# Imputing for Extraordinary Sample Events

## *A Story of Targeted Donor Pools and Administrative Data*

**Bradley Rhein**
Supervisory Statistician
Current Employment Statistics, State and Area
Federal Committee on Statistical Methodology
October 25, 2023

# Disclaimer

- *Any opinions expressed in this paper are those of the author and do not constitute policy of the Bureau of Labor Statistics*

# Overview

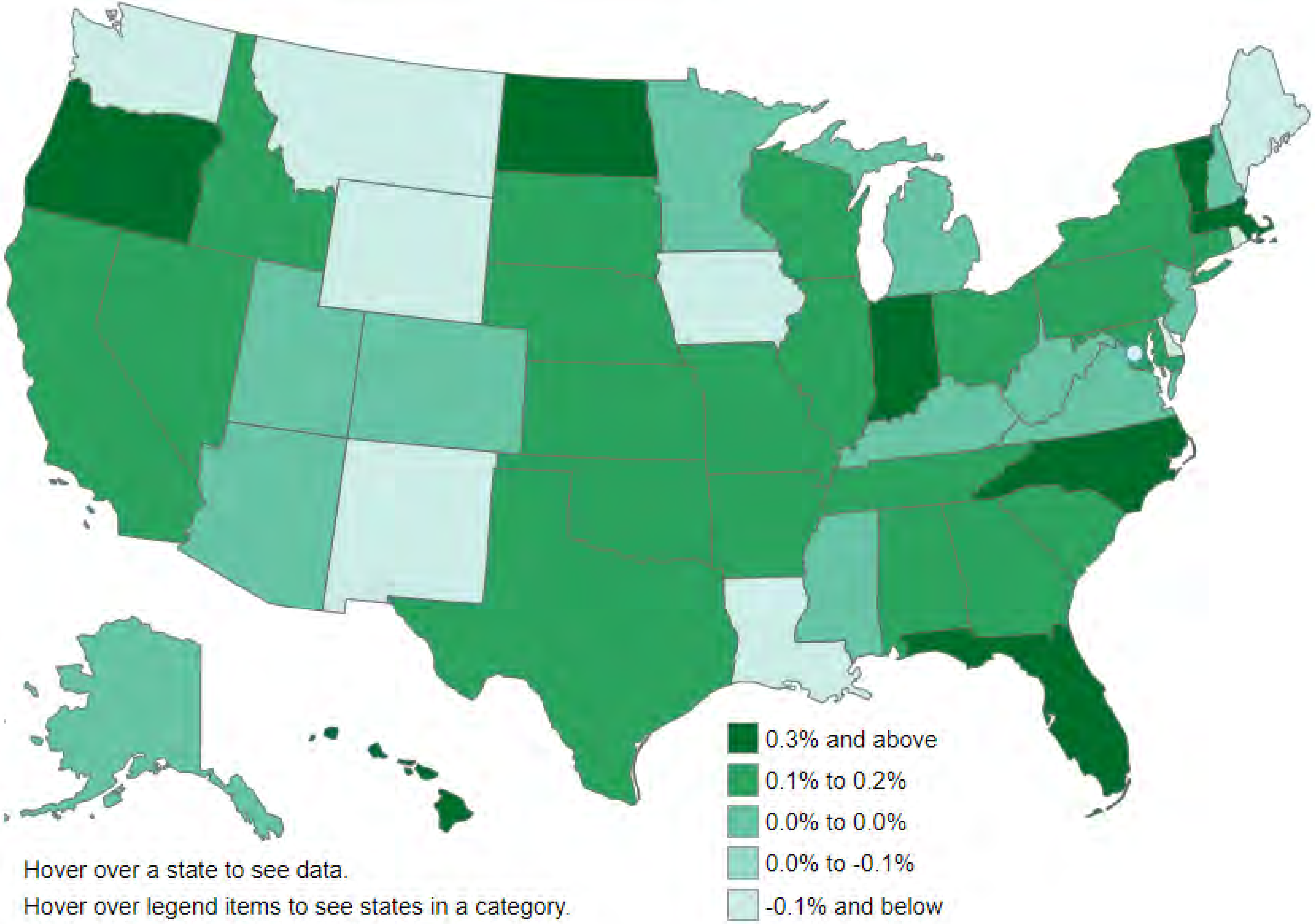
- Introducing the Current Employment Statistics (CES) program
- Extraordinary Events and current methods
- Setting up the imputation option
- Imputation methods of interest
- Examples and results
- Conclusion
- Future work

# CES State and Area

- State (50+ DC, VI, PR), area (450 metropolitan), industry detail (as the sample allows) for non-farm payrolls
  - Employment
  - Hours/Earnings
- Two closings (preliminary and finals)
- September preliminary estimates published on October 20
- Benchmarked data releases in mid-March, updates employment levels up to September of the previous year

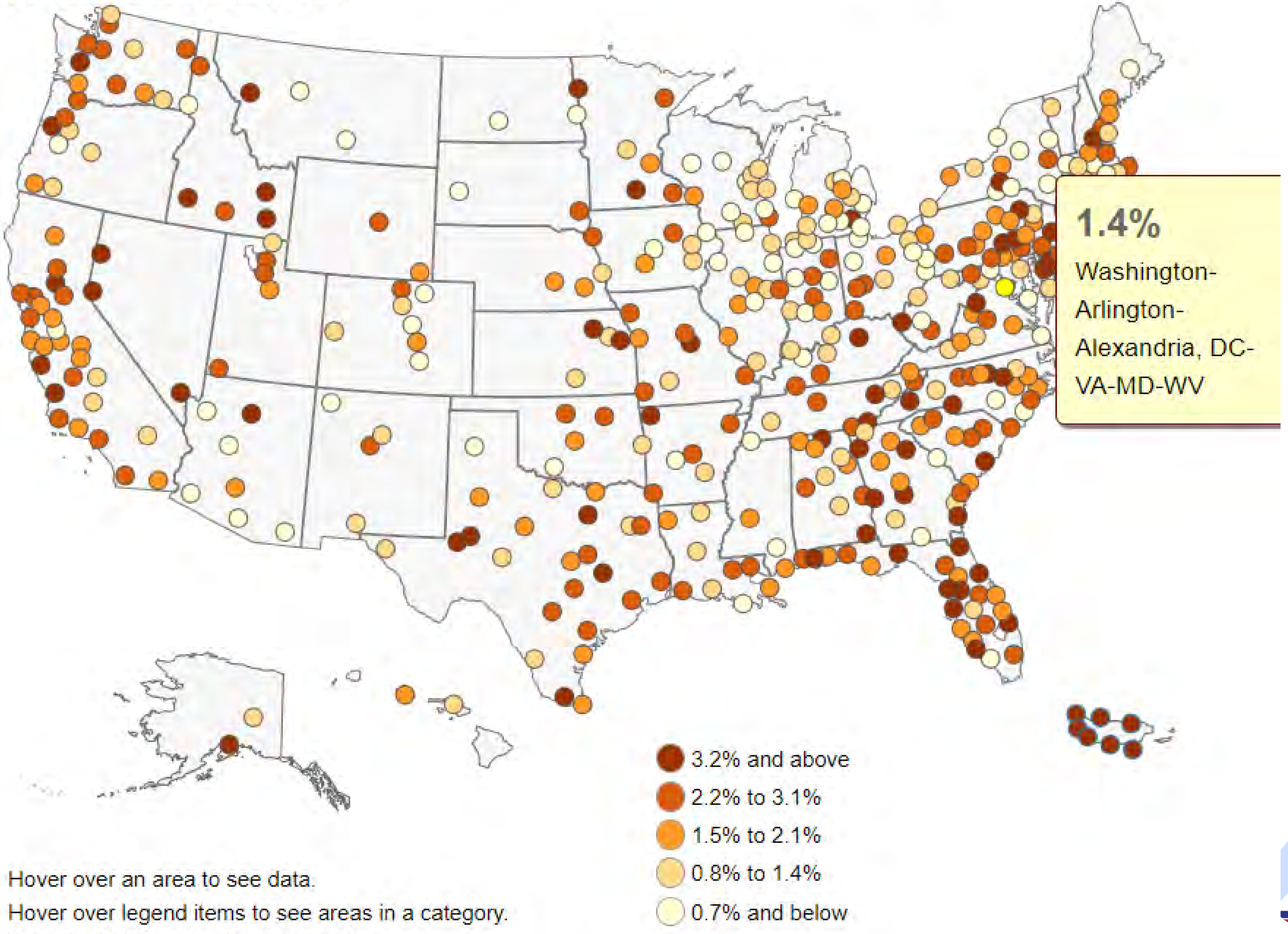
# One-month percent change in industry employment by state, June 2023 to July 2023, seasonally adjusted

