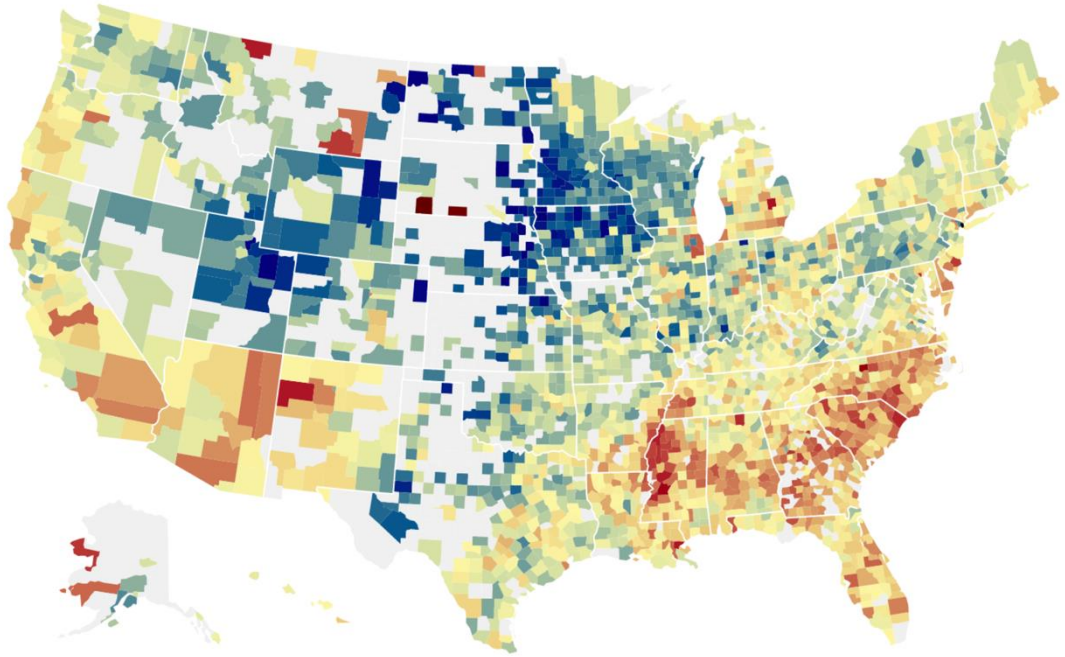
# Measuring Income Mobility Under Differential Privacy

Brett Mullins & Miguel Fuentes

University of Massachusetts Amherst

# Motivation

- Understanding economic mobility is essential for policy evaluation
- Analysts require access to administrative datasets containing sensitive information about individuals
- **Challenge**: how do we publish useful mobility statistics while preserving individual privacy?



# Income Data

## Data Sources

- Tax returns (IRS 1040)
- UI Wage Records (ES-202)
- Program Records (SNAP/TANF)

## Structure

- Longitudinal tabular data

| ID | 2017 | 2018 | … | 2022 |
|---|---|---|---|---|
| 1 | $52,000 | $44,000 | … | $53,000 |
| 2 | $0 | $30,000 | … | $0 |
| 3 | $87,000 | $93,000 | … | $110,000 |
| 4 | $20,000 | $20,000 | … | $20,000 |

