Strategies for Subsampling Nonrespondents for Economic Programs

Stephen J. Kaputa, Laura Bechtel, and Katherine Jenny Thompson

Motivation

- Surveys are experiencing…
  - Declining response rates
  - Severe reductions in funding

- But, are encouraged to…
  - Maintain quality
  - Collect more data items
  - Publish more statistics

Motivation

- Consequently, adaptive collection design strategies are being investigated to mitigate…
  - Cost
  - Imprecision from reduced size sample
  - Nonresponse bias

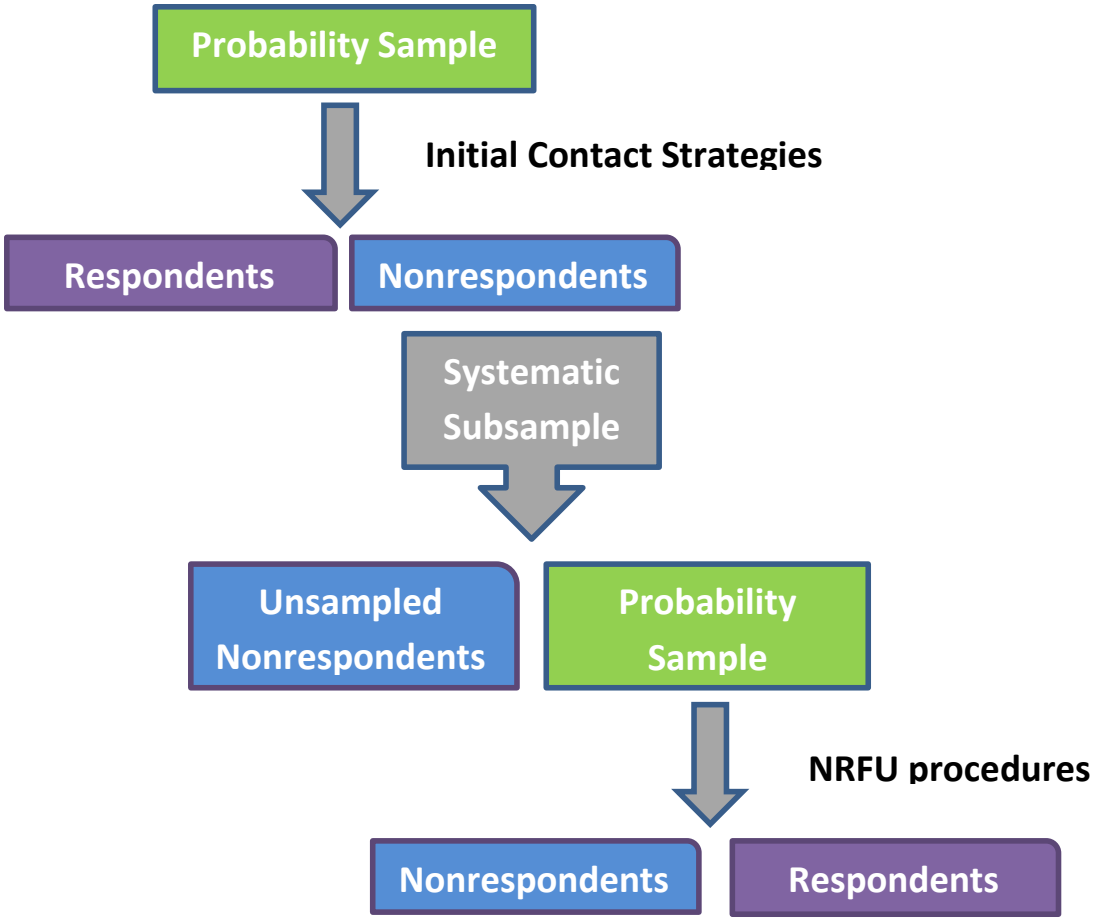
Objective

- Objective…

  Obtaining a set of respondents that are a random subsample of nonrespondents

- Why…
  - Focusing contact efforts on the subsample, we hope to decrease the effects of nonresponse bias on the estimated totals by obtaining data from all types of nonresponding units

Survey Design



Economic Programs

- Nonresponse follow-up (NRFU) varies by type of unit
  - Focusing on obtaining respondent data from the largest or most difficult to impute cases
  - Large units are expected to contribute substantially to the survey total
  - Small units rarely receive personal contact

Administrative Constraints

- Large units are excluded from consideration due to their high expected contribution to industry total

- Mandated lower bound on the unit response rate, along with targeted industry-specific response rates

Nonresponse Subsample Allocation

- Equal probability (1-in-K) sampling
- Optimal Allocation that…
  - minimizes deviation between industry unit response rates
  - Or, minimizes deviation between industry sampling intervals

Constant-K

- "Across the board" systematic sample of nonrespondents
  - Sorted by unit size
  - Helps obtain a subsample that better resembles the originally designed sample

Optimal Allocation

- Unit response rates are computed with respect to all units, but only small units are subsampled
- Nonresponse conversion probability is assumed constant for optimal allocations
- We formulate the allocation program as a quadratic program with linear constraints

Min-URR

Objective Function	Minimize (squared) deviation between stratum response rate and overall target response rate
Constraints	Overall sampling rate for subsample cannot be larger than the 1-in-K subsample
	Cannot subsample more units than present in domain
	Cannot take a negative subsample

Note: Unit response rates must be estimated

- Total number of sampled units in denominator
- Numerator = respondents (not subsampled) + estimated respondents from subsample

(Estimated respondents requires an assumed nonrespondent conversion rate)

