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

Justin Gallagher
Montana State University

Vanderbilt University, April 7, 2023

Ceci n'est pas un désastre

Source: WHAS11 (EF3 Crawford Tornado, Indiana, May 30, 2004)

Missing Data is Prevalent in Applied Economic Research

- Economists frequently work with large databases built by combining information from a number of separate sources
  - Increases the likelihood that variables suffer from missing data
- Almost all researchers have experiences with missing data issues
  - For example, Abrevaya and Donald (2017) survey papers from 4 top economics journals from 2006-8 and find that almost 40% estimate models with data missingness
- The decision on how to handle observations with missing data is critical for the reliability of a model's results

1st Contribution of this Paper

We show that the Special Hazard Events and Losses Database for the United States (SHELDUS) suffers from a severe and poorly understood missing data problem

- SHELDUS is a monthly county-level (1960-present) weather damage database very popular with researchers
  - Economics (e.g. Barrot et al., QJE, 2016; Dou et al., NBER WP, 2022)
  - Finance (e.g. Bernile et al., JF, 2017)
  - Political Science (e.g. Gasper and Reeves, AJPS, 2011)
- We show that missing SHELDUS obs are frequent and not missing at random
  - Approx. 75% obs missing
  - Missingness correlated with a number of factors
  - Failure to account for the missing data will likely lead to inconsistent model estimates
- Many researchers are unaware of this nonrandom missing data problem
  - For example, Ge (JF, 2021) falsely asserts: "the data set includes every natural hazard event that caused injury, death, or property/farm damage since 1960 in the U.S."

2nd Contribution of this Paper

We demonstrate how an instrumental variables multiple imputation procedure can account for missing data even when missingness is nonrandom

- The method doesn't rely on perfect knowledge for why data are reported (selection equation), and thus avoids a common criticism of data imputation (e.g. DiNardo et al., 2006)
- In essence, the method applies the "design-based" econometric modeling approach of the *Credibility Revolution* (Angrist and Pischke, 2017) to data imputation
- First economics study to use a design-based imputation procedure (to our knowledge)
- We follow the recent applied statistics literature (Galimard et al., 2016; Gomes et al., 2020)

3rd Contribution of this Paper

Reanalyze a seminal study on voting behavior that uses weather damage from SHELDUS

1. **Retrospective Voting**
   - Large empirical literature examines how voters evaluate political performance (e.g. Anderson, 2007; Healy and Malhotra, 2013)
   - A key topic in the literature is whether voters distinguish between random events (naive retrospection) and the political response to these events (voters are "attentive")
   - Early empirical literature focuses on how the electorate responds to economic conditions
   - Gasper and Reeves (2011) among the first to use weather damage and the political response to the damage as a quasi-experiment
   - Gasper and Reeves (2011) find: the negative vote share impact of a natural disaster "is dwarfed by the response of attentive electorates to the actions of their officials"
2. **We Show**
   - The findings of Gasper and Reeves (2011) are mostly reversed once accounting for missing data using our instrumental variables multiple imputation procedure

Plan for Rest of Talk

1. Demonstrate the missing data issue with SHELDUS
2. Present the Gasper and Reeves (2011) model
3. Show how an instrumental variable multiple imputation procedure accounts for the missing data in SHELDUS
4. Reevaluate Gasper and Reeves (2011)

Missing SHELDUS Weather Damage Information

We will show 3 things:

1. Missing SHELDUS data is prevalent
2. Missing data are not completely random
3. National Weather Service (NWS) weather reporting structure can be used as an instrument in a selection model to predict if observations are missing

SHELDUS Background

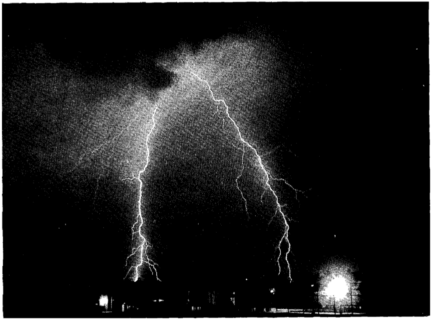
- SHELDUS is a loss and hazard database currently maintained by the Center for Emergency Management and Homeland Security at Arizona State University
- The weather damage information in SHELDUS are from *Storm Data*:
  - *Storm Data* is a monthly publication by the National Centers for Environmental Information (formerly National Climatic Data Center)
  - Includes a list of weather-related deaths & injuries, and estimated damage costs for counties in each issue
  - The weather damage information is **voluntarily reported** by local NWS offices and each issue only includes a partial list of US Counties
  - 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**"

