Differential Privacy, Differential Accuracy?: A Use Case on Data Segmentation by Tract-level Tenure Majority

Leslie Reynolds and Jan Vink
Program on Applied Demographics
Cornell University

Introduction

- People in areas with mostly renter-occupied units have different characteristics than those in areas with mostly owner-occupied housing units
    - Rental-majority areas tend to be more socially and economically disadvantaged
- Research on differentials by tenure is essential for many government/social programs
    - e.g. HUD rental assistance program requirements are determined by research on tenure by household characteristics
- Nationally aggregated Census 2010 differential privacy demonstration data differs little from original 2010 Summary File 1
    - ***But:***
        - Areas with mostly renter-occupied and mostly owner-occupied units are geographically scattered
        - Don’t fit within the top-down approach of differential privacy

Research Questions

1. Does the accuracy of the differentially private demographic and housing characteristics (DHC) data vary when segmented by tenure majority?
   - Are tenure types disproportionately impacted by the application of differential privacy?

2. Do tracts aggregated to sub-state levels exhibit the same patterns of error when separated by tenure as higher aggregate levels?

3. Did disparities in error by tenure change between the March and August versions of the demonstration data (i.e. due to the DAS alterations made)?

Data and Methods

- Two most recent versions of the 2010 Summary File 1 and 2010 Differential Privacy Demonstration data file retrieved from IPUMS (Van Riper et al., 2022)
  - March 29, 2022 (original allocation of privacy loss budget- different than planned)
  - August 25, 2022

- Census tract Housing Unit files
  - Analysis of block groups showed the same patterns as tracts, with more extreme errors especially in rental majority areas
  - Excluded Puerto Rico
  - Only tracts with at least 200 households (to exclude special purpose areas)

Analytical sample of…

Table 1: Number of Census tracts in each Aggregate Geographic Level

	U.S.	New York	Onondaga County (Syracuse)	Monroe County (Rochester)
Tracts	**71,842**	**4,772**	**139**	**189**

Data and Methods (cont.)

- Featured Measures: “large” households (5+ residents) and households with children (under 18)

- Analytical Variable: dominant tenure in a tract
  - Majority owned (>=80% owned households)
  - Majority rented (<=20% owned households)
  - Mixed tenure (> 20% & < 80% owned households)

- Percent owned Calculated as:

[(owned + owned with mortgage)/Total occupied households]*100

Table 2: Number of Tracts in the U.S., by Tenure Type

	Mixed	Majority Rented	Majority Owned
Tracts	**46,641**	**3,634**	**21,567**

The number of tracts in each tenure type did not change between the March and August demo data

Metrics of Error

- Average difference between the demonstration and original data values (Bias)
  - Mean Error: $\frac{1}{N}\sum(X_{dp} - X_{sf})$

- Accuracy of the differentially private data to the original Summary File 1
  - Median Absolute Percent Error (MdAPE): Median [100%* $\sum \left| \frac{X_{dp} - X_{sf}}{X_{sf}} \right|$]
  - Median rather than mean for robustness to outliers

- Share of tracts with “big errors” (Usability of estimates)
  - Mean Absolute Error >=10 & Mean Absolute Percent Error >=10%

Results: Mean Error (Bias)

Table 3: Mean Error for Selected Households, by DAS Version and Aggregate Level

		March Demonstration			August Demonstration		
Measure	Geography	Mixed	Rented	Owned	Mixed	Rented	Owned
5+ Person Households	United States	**-0.15***	**2.60***	-0.02	**-0.66***	0.47	**1.34***
	New York State	-0.17	**-1.82***	**1.35***	-0.37	**-2.73***	**2.25***
	Monroe County	-0.92	2	-0.45	-0.92	-3.38	**3.02***
	Onondaga County	-1.23	5.79	-0.33	**-2.95***	3.00	**2.71***
Households with children	United States	**0.64***	**6.40***	**-1.94***	**-0.36***	**3.34***	0.13
	New York State	0.67	**2.77***	-0.79	0.05	0.34	0.39
	Monroe County	-1.76	10.46	-3.76	**-3.48***	0.31	0.47
	Onondaga County	-3.96	11	-2.04	**-5.51***	6.50	2.35

