



# **APPROACHES FOR COMPUTING MARGINS OF ERROR FOR USER DEFINED VARIABLES BASED ON THE AMERICAN COMMUNITY SURVEY**

David C. Folch  
Northern Arizona University

*ACS Data Users Conference*

*May 19, 2021*

# ACS MARGINS OF ERROR

	Arizona	
Label	Estimate	Margin of Error
▼ Total population	7,278,717	*****
One race	6,998,143	±12,629
Two or more races	280,574	±12,629
▼ One race	6,998,143	±12,629
White	5,701,810	±20,005
Black or African American	343,729	±7,680
▼ American Indian and Alaska Nati...	332,273	±8,619
Cherokee tribal grouping	1,639	±633
Chippewa tribal grouping	1,363	±1,052
Navajo tribal grouping	160,420	±7,154
Sioux tribal grouping	1,956	±1,034

Coefficient  
of Variation

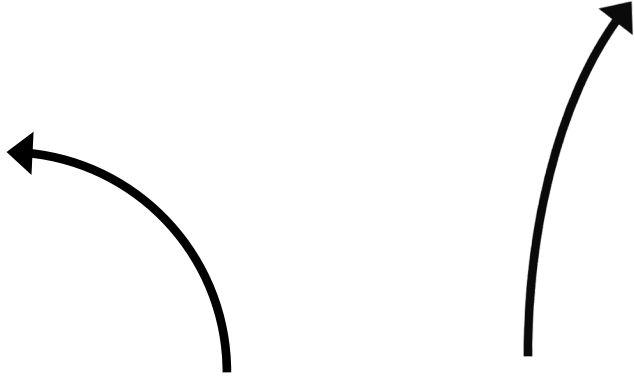
0.469

0.027

# COMBINING ACS ESTIMATES

Unpublished MOE: ??

Unpublished sum: 161,783



The diagram consists of two curved arrows. One arrow originates from the 'Unpublished sum: 161,783' text and points to the 'Chippewa tribal grouping' row of the table. The other arrow originates from the 'Unpublished MOE: ??' text and points to the '±7,154' value in the 'Navajo tribal grouping' row of the table.

Chippewa tribal grouping	1,363	±1,052
Navajo tribal grouping	160,420	±7,154

# STANDARD FORMULAS FOR ACS MOE

## Summation

$$\text{MOE}(\hat{X}_1 + \hat{X}_2 + \dots + \hat{X}_n) = \pm \sqrt{[\text{MOE}(\hat{X}_1)]^2 + [\text{MOE}(\hat{X}_2)]^2 + \dots + [\text{MOE}(\hat{X}_n)]^2}$$

## Proportion

$$\text{MOE}(\hat{P}) = \frac{1}{\hat{Y}} \sqrt{[\text{MOE}(\hat{X})]^2 - (\hat{P}^2 * [\text{MOE}(\hat{Y})]^2)}$$

## Ratio

$$\text{MOE}(\hat{R}) = \frac{1}{\hat{Y}} \sqrt{[\text{MOE}(\hat{X})]^2 + (\hat{R}^2 * [\text{MOE}(\hat{Y})]^2)}$$

## Product

$$\text{MOE}(\hat{X} * \hat{Y}) = \sqrt{(\hat{X}^2 * [\text{MOE}(\hat{Y})]^2) + (\hat{Y}^2 * [\text{MOE}(\hat{X})]^2)}$$

### Understanding and Using American Community Survey Data

*What All Data Users Need to Know*

Issued September 2020



United States<sup>®</sup>  
**Census**  
Bureau

U.S. Department of Commerce  
U.S. CENSUS BUREAU  
[census.gov](https://www.census.gov)

# CONSIDERATIONS FOR THE STANDARD FORMULAS

- Easy to compute from published data
- Provided formulas do not cover all cases

- Example: Theil Index

$$H = \sum_{i=1}^N \frac{t_i(E - E_i)}{ET}$$

$$E_i = \sum_{r=1}^R \Pi_{ri} \ln(1/\Pi_{ri})$$

$$E = \sum_{r=1}^R \Pi_r \ln(1/\Pi_r)$$

- Covariance
  - Missing from the standard formulas

$$SE(\hat{X}_1 + \hat{X}_2) = \sqrt{SE(\hat{X}_1)^2 + SE(\hat{X}_2)^2} + 2\text{cov}(\hat{X}_1, \hat{X}_2)$$

standard formula                      missing covariance

- Generally not combining independent estimates
  - Problem tends to grow as more estimates enter the computation

