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

Justin Gallagher*

January 20, 2023

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 recent applied statistics literature and demonstrate an imputation procedure that relies on an instrument to estimate missing values. The imputed values are robust to nonrandom selection. 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.

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 2021 Association of Environmental and Resource Economists annual meeting, Montana State University, and UNLV. Isaac Birdwell, Katiana Carestia, Simin Gao, Andrew Hoover, Tristan Jones, Daniel Penoyer 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

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]) and political science (e.g. Conover et al. [1986]; Graham et al. [Forthcoming]) 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 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]). 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]).

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.[1]

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).

One challenge GR face in implementing their research design is how to measure weather damage. Historical weather damage databases that provide detailed coverage for the entire US are generally incomplete. GR use weather damage information from the Spatial Hazard Events and Losses Database (SHELDUS). SHELDUS is a popular data source for researchers

[1] 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].

examining questions related to natural disasters. Studies using SHELDUS are published in top general interest economics (e.g. Barrot and Sauvanat, *Quarterly Journal of Economics*, 2016), finance (e.g. Bernile et al., *Journal of Finance*, 2017), and political science (e.g. Gasper and Reeves, *American Journal of Political Science*, 2011) journals. 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).[2]

SHELDUS suffers from a pervasive missing data problem. We show that approximately 75% of the observations are not reported. We also show that these observations are not missing at random. The reason for the missing data in SHELDUS is due to the underlying reporting process of the primary source data. The main primary source data are compiled from reported weather damage in *Storm Data*, a monthly publication of the National Centers for Environmental Information. A critical feature of *Storm Data* is that the weather damage information is voluntarily reported by regional National Weather Service (NWS) offices. Each issue of *Storm Data* includes the following disclaimer: "due to difficulties inherent in the collection of this type of data, it is not all-inclusive" (Storm Data [1995], p2). Many researchers appear unaware of missing (unreported) data in SHELDUS.[3]

Researchers working with missing data face a choice on how to proceed. One common approach is to discard the missing observations and to use the complete case sample. A disadvantage of using a complete case sample is that model estimates using these data may not generalize to the population of interest. For example, since the county-months reporting non-missing data in SHELDUS are a selected sample, estimates from a model using the reported SHELDUS damage (which is sometimes \$0) are unlikely to reflect the underlying damage relationship for all county-months during the time period. A second disadvantage is that the complete case sample can be considerably smaller than a sample that includes the missing observations. Missing data are so common in SHELDUS that if researchers were to use a complete case SHELDUS sample then many of the purported advantages of the database (long panel, complete coverage of all US counties, and frequent observations) are negated.

A second approach to handle missing data is to assign a value to missing observations. Researchers who use SHELDUS often assume that there is no damage if the county-month

[2]Researchers have increasingly used satellite-based data as a source of weather damage (e.g. Donaldson and Storeygard [2016]). However, satellite-based weather damage information is generally only available for recent years, and often only for major weather events (e.g. Gallagher and Hartley [2017]).

[3]For example, Ge [*Journal of Finance*, 2021] falsely asserts that "the data set includes every natural hazard event that caused injury, death, or property/farm damage since 1960 in the U.S. and provides estimated monetary damages for each event" (p15). Dou et al. [*NBER WP*, 2022] defend the use of SHELDUS by citing numerous papers and stating "SHELDUS has been widely used in recent financial literature" (p33).

observation is missing. More than half the published papers reported on the SHELDUS website, that use the data in a regression model, assume that missingness implies zero dollars in damage (author calculation, see Appendix). GR use all the monthly SHELDUS observations during their panel period and make the assumption that missing observations have zero dollars in damage. This assumption is false. We show, using Federal Emergency Management (FEMA) administrative damage information, that 40% of the county-months for the largest weather damage events have missing observations in SHELDUS, but large amounts of verified FEMA damage.

The assumption that missing SHELDUS observations are zero will generally lead the weather damage parameter estimates in a regression model to be biased, while overestimating the model's precision. Moreover, the parameter estimates for other covariates in the same regression model can also be biased when the covariates are correlated with weather damage. For example, in the GR model, another key independent variable is an indicator for whether a Presidential Disaster Declaration is declared for a county. A Presidential Disaster Declaration is highly correlated with the level of weather damage in a county.

A third approach to handle missing data is to impute the missing observations. Multiple imputation is a standard method frequently used by researchers in a number of fields (e.g. Carpenter and Kenwood [2013]). Broadly speaking, conventional data analysis involving multiple imputation has two parts. First, the researcher specifies the equation used as the imputation model. The imputation model is estimated on the subsample with no missing data. The estimated parameters from the imputation model are then used to generate imputed values for observations with missing data. Second, the researcher estimates the research model of interest using the new dataset, which now includes imputed values in place of the missing values.

Economists are often skeptical of imputation as an approach to correct for missing data due to concerns over misspecification of the imputation model (e.g. Hirsch and Schumacher [2004]; Bollinger and Hirsch [2006]; DiNardo et al. [2006]). The imputation model will be misspecified if the error term in the imputation model is correlated with the error term from the assumed model for whether data are missing (selection equation). It is not possible to know for certain whether the selection equation is specified correctly, unless the researcher has access to information regarding the true selection mechanism (e.g. Raghunathan [2016], p7). A misspecified imputation model can lead to biased coefficient estimates when estimating the research model. In other words, the standard imputation procedure that assumes perfect knowledge of the selection equation suffers from a potential omitted variable bias concern.

We overcome this omitted variable bias concern by imputing the missing SHELDUS data using an instrumental variables multiple imputation procedure that follows the recent applied statistics literature (e.g. Galimard et al. [2016]; Ogundimu and Collins [2019]; Gomes et al. [2020]). Imputation using this approach does not rely on perfect knowledge of the selection equation. Valid imputation, and ultimately valid inference, depends on the instrument. In our setting, a valid instrument is a variable that is correlated with whether SHELDUS damage information is missing, but uncorrelated with the actual amount of weather damage.

We use historical changes to the NWS regional office weather reporting zones as an instrument when imputing missing SHELDUS values. The NWS is responsible for forecasting and reporting weather in the US. The NWS relies on a decentralized organizational structure that includes regional NWS offices and forecast zones. Each office is responsible for a specific reporting (forecast) area that consists of a group of counties. Two major NWS structural reorganizations altered the reporting areas during the past half century. We create indicator variables for the NWS regional office reporting areas. The indicators are identified in the selection equation from the counties belonging to each NWS regional office reporting area in the cross section, and from the changes to the NWS office reporting area geographical boundaries. We show that, conditional on county and time fixed effects, whether a SHELDUS observation is missing is highly correlated with the NWS reporting area in which the county is located. The county fixed effects control for any constant geographic correlation in weather damage at a finer geographic scale 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.

