MAR. The authors use multiple imputation by chained equation, a standard approach to sequentially impute missing data when the data are missing for multiple variables (e.g. Raghunathan et al. [2001]). We only impute a single variable, but the MNAR variable is the independent variable in the model of interest. Second, we estimate a Heckman Selection Model that includes unit fixed effects and apply the Fernandez-Val and Vella [2011] incidental parameter bias correction. Third, we use PMM due to the non-normality of the observed X^*_{cm} distribution.

An alternative and closely related imputation model is a one-step maximum likelihood

[16] The imputation is run in Stata using *mi impute*.

Heckman Selection Model estimator (Galimard et al. [2018]). The one-step estimator is more efficient (Vella [1998]). Ogundimu and Collins [2019] also show, via simulation, that the one-step model performs somewhat better when imputing than the two-step model, especially when the variable with missing data is missing completely at random. We use the two-step model in this study for two reasons. First, we show that the SHELDUS weather damage variable is clearly not missing completely at random, so there is limited benefit to use the one-step model for imputation. Second, our model of interest includes unit fixed effects. Fernandez-Val and Vella [2011] provide a bias correction for the incidental parameter bias problem when using a two-step model (and not a one-step model).

3.4 Simulation Experiment

We simulate data and test the performance of the (two-step) Heckman Selection Model imputation procedure in the appendix. We consider a simple setting where the model of interest is $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2^* + \eta$. A selection equation determines whether X_2 is observed (X_2^*), and we set the parameters so that approximately 50% of the observations of X_2 are missing in our complete case and imputation analyses. We consider four cases determined by whether the selection model depends on Y, and whether X_2^* is MAR or MNAR. We simulate 100 random datasets of 10,000 observations for each case. We estimate three models for each dataset and case: full sample, complete case, imputation using the Heckman Selection Model. Appendix Table 1 shows the average coefficient estimate, standard error, and relative bias for β_2 for each case across the 100 simulated datasets.

The simulation exercise underscores several points. First, complete case estimation when there is missing data for an independent variable can provide estimates that are consistent for the full sample, provided selection does not depend on the dependent variable (e.g. Little and Zang [2011]). Second, the estimated coefficient for an independent variable with missing data is biased in a complete case analysis when selection depends on the dependent variable.[17] In our simulation, the relative bias is approximately twice as large when the independent variable is MNAR than when it is MAR. Third, the imputation procedure leads to a coefficient estimate for β_2 that is statistically indistinguishable from the estimated full sample coefficient, and which has low relative bias in all four cases. Forth, we model X_1 and X_2 as being correlated. In principle, the missing data for X_2 could lead the coefficient for X_1 to also be biased (when selection depends on Y). However, we do not find this to be true

[17] We model the selection as depending directly on the dependent variable (from the model of interest). However, the complete case analysis would still lead to bias if misspecification of the selection equation led the error terms of the selection equation and model of interest to be correlated. This could occur, for example, if a health pandemic leads to an increase in unreported weather damage (e.g. because of worker shortages, or other government priorities), while also impacting the vote share for the president.

in our simulation.

4 Reanalysis of Gasper and Reeves (2011)

The early empirical literature on retrospective voting focuses on how the electorate responds to economic conditions when voting for incumbent politicians or political parties. There are several shortcomings of using information about economic conditions to test theories of retrospective voting. These include the often tenuous link between political actions and economic performance, and the challenge that economic conditions are not randomly assigned (e.g. Alesina et al. [1993]; Carsey and Wright [1998]; Healy and Malhotra [2010]; Healy and Malhotra [2013]).

GR use weather damage as the basis of a quasi-experimental research design to separately test whether voters are (more) responsive or attentive. The causal question of interest is whether voters reward elected politicians for their actions (attentive electorate), or only punish politicians for the random state of the world (responsive electorate). That is, do victims of a natural disaster vote more frequently for presidents and governors who provide federal disaster assistance?[18]

GR is frequently cited as compelling evidence for an attentive electorate. GR has been cited at least 356 times (Google Scholar, July, 2021). A recent survey on retrospective voting highlights GR as a prominent study on retrospective voting outside the "economic domain" (Healy and Malhotra [2013], p295). Moreover, one of authors of GR summarizes the findings in an award-winning book, stating: "while voters punish presidents for the mere occurrence of natural disaster damage, they reward them at much higher levels when they respond with federal aid" (Kriner and Reeves [2015], p87).[19]

GR estimate a linear regression model using Equation 5 and a county-by-year panel

[18] We follow the terminology of GR in our analysis. A responsive electorate is less-likely to vote for incumbent politicians or political parties following a disaster due to the personal costs of the disaster damage. An attentive electorate evaluates politicians based on their response to the disaster and, for example, is more likely to support politicians who help secure federal disaster assistance. Other studies emphasize a more nuanced view of retrospective voting. For example, Woon [2012] and Healy and Malhotra [2013] distinguish between two types of attentive voters: reward-punishment (electoral sanction) and electoral selection. Healy and Malhotra [2010] emphasize that it can be rational for voters to respond to weather damage by voting against incumbent politicians, for example, if voters are information-constrained and conclude that at least some portion of the disaster damage is the consequence of political decisions.

[19] Appendix Table 2 shows that GR continues to be cited by papers published in top, general interest political science journals since 2014. For example, Heersink et al. [2017] write that "Gasper and Reeves [2011] find voters punish incumbent presidents for damage from severe weather but that positive electoral effects of disaster declarations overwhelm the negative effects of the disaster itself" (p261). The results have also been widely covered in the popular media, including by CNN, FiveThirtyEight, Solon, and The Washington Post, as recently as 2017.

dataset.

$$
\begin{aligned}
y_{ct} \;&=\; \beta_1 Damage_{ct} + \beta_2 Disaster_{ct} + \beta_3 Turndown_{ct} + \beta_4 PresVote(Lag)_{ct} \\
&\quad + \beta_5 PresVote(2Lag)_{ct} + \beta_5 Income_{ct} + \alpha_c + \gamma_t + \epsilon_{ct}
\end{aligned} \tag{5}
$$

y_{ct} is the dependent variable and measures the incumbent two-party (Democrat and Republican) vote share in county c in election year t. Special elections are excluded from the panel. Since elections are held in November, the year subscript t is the same for the dependent and independent variables in the model even though the disaster damage occurs before the election. The model is run separately for presidential elections from 1972-2004 and for gubernatorial elections from 1970-2006. The presidential vote share model considers all presidential elections and does not distinguish between voting for an incumbent president and the incumbent president's political party. We focus on a reanalysis of the presidential model in the paper.