# Transition Matrices

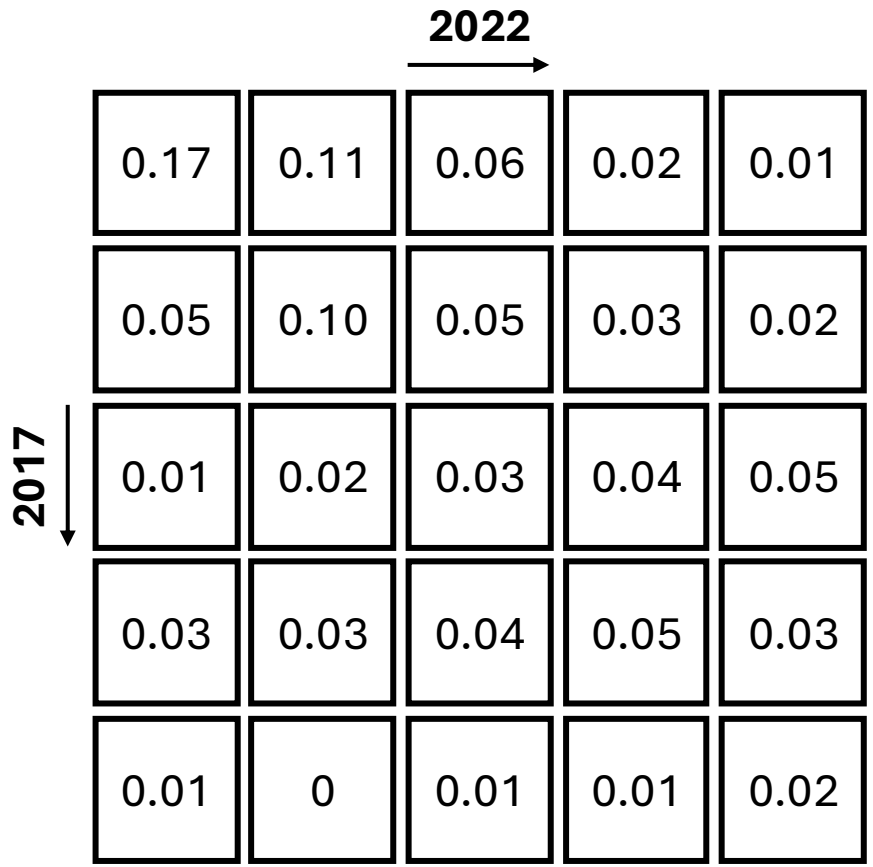
Represent mobility between two years

- Discretize income columns and compute a normalized two-way histogram

Discretization may be data-dependent

- Tax brackets vs. quantiles

Note: Divide rows by row sums to obtain a stochastic matrix

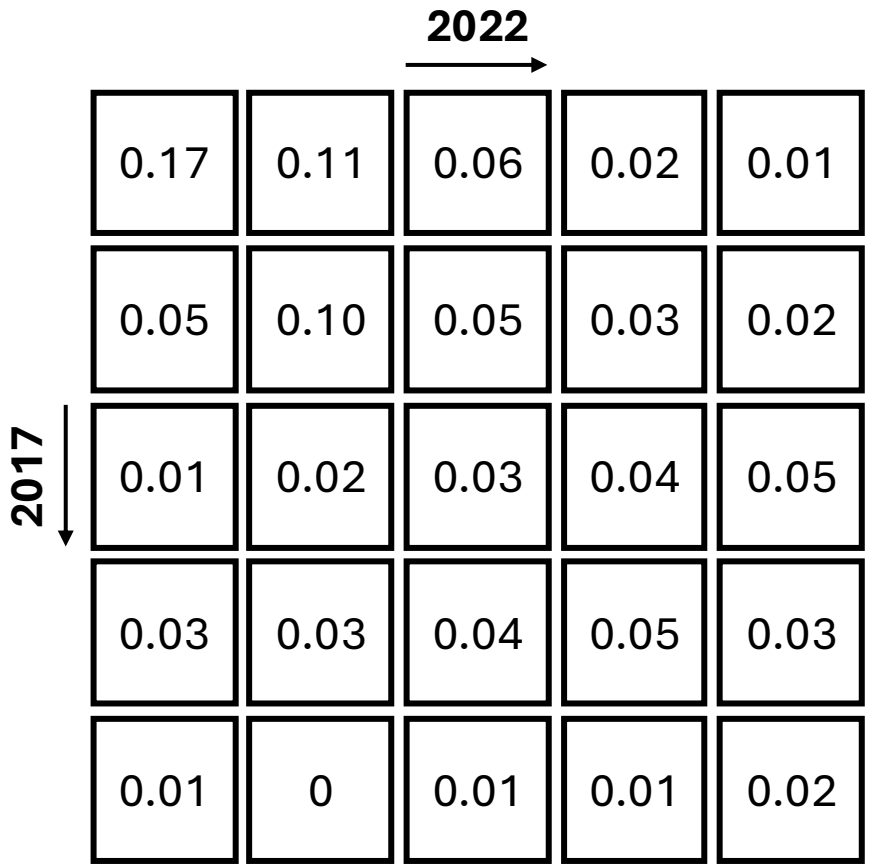


# Mobility Indices

Scalar summaries capturing various facets of mobility

- Analysts typically report a suite of indices

Examples




Katharine Bradbury. “Levels and Trends in the Income Mobility of US Families, 1977- 2012”. In: FRB of Boston Working Paper (2016).

# Mobility Indices

Scalar summaries capturing various facets of mobility

- Analysts typically report a suite of indices