We reanalyze the presidential vote share model in GR after first accounting for the missing data through our instrumental variable multiple imputation procedure. In our reanalysis, the negative effect on vote share from weather damage is three times larger than in GR. We find no evidence in favor of an attentive electorate. Overall, the findings of GR are mostly reversed when we impute the missing data, instead of assuming that all missing observations incurred no damage.

This paper makes three main contributions. First, we document that SHELDUS, a widely used weather damage database, suffers from a severe and poorly understood nonrandom missing data problem. Second, we use an instrumental variables multiple imputation procedure to account for the missing data. The instrument in our model is based on the historical restructuring of the NWS office geographic weather reporting boundaries. We show that the instrument is applicable to a wide range of research settings. Researchers in a number of fields can apply the same imputation procedure to continue to (reliably) use SHELDUS data. Third, our reanalysis of Gasper and Reeves [2011] underscores how the handling of missing data impacts research conclusions. GR is a highly influential study on retrospective voting. The original finding in GR, that voter attentiveness to the actions of a politician outweighs reaction to random events, is reversed when we use our instrumental variables multiple imputation procedure to account for the missing data.

To our knowledge, our paper is the first in economics to use this imputation procedure. A broader goal of this paper is to demonstrate how data imputation, when paired with a credible research design, can be an important tool to address missing data.

The rest of the paper proceeds as follows. Section 2 presents the GR model. We focus on GR because the paper is an influential study on retrospective voting. Presenting the GR model also allows us to clearly articulate when assumptions over data missingness will lead researchers to estimate consistent regression model parameters for the population from which the full sample is drawn. In Section 3 we show that missing data in SHELDUS is rampant, does not imply no damage, and is not missing at random. Section 4 presents our instrumental variable multiple imputation model. Section 5 reanalyzes GR after first accounting for missing SHELDUS observations using our imputation model. Section 6 discusses how our instrumental variables imputation model is widely applicable to other research settings that use SHELDUS data. Section 7 concludes.

2 Retrospective Voting

A large empirical literature in both economics (e.g. Ferrez and Finan [2008]) and political science (e.g. Conover et al. [1986]; Graham et al. [Forthcoming]) examines how voters evaluate political performance. Anderson [2007] and Healy and Malhotra [2013] provide reviews of the retrospective voting literature. Over the past decade a new sub-literature has emerged that uses natural disasters to test theories of retrospective voting. GR are an early, seminal paper that uses weather damage as the basis of a quasi-experimental research design to separately test whether voters are (more) responsive or attentive.

A responsive electorate reacts to a negative event by voting against incumbent politicians or political parties. An attentive electorate does not react to random events, but instead is attuned to the actions that politicians take to deal with the random events. For example, attentive voters will not be more likely to vote against an incumbent politician following a natural disaster that causes damage to personal property when the politician is proactive in managing the situation and assisting the victims.[4]

[4] We follow the terminology of GR in our analysis. 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.

2.1 The Gasper and Reeves [2011] Model

GR estimate a linear regression model using Equation 1 and a county-by-year panel dataset.

$$
\begin{aligned}
Y_{ct} &= \beta_1 Damage^*_{ct} + \beta_2 Disaster_{ct} + \beta_3 Turndown_{ct} + \beta_4 PresVote(Lag)_{ct} \\
&\quad + \beta_5 PresVote(2Lag)_{ct} + \beta_5 Income_{ct} + \alpha_c + \eta_t + \epsilon_{ct}
\end{aligned} \tag{1}
$$

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 disaster response by the politician and the other control variables. A negative 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. SHELDUS 2009 (Version 7.0) is the data source for the weather damage. $Damage^*_{ct}$ includes missing observations. GR assume zero dollars in damage for all missing monthly SHELDUS observations when summing the damage across the six months prior to an election.

$Disaster_{ct}$ is the number of Presidential Disaster Declarations in the county during the six months prior to the election. 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]).

A Presidential Disaster Declaration provides federal assistance to repair public infrastructure. A Presidential Disaster Declaration also typically provides subsidized (Small Business Administration) disaster loans and cash grants (referred to as Individual Assistance) directly to 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 Presidential Disaster Declaration requests during the six months prior to the election. One important limitation of the disaster denial information in GR is that the exact counties considered in the denied requests are unknown. As such, all of the counties in a state have the same value for $Turndown_{ct}$. A negative coefficient estimate for β_3 is support the attentive electorate hypothesis. A negative coefficient is evidence that voters punish presidents for not providing assistance following a destructive weather event. 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 year 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.

GR is frequently cited as compelling evidence for an attentive electorate. GR has been cited at least 448 times (Google Scholar, November, 2022). 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).[5]

2.2 Replication of Gasper and Reeves [2011]

There are three significant limitations to the model estimated in Gasper and Reeves [2011]. Our reanalysis in Section 5 focuses on the severe non-random missing data problem inherent in using the SHELDUS damage information. Our replication in this section addresses two

[5] Appendix Table 1 shows that GR continues to be cited by papers published in top, general interest political science journals since 2014. The results have also been widely covered in the popular media, including by CNN, FiveThirtyEight, Solon, and The Washington Post, as recently as 2017.

other limitations.

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.[6]

The state-by-year spatial correlation in the approval of disaster requests will result in overly precise estimates for the coefficients in Equation 1, unless the spatial correlation is accounted for in the model (e.g. Moulton [1986]; Abadie et al. [2017]). We address spatial correlation by clustering the standard errors at the state-by-year 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 the original GR model, as well as the separate fixed effect and lagged vote share models in our replication.

In Table 1 column 1 we replicate GR's preferred presidential vote share model using the

[6]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 1 column 1) are likely miscoded.

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 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. The coefficients from this model (without allowing for spatial correlation) are the basis of GR's conclusion that the negative vote share impact of a natural disaster "is dwarfed by the response of attentive electorates to the actions of their officials" (p1).[7]

We show estimation results from the lagged vote share and county fixed effect specifications in Table 1 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 1 columns 4-6 estimate the same models as columns 1-3, except that we use updated weather damage information from SHELDUS 2018 (Version 16.0). The monthly damage estimates from SHELDUS 2009 (Version 7.0) are, by user agreement, not posted by GR and no longer available for purchase. We recreate the six month county-level weather damage variable using monthly information from SHELDUS 2018. We follow GR and assume that all missing observations incurred no damage when summing the monthly data across the six months prior to an election. Not surprisingly, the replication results are very similar regardless of which version of SHELDUS we use. The main primary source data are historical NWS-reported weather damage in *Storm Data*. There is no change to these data across the two versions of SHELDUS.