The attentive and responsive electorate hypotheses are captured by three coefficients in the model. β_1 estimates the correlation between vote share and disaster damage, after adjusting for the political response and the other control variables. A positive coefficient estimate for β_1 is support for the responsive electorate hypothesis. $Damage_{ct}$ is defined as the natural log of the county-level weather damage for the six months prior to the election per 10,000 county residents.

$Disaster_{ct}$ is the number of Presidential Disaster Declarations in the county during the six months prior to the election. A Presidential Disaster Declaration provides federal assistance to repair public infrastructure, as well as, cash grants and subsidized loans directly to individual residents. A positive coefficient estimate for β_2 is support for the attentive electorate hypothesis. The source of the Presidential Disaster Declaration information is FEMA.

$Turndown_{ct}$ is the number of denied disaster requests during the six months prior to the election. A negative coefficient estimate for β_3 in the presidential vote share model is support the attentive electorate hypothesis. One important limitation of the disaster denial information in GR is that the exact counties considered in the denied requests are unknown. All of the counties in a state have the same value for $Turndown_{ct}$ if there is one or more denied Presidential Disaster Declaration request for the state in the six months prior to an election. The source of the $Turndown_{ct}$ information is the Public Entity Risk Institute.

The model includes several control variables. $Income_{ct}$ is the median household income as reported in the last decennial US Census prior to the disaster. $PresVote(Lag)_{ct}$ and $PresVote(2Lag)_{ct}$ are the lagged and twice lagged two-way vote share for the presidential candidate of the governor's party in the previous two presidential elections. α_c are county fixed effects and control for county-specific factors that are constant over the data panel (e.g. geography). γ_t are time fixed effects and control for common yearly factors that impact all counties (e.g. an economic recession). The model assumes that the classical OLS assumptions regarding the distribution of the conditional variance of the error term are valid. No adjustments are made to account for spatial correlation.

4.1 Replication of Gasper and Reeves (2011)

There are three significant limitations to the model estimated by GR. Our focus is on the severe non-random missing data problem inherent in using the SHELDUS damage information. We also address two other limitations in implementing our reanalysis.

First, there is spatial correlation in the level of disaster damage. Hurricanes, floods, and other natural disasters can cause tremendous weather damage to personal property and public infrastructure in counties impacted by the disaster. The correlation in weather damage is greater between counties affected by the same natural disaster, than it is between a disaster-affected county and a non-affected county.

There is also spatial correlation in the decision to approve or deny a Presidential Disaster Declaration request. A governor must submit a written letter to FEMA that lists the proposed disaster counties in the state. The US president approves Presidential Disaster Declaration requests state by state. Frequently, all of the proposed counties in the governor's request will be approved or denied federal assistance. Further, due to data limitations and a coding decision, GR assign *all* counties in a state a denied disaster request if there is a denied request for *any* county in the state. $Turndown_{ct}$ is perfectly correlated for counties in the same state during the same year. Apart from exacerbating spatial correlation, this coding decision also raises the question of how we should interpret the $Turndown_{ct}$ variable. Only approximately 9% of the counties in a state coded as having a turndown actually had a denied disaster request during the year.[20]

The state-by-year spatial correlation in the approval of disaster requests will result in overly precise estimates for the coefficients in Equation 5, unless the spatial correlation is accounted for in the model (e.g. Moulton [1986]; Abadie et al. [2017]). We address spatial

[20] Denied disaster requests, on average, involve a less severe weather event, and generally include far fewer proposed disaster counties than do approved disaster requests. We obtained the list of proposed disaster counties from 102 turndowns via a series of Freedom of Information Act requests. The median number of counties included in a request is two. The average turndown only includes 9% of the counties in a state. The estimated turndown coefficient in GR is identified off of a group of counties where the vast majority were never proposed by the governor for federal disaster assistance. More than 90% of the 4,698 turndown observations in the GR replication (Table 4 column 1) are likely miscoded.

correlation by clustering the standard errors at the state-by-year (or state-by-month) level in our reanalysis.

Second, the preferred model in GR includes lagged values of the county vote share and county fixed effects. Coefficient estimates for the parameters of interest are inconsistent when both the lagged dependent variable and unit fixed effects are included as control variables (Nickell [1981]). The most straightforward solution is to estimate the model with either lagged vote share or county fixed effects (Angrist and Pischke [2008]). The fixed effect model is appropriate if we view the unobserved factors that affect voting as being mostly constant across elections. The lagged vote share model is preferred if there are important time-varying factors that affect voting preferences. We estimate both models throughout most of our reanalysis.

In Table 4 column 1 we replicate the preferred presidential vote share model using the datafile posted by the authors. A disaster declaration in the six months before an election increases the vote share for the party of the incumbent president by 0.48 percentage points, while a turndown decreases the vote share by 0.95 percentage points. The weather damage coefficient is negative. The table reports standard errors (in parentheses) that are robust to state-by-year spatial correlation. None of the coefficient estimates are statistically different from zero at conventional significance levels. The standard errors are approximately three to seven times smaller if we do not allow for the state-by-year spatial correlation.[21]

We show estimation results from the lagged vote share and county fixed effect specifications in Table 4 columns 2 and 3, respectively. The weather damage and turndown coefficients are somewhat smaller in magnitude in both specifications, relative to column 1. The disaster declaration coefficient is more stable, but less precisely estimated in the fixed effect model than in the lagged vote share model.