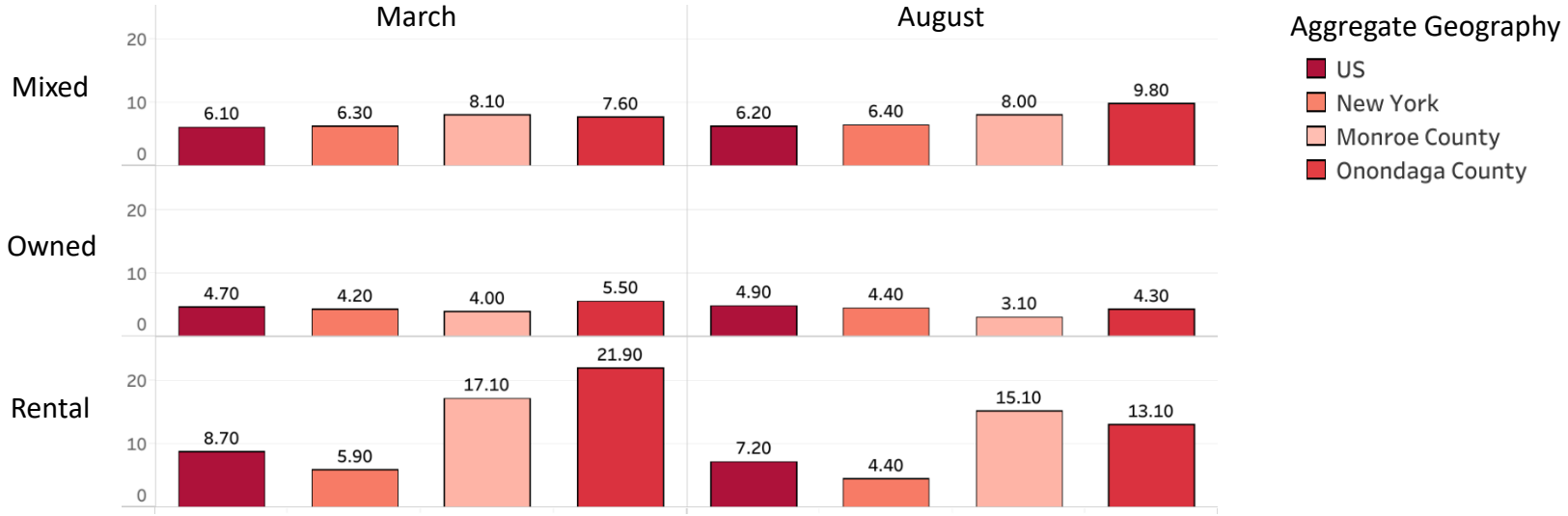
*Errors significantly different from 0 at 95% confidence

Significant mean errors remained in the August demonstration data for both groups

- Large households:
  - Positive bias (overestimation) in owner majority areas at all aggregate levels
  - Negative bias (underestimation) in rental majority areas in New York State, and mixed areas in Onondaga county
- Households with Children:
  - Negative bias in mixed tenure areas at all levels except New York State
  - Positive bias in rental majority areas in the national level only

Results: Median Absolute Percent Error (accuracy)

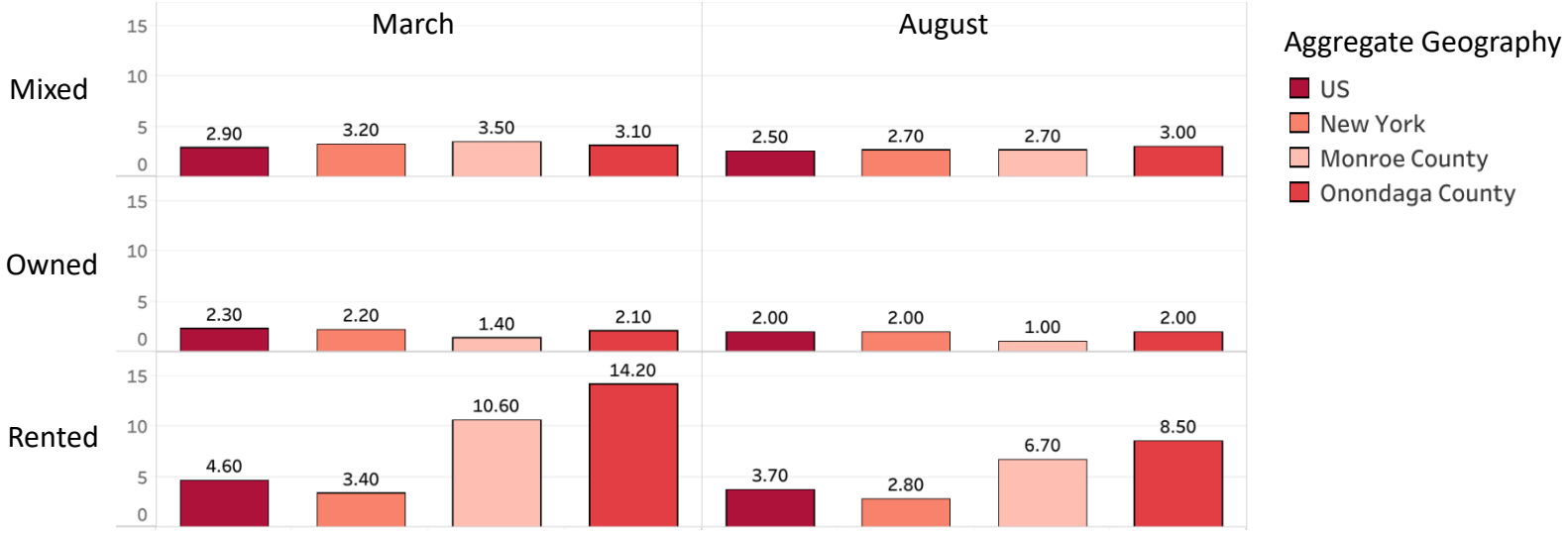
Figure 1: MdAPE for Large Households by Tract-level Tenure Majority