The reason we replicate GR using updated SHELDUS data is to verify that the model results are similar regardless of whether we use SHELDUS Version 7.0 or SHELDUS Version 16.0. Our reanalysis in Section 5 investigates how various assumptions over the missing data affect the model results. We need to use monthly SHELDUS data to conduct this reanalysis.

The bottom panel in Table 1 reports the number of disaster and denied disaster request (turndown) observations. We also list the number of disaster and turndown observations

[7] 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.

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). Natural disasters occurred in these counties, but based on the GR panel, caused no weather damage.

Table 2 shows replication results for a county-by-month panel using the updated SHELDUS data and the same models as in Table 1. Estimates from the original GR model with both county FE and lagged presidential vote share are very similar, regardless of whether we use the county-by-month panel or the aggregated yearly panel (Table 1, column 4). The same is true for the county fixed effects model. The fixed effects model, for reasons that will be clear in Section 5, is the focus of our reanalysis. Oddly, we estimate that weather damage has a positive impact on vote share in the lagged vote share model using the county-by-month panel. This finding is not consistent with the original GR findings or the other replication results.

2.3 Missing Data and Model Parameter Consistency

Equation 2 is a model for weather damage. $\boldsymbol{X_{cm}}$ are the independent variables from Equation 1 (except $Damage^*_{cm}$). $\boldsymbol{W_{cm}}$ are variables that are correlated with weather damage, but not in the economic model of interest. We include the dependent variable from our economic model as an explanatory variable. We do not presume that this model is correctly specified. In practice, the error term υ_{cm} will include important variables correlated with the amount of weather damage.

$$Damage_{cm} = \boldsymbol{X_{cm}\gamma_1} + \boldsymbol{W_{cm}\gamma_2} + \gamma_3 Y_{cm} + \upsilon_{cm} \tag{2}$$

Equation 3 is a selection equation for whether $Damage^*_{cm}$ is observed in the GR model using a county-by-month panel. $Damage^*_{cm}$ is observed and equal to $Damage_{cm}$ (from Equation 2) when $S_{cm} \geq 0$. $Damage^*_{cm}$ is missing when $S_{cm} < 0$.

The same variables from the damage model are included in the selection equation. The selection equation also potentially includes $\boldsymbol{Z_{cm}}$. These variables are excluded from Equation 2 because they are (conditionally) uncorrelated with actual damage, but included in Equation 3 as they predict whether $Damage^*_{cm}$ observations are missing. We do not assert that the selection equation includes all the independent variables that determine whether $Damage^*_{cm}$ is missing. The error term ζ_{cm} is likely to include key factors correlated with missingness.

$$S_{cm} = \boldsymbol{X_{cm}\omega_1} + \boldsymbol{W_{cm}\omega_2} + \omega_3 Y_{cm} + \boldsymbol{Z_{cm}\omega_4} + \zeta_{cm} \tag{3}$$

Missing data can broadly be categorized into one of three cases (e.g. Little and Rubin [2020]). The data are missing completely at random (MCAR) if the observations without missing data, the complete case sample, is a random subsample of the full sample. The data are missing at random (MAR) when the missing observations are random conditional on the other covariates. That is, missing $Damage^*_{cm}$ observations are independent of the level of $Damage_{cm}$ after conditioning on the right-hand-side variables in Equation 2. The MAR assumption is violated when $corr(\upsilon_{cm}, \zeta_{cm}) \neq 0$. Missing not at random (MNAR) is the third missing data case. The probability that $Damage^*_{cm}$ is missing is correlated with the level of $Damage_{cm}$, even after conditioning on the right-hand-side variables in Equation 2. The observations with complete data are a selected subsample and the selection mechanism is unknown.

In the county-by-month GR model there are missing observations for a single independent variable, $Damage^*_{cm}$. Whether or not $\hat{\beta}_1$ is a consistent estimate of the $Damage^*_{cm}$ parameter for the full population of US counties during the sample period depends on three things: the missing data case, if missingness is correlated with the dependent variable, and the researcher's choice on how to estimate the model.

2.3.1 Complete Case Subsample

Regression using Equation 1 and the complete case subsample of county-months provides consistent estimates of β_1 for the full population of county-months when, conditional on the other independent variables in Equation 1, data missingness is independent of Y_{cm} (e.g. Little and Rubin [2020], p49). Data missingness could be MNAR and estimates of β_1 will still be consistent. However, it not possible to verify that missingness is independent of Y_{cm} based only on the observed data.[8] Complete case analysis is also typically less efficient than if there were no missing observations and will generally lead to larger standard errors (e.g. White and Carlin [2010]).

2.3.2 Assigning Zeros to Missing Observations

In our survey of published papers that use SHELDUS, we find that more than half of the papers that estimate a regression model assume that counties incur no weather damage when

[8]External information about the institutional setting can offer support for assumptions over the missing data mechanism. Researchers have proposed tests to provide evidence for whether a missing data assumption is valid in some settings (e.g. Little [1988a]), but ultimately the assumption is unverifiable using only the observed data.

the observation is missing. In these papers, the authors assign zeros for missing $Damage^*_{cm}$ observations and then estimate the regression model on the full sample. We show in Section 3 that the assumption that missingness implies no damage is false.

Assigning missing $Damage^*_{cm}$ observations zeros affects the estimation of Equation 1 in three ways. First, $\hat{\beta}_1$ will be an inconsistent parameter estimate for the full population of US counties during this panel period when the model is estimated after assigning zeros. Second, the estimated variance of $\hat{\beta}_1$ will be underestimated (e.g. Little and Rubin [2020], p81). Third, parameter estimates for other independent variables can be biased (e.g. Carpenter and Kenwood [2013], p32).

2.3.3 Imputing Missing Observations using a Regression Model

Imputation of missing values using a regression model is standard practice in a number of research fields (e.g. Carpenter and Kenwood [2013]). We provide details on conventional multiple imputation and instrumental variables multiple imputation (e.g. Galimard et al. [2016]) methods in Section 4. In this section, we outline when multiple imputation will lead to consistent parameter estimates in Equation 1 for the population of interest. We assume that the full sample is randomly drawn from the population. We also assume that the economic model is specified correctly and that β_1 is consistently estimated for the full population of county-month observations using OLS when there is no missing data.