Table 4 columns 4-6 estimate the same models as columns 1-3, except that we use updated weather damage information from SHELDUS. Specifically, we recreate the six month county-level weather damage variable using monthly information from the 2018 SHELDUS (version 16). Using the updated SHELDUS information has little effect on the estimated coefficients in the lagged vote share and county fixed effect regressions. The monthly damage estimates from SHELDUS 2009 are (by user agreement) not posted by GR, and are no longer available for purchase. We use the updated 2018 SHELDUS information to define the weather damage

[21] The coefficient estimates are close to, but not identical to those in GR Table 2 column 3. The reason is that we correct for two errors in the posted datafile. First, there are 1,852 repeated observations (i.e. rows of data) in the panel. Each repeated county-year observation has identical information for all variables as its duplicate. The panel includes 27,894 unique county-by-year observations after we drop the repeated observations. Second, approximately 5% of the disaster observations are incorrectly coded. We recode these observations. The Appendix provides more details.

variable in our reanalysis.

The bottom panel in the table reports the number of disaster and denied disaster request (turndown) observations. We also list the number of disaster and turndown observations where the weather damage variable is zero. The weather damage variable is zero if there is no positive damage information reported in SHELDUS during the six months before the election. Notably, the weather damage variable is zero for nearly one third of the disaster observations and one half of the turndown observations in the GR replication panel (column 1).

4.2 Complete Case Reanalysis

GR estimate their model using a county-year panel. County vote share and income both vary by year. Disaster damage, the number of Presidential Disaster Declarations, and the number of denied disaster requests are all summed across the six months preceding the November elections. The missing damage information is monthly. An important consideration when imputing missing data is to maintain the same covariate structure in the data for both the imputation model and the model of interest (e.g. Raghunathan [2016], p101).

Table 5 columns 1 and 2 reestimate the GR model using the same data and a county-month panel. Recall that GR assume that all missing values are equal to zero dollars in damage. The point estimates are similar to those from the county-year models (Table 4 columns 5 and 6). We use the county-month panel in the rest of the paper, as this facilitates both complete case and multiple imputation reanalyzes.

Table 5 columns 3 and 4 estimate the GR model using a complete case sample of observations with non-missing SHELDUS damage. The complete case sample is a selected sample and not representative of the full GR panel. The weather damage estimates are negative, and substantially larger in magnitude than the estimates in columns 1 and 2. The weather damage coefficient in column 3 is statistically different from zero (probability value 0.001), and implies a 1.5 percentage point reduction in the presidential vote share for the disaster county with the median level of damage. The disaster declaration and turndown coefficients are marginally statistically significant in the lagged vote share model (probability values 0.083 and 0.064, respectively). However, it is difficult to know how to interpret the turndown coefficient as it is identified at the state level. The vast majority of the counties with $Turndown_{ct}$ equal to one were not actually part of a potential disaster declaration. All three estimated coefficients are smaller in magnitude and none are statistically different from zero in the county fixed effects model.

4.3 Reanalysis after Imputing Missing SHELDUS Observations

Table 5 columns 5 and 6 estimate the model after first imputing the missing weather damage observations using a MAR imputation model. The imputation model variables include those in Table 3 column 2, as well as, presidential vote share. We replace the county fixed effects with variables for the (twice) lagged vote share in the imputation model in column 5. The weather damage point estimates are negative and of a similar magnitude as to the complete case estimates, except that they are estimated with greater precision (statistically different from zero with probability values 0.000 and 0.014, respectively). The disaster declaration coefficient in the lagged model implies that there is an increase in the vote share for an incumbent of just over 1 percentage point (probability value 0.008). The disaster declaration coefficient in the fixed effects model and both turndown coefficients are imprecisely estimated.

Table 5 column 7 estimates a fixed effects model that first imputes the missing SHELDUS weather damage observations using an imputation model that is robust to a MNAR selection process.[22] The weather damage coefficient estimate is negative, statistically significant from zero (probability value 0.010), and approximately three times as large as in the GR fixed effects model (column 2). The disaster declaration and turndown coefficients are imprecisely estimated and similar to those in GR model.

4.4 Discussion

Recall that GR summarize their results as: "We find that electorates punish presidents and governors for severe weather damage. However, we find that these effects are dwarfed by the response of attentive electorates to the actions of their officials" (p1). None of the estimated weather damage, disaster declaration, and turndown coefficients are statistically different from zero when we replicate the GR model on the county-year panel, after allowing for the documented spatial correlation in the disaster declaration process and in the location of weather damage (Table 4). The GR model coefficients are of a similar magnitude, but more precisely estimated, when we switch to the county-month panel (Table 5 columns 1-2). The probability values for the weather damage and disaster declaration variables are 0.038 and 0.081 in the lagged vote share model.

Figure 2 shows the overall estimated vote share impact for disaster counties based on our estimated weather damage and disaster declaration coefficient estimates in Table 5 columns 1-6. Each subfigure calculates the implied vote share effect and the 95% confidence interval for a county with a Disaster Declaration at the 10th, 50th, and 90th points in the weather

[22]We do not estimate a lagged dependent variable model, as the Fernandez-Val and Vella [2011] bias correction (to our knowledge) is only applicable for unit fixed effects.

damage distribution.[23] The composite vote share impact is not different from zero across the damage distribution in the GR replication (panel A), complete case (panel B), or MAR imputation (panel C) fixed effects models.

The lagged vote share models tell a different story. The 95% confidence interval is nearly above zero in the GR replication for those counties in the lower half of the weather damage distribution. The subfigure in panel A offers weak support for the GR assertion that attentiveness to the actions of a politician can outweigh knee-jerk voter reaction to random, negative events. However, keep in mind that most of the weather damage distribution below the median is comprised of counties that have missing weather damage that is imputed as zero. There is no evidence in favor of the GR conclusion if we consider the subsample with non-missing information, or use a model to impute plausible values for the missing observations (instead of assuming zero). The complete case and MAR imputation lagged vote share models suggest that the negative weather damage impact dominates for those counties in the upper half of the weather damage distributions.

Overall, the FE model estimates that use an imputation procedure that is robust to the misspecification of the true selection process (Table 5, column 7) provide similar estimates as those from the MAR imputation model (Table 5, column 6). The 95% confidence interval for the weather damage coefficient under the imputation process that is robust to selection model misspecification contains somewhat larger negative values, but the coefficient estimates are not statistically different. There are (at least) two reasons why this could be the case.