Choose an industry: Total nonfarm




Hover over a state to see data.
Hover over legend items to see states in a category.
Source: U.S. Bureau of Labor Statistics.

## Change in nonfarm employment by metropolitan area, August 2022 to August 2023, not seasonally adjusted




Hover over an area to see data.
Hover over legend items to see areas in a category.
Source: U.S. Bureau of Labor Statistics.

# CES Sample Design

- Frame: Quarterly Census of Employment and Wages (QCEW)
- Stratified simple random sample of 122,000 UI accounts, sampling rates by optimum allocation
  - Strata are defined by state, industry (13, defined by NAICS), and employment size (8 size classes)
  - Certainty units
- Probability of selection is n/N for any UI, within a strata
- Design weights are the inverse probability of selection

# CES State and Area Estimators

- Sample-based
- Model
- Features: outlier detection/adjustment, birth-death forecast, non-sample events, seasonal adjustment, and key non-respondent (KNR) imputation

# Quite Extraordinary

# List of Extraordinary Events

- **Hurricanes**
- Tornados
- Wildfires
- Flooding
- Snowstorms
- Extreme heat?
- Other…

# Current ways of capturing effects

- Monitor response rates

- Monitor microdata received before closing (and re-calculating estimates near the deadline to maximize response)

| Sunday | Monday | Tuesday | Wednesday | Thurs | Friday | Saturday |
|---|---|---|---|---|---|---|
| Week 1 | | | | Calculate estimates | | |
| Week 2 | | | Closing of estimates | | | |

# Current ways of capturing effects (2)

- Use the model feature that isolates certain characteristics in the grouping mechanism (determine isolated groups by response rates or disaster zones, etc)
- Possibly rely on the sample data more (or entirely) in model situations

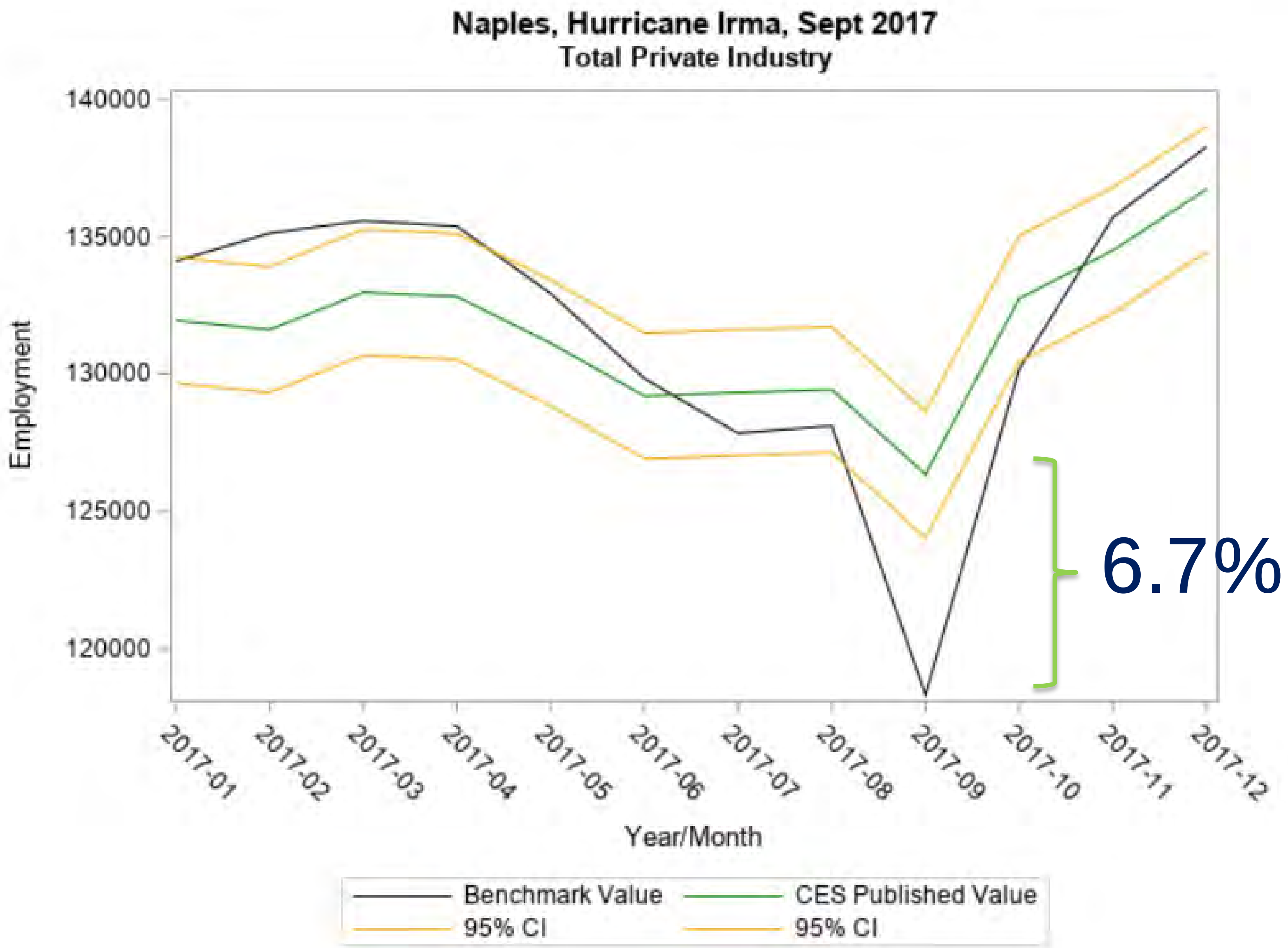
# Commentary on the current methods…

- Current methods do not directly address the non-response due to extraordinary events
  - Some non-response could be assumed to be not missing at random (MAR), but missing due to the extraordinary event
  - Using the relationship between a current and past circumstance by targeting non-respondents and donor pools

- Turning to imputation as a possible solution!