Conventional multiple imputation uses Equation 2 to impute missing $Damage^*_{cm}$ observations and then estimates Equation 1 on the full panel. Conventional multiple imputation will lead to a consistent estimate of β_1 in Equation 1 when missingnesss is MAR (e.g. Little [1992]). However, it is not possible to know whether the MAR assumption is valid based only on the observed data (e.g. Abayomi et al. [2008]).

Instrumental variables multiple imputation estimates both Equations 2 and 3 as part of the imputation procedure. $\boldsymbol{Z_{cm}}$ is the instrument. $\boldsymbol{Z_{cm}}$ is a valid instrument if the variables are correlated with whether $Damage^*_{cm}$ is missing, but uncorrelated with the level of $Damage_{cm}$ in Equation 2 (the exclusion restriction). The main advantage of this imputation approach, relative to conventional imputation, is that $\hat{\beta}_1$ is consistent when the data are missing not at random, provided $\boldsymbol{Z_{cm}}$ is a valid instrument. The reliability of the model's results does not depend on the assumption that the missing data mechanism is MAR as in conventional multiple imputation, nor that missingness is independent of Y_{cm} as in complete case analysis.

3 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 main primary source data in SHELDUS are from *Storm Data*, a the monthly publication by the National Centers for Environmental Information (formerly National Climatic Data Center). The first issue of *Storm Data* was published in 1959. Each issue includes estimated weather-related property and crop damage for a partial list of US counties. The weather information in *Storm Data* is voluntarily reported by regional NWS offices. All issues 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.[9]

Table 3 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-quarters of the SHELDUS database in the 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 3 considers "yearly" observations for each county. 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 of the six months during the year. Approximately, one-third of the counties contain only missing weather damage for the six months.

The table also provides initial evidence for why it is incorrect to assume that counties with missing weather damage incur no 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 4 uses administrative data from FEMA to confirm that missing weather damage information in SHELDUS should not be interpreted as no 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

[9] Digital copies of *Storm Data* can be accessed at the National Centers for Environmental Information website.

are authorized by 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.

We compare the county-specific FEMA disaster information to SHELDUS for counties with a Presidential Disaster Declaration. 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.[10]

3.1 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 NWS offices may be more likely to report damage information to the National Climatic Data Center when there is a large storm that causes damage to one or more counties in the reporting region. Table 3 suggests that this is indeed the case, as 88% of non-missing damage observations report positive damage.

Table 5 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 3 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. 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

[10]The latter portion of the overlap 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).

column adds NWS forecast zone fixed effects. We argue in the next section 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.

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 (p-value < 0.01 in all specifications), and 10 percentage point or 13% less likely to be missing the month before a Disaster Declaration (p-value < 0.01 in all specifications). The latter correlation is explained by the fact that there is sometimes a delay in declaring a disaster, and because the disaster may be part of a prolonged severe weather pattern. Second, missingness is correlated with demographics. Damage is less likely to be missing in counties with larger populations and in counties where residents have higher incomes (p-value < 0.01 in all specifications). Damage is less likely to be missing when there is a higher fraction of older residents (p-value < 0.01) and more likely to be missing when there is a higher fraction of African Americans (p-value < 0.05) in the specification that does not include county and NWS forecast zone fixed effects. 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.

The results in Table 5 show that, at a minimum, damage observations are not missing completely at random. In our view, the mostly likely missing data case is MNAR: the probability that SHELDUS observations are missing depends on the level of damage (even after conditioning on the independent variables in Equation 2). It seems reasonable to assume that SHELDUS observations are more likely to be missing when the actual level of weather damage is lower.

3.2 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.[11] 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 5 column 3 includes the NWS forecast office indicators 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

[11]We code the NWS forecast area indicators as follows. If the location of the NWS regional office is the same (between the 1st and 2nd, or 2nd and 3rd reorganizations), then we maintain the same name (indicator) for the office and simply assign the correct counties to that office's area of responsibility. If a new office is opened, then we create a new NWS indicator.

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 5 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 5 column 3). The F-statistic from a hypothesis test that each NWS indicator has an estimated coefficient equal to zero is 7.53.[12]

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 critical for the validity of the exclusion restriction because there is spatial 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 scale 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. This assumption is valid by construction if no counties switch NWS zones throughout the panel period.

The exclusion restriction could be violated if there are county-specific trends in weather damage for switching counties. If this were the case, the NWS zone indicators may correlate with county-level weather damage because the indicators partially reflect the historical change in damage that is not captured by the county fixed effects. However, we emphasize that the NWS reorganizations were an effort to improve the accuracy and dissemination of weather forecasts. Advancements in available technology, lengthy bureaucratic planning, and multi-year budget processes determined the timing of the two NWS regional office reorganizations. There is no mention of changing weather conditions as a factor in the reorganizations in the documents we reviewed (e.g. NWS [1985]; National Research Council [1991]; National Research Council [2012]). Moreover, we test for the role that county-specific trends have on the likelihood to report damage information in SHELDUS by adding county-specific linear trends to the specification in Table 5 column 3. There is no evidence for county-specific trends in the reporting of weather damage, after conditioning on overall time trends (year fixed effects), seasonality (month fixed effects), and persistent county risk (county fixed effects). The F-statistic for the null hypothesis that the county-specific trends are jointly equal to zero is very low (F = 1.45) in the specification that does not cluster the standard errors (and thus over-states statistical significance). Degrees of freedom restrictions prevent

[12] 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. The standard errors are clustered at the NWS zone-by-year level and allow for correlation of reporting both within zones and across months in the same calendar year (e.g due to a prolonged weather pattern). When we assume that all observations are independent the F-statistic is 22.17.

us from using our preferred NWS-by-year clustering for this test.

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, it is still important to control for geographic fixed effects since some geographic regions have higher underlying levels of HIV infection in the population.

4 Imputation of Missing Data

4.1 Conventional Multiple Imputation

Standard missing data imputation approaches assume that the missing data are MAR. A MAR assumption implies that the error terms from the assumed damage and selection models (Equations 2 and 3) are uncorrelated. The MAR assumption, barring researcher knowledge of the actual selection mechanism (e.g. a database rule on when to report or withhold information), is "inherently" untestable using observed data (Abayomi et al. [2008], p273). Imputation diagnostic checks can only provide evidence that the missingness assumptions are reasonable given the observed data, but can not rule out that the data are missing not at random (e.g. Nguyen et al. [2017]).