First, selection (Equation 3) may not depend on the Presidential vote share. It is not too surprising that missing weather damage would have no direct correlation with vote share. The larger worry is that an omitted variable could affect vote share and influence if weather damage is missing (i.e. $corr(\eta_{cm}, \upsilon_{cm}) \neq 0$). Our findings suggest that this is not a major concern.

Second, it is possible that selection does depend on vote share (again possibly because $corr(\eta_{cm}, \upsilon_{cm}) \neq 0$), but that missing weather damage is MAR (i.e. $corr(\epsilon_{cm}, \upsilon_{cm}) \neq 0$ in Equations 2 and 3). The simulation exercise indicates that we might expect a small amount of bias when the dependent variable partially determines selection when missing observations are MAR.

[23]Specifically, we calculate a t-test for the following linear combination using the weather damage at each of the damage percentiles after estimating the county-month version of Equation 5: (weather damage * $\hat{\beta}_1$) + $\hat{\beta}_2 = 0$. The figure plots the coefficient estimates from the t-test along with the 95% confidence intervals. In panel C, we test the linear combination using the average weather damage across the 20 imputed datasets. We conduct the t-test for each of the 20 imputations and use Robin's Rules to combine the estimates.

5 Conclusion

This paper makes three contributions. First, we show that a popular weather damage database, SHELDUS, suffers from a severe and poorly understood missing data problem. Historically, the weather damage data that comprise SHELDUS relied on the self reporting of county-by-month damage information from regional NWS offices to the National Climatic Data Center. Approximately, three-quarters of the observations are missing. The missing observations are correlated with a number of factors, including the month of the year and if there is a Presidential Disaster Declaration. We do not attempt to definitely explore all factors that correlate with the likelihood that weather damage is missing.

Second, the missing weather damage information is likely to lead to biased coefficient estimates in many settings. We follow the recent applied statistics literature and show that imputation using a selection model, combined with an instrument, can lead to unbiased model estimates for an independent variable with missing data, even without full knowledge of the selection equation. We leverage the historical regional office structure of the NWS as an instrument to impute missing weather damage information. The instrument and imputation procedure allow for researchers studying any topic to use the SHELDUS database, while avoiding bias due to selective non-reporting.

Third, we reanalyze Gasper and Reeves [2011], a seminal study on voting behavior that relies on the weather damage information from SHELDUS. Gasper and Reeves [2011] assume that all of the counties with missing SHELDUS weather damage observations incurred no damage in those months. In our reanalysis, we allow for the non-random selection of the reported weather damage. The findings of Gasper and Reeves [2011] are mostly reversed once accounting for the missing data. Overall, the negative vote response of the electorate to weather damage either cancels out or dominates the positive, attentive influence attributed to the post-disaster actions of a politician.

6 References

Operations of the national weather service. Technical report, National Weather Service, January 1978.

Operations of the national weather service. Technical report, National Weather Service, January 1985.

Storm data. Technical Report 5, National Oceanic and Atmospheric Administration, May 1992. URL https://www.ncdc.noaa.gov/IPS/static/images/sdsample.pdf.

Storm data. Technical Report 12, National Oceanic and Atmospheric Administration, December 1995. URL https://www.ncdc.noaa.gov/IPS/static/images/sdsample.pdf.

The National Weather Service Modernization and Associated Restructuring: A Retrospective Assessment. The National Academies Press, 2012.

Alberto Abadie, Susan Athey, Guido W. Imbens, and Jeffery Wooldridge. When should you adjust standard errors for clustering? *National Bureau of Economic Research Working Paper*, 24003, 2017.

Kobi Abayomi, Andrew Gelman, and Marc Levy. Diagnostics for multivariate imputations. *Journal of the Royal Statistical Society*, 57, 2008.

Jason Abrevaya and Stephen G. Donald. A gmm approach for dealing with missing data on regressors. *The Review of Economics and Statistics*, 99, 2017.

Alberto Alesina, John Londregan, and Howard Rosenthal. A model of the political economy of the united states. *American Political Science Review*, 87(1), 1993.

Christopher J. Anderson. The end of economic voting? contingency dilemmas and the limits of democratic accountability. *Annual Review of Political Science*, 10, 2007.

Joshua D Angrist and Jörn-Steffen Pischke. *Mostly harmless econometrics: An empiricist's companion.* Princeton university press, 2008.

Joshua D. Angrist and Jorn-Steffen Pischke. Undergraduate econometrics instruction: Through our classes, darkly. *Journal of Economic Perspectives*, 31, 2017.

Till Barnighausen, Jacob Bor, Speciosa Wandira-Kazibwe, and David Canning. Correcting hiv prevalence estimates for survey nonparticipation using heckman-type selection models. *Epidemiology*, 22(1):27–35, 2011.

Jean-Noel Barrot and Julien Sauvanat. Input specificity and the propogation of idiosyncratic shocks in production networks. *Quarterly Journal of Economics*, 131(3):1543–1592, 2016.

Michael M. Bechtel and Jens Hainmueller. How lasting is voter gratitude? an analysis of the short- and long-term electoral returns of beneficial policy. *American Journal of Political Science*, 55 (4), 2011.

Gennaro Bernile, Vineet Bhagwat, and P. Raghavendra Rau. What doesn't kill you will only make you more risk-loving: Early-life disasters and ceo behavior. *Journal of Finance*, 72(1), 2017.

Christopher R. Bollinger and Barry T. Hirsch. Match bias from earnings imputation in the current population survey: The case of imperfect matching. *Journal of Labor Economics*, 24(3):483–519, 2006.

Shawn Bushway, Brian D. Johnson, and Lee Ann Slocum. Is the magic still there? the use of the heckman two-step correction for selection bias in criminology. *Journal of Quantitative Criminology*, 23:151–178, 2007.

James R. Carpenter and Michael G. Kenwood. *Multiple Imputation and its Application*. Wiley, 2013.

Thomas M. Carsey and Gerald C. Wright. State and national factors in gubernatorial and senatorial elections. *American Journal of Political Science*, 42(3), 1998.

Jowei Chen. Voter partisanship and the effect of distributive spending on political participation. *American Journal of Political Science*, 57(1), 2013.