# Not Missing-at-Random Clarifications

- Assuming that, at minimum, the act of submitting data has been affected by the extraordinary event
- Employment *may or may not* have been affected by the event
  - Establishment could be closed
  - Or not able to operate
  - Or everyone was evacuated
  - But still paying employees
  - Or could be open but too busy to respond


Naples, Hurricane Irma, Sept 2017
Total Private Industry
140000
135000
130000
125000
120000
Employment
2017-01
2017-02
2017-03
2017-04
2017-05
2017-06
2017-07
2017-08
2017-09
2017-10
2017-11
2017-12
Year/Month
Benchmark Value
CES Published Value
95% CI
95% CI
6.7%

# An Imputation Option

- Goal: to impute for non-respondents possibly affected by the extraordinary event using donors from a similar circumstance
- Carefully define the “recipient” – a non-respondent that will receive an imputed value
- Thoughtfully determine which donor pool to use
- Use imputed values to calculate an estimate adjustment

# Recipients: The Usuals

- Assuming The Usuals are establishments that would have submitted data if the disaster had not occurred
  - Establishments that reliably ("usually") submit data for the first closing
  - Not known to have significant reporting errors in the past

# Possible donor source: the QCEW

- Administrative data for employment, hours, and earnings via unemployment insurance reporting (on a 6-month lag, about 95% coverage of US jobs)
  - Monthly, county-level, full NAICS designations, ownership
- Some concerns – revisions, errors in reporting, non-covered employment, other
- However, still very useful source of information and we have access to the microdata!

# Targeted Donor Pool Assumptions

- Assuming past event's effects will be similar to a current event
  - Example: effects of Hurricane Irma (2017) approximate the effects of Hurricane Ian (2022)
    - Area size/location
    - Distance from landfall
    - Hurricane magnitude
    - Industry make-up
    - Geographic details

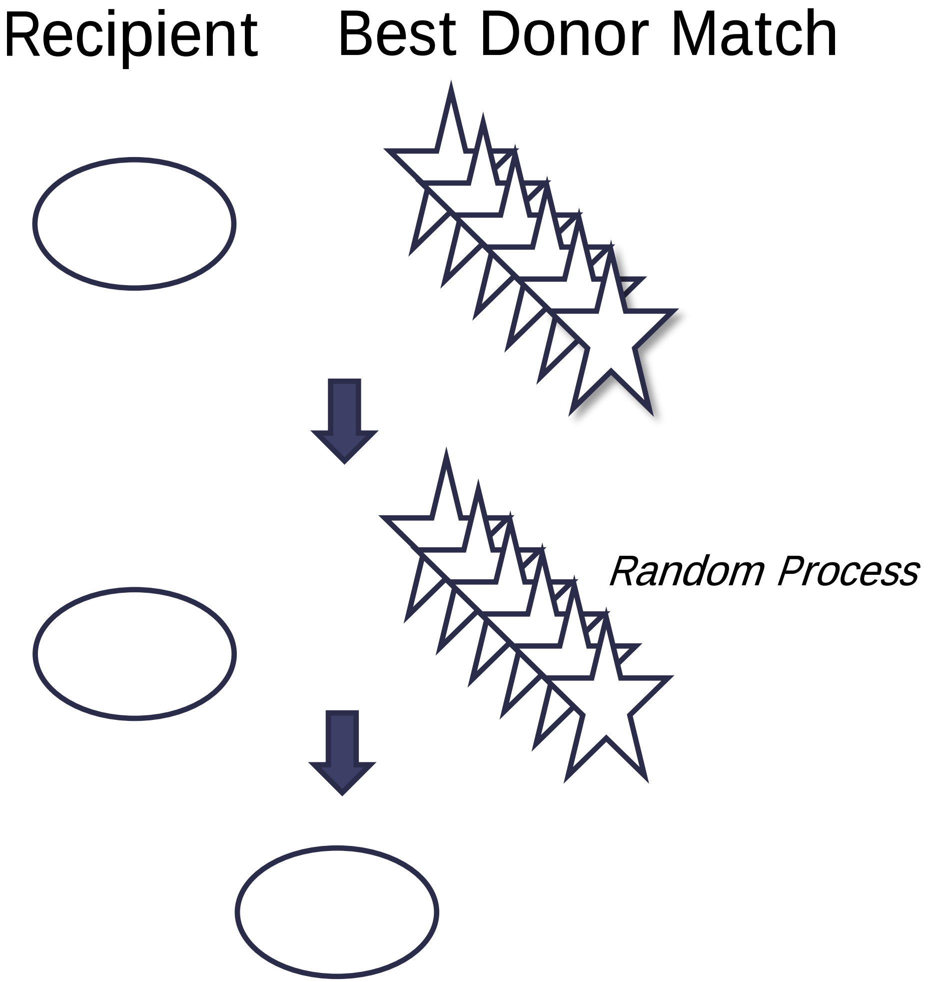
# Hurricane Donor Pool Examples

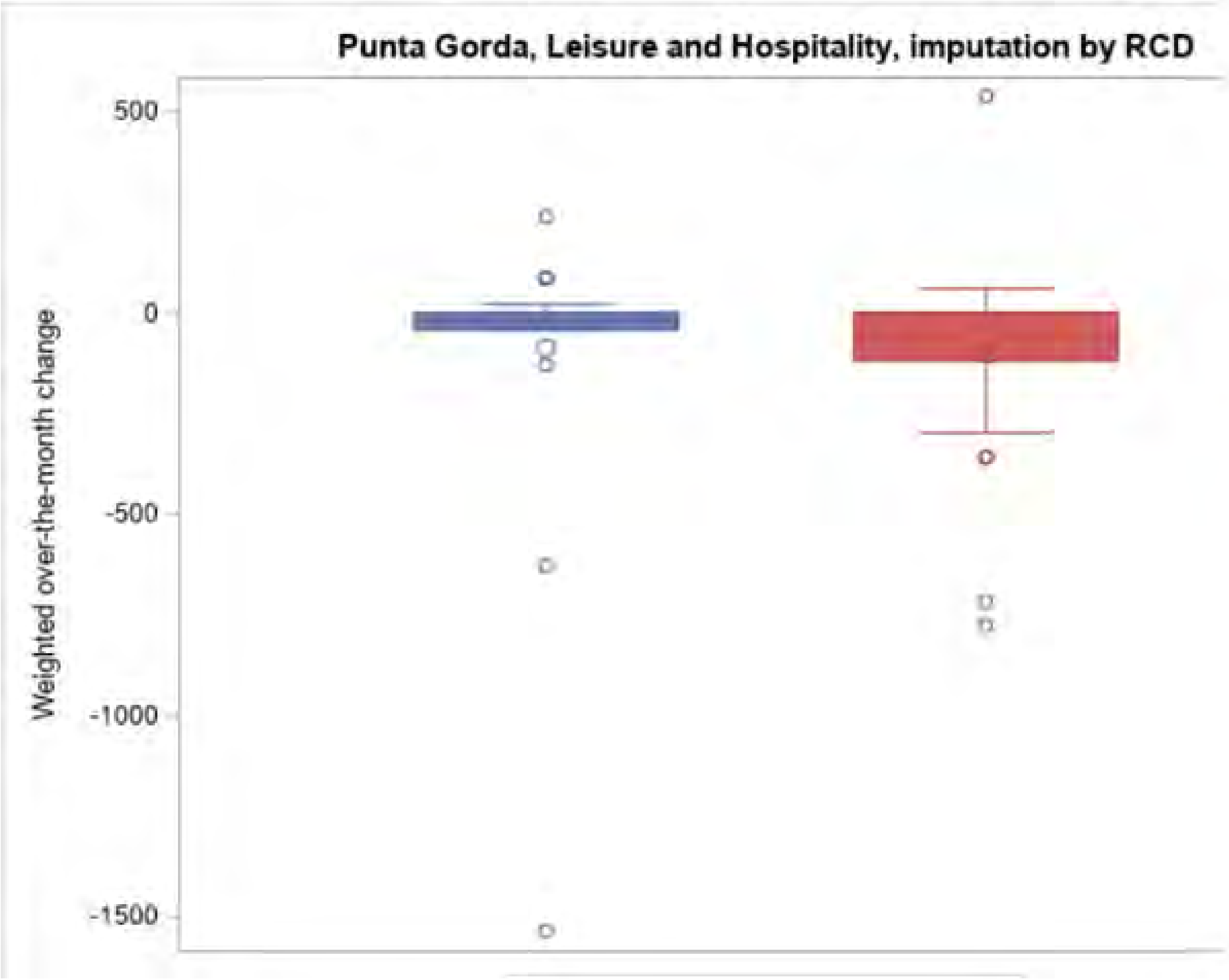
| Donor Pool | Hurricane | Category | Approx. Emp. (thousands) |
|---|---|---|---|
| Naples | Irma | 4 | 165 |
| FL | Irma | 4 | 9668 |
| | | | |
| Corpus Christi | Harvey | 5 | 195 |
| Houston | Harvey | 5 | 3350 |
| Victoria, TX | Harvey | 5 | 41 |
| Beaumont, TX | Harvey | 5 | 160 |
| Lafayette | Harvey | 5 | 204 |
| | | | |
| Panama City | Michael | 5 | 88 |
| Pensacola | Michael | 5 | 196 |
| Tallahassee | Michael | 5 | 195 |
| | | | |
| Jacksonville | Florence | 1 | 800 |