In our setting, there is a single independent variable with missing data ($Damage^*_{cm}$). Imputation of the missing data and estimation of the model of interest can be divided into four steps.[13] First, estimate Equation 2 on the subsample with no missing data. Calculate the 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 from step one, and then using Equation 2 to estimate an imputed value for each missing observation ($\widehat{Damage}_{cm}$). When imputing each observation, an error term is randomly drawn using the estimated distribution from the first step. Third, estimate the model of interest (Equation 1). Fourth, we repeat steps 2-3 one hundred times. We combine the one hundred estimates for each parameter using Rubin's rules (Rubin [1987]), which is essentially an averaging of the coefficients across each imputation.

[13] Rubin [1987] and Carpenter and Kenwood [2013] provide detailed discussions of multiple imputation. The imputation is run in Stata using *mi impute*.

4.2 Instrumental Variable Multiple Imputation

Economists are often skeptical of imputation as an approach to correct for missing data because of the need to make the strong and unverifiable assumption that the data are missing at random (e.g. Hirsch and Schumacher [2004]; Bollinger and Hirsch [2006]; DiNardo et al. [2006]). Underpinning this assumption is that the researcher has complete knowledge of the selection equation. The instrumental variables multiple imputation approach, by contrast, does not require complete knowledge of the selection equation. Instead, the credibility of the instrumental variables imputation procedure hinges on the validity of the exclusion restriction. Instrumental variable multiple imputation is a move towards a more "design-based" imputation approach that does not attempt to specify the "true" selection equation, but rather focuses on credible modeling of a single causal relationship (Angrist and Pischke [2017]).

The recent innovation in the applied statistics literature is to use a Heckman-style selection model as part of an imputation procedure to address missing data for an independent variable (e.g. Galimard et al. [2016]; Ogundimu and Collins [2019]; Gomes et al. [2020]). Galimard et al. [2016] demonstrate that imputation based on a two equation selection model can provide consistent estimates for the model of interest when the missing variable is missing not at random. Specifically, the estimation of a two equation selection model replaces the estimation of Equation 2 in conventional multiple imputation (the first step outlined in Section 4.1).

Heckman [1979] shows that potential bias from estimating Equation 2 on a selected sample can be reformulated as an omitted variable problem. Consistent parameter estimates for the full sample can be estimated using a two-step estimation procedure. First, estimate the selection equation on the full sample using a probit model and calculate the estimated Inverse Mills Ratio. Second, estimate a version of Equation 2 that also includes the estimated Inverse Mills Ratio.[14] The original Heckman [1979] formulation does not include $\boldsymbol{Z_{cm}}$ in the selection equation, and identifies the Inverse Mills Ratio from the assumed non-linearity of the selection equation. Modern, credible applications use the instrument $\boldsymbol{Z_{cm}}$ for identification (Vella [1998]).

We follow Galimard et al. [2016] and expand on their approach in two ways. First, the independent variable in our model of interest is allowed to be missing not at random (rather

[14] 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. Standard selection model assumptions include that $(\upsilon_{cm}, \zeta_{cm})$ is independent of $(\boldsymbol{W_{cm}}, \boldsymbol{Z_{cm}}, Y_{cm})$ with mean zero, $\zeta_{cm} \sim$ Normal(0,1), and linearity of the population regression of υ_{cm} on ζ_{cm}. Of note, we do not need to assume that the error term in Equation 2 is distributed normally (e.g. Wooldridge [2002], p562).

than the dependent variable).[15] Second, we estimate the selection equation using a fixed effects logit model. We then follow Lee [1982] to construct the Inverse Mills Ratio, even though we do not assume that the selection equation follows a normal distribution.[16]

An alternative and closely related instrumental variables imputation model uses a one-step maximum likelihood selection model estimator (e.g. Galimard et al. [2018]). 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.

Researchers have highlighted the difficulty in finding an instrument for the probability of missingness that satisfies a selection model exclusion restriction (e.g. DiNardo et al. [2006]; Bushway et al. [2007]). 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 instrument 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 the economic model. Moreover, the bounds from these methods are typically large and often make data inference impractical.

5 Reanalysis of GR Model

GR estimate their model using a county-year panel. Our reanalysis of the GR model uses a county-month panel. The reason is the data missingness is at the county-month panel. We estimate the GR fixed effects model on the complete case sample and on the full sample after imputing the missing $Damage^{*}_{cm}$ observations. It is infeasible to estimate the complete case sample using a county-year panel because fewer than one percent of the counties have

[15] Little and Rubin [2020] discuss the Heckman-style selection model as a method when the dependent variable is missing not at random, but write that the approach "could also be used to model a *predictor* variable with missing values" (p362) [emphasis added].

[16] Estimating a probit model with fixed effects suffers from the incidental parameters problem (Neyman and Scott [1948]; Lancaster [2000]). Lancaster [2000] 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). Another approach to circumvent this problem would be to estimate a probit model with unit fixed effects and then apply the Fernandez-Val and Vella [2011] incidental parameter bias correction (e.g. Fell and Kaffine [2018]).

non-missing weather damage reported in SHELDUS for each of the six months preceding an election (see Table 3). Imputation is likewise infeasible in the county-year panel when there are so few non-missing observations.[17]

Our reanalysis should be compared to the GR fixed effects model that uses the updated SHELDUS data and a county-month panel. We reproduce these estimates in Table 6 column 1. Recall that missing values are assigned zeros. None of the point estimates are statistically different from zero in column 1. We emphasize that these point estimates are very similar to those from the original GR model (Table 1 column 1). A key difference from the published GR results is that our standard errors reflect the spatial correlation in both the disaster damage and the political assistance. Our focus is on the county fixed effects model because estimates from a model with both county fixed effects and lagged dependent variables are inconsistent (Nickell [1981]), and because the validity of our exclusion restriction when using instrumental variables imputation is conditional on county fixed effects.

Model estimates for the complete case subsample are shown in Table 6 column 2. The point estimate for the weather damage variable is negative and twice as large in magnitude as that in column 1. However, the standard error is also twice as large. This reflects the fact that 75% of the observations have missing damage information and are dropped from the analysis. None of the coefficients in column 2 are statistically different from zero. Column 3 shows estimates for the full panel after we first impute the missing damage information using the conventional imputation approach. The point estimate for the weather damage variable is negative and statistically significant (p-value = 0.02).