Shawn Cole, Andrew Healy, and Eric Werker. Do voters demand responsive governments? evidence from indian disaster relief. *Journal of Development Economics*, 97(2), 2012.

Pamela Johnston Conover, Stanley Feldman, and Kathleen Knight. Judging inflation and unemployment: The origins of retrospective evaluations. *The Journal of Politics*, 48(3):565–588, 1986.

Ronald J. Daniels and Michael J. Trebilcock. Rationales and instruments for government intervention in natural disasters. *University of Pennsylvania Scholarly Commons*, January 2006. URL http://repository.upenn.edu/law_series/19.

John DiNardo, Justin McCrary, and Lisa Sanbonmatsu. Constructive proposals for dealing with attrition: An empirical example. *NBER Working Paper*, 2006.

C Christine Fair, Patrick Kuhn, Neil A Malhotra, and Jacob Shapiro. Natural disasters and political engagement: evidence from the 2010–11 pakistani floods. 2017.

Ray C. Fair. The effect of economic events on votes for president. *The Review of Economics and Statistics*, LX(2), 1978.

Harrison Fell and Daniel T. Kaffine. The fall of coal: Joint impacts of fuel prices and renewables on generation and emissions. *American Economic Journal; Economic Policy*, 10(2):90–116, 2018.

Ivan Fernandez-Val and Francis Vella. Bias corrections for two-step fixed effects panel data estimators. *Journal of Econometrics*, 163:144–162, 2011.

Claudio Ferrez and Federico Finan. Exposing corrupt politicians: the effect of brazil's publicly released audits on electoral outcomes. *Quarterly Journal of Economics*, 123(2):703–45, 2008.

Claudio Ferrez and Federico Finan. Electoral accountability and corruption in local governments: evidence from audit reports. *American Economic Review*, 101(4):1274–1311, 2011.

Jacques-Emmanuel Galimard, Sylvie Chevret, Camelia Protopopescu, and Matthieu Resche-Rigon. A multiple imputation approach for mnar mechanisms compatible with heckman's model. *Statistics in Medicine*, 35:2907–2920, 2016.

Jacques-Emmanuel Galimard, Sylvie Chevret, Emmanuel Curis, and Matthieu Resche-Rigon. Heckman imputation models for binary or continuous mnar outcomes and mar predictors. *BMC Medical Research Methodology*, 18, 2018.

Justin Gallagher. Learning about an infrequent event: Evidence from flood insurance take-up in the us. *American Economic Journal: Applied Economics*, 6, 2014.

John T. Gasper and Andrew Reeves. Make it rain? retrospection and the attentive electorate in the context of natural disasters. *American Journal of Political Science*, 55(2), 2011.

Manuel Gomes, Michael G. Kenward, Richard Grieve, and James Carpenter. Estimating treatment effects under untestable assumptions with nonignorable missing data. *Statistics in Medicine*, 39: 1658–1674, 2020.

Andrew Healy and Gabriel S. Lenz. Substituting the end for the whole: Why voters respond primarily to the election-year economy. *American Journal of Political Science*, 58:31–47, 2014.

Andrew Healy and Neil Malhotra. Random events, economic losses, and retrospective voting: Implications for democratic competence. *Quarterly Journal of Political Science*, 5(2), 2010.

Andrew Healy and Neil Malhotra. Retrospective voting reconsidered. *Annual Review of Political Science*, 16, 2013.

James J. Heckman. The common structure of statistical models of truncation, sample selection, and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement*, 5:475–492, 1976.

James J. Heckman. Sample selection bias as a specification error. *Econometrica*, 47(1):153–161, 1979.

Boris Heersink, Brenton D Peterson, and Jeffery A Jenkins. Disasters and elections: Estimating the net effect of damage and relief in historical perspective. *Political Analysis*, 25(2):260–268, 2017.

Boris Heersink, Michael P Olsen, Brenton D Peterson, and Jeffery A Jenkins. Natural disasters, 'partisan retrospection,' and u.s. presidential elections. *Political Behavior*, 2020.

Barry Hirsch and Edward Schumacher. Match bias in wage gap estimates due to earnings imputation. *Journal of Labor Economics*, 22(3):689–722, 2004.

Joel L. Horowitz and Charles F. Manski. Nonparametric analysis of randomized experiments with missing covariate and outcome data. *Journal of the American Statistical Association*, 95(449): 77–84, 2000.

V.O. Key. *The Responsible Electorate: Rationality in Presidential Voting, 1936-1960*. Harvard University Press, Cambridge, MA, 1966.

Patrick Kline and Andres Santos. Sensitivity to missing data assumptions: Theory and an evaluation of the u.s. wage structure. *Quantitative Economics*, 4:231–267, 2013.

Douglas L. Kriner and Andrew Reeves. *The Particularistic President: Executive Branch Politics and Political Inequality*. Cambridge University Press, New York, NY, 2015.

Tony Lancaster. The incidental parameter problem since 1948. *Journal of Econometrics*, 95: 391–413, 2000.

David S. Lee. Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, 76:1071–1102, 2009.

Roderick J. Little and Nanhua Zang. Subsample ignorable likelihood for regression analysis with missing data. *Journal of the Royal Statistical Society*, 60, 2011.

Roderick J.A. Little and Donald B. Rubin. *Statistical Analysis with Missing Data*. Wiley, Hoboken, NJ, 3 edition, 2020.

Eric Loken and Andrew Gelman. Measurement error and the replication crisis. *Science*, 355, 2017.

Ian K. McDonough and Daniel L. Millimet. Missing data, imputation, and endogeneity. *Journal of Econometrics*, 199, 2017.

Brent Moulton. Random group effects and the precision of regression estimates. *Journal of Econometrics*, 32, 1986.

J. Neyman and Elizabeth L. Scott. Consistent estimation from partially consistent observations. *Econometrica*, 16:1–16, 1948.

Cattram D. Nguyen, John B. Carlin, and Katherine J. Lee. Model checking in multiple imputation: an overview and case study. *Emerging Themes in Epidemiology*, 14(8), 2017.

Stephen Nickell. Biases in dynamic models with fixed effects. *Econometrica*, 49, 1981.