# Imputation Methods of Interest

- Since we have detailed, plentiful administrative data, not that interested in modeling or using complicated methods to determine donor values
  - Random cold deck (RCD) - drawing from QCEW, with multiple imputations, specifying area, 6-digit NAICS, establishment-level employment size
  - Nearest neighbor (NN), similar specifications, with ties determined randomly and limiting donors to one use

# Illustration of nearest neighbor ties

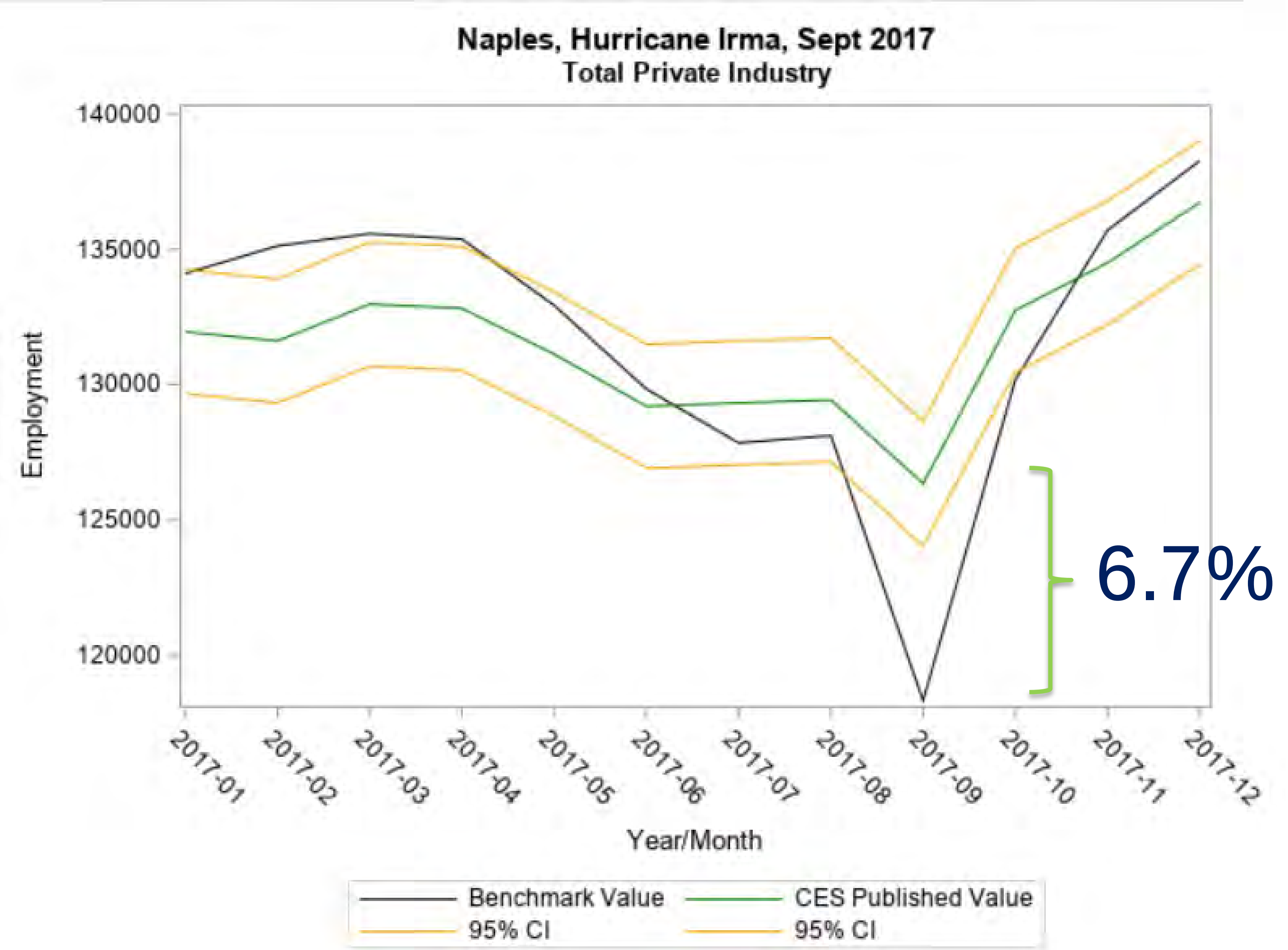


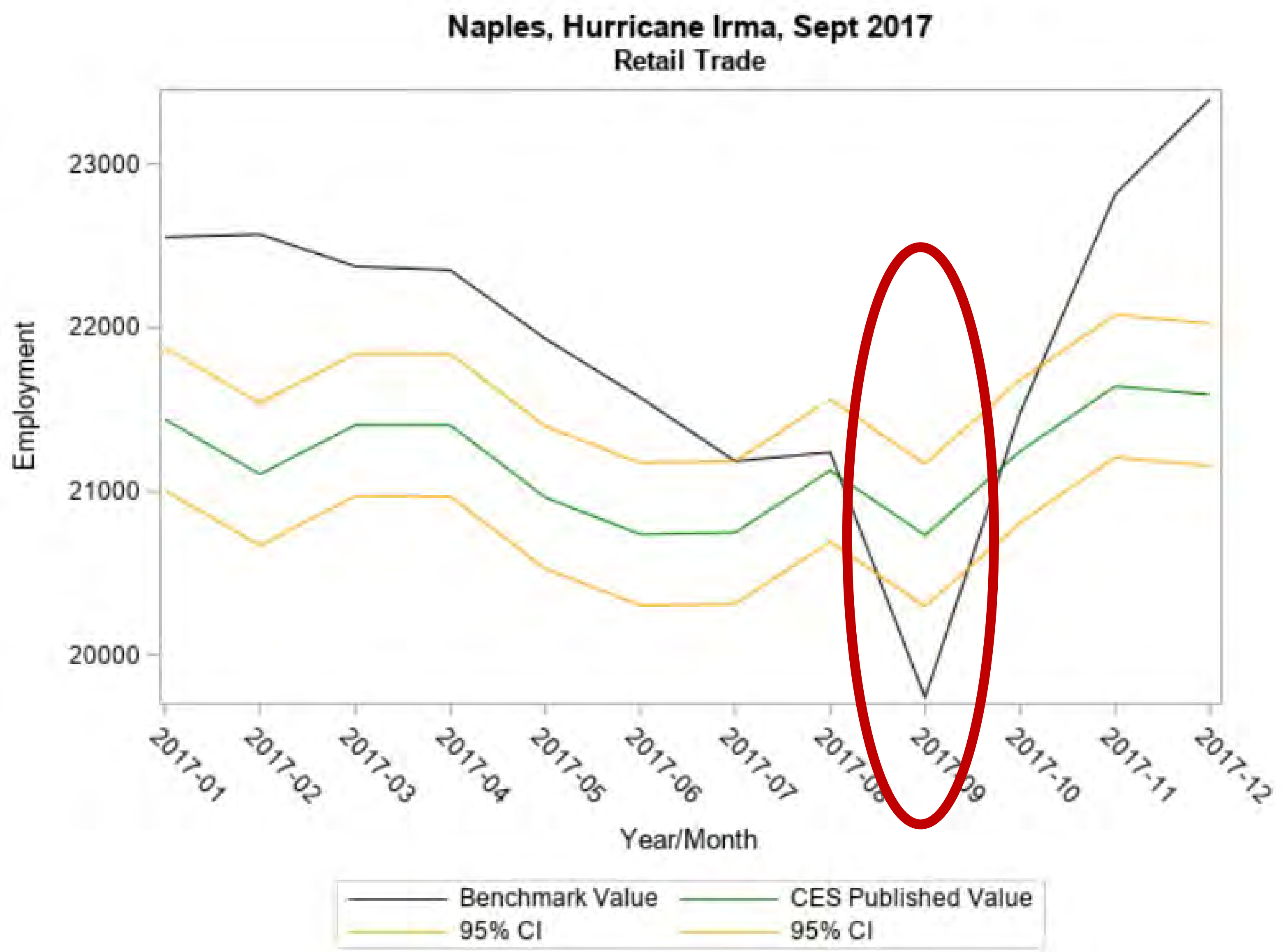


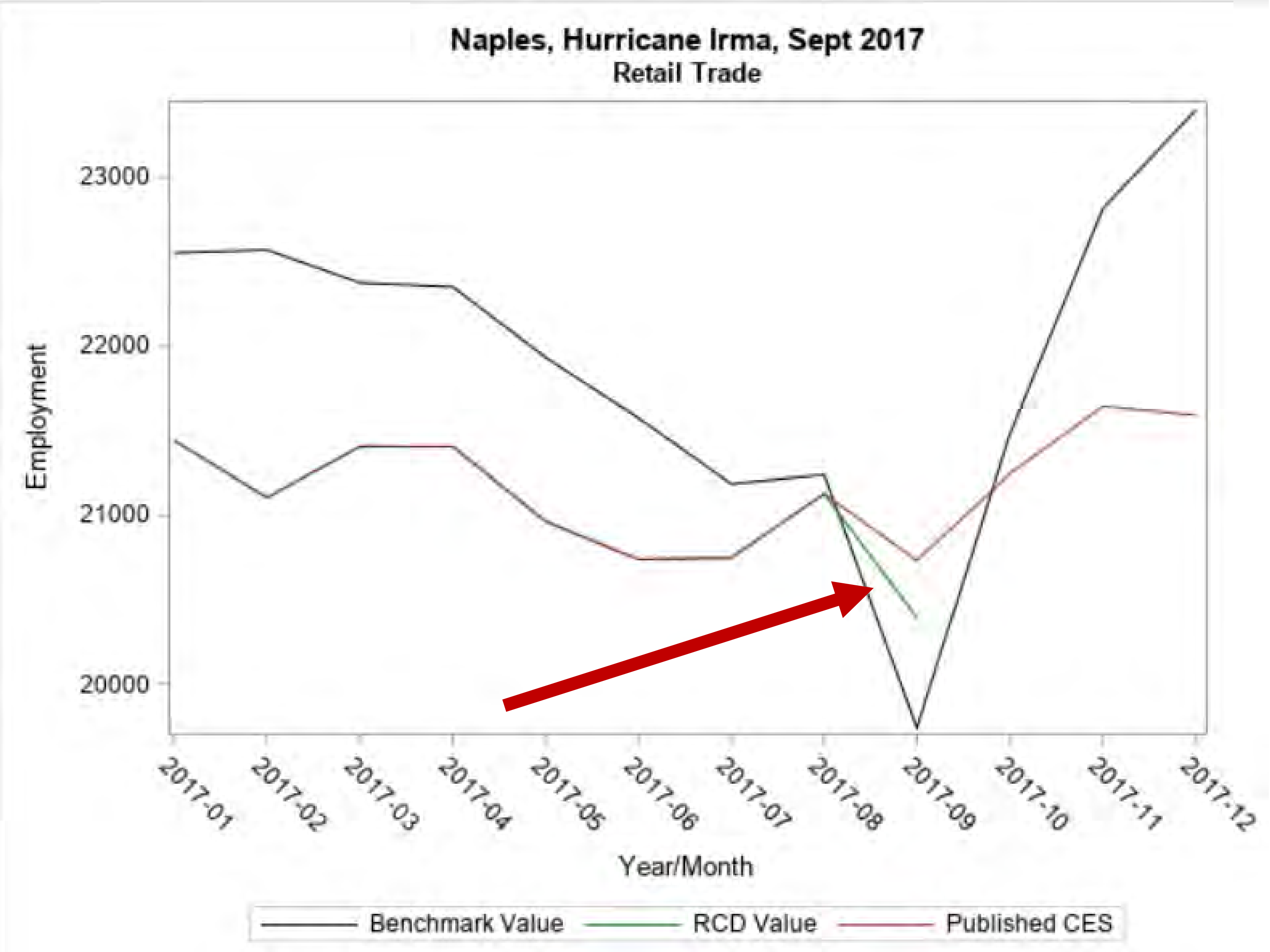

Company 1 Company 2