Example Issue of *Storm Data*

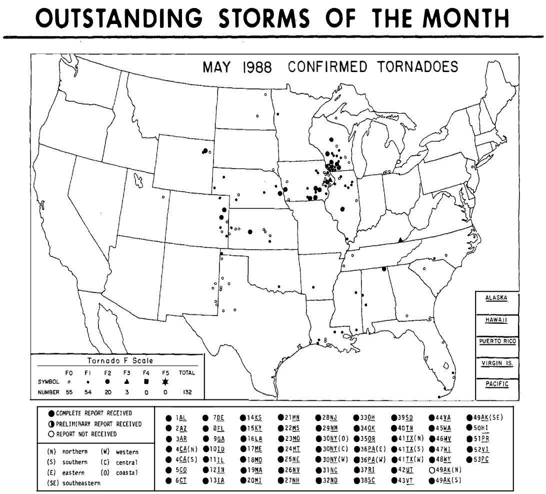
MAY 1988 VOLUME 30 NUMBER 5

STORM DATA

DEPARTMENT OF COMMERCE UNITED STATES OF AMERICA



Example Issue of *Storm Data*



Example Issue of *Storm Data*

PLACE	DATE	TIME – LOCAL STANDARD	LENGTH OF PATH (MILES)	WIDTH OF PATH (YARDS)	NO. OF PERSONS		ESTIMATED† DAMAGE		CHARACTER OF STORM
					KILLED	INJURED	PROPERTY	CROPS	
TENNESSEE									
Knox County	10	1230CST			0	0	4	0	Lightning
	Lightning struck numerous power transformers throughout the county, doing serious damage to some of them and leaving thousands of people without electricity for up to six hours. Estimates of the damage were $6000.00.								
Marian County, Jasper	14	1355CST			0	0	2	0	Hail (1.75)
McMinn County	14	1359CST			0	0	1	0	Hail (0.75)
Hamilton County, Suck Creek	14	1421CST			0	0	2	0	Hail (1.75)
Meigs County	14	1500CST			0	0	1	0	Hail (1.00)
	Thunderstorms produced large hail in a number of communities in the four county area during the afternoon.								
Hickman County, Bucksnort	16	1426CST			0	0	0	0	Hail (1.75)
Perry County, Linden	16	1440CST			0	0	1	0	Hail (0.75)
Clay County	16	1601CST			0	0	2	0	Hail (1.75)
Livingston County, Alpine	16	1645CST			0	0	1	0	Hail (0.75)

Legend Notes: "1" $<$ \$50; "2" \$50 to \$500; "3" \$500 to \$5,000; "4" \$5,000 to \$50,000; "5" \$50,000 to \$500,000; "6" \$500,000 to \$5 Mil; "7" \$5 Mil to \$50 Mil; "8" \$50 Mil to \$500 Mil

Missing Weather Damage Data in SHELDUS is Prevalent

Missing data statistics for a sample of county-month observations from May-October 1972-2004 during US Presidential election years (matches Gasper and Reeves, 2011)

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%)

1. Demonstrates magnitude of the missing data problem: 75% of the obs missing!
2. Initial evidence that missing data are not the same as no damage: $0 is reported

FEMA Disaster Information Confirms Missing SHELDUS Observations are Not Equivalent to $0 Damage

1. We collected data via a FOIA and FEMA's website on county-level disaster damage verified by FEMA as part of grant assistance following Presidential Disaster Declarations

2. We have FEMA data for just under half of the Presidential Disaster Declaration county-months in the SHELDUS sample (skewed towards more recent months)

3. **40% of the disaster counties in this FEMA-SHELDUS overlap sample have missing data in SHELDUS**
   (1) Missing SHELDUS data likely more prevalent for Disaster counties in earlier years

   (2) We are not able to directly gauge missingness for smaller weather damage events, but seems likely that non-reported weather damage is even greater when there is not a Disaster Declaration

SHELDUS Damage is Not Missing Completely at Random

- Missing damage information in SHELDUS does not imply no damage
- Still, purpose of *Storm Data* is to promulgate information related to weather events
- 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
- A linear probability model shows a number of factors are correlated with the likelihood that SHELDUS damage information is missing
  (1) We did not attempt to examine a comprehensive list of potential variables that might correlate with missingness
  (2) Goal is to simply show that missingness is not completely random