Table 6 column 4 shows estimates for the full panel after we first impute the missing damage information using our instrumental variables imputation model.[18] The point estimate for the weather damage variable is negative and statistically significant (p-value = 0.02). The point estimate is slightly larger in magnitude than when we use conventional imputation, but we can not reject the null hypothesis that the two point estimates are the same. The point estimate implies a 0.46 percentage point reduction in the vote share for a county with the median level of damage. The natural experiment that tests how the electorate trades off weather damage and a politician's response to the damage is sharpest for counties that have a disaster declaration. There is an approximate 0.72 percentage point reduction in the vote share for a disaster county that sustains the median level of damage. The point estimates

[17]Another approach would be to impute the missing SHELDUS observations using a county-month panel, aggregate the (now complete) county data to the original county-year panel, and then estimate the economic model of interest. However, the imputation literature emphasizes that best practice 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).

[18]The Inverse Mills Ratio is highly statistically significant from zero (p-value < 0.001) in the imputation model.

for the political response variables are again imprecisely estimated.

Our reanalysis is in sharp contrast to the original GR conclusion that, on balance, the attentiveness of the electorate to actions of a politician outweighs the negative vote share response to random weather damage. We find no evidence that a politician's response affects vote share. The negative effect of weather damage on vote share is three times larger in our reanalysis when we use the instrumental variables imputation model, than in the same GR model that incorrectly assigns zeros for missing damage observations.

6 Discussion

SHELDUS is a very common database used by researchers across a number of fields who estimate models using historical weather damage information. The database is popular because there are few alternatives that provide the same spatial detail, observation frequency, and historical panel. However, SHELDUS suffers from an under-recognized, severe, and non-random missing data problem. We show that missingness is correlated with a number of factors. These include population demographics, the year of the observation, and whether there is a Presidential Disaster Declaration. The true selection model is unknown.

Researchers who wish to include weather damage information in their statistical model face a choice on how to proceed. The instrumental variables model in this paper is applicable to a wide range of research settings. Researchers in a number of fields can apply the same imputation procedure and reliably use SHELDUS. The model requires two key assumptions. These assumptions are more transparent and less restrictive than the assumptions required for alternative strategies to handle missingness.

First, the NWS field office reporting zone must predict whether a SHELDUS damage observation is missing, conditional on the other variables in the estimated selection model. This assumption is testable. Figure 2 plots the F-statistics for a test of the null hypothesis that the NWS zone indicators are jointly statistically different from zero using a very general linear probability model for SHELDUS missingness. The model includes month, year, county, and NWS field office fixed effects as independent variables. We vary the starting year of the panel. The longest panel is 1970-2016 and the shortest panel is 2000-2016. We show results when we do not cluster the standard errors (left side) and when we cluster at the NWS zone-by-year level (right side). The dashed vertical lines indicate the start year for three panels: the longest panel (1970), a panel that excludes years before the first NWS zone restructuring (1985), and a panel that only includes years after the second restructuring (1997). The NWS field office reporting zone is highly predictive of missingness for nearly all panel lengths. For example, the F-statistics for the 1970, 1985, and 1997 start date panels: 95.1, 75.9, and 17.1

(without clustering), and 10.3, 6.7, and 12.2 (clustering zone-by-year).

Second, the NWS office reporting zone must be uncorrelated with the actual level of damage and excludible from the economic model, after conditioning on county fixed effects. The reason that the NWS zone indicators must be excluded from a researcher's model of interest is because the independent variables from the model of interest are also included as regressors in the damage imputation model (e.g. Equation 2). Thus, our instrumental variable imputation procedure is not applicable to research questions where the NWS zone is a variable of interest in the economic model. However, researchers using SHELDUS have historically not estimated models that consider the NWS offices as an independent variable. We are not aware of any existing study. Our instrumental variable imputation model would also be invalid if the NWS office indicators correlate with disaster damage, after controlling for the more spatially precise county fixed effects. We view the violation of this assumption as highly unlikely (see Section 3.2).

We discuss three other approaches to handling the missing data. More than half of the published papers reported on the SHELDUS website that use the data in a regression model incorrectly assume that missingness implies zero dollars in damage. The weather damage parameter estimate from a regression model that uses the complete SHELDUS sample for a panel period, after assigning zeros for missing values, will be inconsistent for the population parameter. The parameter estimate will often be biased towards zero.[19] The standard error for the damage coefficient will be underestimated. As such, models that use SHELDUS and assume that missing damage implies no damage are likely to underestimate the role of weather damage.

Two standard approaches to handle missing data are to use the complete case subsample, or to impute missing values using conventional multiple imputation. These approaches require the researcher to make strong and unverifiable assumptions regarding the missing data. Estimation of a regression model on the complete case subsample will provide a consistent weather damage parameter estimate for the full population of county-months when, conditional on the other independent variables in the regression model, data missingness is independent of the dependent variable in the model of interest. Conventional multiple imputation will lead to a consistent weather damage parameter estimate when missingness is MAR.

[19] Assigning missing values is a type of measurement error. A common assumption is that measurement error attenuates the value of the estimated slope coefficient in a regression model. This is not always the case (e.g. Loken and Gelman [2017]). However, our simulations that vary the missingness mechanism, and effect size and sign of the true slope coefficient, suggest that the estimated damage coefficient is often biased towards zero.

7 Conclusion

This paper makes three main contributions. First, we document that SHELDUS, a widely used weather damage database, suffers from a severe and poorly understood nonrandom missing data problem. Second, we use an instrumental variables multiple imputation procedure to account for the missing data. The instrument in our model is based on the historical restructuring of the NWS office geographic weather reporting boundaries. We show that the instrument is applicable to a wide range of research settings. Researchers in a number of fields can apply the same imputation procedure to continue to (reliably) use SHELDUS data. Third, our reanalysis of Gasper and Reeves [2011] underscores how the handling of missing data impacts research conclusions. GR is a highly influential study on retrospective voting. The original finding in GR, that voter attentiveness to the actions of a politician outweighs reaction to random events, is reversed when we use our instrumental variables multiple imputation procedure to account for the missing data.

8 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.

Toward a new national weather service, a first report. Technical report, Committee on National Weather Service Modernization, National Research Council, March 1991.

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.

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.

Dave Donaldson and Adam Storeygard. The view from above: Applications of satellite data in economics. *Journal of Economic Perspectives*, 30(4), 2016.

Winston Wei Dou, Leonid Kogan, and Wei Wu. Common fund flows: Flow hedging and factor pricing. *NBER Working Paper*, (30234), July 2022.

