Assessing the Relationship Between Balanced Sample and Sample Representativity[*]

Yonil Park, John Chesnut

Demographic Statistical Methods Division
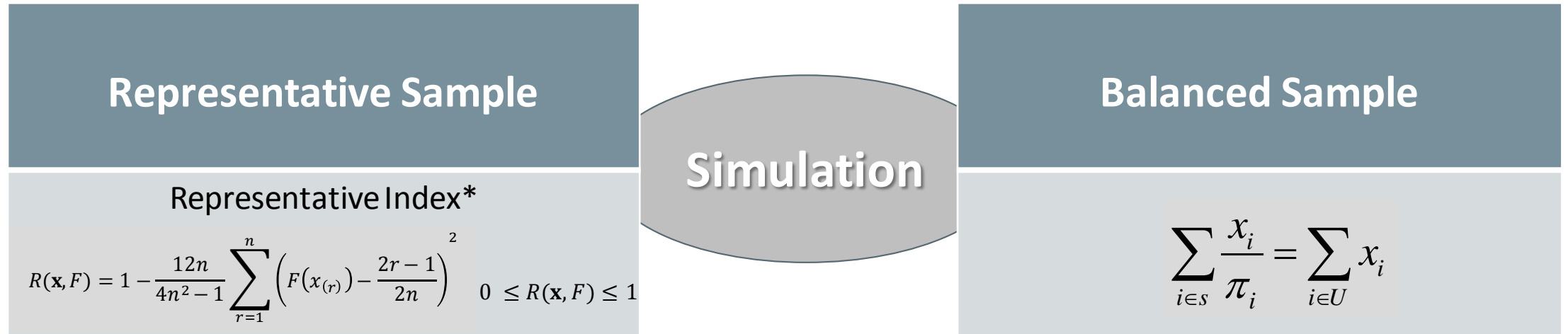
U.S. Census Bureau

Motivation

2010 Sample Redesign Sample Databases for Demographic Surveys
- Current Population Survey
- Survey of Income and Program Participation
- American Housing Survey

Representative Sample	Balanced Sample
Conventional Term	Mathematical Definition

"No supporting evidence to the relationship
between Representative Sample and Balanced Sample"

United States™ Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Methodology

Representative Sample		Balanced Sample
Representative Index*	**Simulation**	

$$R(\mathbf{x}, F) = 1 - \frac{12n}{4n^2 - 1} \sum_{r=1}^{n} \left(F(x_{(r)}) - \frac{2r-1}{2n} \right)^2 \qquad 0 \le R(\mathbf{x}, F) \le 1$$

$$\sum_{i \in s} \frac{x_i}{\pi_i} = \sum_{i \in U} x_i$$
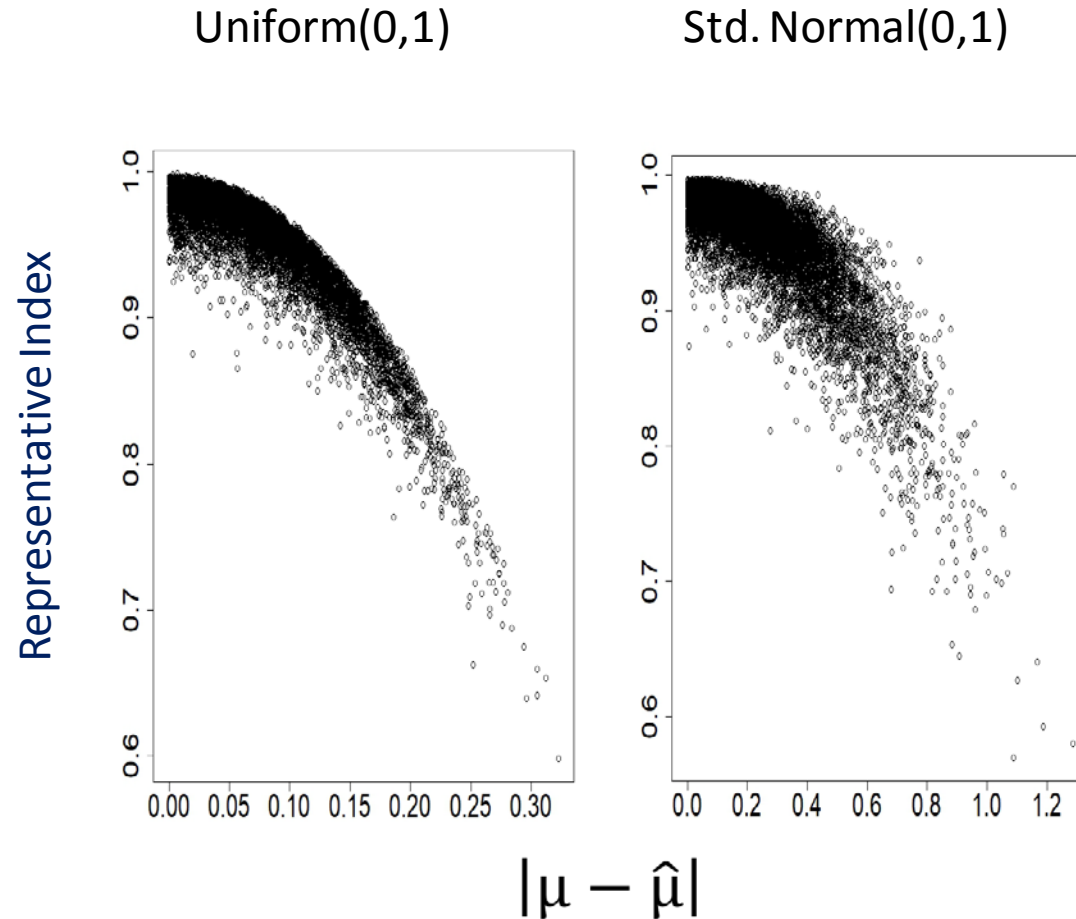
* Bertino, S. (2006), A Measure of Representativeness of a Sample for Inferential Purposes. International Statistical Review, 74: 149-159.