# SIMULATION APPROACH

- Assume each published ACS estimate comes from a normal distribution
  - Mean: ACS estimate
  - Standard deviation: ACS standard error  $\rightarrow SE(\hat{X}) = MOE(\hat{X})/1.645$
- Simulate combined estimates
  - Define a normal distribution for each input variable
  - Make one draw from each distribution
  - Compute the combined estimate (e.g., summation, Theil index, etc.) from these draws
  - Repeat the process many times (e.g., 1000)
- Compute MOE of combined estimate
  - Define an ACS style (90%) confidence interval
  - Difference between the 5<sup>th</sup> and 95<sup>th</sup> percentiles of the combined estimates
  - Half of that difference

Sources:

Brault (2014)

Napierala and Denton (2017)

# CONSIDERATIONS FOR THE SIMULATION APPROACH

- Uses standard ACS data
- Complicated computational process
- Flexible → can be applied to any combined estimate
- Does not consider covariance

# REPLICATES APPROACH

- Published MOEs are built using a replicates approach
- 80 replicates of the entire ACS are built from the original surveys
- Each replicate is internally consistent
- Recently, replicate tables for ACS estimates have been published
- Compute combined estimates
  - Compute the combined estimate (e.g., summation, Theil index, etc.) for one replicate
  - Repeat the process 80 times
- Compute MOE of combined estimate

$$\text{MOE}(\hat{\theta}_0) = 1.645 \sqrt{\text{var}(\hat{\theta}_0)} = 1.645 \sqrt{\frac{1}{80} \sum_{r=1}^{80} (\hat{\theta}_r - \hat{\theta}_0)^2}$$

combined estimate  
computed from  
replicate data

combined estimate  
computed from  
original data



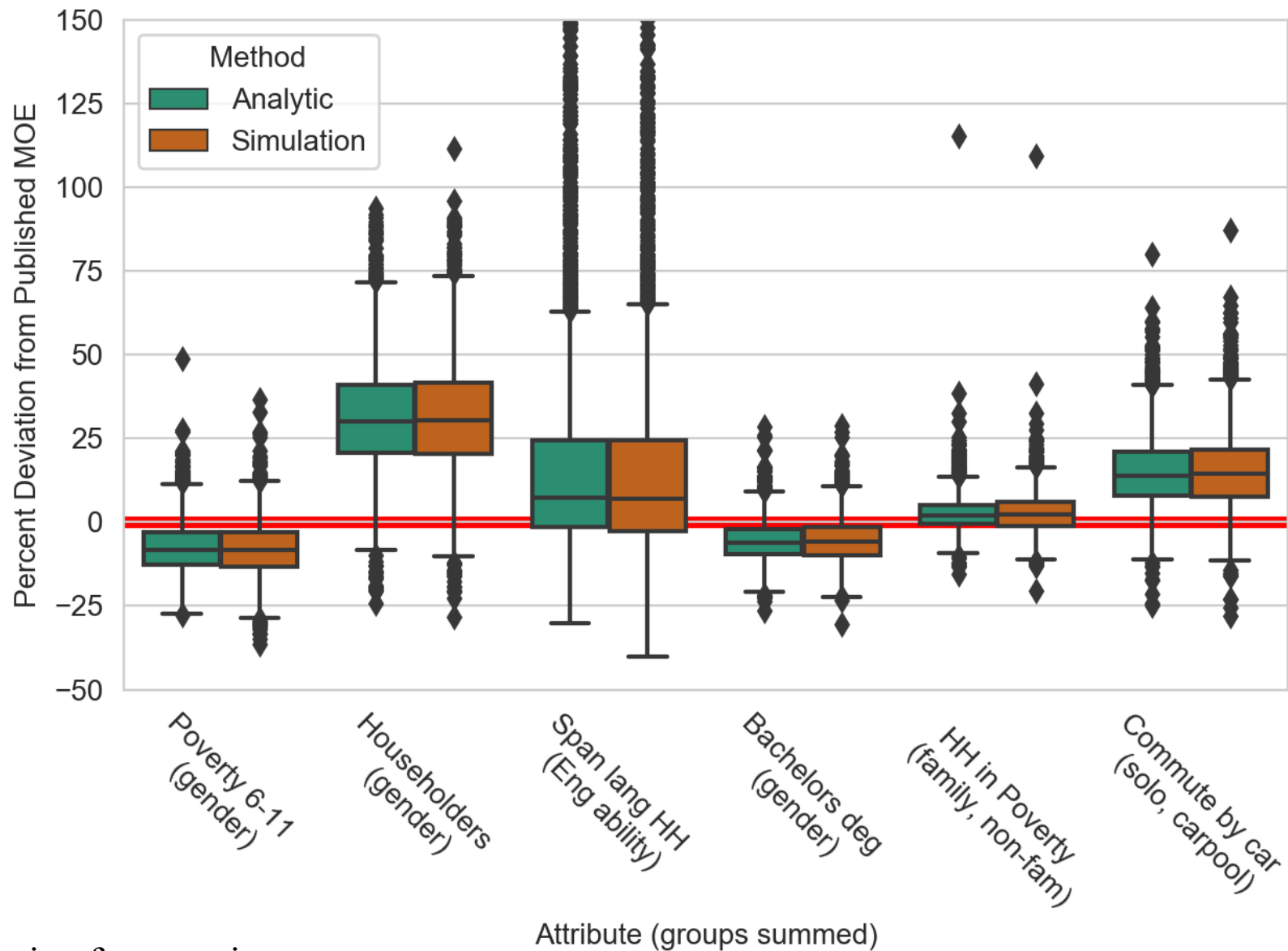
# CONSIDERATIONS FOR THE REPLICATES APPROACH