Examples

- Prais-Bibby index measures the proportion of sample that changed bins


Katharine Bradbury. “Levels and Trends in the Income Mobility of US Families, 1977- 2012”. In: FRB of Boston Working Paper (2016).

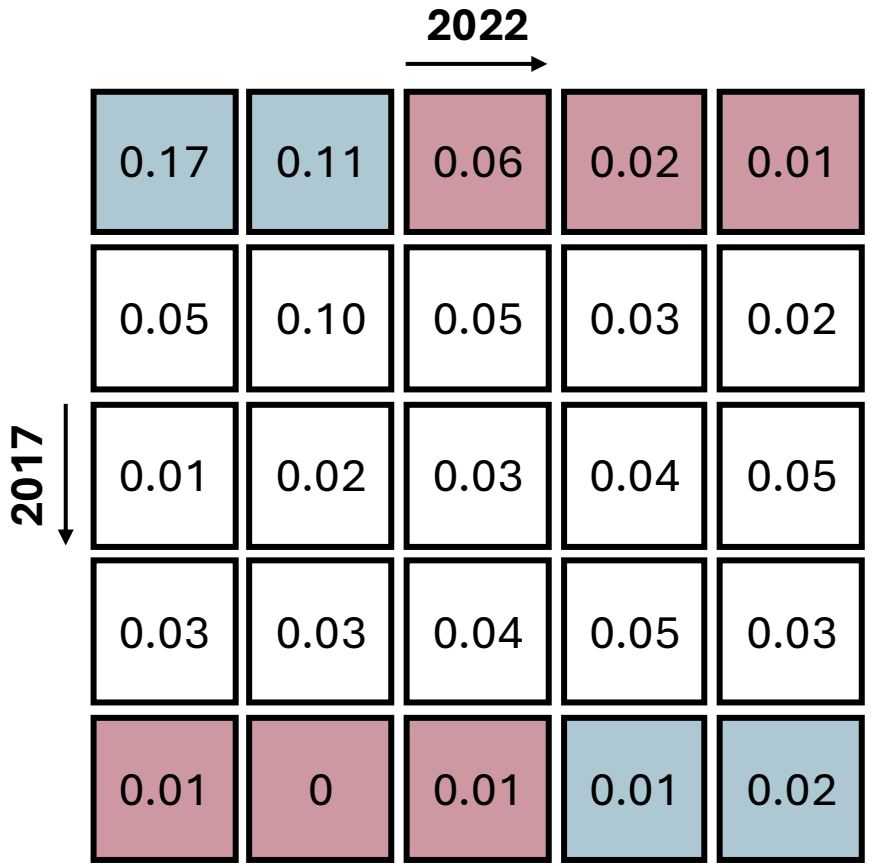
# Mobility Indices

Scalar summaries capturing various facets of mobility

- Analysts typically report a suite of indices

Examples

- Prais-Bibby index measures the proportion of sample that changed bins
- Origin-specific indices measure the proportions that "escaped the bottom" and "fell from the top"




Katharine Bradbury. “Levels and Trends in the Income Mobility of US Families, 1977- 2012”. In: FRB of Boston Working Paper (2016).