# Hurricane Test Cases

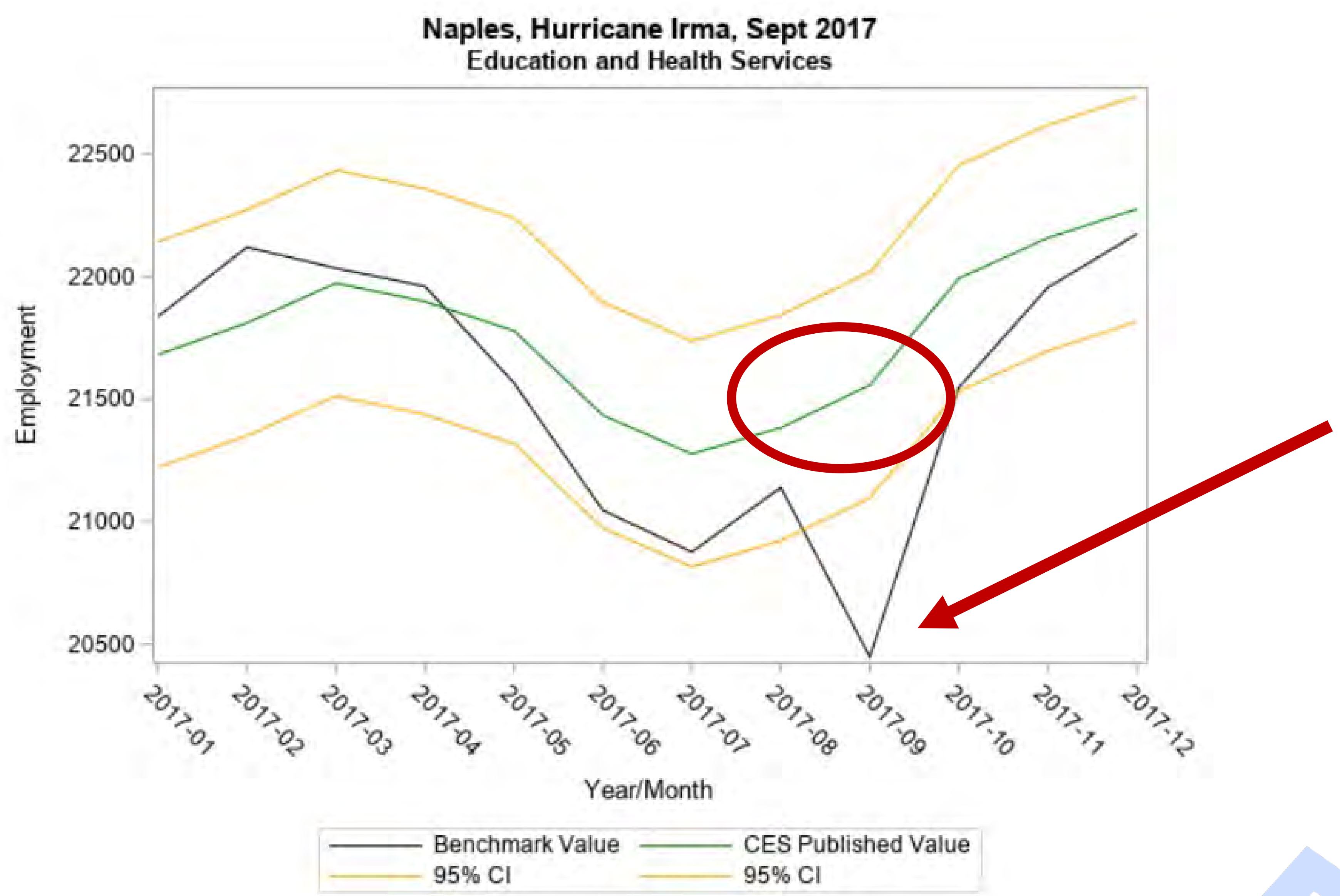
| Hurricane | Category | Month | Year | Area Affected | Revision |
|---|---|---|---|---|---|
| Irma | 4 | September | 2017 | Naples | -6.70% |
| Irma | 4 | September | 2017 | Florida | -1.20% |
| | | | | | |
| Ida | 4 | September | 2021 | New Orleans | -1.04% |
| Ida | 4 | September | 2021 | Lafayette | -0.73% |
| | | | | | |
| Ian | 5 | October | 2022 | Cape Coral** | -2.80% |
| Ian | 5 | October | 2022 | Punta Gorda** | -3.60% |
| | | | | | |
| Elsa | 1 | July | 2021 | Tallahassee | 0.10% |
| | | | | | |
| Idalia | 3 | September | 2023 | | |