- Limited to the estimates, years and geographic scales with published replicate tables
- Moderately complicated computational process
- Flexible → can be applied to any combined estimate
- Includes covariance

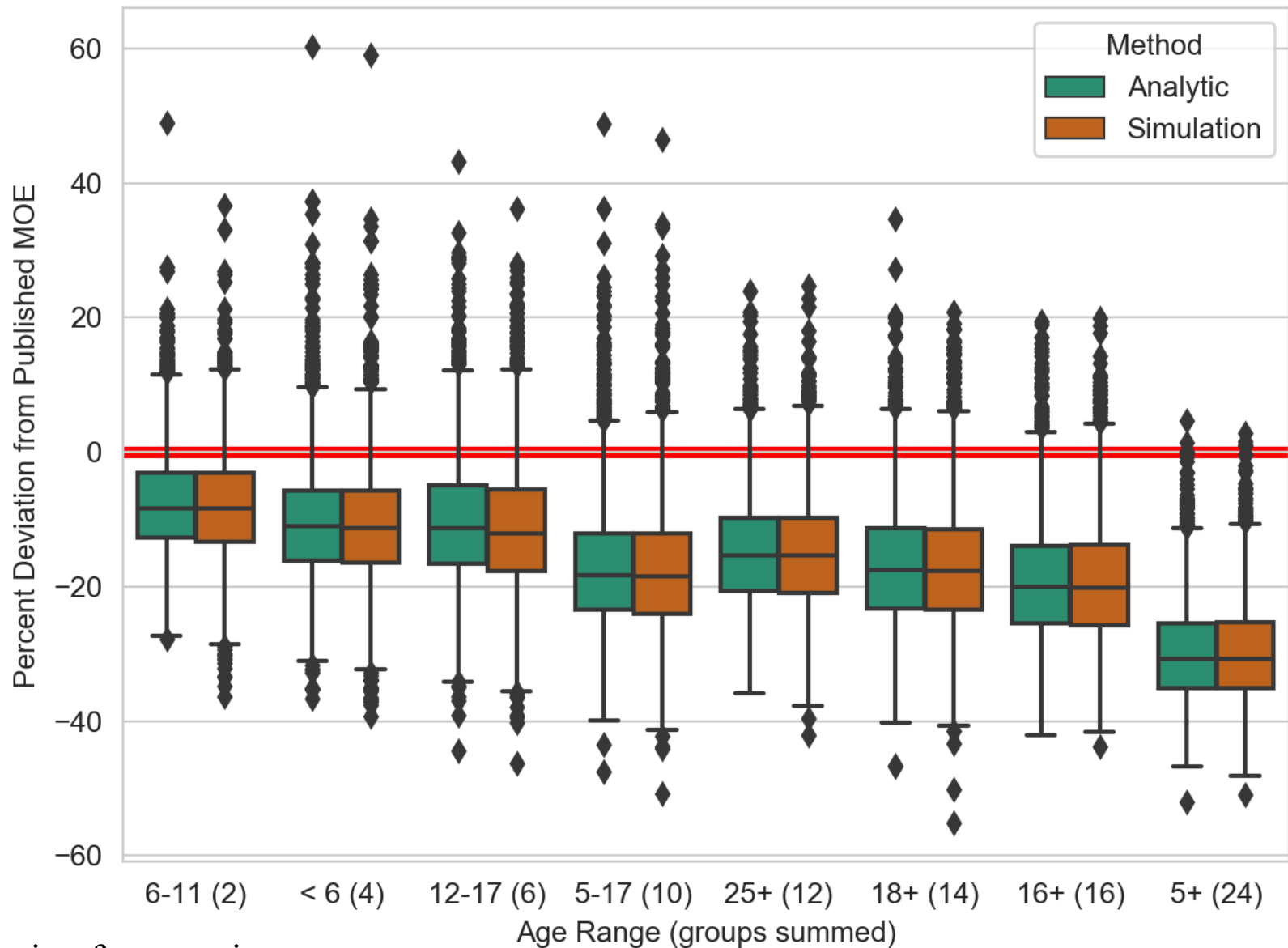
# RECONSTRUCTION APPROACH FOR TESTING

- How do user calculated MOEs differ from published MOEs?
- Testing requirements
  - Published inputs and outputs (estimates and MOEs)
  - Data available in replicate tables
- 5-year estimates (2011-2015)
- All counties
- Zero estimates
  - The published MOEs on zero estimates follows an ad hoc approach
  - We exclude these cases in the following examples
- User implemented replicate approach matches published values
  - Only presenting results for the analytic and simulation approaches