# Differential Privacy

## Informal Definition

- An algorithm is *differentially private* if its distribution over outputs does not change much when swapping the values of a single data record

## Achieved by injecting carefully calibrated randomness

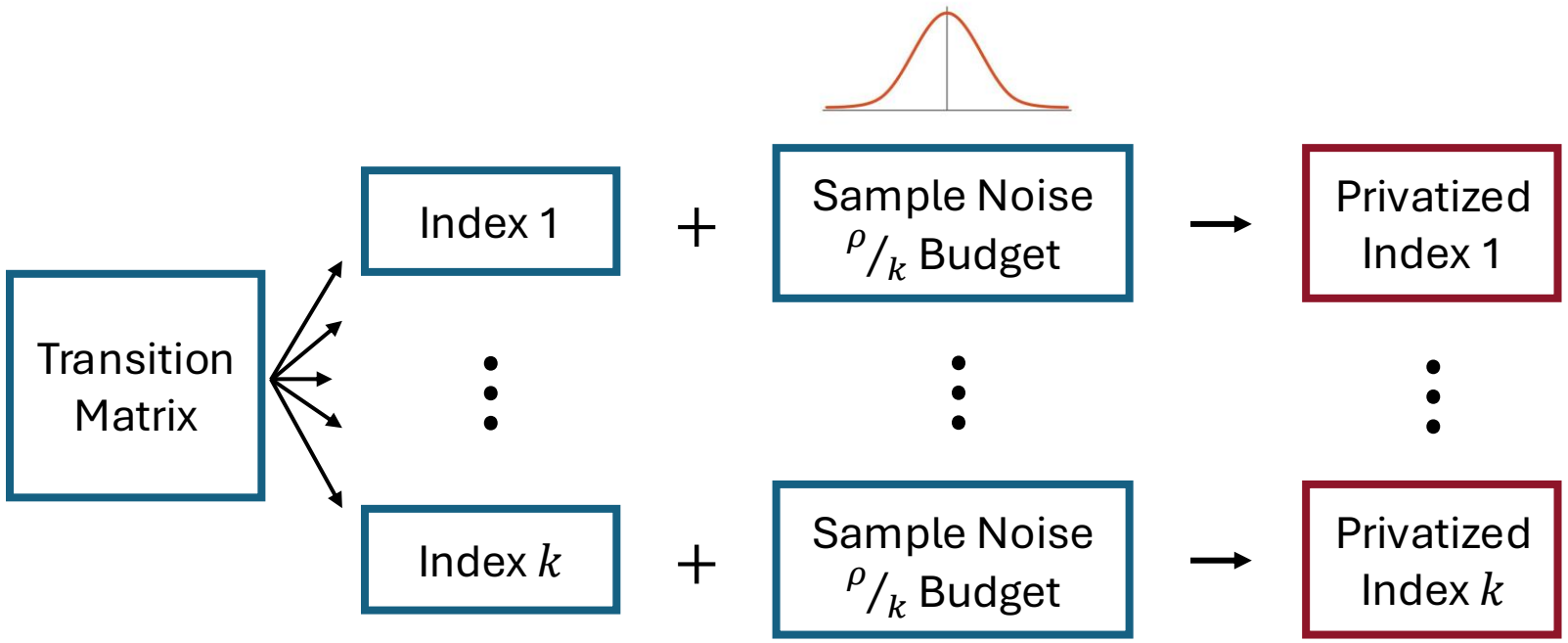
- Adding Gaussian noise (Gaussian mechanism)
- Privacy budget $\rho$ controls the privacy-accuracy tradeoff
- Lower $\rho$ implies stronger privacy (more noise)