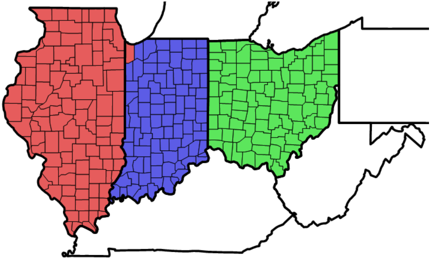
SHELDUS Damage is Not Missing Completely at Random

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

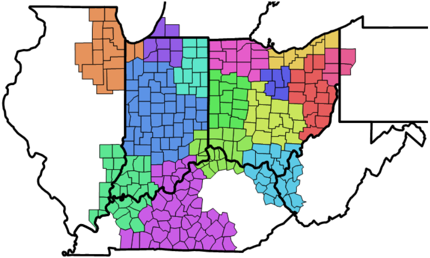
NWS Forecast Zones as Exogenous Predictors for Missing Damage

- NWS is responsible for forecasting weather in the US
- NWS relies on a decentralized organizational structure that includes regional NWS offices and forecast zones
- 2 major NWS structural reorganizations impacted the reporting of weather damage from 1972-2004
  1. **1972 - July 1984**:
     - 52 Weather Service Forecast Offices
     - Each office responsible for forecasting weather and reporting on weather conditions for a specific area of responsibility
  2. **Aug 1984 - 1997**:
     - Local forecasts and severe weather information now primary responsibility of approx. 200 Weather Service Offices (WSOs)
     - Each WSO covers a collection of counties (median is 11 counties)
  3. **1997 - 2004**:
     - Referred to as "modernization and restructuring"
     - Reduced the number of WSOs to approx. 130 (median is 23 counties)

Historical NWS Forecast Zones for Indiana and Ohio

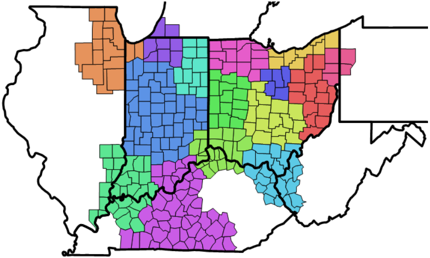


1972 - July 1984

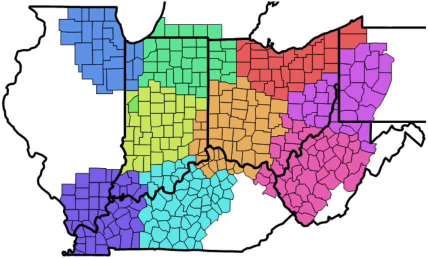


August 1984 - 1996

Historical NWS Forecast Zones for Indiana and Ohio (cont.)



August 1984 - 1996



August 1997 - 2004

NWS Forecast Zones as Exogenous Predictors for Missing Observations

- The decentralized and changing structure of the NWS offices is the basis of our instrument
- We create a vector of NWS indicator variables
- To act as a valid instrument in the instrumental variable multiple imputation estimation procedure the NWS indicators must:

(1) Predict if a weather observation is missing in SHELDUS

- A simple LPM that *only* includes the NWS indicators has a R-squared = 0.093, which explains 50% more of the variation than a specification with the 8 weather/demographic variables, year FE, and month FE
- Adding the NWS to a specification that already includes these variables and county FE still explains substantially more variation in the missing weather data (F-stat = 7.53)

(2) Be uncorrelated with the actual level of weather damage in a county

- Assumption is that there is no correlation between weather damage and the NWS indicators after including county FE in the model
- NWS zones are a collection of counties; County FE control for any fixed geographic differences in weather damage do to topography, weather patterns etc.

Literature on Retrospective Voting

- Models of voting behavior often assume that the electorate is retrospective (e.g. Key, 1966; Wittman, 1989; Persson, Roland, Tabellini, 1997)
- Large empirical literature in both economics and political science examines how voters evaluate political performance and react to different types of information (e.g. Healy and Malhotra, 2013)
- Early empirical literature on retrospective voting focuses on how the electorate responds to economic conditions (e.g. Anderson, 2007)
- Problems with this approach include:
  (1) Tenuous link between political actions & economic performance
  (2) Challenge that economic conditions are not randomly assigned
- Whether voters hold incumbents responsible for random events (*naive* retrospection), or for the political response (*attentive* retrospection) is a key topic in the literature