# TWO ESTIMATE SUMMATION ON DIFFERENT DOMAINS

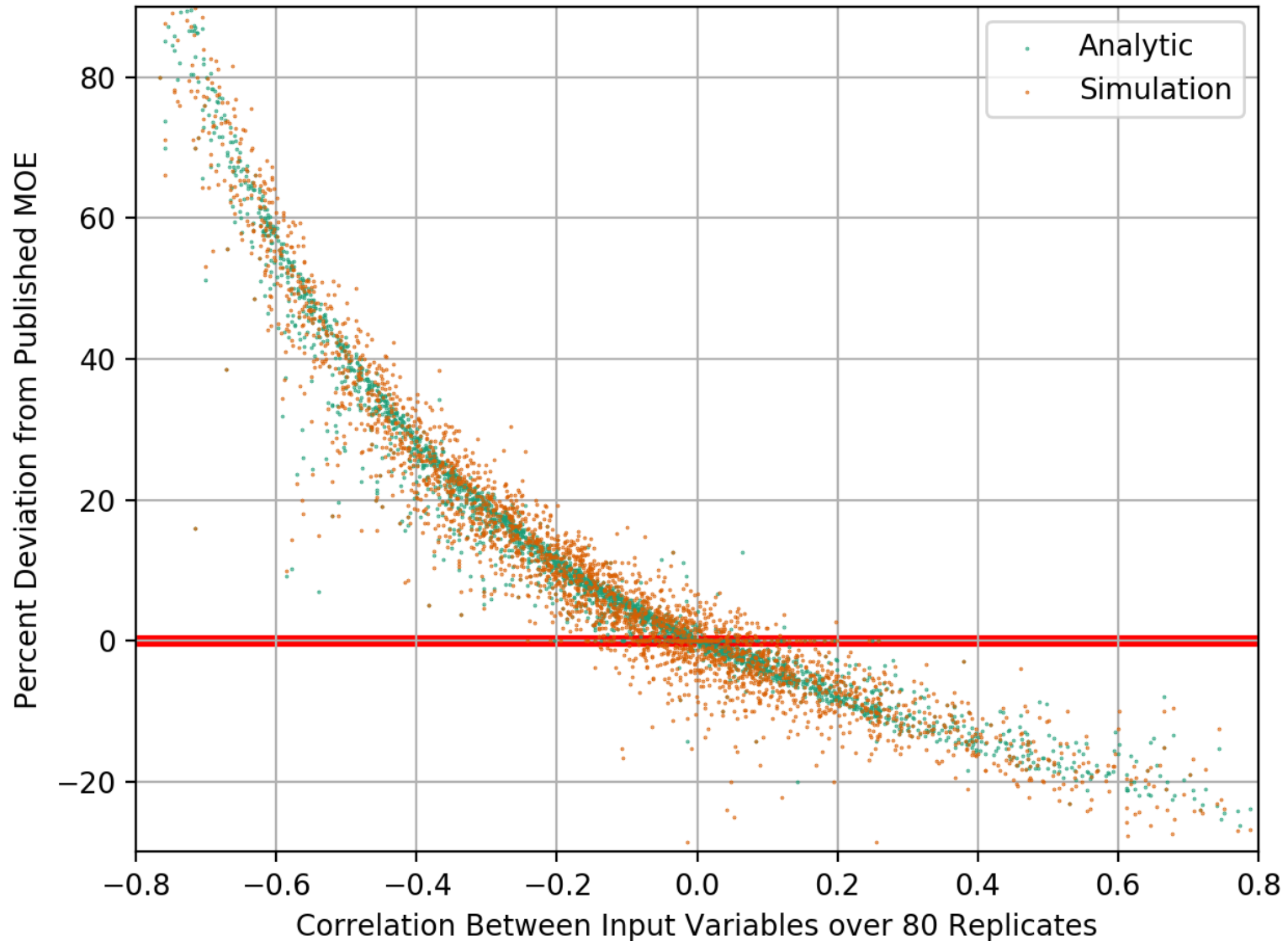


# SUMMATION ON DIFFERENT NUMBERS OF POVERTY ESTIMATES



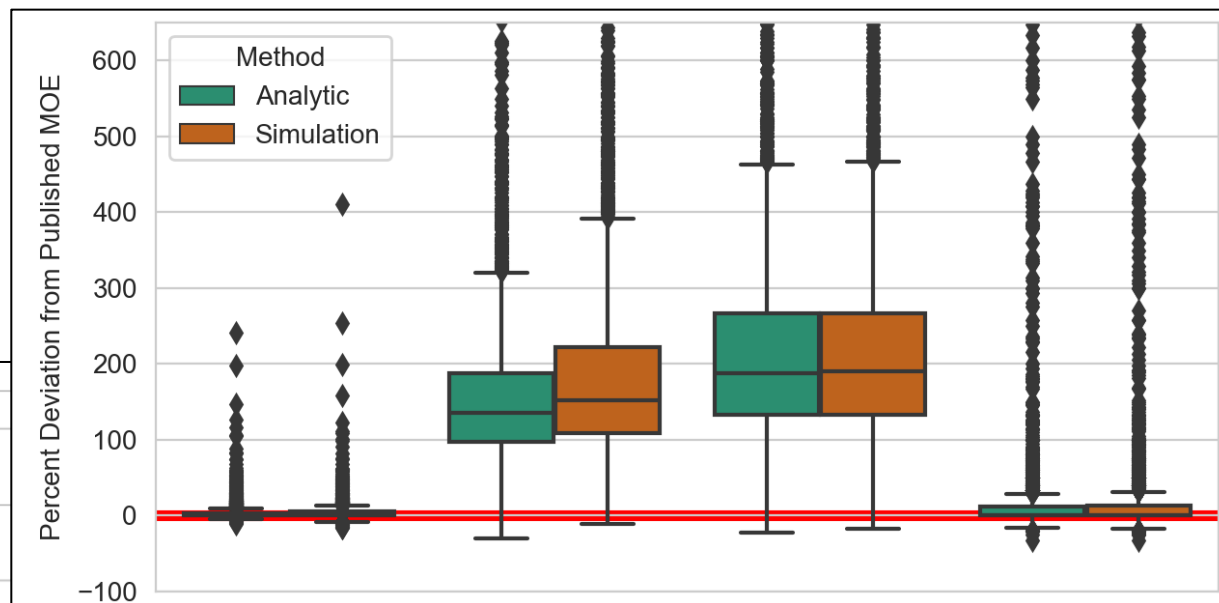
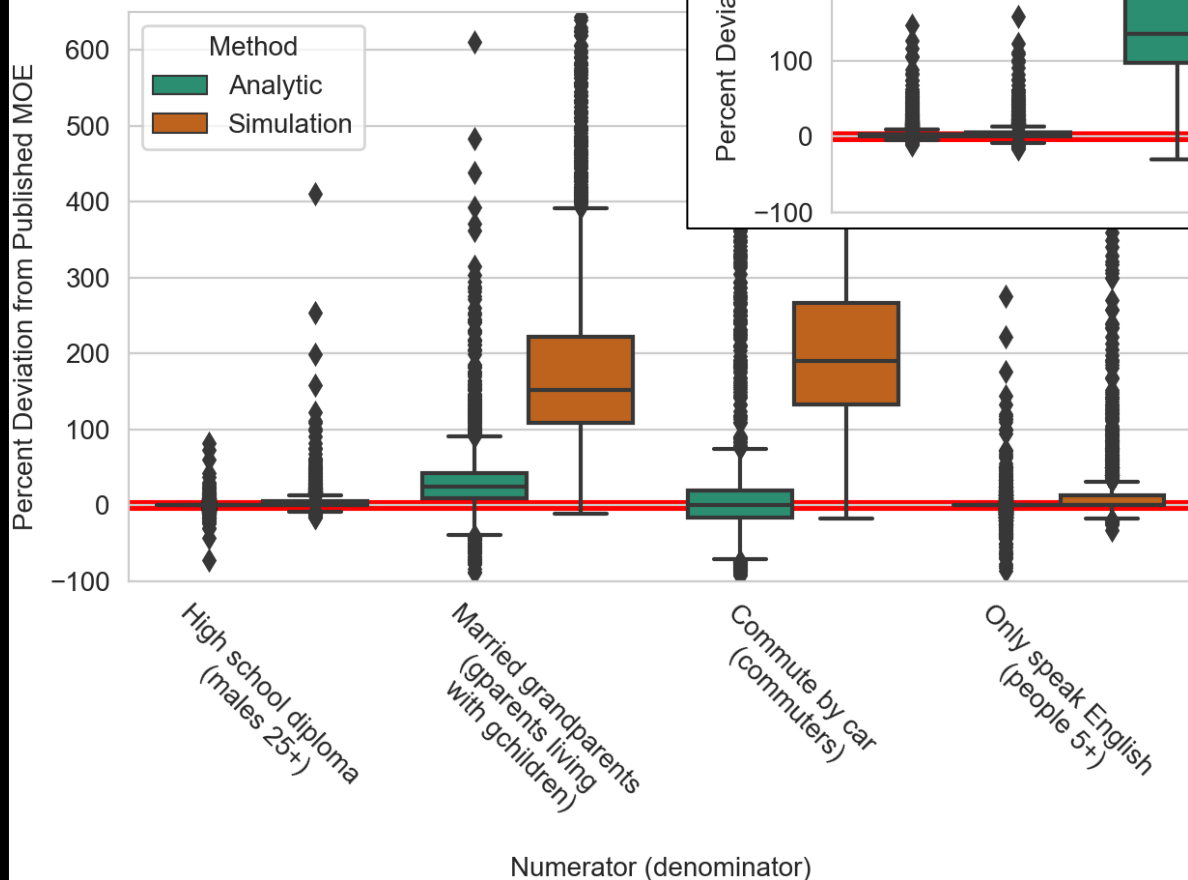
Distribution for counties

# REPLICATE CORRELATION VS. PUBLISHED MOE



# PROPORTIONS ON DIFFERENT DOMAINS

Proportion analytic formula



Distribution for counties

# RECOMMENDATIONS

- Replicates approach is always the best
  - Can handle most any user defined combination
  - Not always possible due to limited replicate table availability
- Analytic and simulation approaches give similar results on summations
  - Analytic is much simpler to compute
- Analytic approach produces better results than simulation approach for proportions
  - Assuming the use of the Census Bureau's recommended formula
- Ratios were not directly tested
  - Other results indicate that analytic and simulation likely give similar results for ratios
- Simulation for everything else
  - Complex user-defined combination and
  - Replicate tables are not available

# COLLABORATORS

- Seth E. Spielman, University of Colorado Boulder
- Molly Graber, Department of City Planning, City of New York





**APPROACHES FOR COMPUTING  
MARGINS OF ERROR FOR USER  
DEFINED VARIABLES BASED ON THE  
AMERICAN COMMUNITY SURVEY**

David C. Folch  
Northern Arizona University  
[david.folch@nau.edu](mailto:david.folch@nau.edu)