Brendan Nyhan. Media scandals are political events: How contextual factors affect public controversies over alleged misconduct by us governors. *Political Research Quarterly*, 70(1):223–236, 2017.

Emmanuel O. Ogundimu and Gary S. Collins. A robust imputation method for missing responses and covariates in sample selection models. *Statistical Methods in Medical Research*, 28(1), 2019.

Torsten Persson, Gerard Roland, and Guido Tabellini. Separation of powers and political accountability. *The Quarterly Journal of Economics*, 112(4), 1997.

Trivellore Raghunathan. *Missing Data Analysis in Practice*. CRC Press, Boca Raton, FL, 2016.

Trivellore E. Raghunathan, James M. Lepkowski, Jon Van Hoewyk, and Peter Solenberger. A multivariate techniqu for multiply imputing missing values using a sequence of regression models. *Survey Methodology*, 27(1), June 2001.

Jose Maria Rodriguez-Valadez and Cesar Marinez-Alvarez. Natural disasters, electoral performance, and social policy: Evidence from mexico. *Unpublished Manuscript*, 2021.

Donald B. Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976.

Donald B. Rubin. *Multiple Imputation for Nonresponse in Surveys*. Wiley, 1987.

Francis Vella. Estimating models with sample selection bias: A survey. *The Journal of Human Resources*, 33(1), 1998.

Donald Wittman. Wy democracies produce efficient results. *Journal of Political Economy*, 97(6): 1395–1424, 1989.

Jeffery M Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA, 2002.

Jonathan Woon. Democratic accountability and retrospective voting: A laboratory experiment. *American Journal of Political Science*, 56(4), 2012.

Table 1: **Missing Weather Damage Data in SHELDUS**

I. SHELDUS Weather Damage Data	
Total (Percent) Monthly Observations	167,364
Missing Observations	125,680 (75%)
Observations Reporting Damage	41,590 (25%)
Reporting Positive Damage	36,604 (88%)
Reporting Zero Damage	4,986 (12%)
II. SHELDUS Yearly Observations (6 Months May-Oct)	
Total (Percent) Observations	27,894
Reporting Damage All 6 Months	241 (0.86%)
Missing Damage All 6 Months	8,568 (31%)
Missing and Non-Missing Damage	19,085 (68%)
Zero Damage and Missing Damage	3,722 (20%)
Exclussively Positive Damage and Missing Damage	15,356 (80%)
III. County SHELDUS Statistics	
Total (Percent) Counties	3,102
Report Zero Monthly Damage	2,137 (69%)
Never Report Zero Monthly Damage	965 (31%)

Each weather damage observation in Gasper and Reeves (2011) is the sum of six underlying SHELDUS monthly observations. Number of monthly SHELDUS observations are equal to the number of county-year panel observations multiplied by six. We use as the number of county-year panel observations from the corrected Gasper and Reeves (2011) panels that drop repeat observations (see Appendix Section 2.1). We use the 2018 SHELDUS (version 16) to calculate the statistics in this table. Gasper and Reeves (2011) use the 2009 SHELDUS. Appendix Section 1.1 discusses how SHELDUS updates the database over time. Data sources: Special Hazards and Losses Database for the United States (SHELDUS).

Table 2: **FEMA Damage Information Contradicts Assumption that Missing Weather Damage Data in SHELDUS Implies Zero Weather Damage**

Total Monthly Disaster Observations	3,386
Observations (Percent) with FEMA County-Specific Damage Information:	1573 (46%)
SHELDUS Missing for Disaster Month & FEMA Reports Damage	624 (40%)
SHELDUS Missing for Disaster Month and Previous Month & FEMA Reports Damage	392 (25%)
Damage Statistics, Mean (Median):	
Natural Log of FEMA Damage per 10,000 Residents	12.30 (12.25)
Natural log of SHELDUS Damage per 10,000 Residents	10.90 (11.81)

We use the 2018 SHELDUS (version 16) to calculate the statistics in this table. The FEMA damage is for those disaster counties for which we have both Public Assistance and Individual Assistance data. The SHELDUS mean and median are taken over all disaster observations with non-missing SHELDUS damage data. We convert the damage to real 2005$ before taking the natural log per 10,000 county residents. We add one to each SHELDUS damage observation before taking the natural log. Data sources: Federal Emergency Management Agency (FEMA), Special Hazards and Losses Database for the United States (SHELDUS).

Table 3: **SHELDUS Weather Damage is Not Missing Completely at Random**

Dependent Variable: Pr(Damage Variable Missing)			
	(1)	(2)	(3)
Weather Event Variables			
Disaster Declaration	-0.220	-0.210	-0.204
	(0.044)	(0.047)	(0.046)
Disaster Declaration Next Month	-0.134	-0.114	-0.097
	(0.038)	(0.038)	(0.041)
Size of Disaster (No. Counties)	-0.001	-.001	-0.001
	(0.001)	(0.001)	(0.001)
Turndown	0.032	0.039	0.045
	(0.035)	(0.034)	(0.033)
Demographic Variables			
Median Income (log)	-0.069	-0.037	-0.095
	(0.022)	(0.036)	(0.025)
Population Size (log per 1,000)	-0.040	-0.088	-0.055
	(0.004)	(0.031)	(0.021)
African American Population (%)	0.001	0.002	0.001
	(0.000)	(0.002)	(0.001)
Older Population (%)	-0.005	-0.001	-0.002
	(0.002)	(0.002)	(0.001)
Year FE	X	X	X
Month FE	X	X	X
County FE		X	X
NWS Forecast Zone FE			X
R-squared	0.065	0.136	0.173
Observations	167,340	167,340	167,010

The table estimates a linear probability model for whether the SHELDUS weather damage variable is missing. The dependent variable equals one if the monthly SHELDUS damage is unreported (missing) for a county and zero otherwise. The 1972-2004 sample is the same as Table 2. Data sources: Federal Emergency Management Agency, National Weather Service, Public Entity Risk Institute, Special Hazards and Losses Database for the United States, US Decennial Census.