Simulation Study Results

10,000 runs with each
sample size = 10

Uniform(0,1)   Std. Normal(0,1)

Representative Index

$$|\mu - \hat{\mu}|$$

Evaluation of the 2010 Sample Redesign PSUs

American
Housing Survey
(Tenure)

Current Population
Survey
(Unemployment)

Survey of Income and
Program Participation
(Poverty)

Representative Index
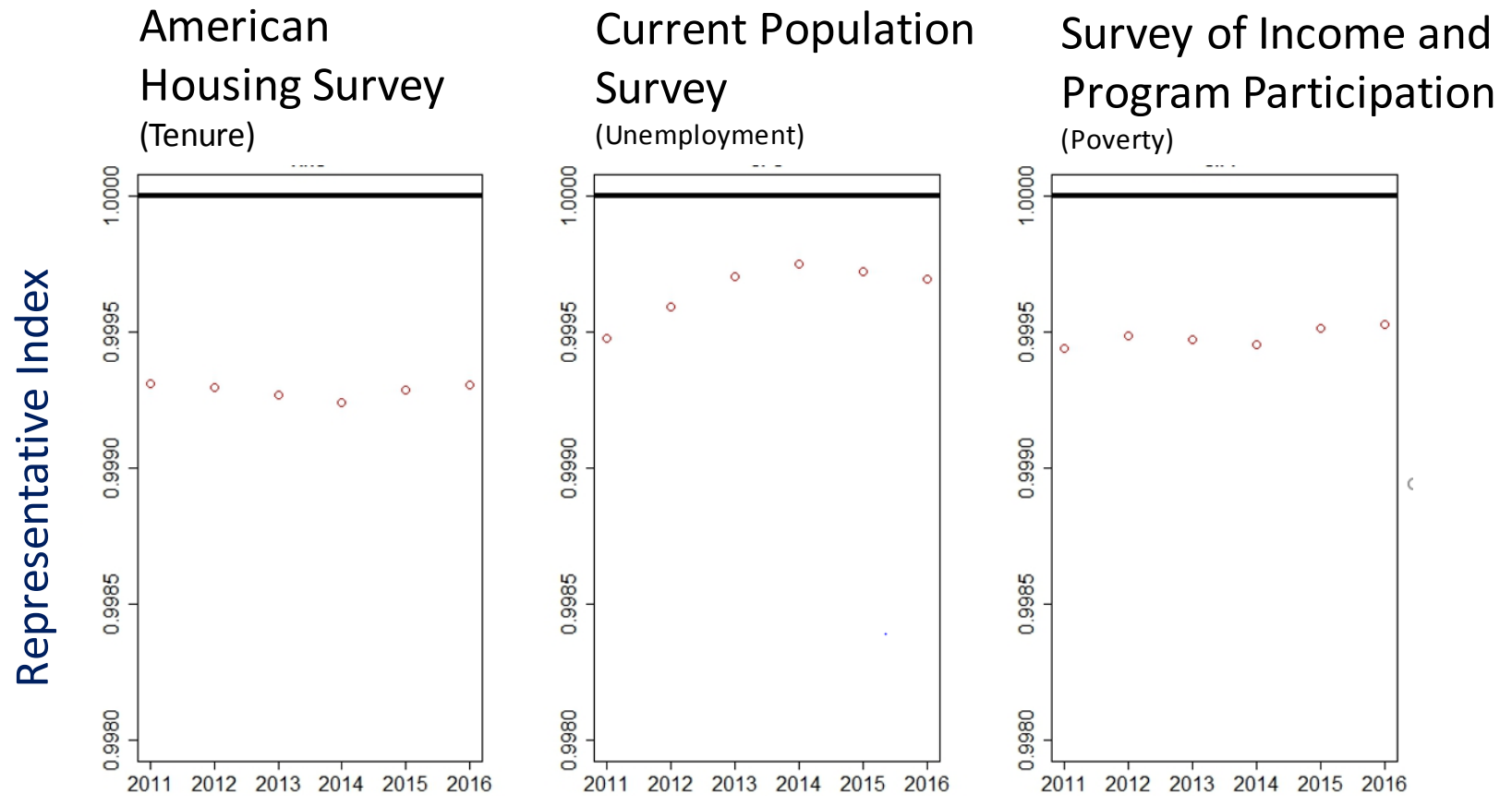
o : probability proportional to size (Sample Redesign PSUs)

Conclusion

➢ There is a strong positive relationship between balanced sample and representative sample.

➢ 2010 Sample Redesign PSUs showed very high Representative Index (RI).

➢ Length-biased correction to RI has a potential to improve the utility of the RI for a simple random sampling.