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.

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.

Justin Gallagher and Daniel Hartley. Household finance after a natural disaster: The case of hurricane katrina. *American Economic Journal: Economic Policy*, 9(3), 2017.

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

Shan Ge. How do financial constraints affect product pricing? evidence from weather and life insurance premiums. *Journal of Finance*, 77(1), 2021.

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.

Matthew H. Graham, Gregory A. Huber, Neil Malhotra, and Cecilia Hyunjung Mo. Irrelevant events and voting behavior: Replications using principles from open science. *Journal of Politics*, Forthcoming.

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. 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.

Lung-Fei Lee. Some approaches to the correction of selectivity bias. *The Review of Economic Studies*, 49:355–372, 1982.

Roderick J. A. Little. A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, 83, 1988a.

Roderick J. A. Little. Missing-data adjustments in large surveys. *Journal of Business and Economic Statistics*, 6, July 1988b.

Roderick J. A. Little. Regression with missing x's: A review. *Journal of the American Statistical Association*, 87, 1992.

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.

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

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.

Ian R. White and John B. Carlin. Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. *Statistics in Medicine*, 29, 2010.

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

Replication of Gasper and Reeves (2011)

Panel: **County-by-Year**						
Damage Data:	**SHELDUS 2009**			**SHELDUS 2018**		
Specification:	**GR Replication**	**Lagged Vote Share**	**County Fixed Effects**	**GR Replication**	**Lagged Vote Share**	**County Fixed Effects**
	(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
Observations	27,894	27,894	27,894	27,894	27,894	27,894
Disaster Obs.	3,132	3,132	3,132	3,132	3,132	3,132
Disaster Obs. with Damage = 0	1,017	1,017	1,017	687	687	687
Turndown Obs.	4,698	4,698	4,698	4,698	4,698	4,698
Turndown Obs. with Damage = 0	2,343	2,343	2,343	1,765	1,765	1,765
R-squared	0.816	0.793	0.415	0.816	0.793	0.415

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: FEMA, Public Entity Risk Institute, SHELDUS, US Decennial Census.

Table 2: **Effect of Severe Weather and Disaster Assistance on Incumbent Presidential Vote Share**

Replication of Gasper and Reeves (2011)

Panel: **County-by-Month**
Damage Data: **SHELDUS 2018**

Specification:	**GR Replication** (1)	**Lagged Vote Share** (2)	**County Fixed Effects** (3)
Weather Damage	-0.045	0.097	-0.020
	(0.020)	(0.056)	(0.028)
Disaster Declaration	0.576	1.789	0.411
	(0.397)	(0.970)	(0.536)
Turndown	-0.874	-2.862	-0.729
	(0.571)	(1.827)	(0.816)
Lagged Vote Share	X	X	
County Fixed Effects	X		X
Income	X	X	X
Year Fixed Effects	X	X	X
Observations	167,148	167,148	167,148
R-squared	0.815	0.405	0.415

The table shows replication results for a county-by-month panel using the same models as in Table 1 columns 4-6. Standard errors that allow for state-by-year spatial correlation are in parentheses. Data sources: FEMA, Public Entity Risk Institute, SHELDUS, US Decennial Census.

Table 3: **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: SHELDUS.

Table 4: **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: FEMA, SHELDUS.

Table 5: **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	-0.205
	(0.038)	(0.038)	(0.038)
Disaster Declaration Next Month	-0.134	-0.115	-0.097
	(0.035)	(0.037)	(0.038)
Size of Disaster (No. Counties)	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)
Turndown	0.032	0.039	0.045
	(0.028)	(0.027)	(0.027)
Demographic Variables			
Median Income (log)	-0.070	-0.036	-0.093
	(0.020)	(0.032)	(0.022)
Population Size (log per 1,000)	-0.040	-0.088	-0.061
	(0.003)	(0.029)	(0.019)
African American Population (%)	0.001	0.002	0.001
	(0.000)	(0.002)	(0.001)
Older Population (%)	-0.005	-0.001	-0.002
	(0.001)	(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	0.066	0.136	0.173
Observations	167,124	167,124	167,124

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 4. Standard errors that allow for state-by-year spatial correlation are in parentheses. Data sources: FEMA, NWS, Public Entity Risk Institute, SHELDUS, US Decennial Census.

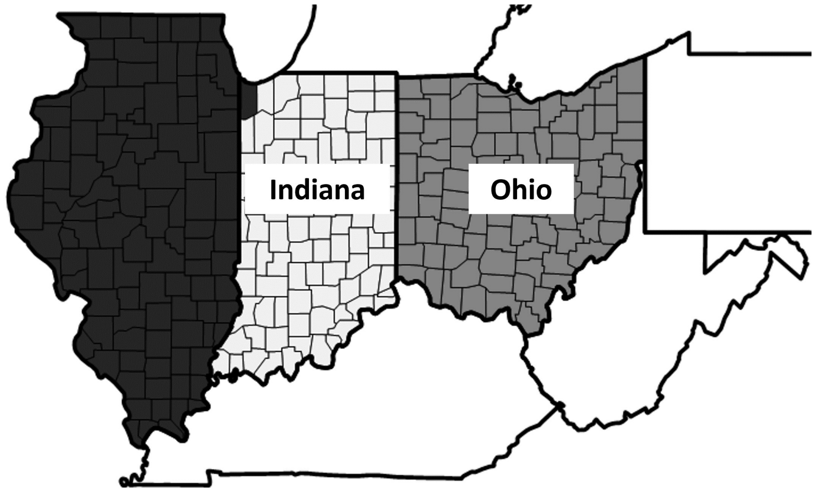
Table 6: **Effect of Severe Weather and Disaster Assistance on Incumbent Presidential Vote Share**

Reanalysis of Gasper and Reeves (2011)