## Use zCDP version of differential privacy

# Output Perturbation

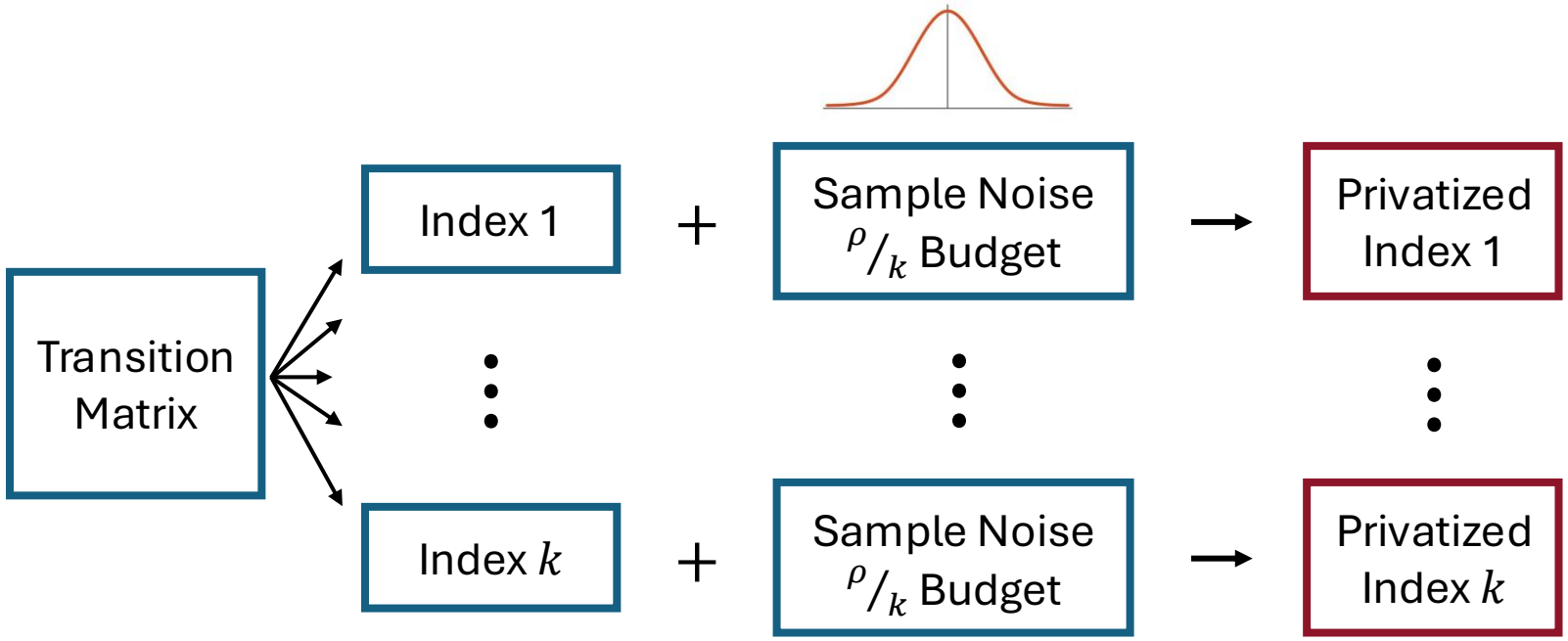
## Approach

- Compute each of $k$ indices from the raw data
- Add Gaussian noise calibrated to privacy budget $\rho/k$ to each index