Gasper and Reeves (2011)

- Gasper and Reeves (2011)–hereafter GR–use weather damage as the basis of a quasi-experimental research design
- 1st goal: Overcome 2 shortcomings in the literature
  (1) Provide clear link between (proxies for) political action & timing of a negative event
  (2) Weather damage from a natural disaster, unlike economic conditions, can (mostly) be thought of as a random negative event
- 2nd goal: Separately compare how a negative random event and the endogenous political response affect presidential two-party vote share
- GR is a seminal study in the literature
  - Frequently cited as evidence for an attentive electorate (Google Scholar: 448 cites)
  - Spawned sub-literature on retrospective voting that examines natural disasters
  - Widely covered in popular press

GR Econometric Model

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

- Panel: 1972-2004 (9 presidential election years; 6 underlying months each year)
- y_{ct}: vote share for incumbent president's party in county c, year t
- $Damage^*_{ct}$: log weather damage in 6 months before election
  - $\rightarrow$ Includes missing obs; GR assume missing are \$0
- $Disaster_{ct}$: number of Presidential Disaster Declarations in 6 months before election
- $Turndown_{ct}$: denied Disaster requests in 6 months pre-election
  - $\rightarrow$ GR only know if there is a denied Declaration during a month for a state
  - $\rightarrow$ GR code <u>all</u> counties in the state that year as having a turndown
  - $\rightarrow$ On average, only about 9% of a state's counties are included in a turndown
- Control variables: lagged and twice lagged vote share, income, county and year FEs
- Standard errors <u>do not</u> allow for spatial correlation

Limitations of Gasper and Reeves (2011)

- There are 3 significant limitations to the model estimated by GR
- Our focus is on the severe, non-random weather damage missing data problem
- However, we also address 2 other limitations and a modest data coding issue:
  1. Spatial Correlation: We cluster SE's at state-by-year
     - Correlation in weather damage is greater between counties affected by the same disaster
     - Correlation in a Presidential Disaster Declaration requests
  2. Model includes FE and Lagged DV: We run separate models
     - Coefficient estimates inconsistent when both FE and lagged DV (Nickell, 1981)
     - Estimate a model with either FE or lagged DV (Angrist and Pischke, 2008)
  3. Coding mistakes in GR panel: We correct these
     - 1,852 repeated observations (6% of panel)
     - Approx. 5% of Disaster observations miscoded

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

1. Coefficient point estimates all have the predicted sign
2. None are statistically significant at conventional levels after allowing for spatial correlation (SE's are 3-7 times smaller if we do not cluster)

Replication of Gasper and Reeves (2011): County-by-Month Panel

Panel: **County-by-Month**		
Damage Data: **SHELDUS 2018**		
Specification:	**GR Replication**	**County Fixed Effects**
	(1)	(2)
Weather Damage	-0.045	-0.020
	(0.020)	(0.028)
Disaster Declaration	0.576	0.411
	(0.397)	(0.536)
Turndown	-0.874	-0.729
	(0.571)	(0.816)
Lagged Vote Share	X	
County Fixed Effects	X	X
Income	X	X
Year Fixed Effects	X	X
Observations	167,148	167,148
R-squared	0.815	0.415

- We switch to a county-by-month panel in our reanalysis
  (1) Monthly panel facilitates a complete case analysis
  (2) Impute using a monthly panel; Best to estimate model of interest with monthly panel

Missing Data and Model Parameter Consistency

Model for Weather Damage:

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

Selection Equation for if $Damage^*_{ct}$ is observed:

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

- $Damage^*_{cm}$ observed and equal to $Damage_{cm}$ when $S_{cm} \geq 0$ (missing when $S_{cm} < 0$)
- <u>Independent Variables</u>
  - $\boldsymbol{X_{cm}}$: RHS variables from economic model (Eq 1), except $Damage^*_{cm}$
  - $\boldsymbol{W_{cm}}$: Variables correlated with weather damage, but not in economic model
  - $\boldsymbol{Z_{cm}}$: Variables included in Eq 3 because they predict if $Damage^*_{cm}$ is missing; Excluded from Eq 2 because (conditionally) uncorrelated with actual damage

    $\rightarrow$ NWS indicators
- Importantly, we do not assume perfect knowledge of either Eq 2 or Eq 3