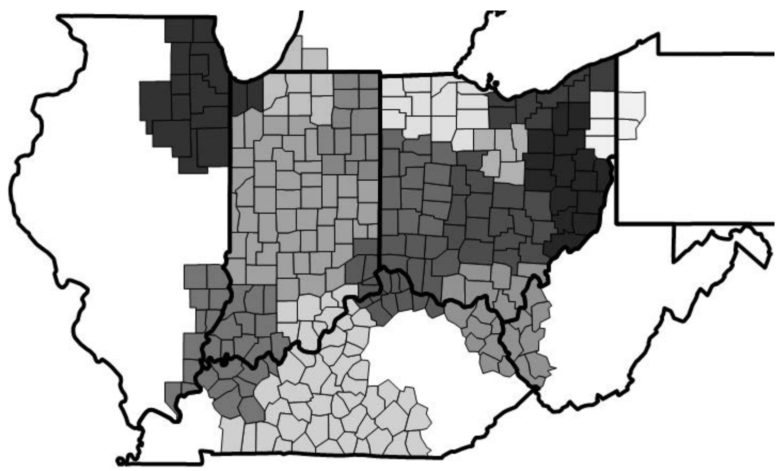
Missing Data Approach:	Assume Zeros (1)	Complete Case (2)	Conventional Imputation (3)	Instrumental Variables Imputation (4)
Weather Damage	-0.020	-0.047	-0.058	-0.062
	(0.028)	(0.066)	(0.026)	(0.025)
Disaster Declaration	0.411	0.103	0.546	0.552
	(0.536)	(0.613)	(0.535)	(0.619)
Turndown	-0.729	-1.369	-0.756	-0.757
	(0.816)	(1.105)	(0.818)	(0.817)
Income	X	X	X	X
County Fixed Effects	X	X	X	X
Year Fixed Effects	X	X	X	X
Observations	167,148	41,656	167,132	167,124

All models use a county-month data panel and SHELDUS 2018 (Version 16.0) damage information. Standard errors are clustered at the state-by-year level. Column 1 reproduces the GR fixed effects model results from Table 2 column 3. Missing damage observations are imputed as zeros. Column 2 estimates the same model on the complete case subsample. Column 3 imputes the missing damage observations using the conventional multiple imputation approach. Column 4 imputes the missing damage observations using the instrumental variables imputation approach discusses in Section 4.2. Data sources: FEMA, NWS, Public Entity Risk Institute, SHELDUS, US Decennial Census.

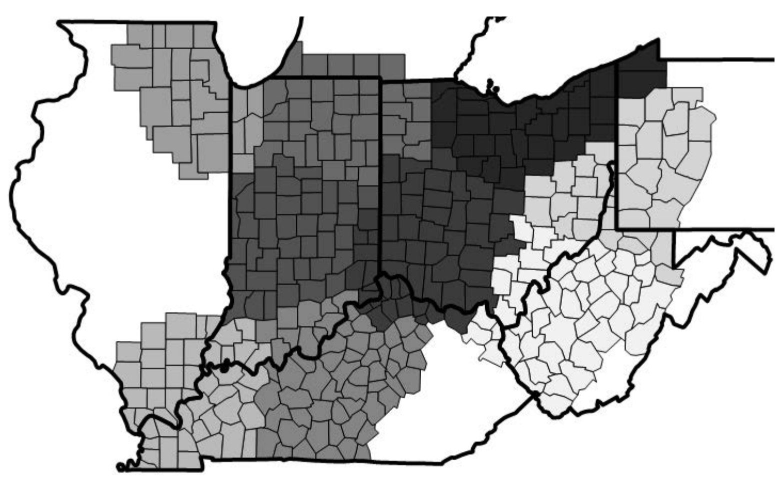
Figure 1: **Historical NWS Forecast Zones for Indiana and Ohio**



1972 - July 1984



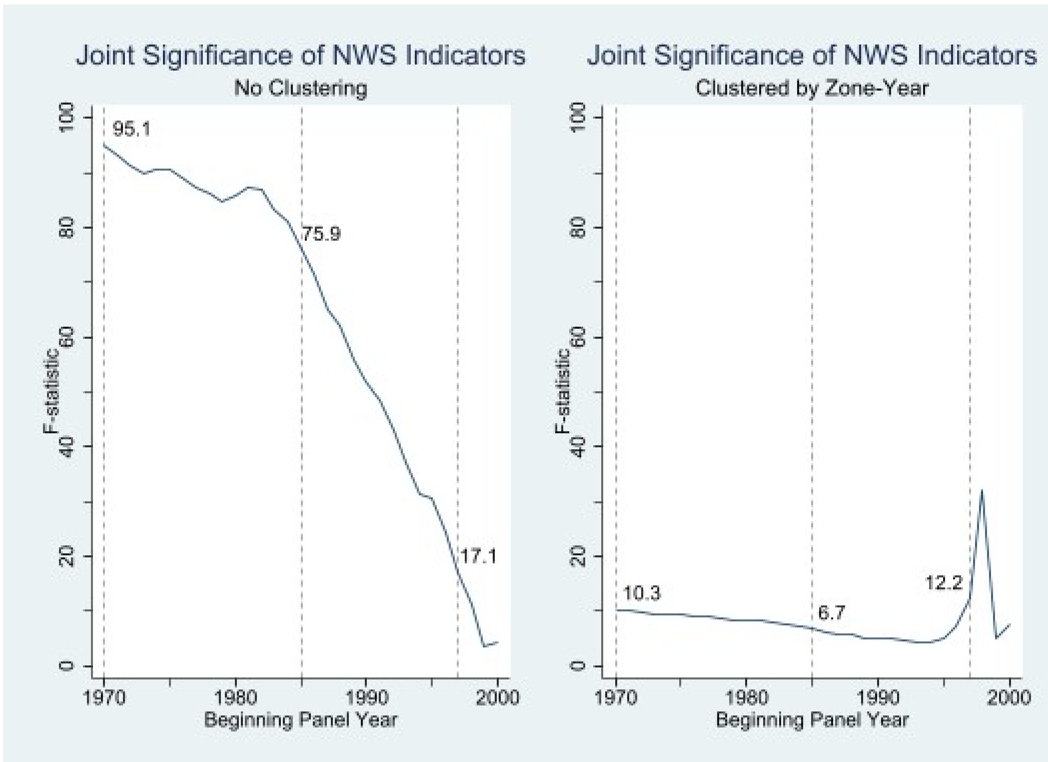
August 1984 - 1996



1997 - 2004

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). Data source: NWS.

Figure 2: **SHELDUS Panel Length and NWS Zone Instrument Strength**



The figure shows the F-statistic for joint significance of the NWS zone indicators in a model that includes month, year, county, and NWS field office fixed effects as independent variables. We plot the F-statistics for varying panel lengths. The longest panel is 1970-2016 and the shortest panel is 2000-2016. We show results when we do not cluster the standard errors (left side) and when we cluster at the NWS zone-by-year level (right side). The dashed vertical lines indicate the start year for three panels: the longest panel (1970), a panel that excludes years before the first NWS zone restructuring (1985), and a panel that only includes years after the second restructuring (1997). Data source: NWS.