Survey Weighting with Differentially Private Releases of Population Data

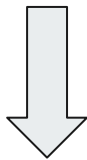Bhaven Patel
Anthony Rentsch

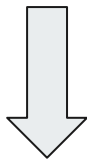# Motivation

Census + differential privacy

⬇

Effect on statistical calibration methods that rely on population data

⬇

Case study: survey weighting

Data

- Cooperative Congressional Election Study (CCES)
  - Large-N political survey fielded during election years
- American Community Survey
  - Census data product widely considered to be a reliable population benchmark
  - Public Use Microsample (PUMS) data that we worked with in class is a sample of the ACS
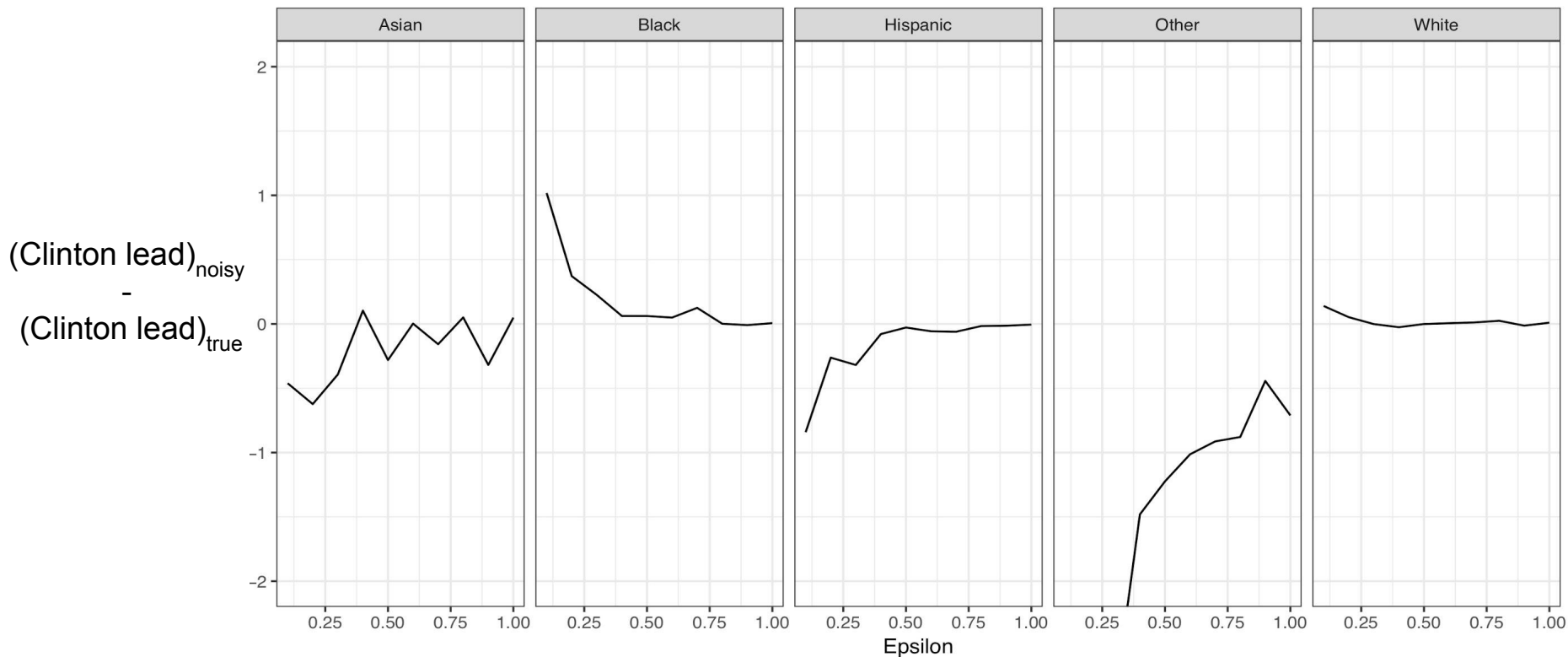
ACS data (PUMS)