# Limitations

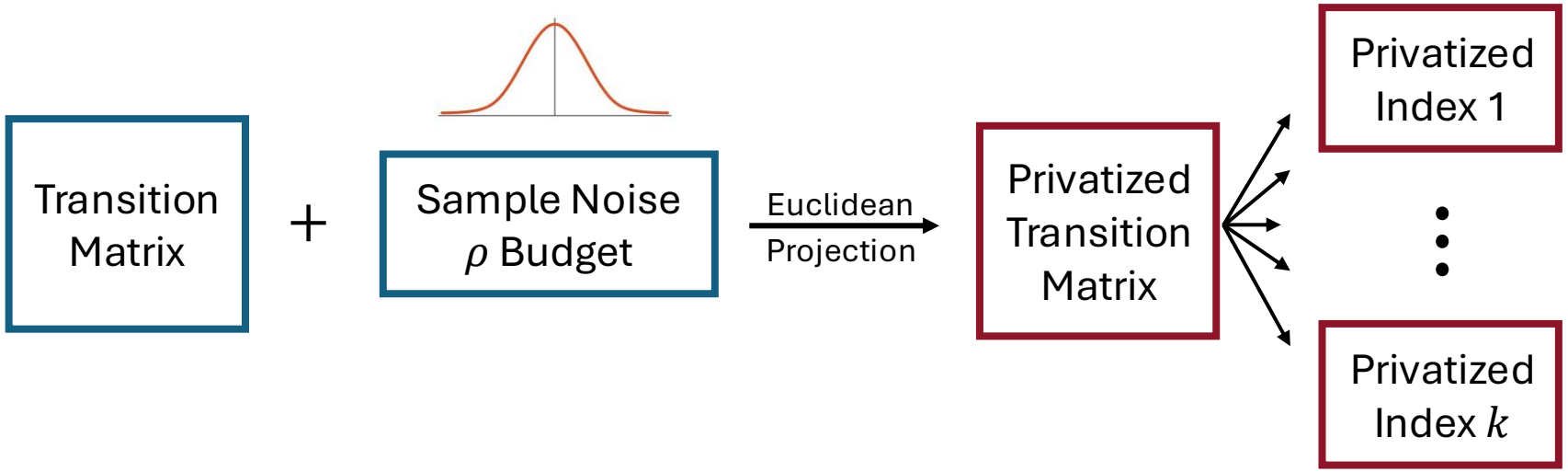
- Released indices may be **mutually inconsistent**
- Must **commit to indices** before accessing data
- **Budget fragments** as more indices are measured
- **Limited** to well-behaved indices



# Perturb and Project

## Approach

- Add Gaussian noise calibrated to privacy budget $\rho$ to transition matrix
- Project noisy transition matrix to a valid probability distribution (Euclidean projection)
- Compute indices from privatized transition matrix