Table 4: **Effect of Severe Weather and Disaster Assistance on Incumbent Presidential Vote Share**

Replication of Gasper and Reeves (2011)

Damage Data:	SHELDUS 2009			SHELDUS 2018		
Specification:	GR Replication	Lagged Vote Share	County F.E.	GR Replication	Lagged Vote Share	County F.E.
	(1)	(2)	(3)	(4)	(5)	(6)
Weather Damage	-0.028	-0.021	-0.013	-0.039	-0.028	-0.014
	(0.032)	(0.034)	(0.048)	(0.030)	(0.033)	(0.046)
Disaster Declaration	0.483	0.548	0.415	0.503	0.564	0.415
	(0.469)	(0.393)	(0.662)	(0.462)	(0.387)	(0.651)
Turndown	-0.949	-0.651	-0.799	-0.963	-0.662	-0.803
	(0.657)	(0.566)	(0.937)	(0.657)	(0.567)	(0.938)
Lagged Vote Share	X	X		X	X	
County Fixed Effects	X		X	X		X
Income	X	X	X	X	X	X
Year Fixed Effects	X	X	X	X	X	X
R^2	0.816	0.793	0.415	0.816	0.793	0.415
Observations	27,894	27,894	27,894	27,894	27,894	27,894
Disaster	3,132	3,132	3,132	3,132	3,132	3,132
Disaster, No Damage	1,017	1,017	1,017	687	687	687
Turndown	4,698	4,698	4,698	4,698	4,698	4,698
Turndown, No Damage	2,343	2,343	2,343	1,765	1,765	1,765

The bottom panel reports the number of Presidential Disaster Declaration observations and denied Presidential Disaster Declaration observations (Turndowns) where the six month weather damage variable is zero (i.e. all six months have non-reported information or report zero damage). Standard errors that allow for state-by-year spatial correlation are in parentheses. Data sources: Federal Emergency Management Agency, Public Entity Risk Institute, Special Hazards and Losses Database for the United States, US Decennial Census.

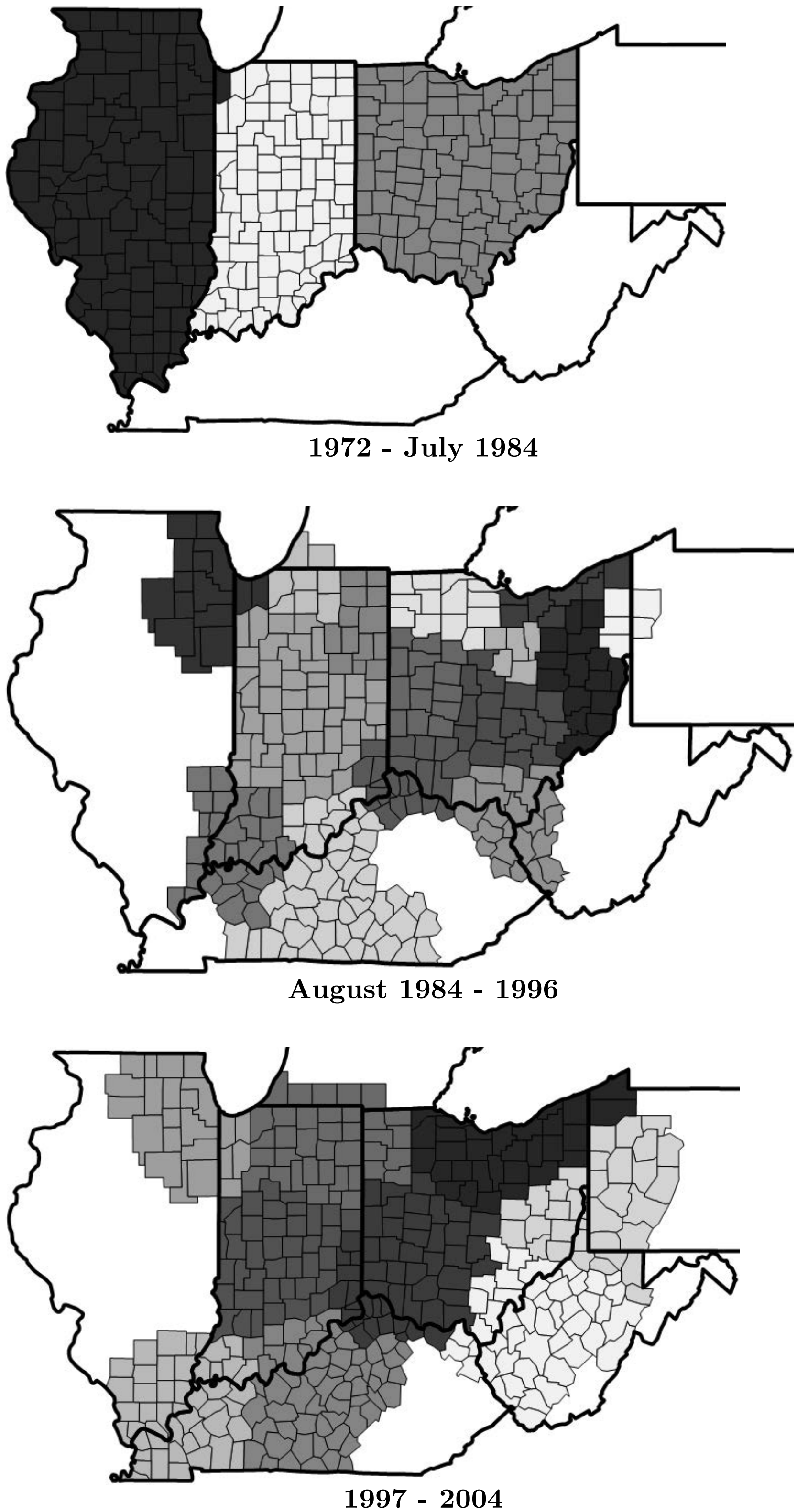
Table 5: **Effect of Severe Weather and Disaster Assistance on Incumbent Presidential Vote Share**

Reanalysis of Gasper and Reeves (2011)