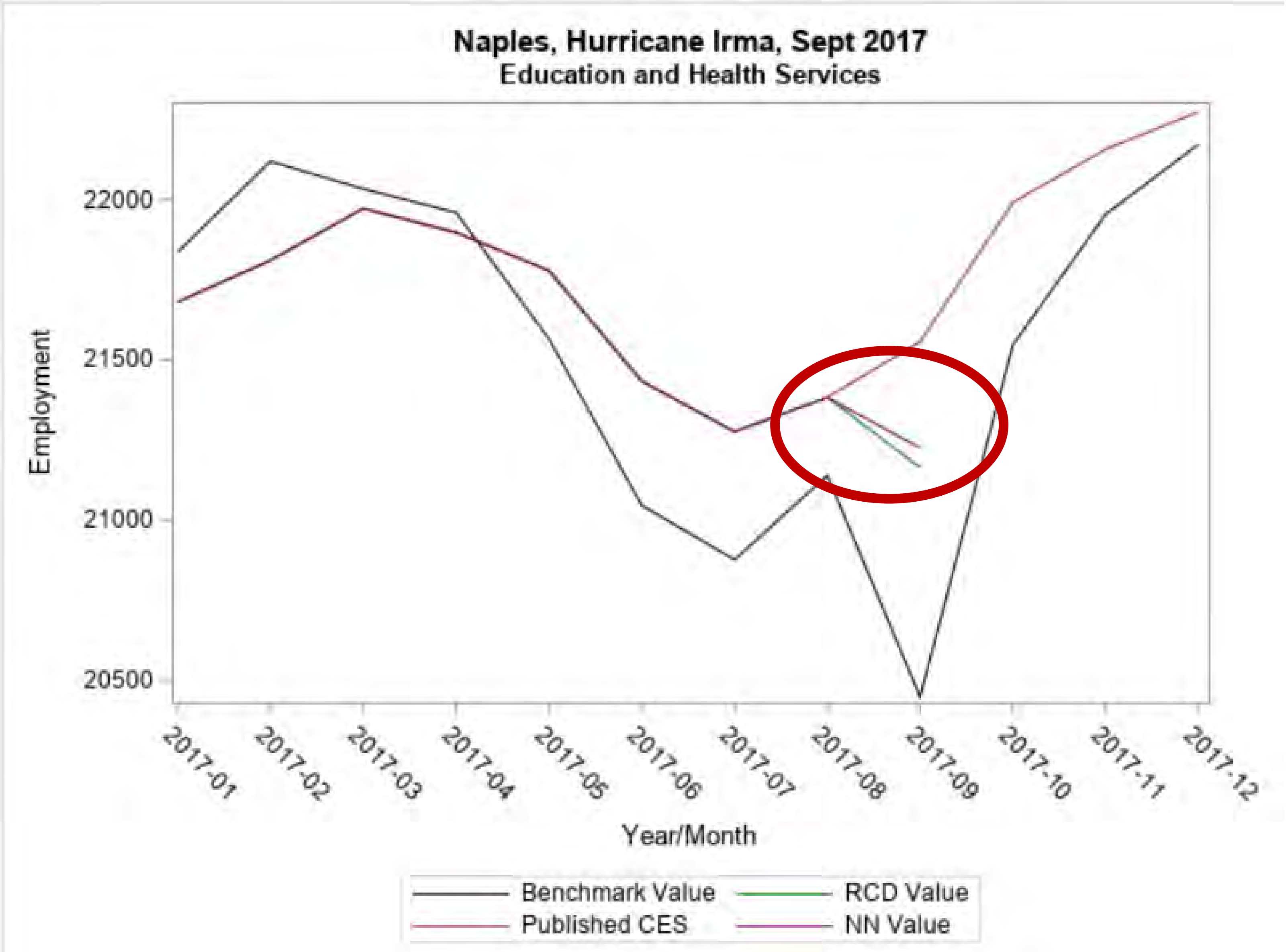

Naples, Hurricane Irma, Sept 2017
Total Private Industry
Employment
140000
135000
130000
125000
120000
2017-01
2017-02
2017-03
2017-04
2017-05
2017-06
2017-07
2017-08
2017-09
2017-10
2017-11
2017-12
Year/Month
Benchmark Value
CES Published Value
95% CI
95% CI
6.7%


Naples, Hurricane Irma, Sept 2017
Retail Trade
Employment
23000
22000
21000
20000
2017-01
2017-02
2017-03
2017-04
2017-05
2017-06
2017-07
2017-08
2017-09
2017-10
2017-11
2017-12
Year/Month
Benchmark Value
CES Published Value
95% CI
95% CI