# Perturb and Project

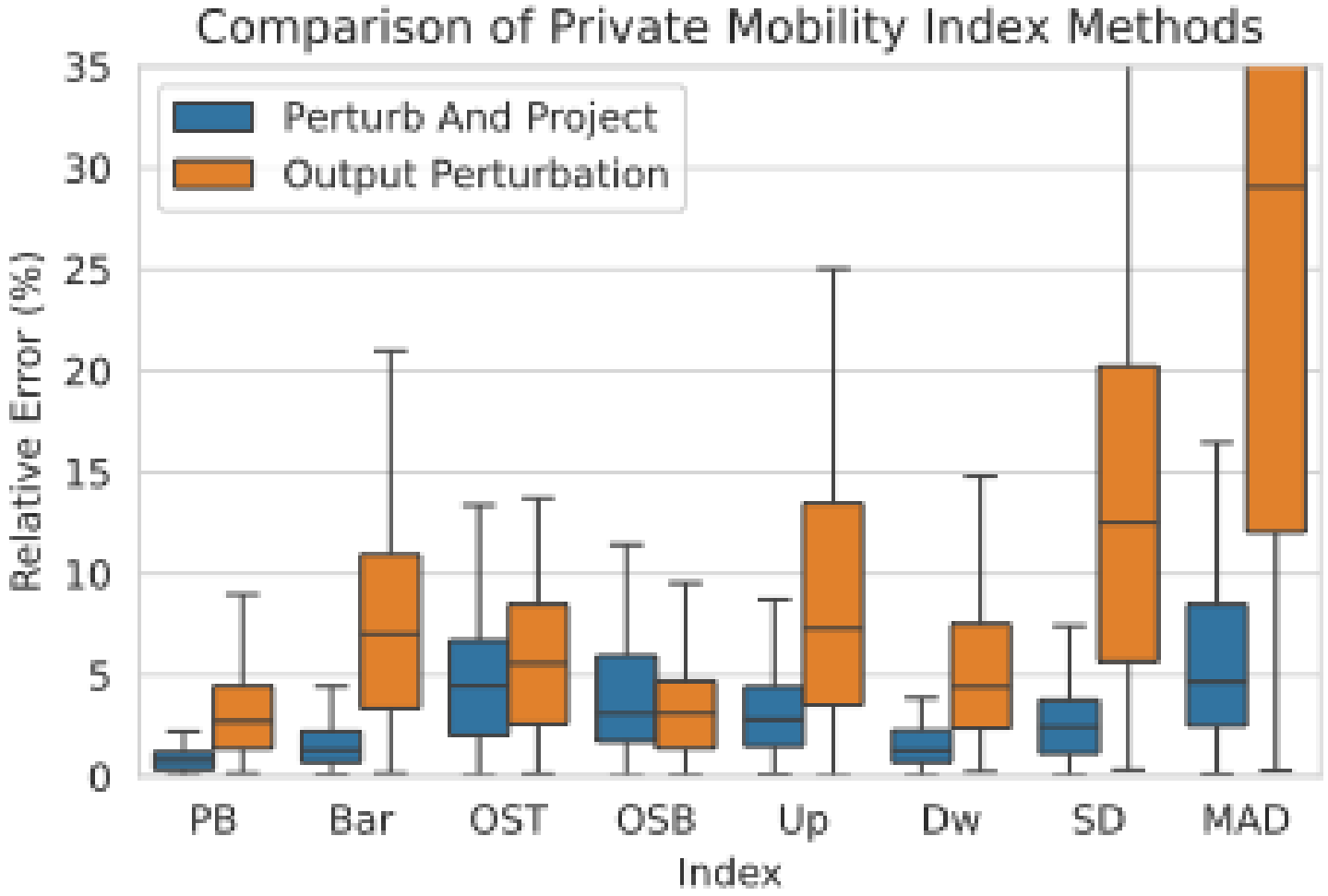
## Addresses limitations of Output Perturbation

- **Consistency**: all indices released are mutually consistent
- **Flexibility**: new indices can be released without incurring additional privacy cost
- **Robustness**: supports all indices, e.g., indices with unbounded sensitivity

## Additional Benefits

- **Transparency**: privatized transition matrix can be released

# Empirical Evaluation




Brett Mullins, David L Sjoquist, and Sally Wallace. “Earnings mobility and the Great Recession” Social Science Quarterly (2021)

# Extensions

## Transition matrices using quantiles

- Allocate part of privacy budget to estimate quantiles (e.g. JointExp)
- Preliminary experiments suggest allocating ~10% of budget to quantile estimation

## Multi-period mobility

- Track indices over time (2017-2018, 2017-2019, …)
- Sliding window (2017-2020, 2018-2021, …)
- Adapting Perturb and Project requires sophisticated query answering techniques
  - e.g. ResidualPlanner, GReM-LNN
- Preliminary results suggest even greater improvement over Output Perturbation

Jennifer Gillenwater, Matthew Joseph, and Alex Kulesza. “Differentially private quantiles” ICML (2021)
Xiao, et al. "An optimal and scalable matrix mechanism for noisy marginals under convex loss functions" NeurIPS (2023)
Mullins, Fuentes, et al. "Efficient and private marginal reconstruction with local non-negativity" NeurIPS (2024)

# Takeaways

- Practitioners can publish mobility indices with a formal privacy guarantee
- Perturb and Project offers a compelling combination of accuracy, transparency, and flexibility
- Provides a practical strategy for statistical agencies working with sensitive administrative records

# Thank You

Questions?

Contact: bmullins@umass.edu