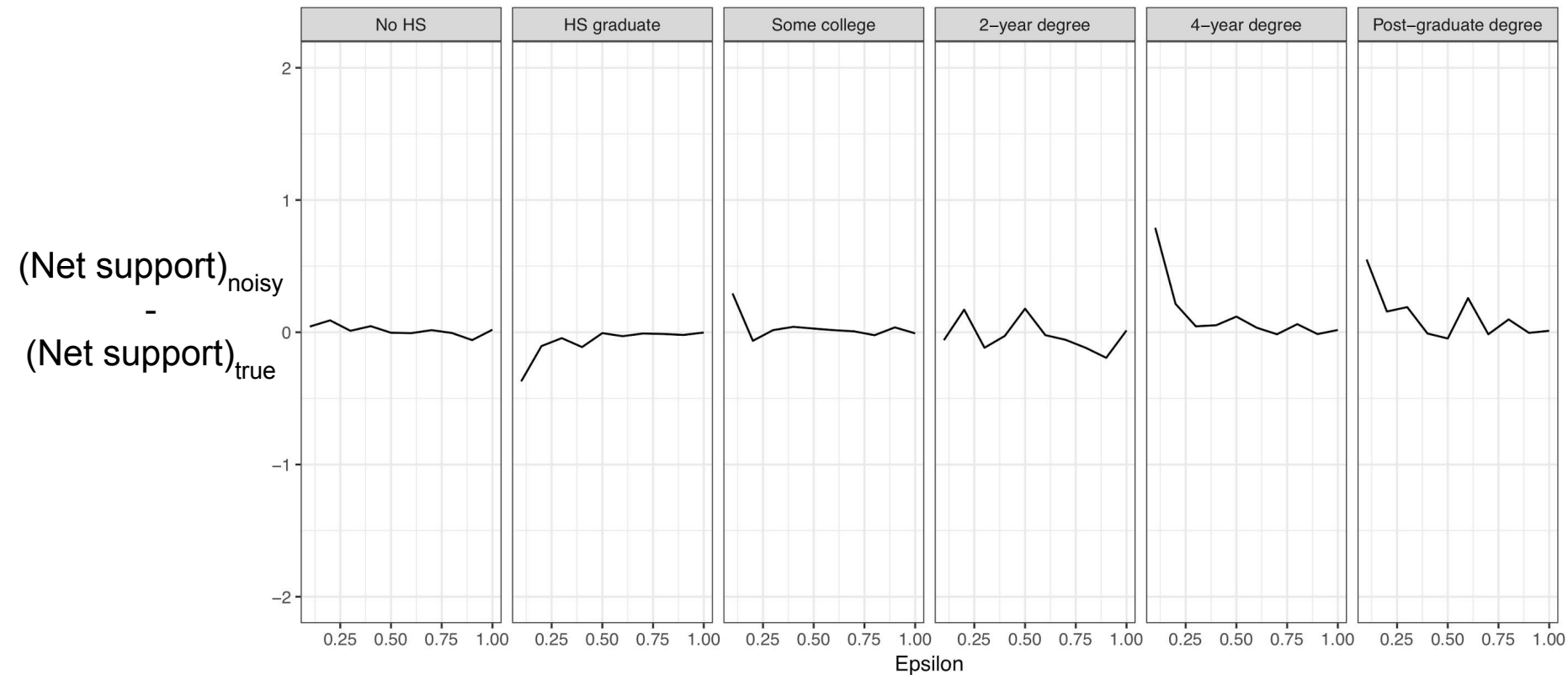| State	Age	Sex	Education	Race	Person weight
CA	32	1	4-year Bachelor	White	1
MA	21	2	HS graduate	Black	20
WY	57	1	No HS	Asian	549
TX	26	2	Some College	Hispanic	87
LA	45	2	Postgraduate degree	Other	65

Performing a DP-release of ACS data

1. Estimate joint distribution of population by aggregating PUMS data
    a. Joint distribution = cell counts of all *state* x *gender* x *race* x *age* x *education* combinations
2. Add Laplace noise to cell counts
3. Compute post-stratification weights for the CCES
4. Estimate quantity of interest from CCES
    a. e.g., Clinton - Trump vote share, net support for assault rifle ban

Results - 2016 election candidate preferences

Results - support for assault rifle ban

$(\text{Net support})_{noisy}$
$-$
$(\text{Net support})_{true}$

No HS | HS graduate | Some college | 2-year degree | 4-year degree | Post-graduate degree

Epsilon

Conclusions

- Small changes in sub-group level estimates once DP implemented
- Present stylized case for understanding how DP-ACS data will affect survey weighting
- Future work
  - Account for DP noise in post-stratification algorithm
  - Try clipping PUMS weights
  - Model ACS synthetic data generation