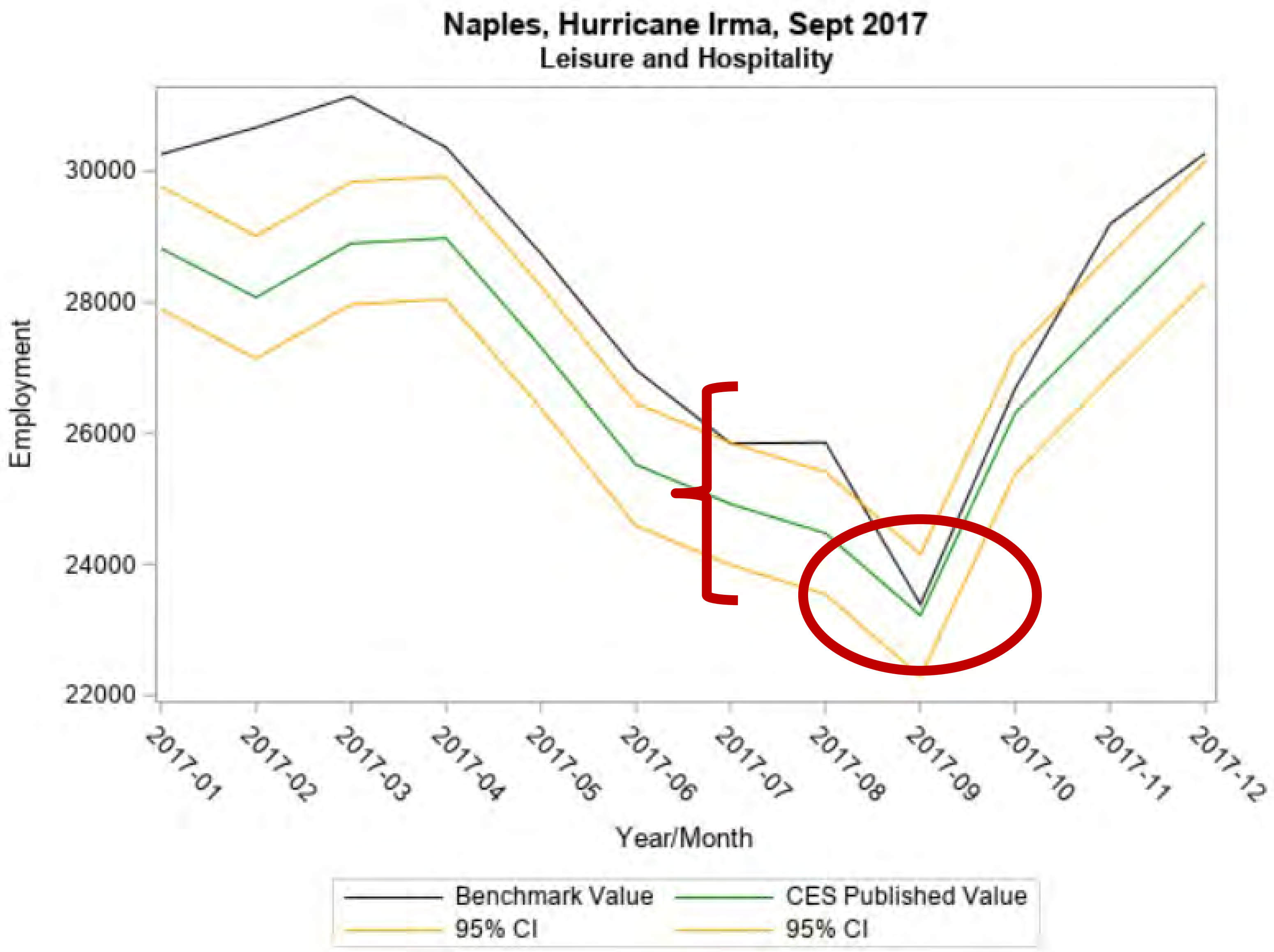


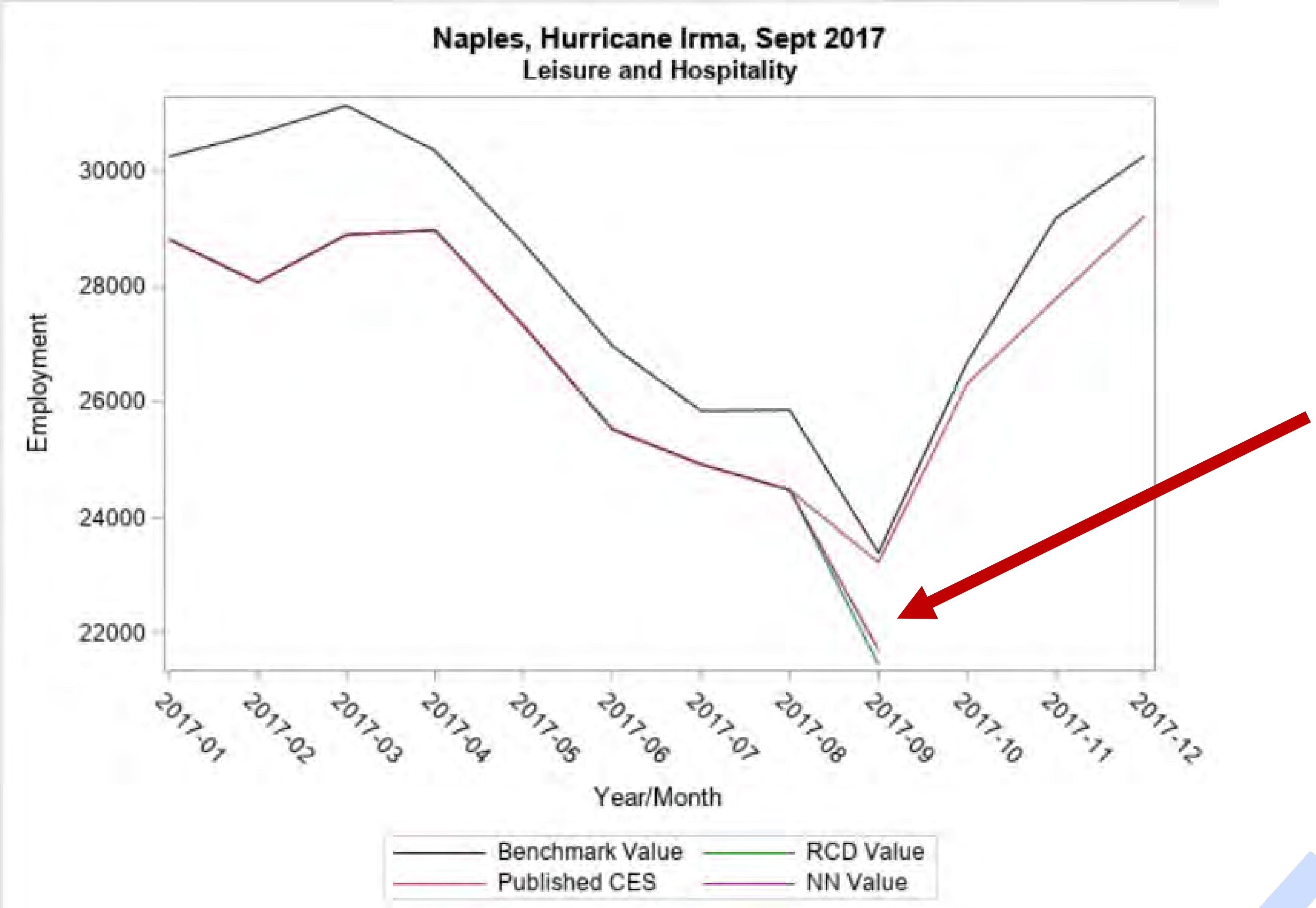
Naples, Hurricane Irma, Sept 2017
Retail Trade
Employment
23000
22000
21000
20000
2017-01
2017-02
2017-03
2017-04
2017-05
2017-06
2017-07
2017-08
2017-09
2017-10
2017-11
2017-12
Year/Month
Benchmark Value
RCD Value
Published CES


Naples, Hurricane Irma, Sept 2017
Education and Health Services
Employment
22500
22000
21500
21000
20500
2017-01
2017-02
2017-03
2017-04
2017-05
2017-06
2017-07
2017-08
2017-09
2017-10
2017-11
2017-12
Year/Month
Benchmark Value
CES Published Value
95% CI
95% CI



Naples, Hurricane Irma, Sept 2017
Education and Health Services
Employment
22000
21500
21000
20500
2017-01
2017-02
2017-03
2017-04
2017-05
2017-06
2017-07
2017-08
2017-09
2017-10
2017-11
2017-12
Year/Month
Benchmark Value
RCD Value
Published CES
NN Value


Naples, Hurricane Irma, Sept 2017
Leisure and Hospitality
Employment
30000
28000
26000
24000
22000
2017-01
2017-02
2017-03
2017-04
2017-05
2017-06
2017-07
2017-08
2017-09
2017-10
2017-11
2017-12
Year/Month
Benchmark Value
CES Published Value
95% CI
95% CI



Naples, Hurricane Irma, Sept 2017
Leisure and Hospitality
Employment
30000
28000
26000
24000
22000
2017-01
2017-02
2017-03
2017-04
2017-05
2017-06
2017-07
2017-08
2017-09
2017-10
2017-11
2017-12
Year/Month
Benchmark Value
RCD Value
Published CES
NN Value

# Results for Naples during Irma

| Event | Area Affected | Donor Pool | Industry | # of Usuals | Employment Level | Orig Revision | RCD Revision | NN Revision |
|---|---|---|---|---|---|---|---|---|
| Irma | Naples | Panama City (Michael) | Retail | 3 | 20,000 | 5.0% | 3.3% | . |
| Irma | Naples | Panama City (Michael) | Professional | 2 | 15,000 | 9.6% | 6.8% | 2.0% |
| Irma | Naples | Panama City (Michael) | Health/Ed | 4 | 20,500 | 5.4% | 3.5% | 3.8% |
| Irma | Naples | Panama City (Michael) | Leisure | 4 | 24,000 | 0.7% | -8.3% | -7.2% |
| Irma | Florida | Houston (Harvey) | All Industries | 1,265 | 9,688,000 | 1.2% | | |

# Other Results

| Event | Area Affected | Donor Pool | Industry | # of Usuals | Employment Level | Orig Revision | RCD Revision | NN Revision |