Categorizing Missing Data

Missing data can broadly be categorized into 3 cases

1. **Missing Completely at Random (MCAR)**: complete case sample is a random subsample of the full sample
2. **Missing at Random (MAR)**: missing obs are random conditional on covariates
   - Implies $Damage^*_{cm}$ is independent of the level of actual damage ($Damage_{cm}$) after conditioning on RHS variables in weather damage model (Eq 2)
   - MAR violated when $corr(\upsilon_{cm}, \zeta_{cm}) \neq 0$
3. **Missing Not at Random (MNAR)**: probability that $Damage^*_{cm}$ is missing is correlated with the level of $Damage_{cm}$ (even after conditioning on RHS vars in Eq 2)
   - In other words, obs with complete data are a selected subsample and the selection mechanism is unknown
   - Importantly, the selection mechanism is almost never known and cases 1 and 2 require an assumption that the researcher knows the selection mechanism

Missing Data and the GR Model

Whether or not $\hat{\beta}_1$ is a consistent estimate of the $Damage^*_{cm}$ parameter in the economic model (Eq 1) for the full population of US counties depends on 3 things

1. The missing data case
2. If missingness is correlated with the dependent variable
3. Researcher's choice on how to estimate the model
   (1) Use complete case subsample
   (2) Assume values for the missing obs (often assigned 0)
   (3) Impute missing obs using a regression model

Estimating the GR Model using a Complete Case Subsample

Researcher Choice 1: Use Complete Case subsample

Provides consistent estimates of β_1 for the full population of county-months when, conditional on the other independent variables in Eq 2, missingness is independent of Y_{cm}

- Missing data case (i.e. MCAR, MAR, or MNAR) doesn't matter

  $\rightarrow$ Advantage: Selection Model assumed by the researcher doesn't matter

- Researcher must assume that missingness is independent of Y_{cm}

  $\rightarrow$ Disadvantage: Not possible to verify that missingness is independent of Y_{cm} based only on the observed data

- Missing data are so common in SHELDUS that using a complete case sample negates the advantages of the database (long panel, all US counties, frequent obs)

  $\rightarrow$ Disadvantage: Larger SEs; May need to alter model to exclude FEs, etc.

Estimating the GR Model by Assuming Zeros

Researcher Choice 2: Assign $0 to Missing Observations

- Very common assumption for SHELDUS
  - → $>$ 50% of the published papers that estimate a regression model assume counties with missing obs incurred no damage
  - → GR assume missing SHELDUS obs are $0
- Assigning missing $Damage^*_{cm}$ obs 0's:
  - (1) $\hat{\beta}_1$ will be inconsistent estimate for the full population of US counties during the sample period in the Economic Model
  - (2) Estimated variance of $\hat{\beta}_1$ will be underestimated
- Models will often report "statistically significant" results that are wrong

Estimating the GR Model via Conventional Multiple Imputation

Researcher Choice 3: Conventional Multiple Imputation

- Standard practice in a number of research fields
- Conventional multiple imputation uses the Damage Model (Eq 2) to impute missing $Damage^*_{cm}$ obs, and then estimates the Economic Model on the full sample
  - → Leads to consistent estimates of β_1 under MAR assumption
  - → Disadvantage: Not possible to know whether MAR assumption is valid based only on observed data
- Economists often skeptical of multiple imputation because of the MAR assumption (e.g. DiNardo et al., 2006)
  - → MAR assumption can be recharacterized as a concern over Omitted Variable Bias

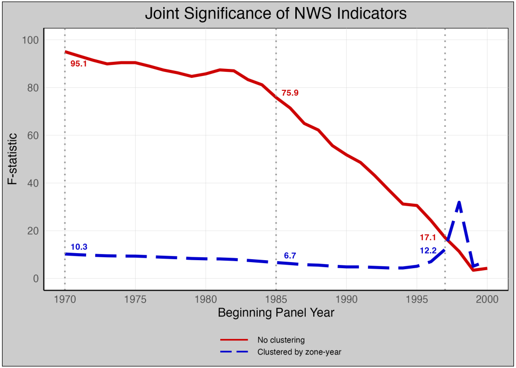
Estimating the GR Model via Instrumental Variables Multiple Imputation

New Researcher Choice 3: Instrumental Variables Multiple Imputation

- Uses both Equations 2 and 3 as part of the imputation procedure
  (1) Estimate Selection Model on full sample
  (2) Calculate the estimated Inverse Mills Ratio
  (3) Impute using estimates of Equation 2 that include the Inverse Mills Ratio
- Heckman (1979) shows that this 2-step estimation procedure can lead to consistent estimates for a model when the *dependent variable* has missing values
- Recent applied statistics innovation embeds this 2-step approach as part of an imputation process for an *independent variable* (e.g. Little and Rubin, 2020)
- Approach valid under any missing data case
  → Addresses the concern over the MAR assumption
- "Design-based" (Angrist and Pischke, 2017) approach consistent with the *Credibility Revolution* in economics