Panel:	GR (Impute Zero)		Complete Case		MAR Imputation		MNAR Imputation
Specification:	Lagged Vote Share	County F.E.	Lagged Vote Share	County F.E.	Lagged Vote Share	County F.E.	County F.E.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Weather Damage	-0.036	-0.020	-0.126	-0.047	-0.145	-0.053	-0.061
	(0.017)	(0.024)	(0.040)	(0.054)	(0.015)	(0.021)	(0.024)
Disaster Declaration	0.683	0.411	0.717	0.103	1.057	0.514	0.101
	(0.392)	(0.559)	(0.414)	(0.588)	(0.395)	(0.414)	(0.593)
Turndown	-0.714	-0.729	-1.526	-1.369	-0.742	-0.749	-0.734
	(0.579)	(0.854)	(0.824)	(1.111)	(0.580)	(0.856)	(0.854)
Lagged Vote Share	X		X		X		
County Fixed Effects		X		X		X	X
Income	X	X	X	X	X	X	X
Year Fixed Effects	X	X	X	X	X	X	X
Observations	167,148	167,148	41,656	41,656	167,148	167,148	167,148

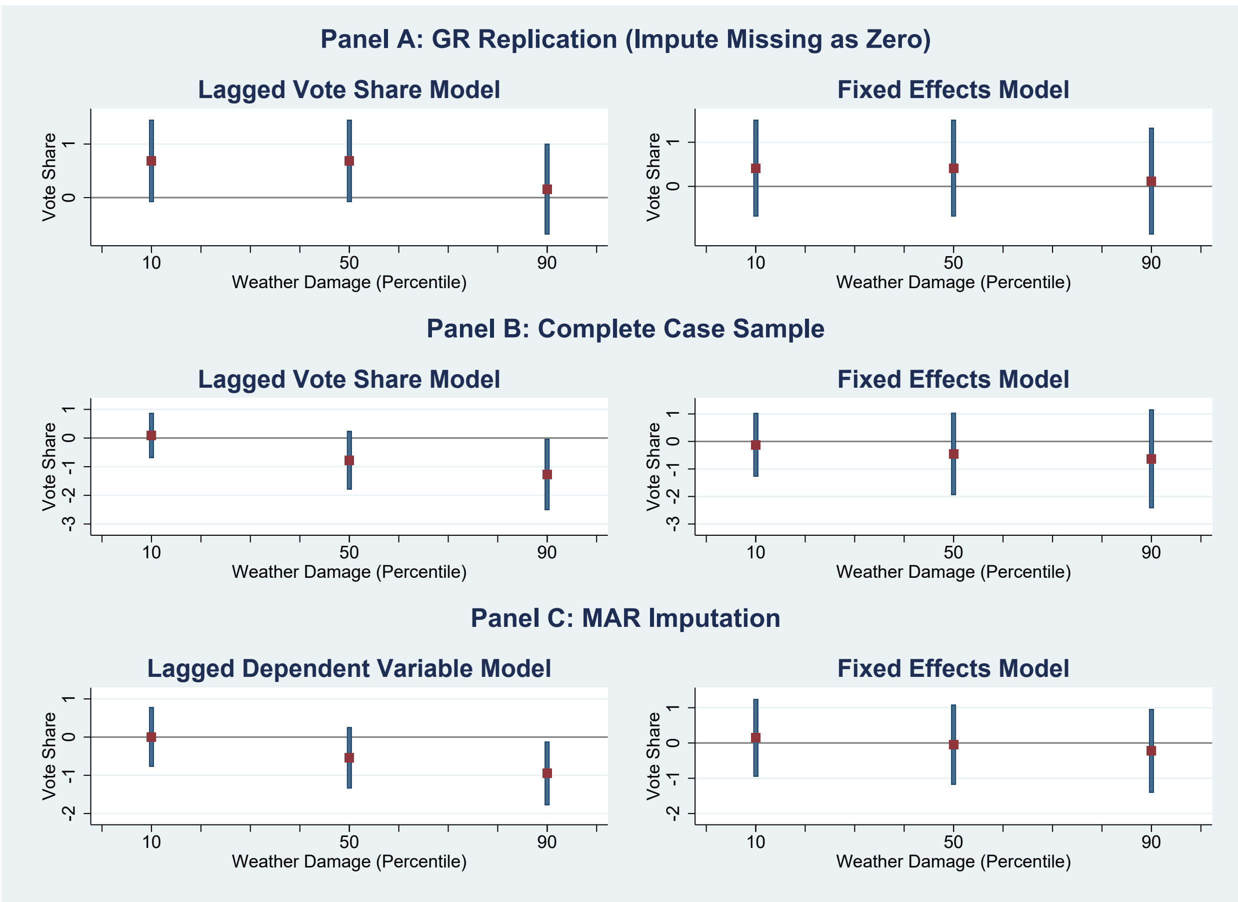
All models use a county-month data panel and cluster the standard errors at the state-by-month level. Columns 1 and 2 estimate the same models as in Table 4 panels 5 and 6, and assume all missing damage observations are equal to zero. Columns 3 and 4 estimate the model on a complete case subsample. Columns 5 and 6 first impute the missing damage data using an imputation model that assumes the data are missing at random. Column 7 first imputes the missing damage data using an imputation model that is robust to the possibility that the data are not missing at random. Data sources: Federal Emergency Management Agency, National Weather Service, Public Entity Risk Institute, Special Hazards and Losses Database for the United States, US Decennial Census.

Figure 1: **Historical NWS Forecast Zones for Indiana and Ohio**



The figure shows county boarders for Indiana and Ohio and for counties in adjacent states that share a NWS forecast area as a county in Indiana or Ohio. The three panels correspond to three different time periods in our sample. There are minor forecast area changes during the last time period (not shown), but the main reconfigurations occur between periods. Data source: National Weather Service.

Figure 2: **Overall Estimated Voteshare Impact from Weather Damage and a Disaster Declaration by Level of Weather Damage**



The figure shows the overall estimated vote share impact for disaster counties based on our estimated weather damage and disaster declaration coefficient estimates in Table 5 columns 1-6. Each subfigure calculates the implied vote share effect and the 95% confidence interval for a county with a Disaster Declaration at the 10th, 50th, and 90th points in the weather damage distribution. Data sources: Federal Emergency Management Agency, Public Entity Risk Institute, Special Hazards and Losses Database for the United States, US Decennial Census.