- Accuracy improved between versions for some aggregate level/tenure-type combinations, but not others
  - Accuracy in Onondaga county worsened for mixed tenure areas; both counties improved in owned and rental majority areas
  - MdAPE still problematic in county level rental majority areas, and now mixed tenure areas of Onondaga county

Results: Median Absolute Percent Error (Accuracy) (cont.)

Figure 2: MdAPE for Households With Children by Tract-level Tenure Majority



- Changes made to the DAS between versions improved accuracy for households with children more than large households
  - MdAPE still largest in rental majority areas for all geographic levels, but smaller differences between tenure types
  - County level MdAPE's no longer over 10

Results: Prevalence of “Big Errors”

Table 4: Share of Tracts with “Big Errors” for Large Households, by DAS version and aggregate level

	March				August			
	U.S.	New York	Monroe County	Onondaga County	U.S.	New York	Monroe County	Onondaga County
Mixed	26.8%	27.8%	36.2%	37.7%	26.9%	26.3%	53.2%	22.0%
Rent	36.8%	23.0%	46.2%	50.0%	28.9%	16.8%	35.7%	46.2%
Own	16.8%	13.6%	14.3%	14.6%	19.2%	19.6%	10.4%	12.2%

- Some differences between versions, but shares of big errors are still quite large
  - About half (46%) of rental majority tracts in Onondaga County
  - Over half (53%) of mixed tenure tracts of Monroe County
- Larger shares of big errors in mixed tenure and majority rental areas for the U.S. than New York State
- Owner majority areas much less impacted by issues of usability than rental and mixed tenure areas

Results: Prevalence of “Big Errors” (cont.)

Table 5: Share of Tracts with “Big Errors” for Households With Children, by DAS version and aggregate level

	March				August			
	U.S.	New York	Monroe County	Onondaga County	U.S.	New York	Monroe County	Onondaga County
Mixed	7.2%	8.7%	8.7%	13.0%	4.9%	5.1%	11.7%	3.9%
Rent	27.4%	15.0%	53.8%	64.3%	20.1%	8.7%	35.7%	46.2%
Own	4.3%	2.3%	0.0%	2.1%	3.2%	2.3%	4.2%	2.0%

- Estimates of households with children in rental majority areas remained most prone to big errors
  - About half (46%) of rental majority tracts in Onondaga County
  - Over one-third (36%) of rental majority tracts in Monroe County
- New York state had smallest share of rental majority tracts with big errors
- As with large households, owner majority areas much less impacted by big errors than rental or even mixed tenure areas

In Conclusion:

Analysis demonstrates that segmenting differentially private data by geographically clustered groups of characteristics results in unique accuracy issues

- Rental majority tracts are most prone to large errors
  - May be due to the concentration of disadvantage
  - Estimates of large households in rental majority tracts are particularly problematic, especially when aggregated to counties
- Estimates of households in owner majority tracts were more accurate, less biased, and more usable (less big errors) than the other tenure types
- Measures of error improved in the newest demonstration data for some tenure types and aggregate levels but worsened for others
- We await the results from the final 2010 demonstration data to show how the PLB ultimately impacts these errors

Thank you!

lr523@cornell.edu
(607) 255-2127