Our Instrumental Variables Model is Applicable to Most Research Settings

- Researchers in a number of fields can apply the same imputation model and reliably use SHELDUS
- Model requires 2 key assumptions
- **Assumption 1 (testable)**:
  NWS field office reporting zone must predict if damage obs is missing



Model includes: month, year, county, and NWS fixed effects. Standard errors clustered at NWS zone-by-year level. Panels run 1970-2016 to 2000-2016.

Our Instrumental Variables Model is Applicable to Most Research Settings

- **Assumption 2**
  NWS office zones: (i) uncorrelated with the actual level of damage and (ii) excludible from the economic model, after conditioning on county fixed effects

Our Instrumental Variables Model is Applicable to Most Research Settings

- **Assumption 2**
  NWS office zones: (i) uncorrelated with the actual level of damage and (ii) excludible from the economic model, after conditioning on county fixed effects

- Part (i) assumes no correlation between weather damage and the NWS indicators beyond the geographic correlation captured by the county fixed effects
  - → Could be violated if a county's underlying historical weather damage risk changes, and the timing of the NWS reorganizations is correlated with this change
  - → We view a violation of part (i) as extremely unlikely:
    1. NWS reorganizations done to improve the accuracy and dissemination of weather forecasts
    2. Advancements in technology, lengthy bureaucratic planning, and multi-year budgeting determined the timing of the reorganizations
    3. No mention of changing weather conditions as a factor in timing or setting of new zones
    4. Moreover, we test for if county-specific time trends are correlated with reporting damage and easily reject F-test

Our Instrumental Variables Model is Applicable to Most Research Settings

- **Assumption 2**
  NWS office zones: (i) uncorrelated with the actual level of damage and (ii) excludible from the economic model, after conditioning on county fixed effects

- Part (i) assumes no correlation between weather damage and the NWS indicators beyond the geographic correlation captured by the county fixed effects
  - → Could be violated if a county's underlying historical weather damage risk changes, and the timing of the NWS reorganizations is correlated with this change
  - → We view a violation of part (i) as extremely unlikely:
    1. NWS reorganizations done to improve the accuracy and dissemination of weather forecasts
    2. Advancements in technology, lengthy bureaucratic planning, and multi-year budgeting determined the timing of the reorganizations
    3. No mention of changing weather conditions as a factor in timing or setting of new zones
    4. Moreover, we test for if county-specific time trends are correlated with reporting damage and easily reject F-test

- Part (ii) requires the NWS indicators to be excluded from a researcher's model of interest
  - → However, we are not aware of any existing study that uses SHELDUS and includes the NWS zone as an independent variable

Gasper and Reeves (2011) Reanalysis

Make it Rain? Retrospection and Attentive Electorate in the Context of Natural Disasters (Gasper and Reeves, 2011, AJPS)

1. We reanalyze Gasper and Reeves (2011)
2. Goal is to evaluate how the their results change after accounting for the missing SHELDUS data

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

1. CC, and imputation reanalyzes suggest the impact of *Weather Damage* on vote share is 2 to 3 times as large as that reported in GR
2. *Disaster Declaration* and *Turndown* estimates are similar to GR
3. Instrumented and conventional imputation lead to similar estimates, suggesting that vote share does not influence if weather damage is missing

Conclusion

This paper makes 3 contributions:

1. We show that a popular weather damage database, SHELDUS, suffers from a severe, non-random missing data problem
2. Demonstrate how an instrumental variables multiple imputation procedure can account for missing data
   - Key advantage: the researcher doesn't need full knowledge of the selection equation
   - We leverage the historical regional office structure of the NWS as an instrument to impute missing SHELDUS obs
   - Our imputation procedure allows researchers studying almost any topic to continue to (reliably) use SHELDUS
3. Reanalyze Gasper and Reeves (2011), a seminal study on voting behavior that relies on SHELDUS
   - Findings are mostly reversed after accounting for missing data
   - Evidence in favor of naive retrospection and no evidence in favor of an attentive electorate

Thanks!

The paper and appendix can be found at `justinpgallagher.com`