|---|---|---|---|---|---|---|---|---|
| Ida | New Orleans | Corpus Christi (Harvey) | All Industries | 69 | 448,000 | 1.2% | | |
| Ida | Lafayette | Lafayette or CC (Harvey) | All Industries | 26 | 195,000 | 0.7% | | |
| | | | | | | | | |
| Ian | Cape Coral | Corpus Christi (Harvey) | Manufacturing | 4 | 7,700 | 4%** | 0.0% | . |
| Ian | Cape Coral | Corpus Christi (Harvey) | Wholesale Trade | 4 | 7,700 | 13.2%** | . | 10.8% |
| Ian | Cape Coral | Corpus Christi (Harvey) | Financial | 7 | 14,000 | 9.4%** | 8.4% | . |
| Ian | Cape Coral | Corpus Christi (Harvey) | Professional | 6 | 40,000 | 12%** | . | 4.9% |
| Ian | Cape Coral | Corpus Christi (Harvey) | Leisure | 9 | 37,000 | 19.8%** | 13.1% | 11.0% |
| | | | | | | | | |
| Ian | Punta Gorda | Corpus Christi (Harvey) | Leisure | 2 | 6,900 | 11.3%** | 5.7% | . |
| Ian | Punta Gorda | Naples (Irma) | Leisure | 2 | 6,900 | 11.3%** | -3.0% | 3.6% |
| Ian | Punta Gorda | Naples (Irma) | Retail | 2 | 9,500 | 10.1%** | 7.1% | . |
| | | | | | | | | |
| Elsa | Tallahassee | Corpus Christi (Harvey) | All Industries | 9 | 182,000 | 0.1% | | |
| | | | | | | | | |
| Idalia | FL, GA, NC | | | | | | | |

# Conclusions

- There appear to be some gains from attempting imputation at the MSA level in a month with a hurricane.
- Random cold deck imputation provides more (and more stable) results compared to nearest neighbor, though there could still be some concerns with added variance
- Improvements could occur over time as more donor pools become available

# Future work

- Use a regression tree or other model/algorithm to further categorize/determine donor pool selection
- Analyze/account for variance effects, both for the multiple imputation and adjusted estimate
- Is there a concern about catching the recovery in the next few months?
- Investigate other uses – wildfires, tornados, flooding, snowstorms, “extreme” cold/heat, etc

# Contact Information

**Bradley Rhein**
Supervisory Statistician
CES State and Area
rhein.bradley@bls.gov