Min-K

Pre-Condition	If overall target response rate cannot be achieved with assumed conversion rate then take full subsample
	If overall target response rate has been achieved then end follow-up
Objective Function	Minimize (squared) deviation between stratum sampling rate and overall sampling rate
Constraints	Overall sampling rate for subsample cannot be larger than the 1-in-K subsample
	Cannot subsample more units than present in domain
	Cannot take a negative subsample
	Overall target response rate must be met

Estimation

$$\hat{Y} = \hat{Y}_R + \hat{Y}_{NR}$$

Unadjusted estimate from original respondents

Adjusted estimate from subsampled nonrespondents

For $\hat{Y}_{NR}$, we tested

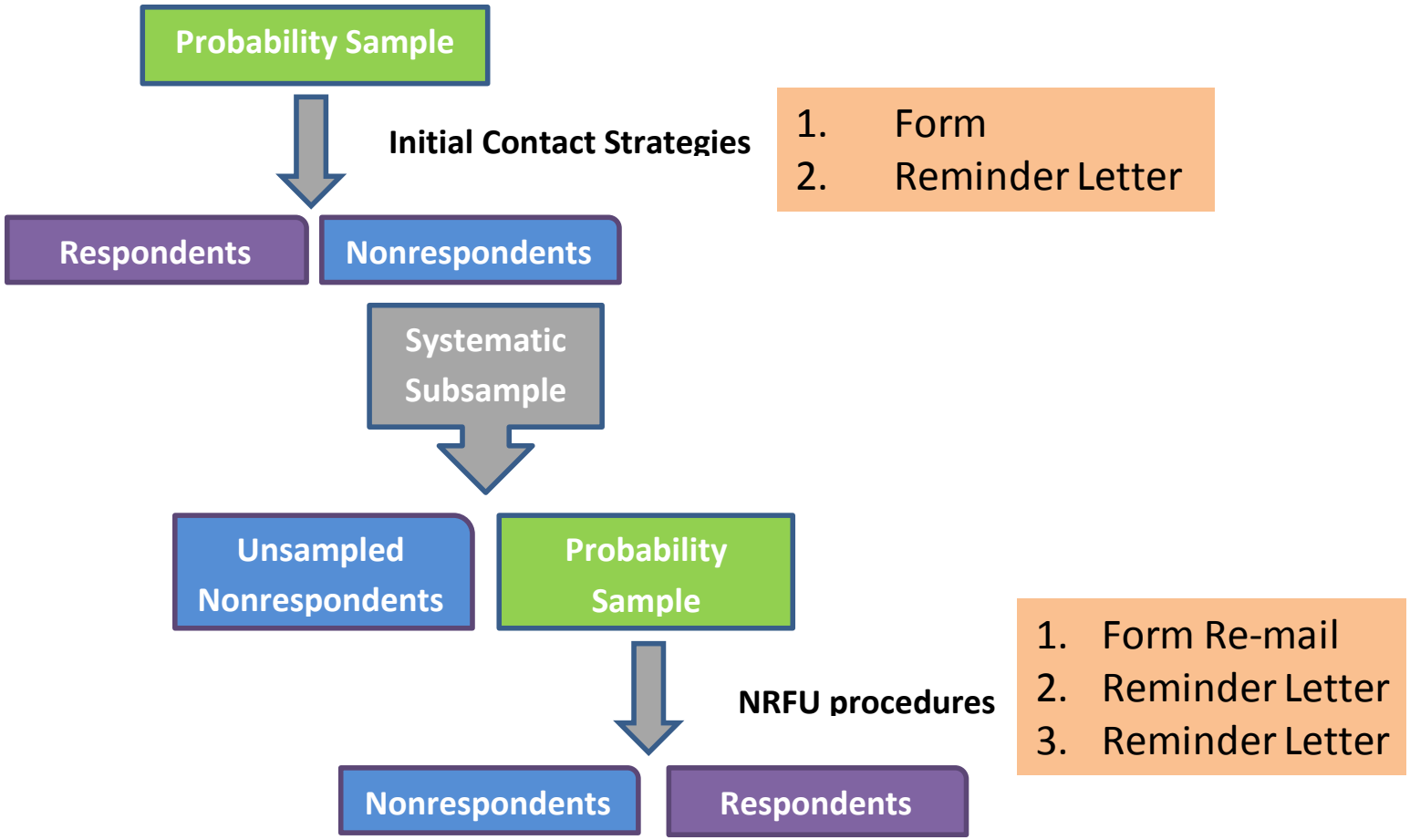
- Simple expansion estimator (reweighted)
- Separate ratio estimator
- Combined ratio estimator (**best results**)

Simulation Study

- Annual Survey of Manufactures (ASM)
  - Pareto-PPS establishment survey designed to produce "sample estimates of statistics for all manufacturing establishments with one or more paid employee(s)"

- The ASM questionnaire is a subset of the manufactures sector questionnaire in the Economic Census (Great for testing)

When to Subsample (ASM)



ASM Allocations

- Optimal allocations were performed once using historic ASM data for different conversion rates
  - Constant “assumed” conversion rates of 0.30, 0.50, and 0.70
  - A single set of estimated stratum level conversion rates

- If an optimal solution could not be obtained, we adjusted the target response rate to obtain a feasible solution (Min-K)

Simulation Study

1. Using historic check-in-rates, randomly induce nonresponse in the complete dataset

2. Sort the remaining nonrespondents by sampling weight

3. Select a stratified systematic sample using the subsampling rates for a given allocation strategy

4. Simulate unit response for each round of NRFU. After assigning response status to each unit compute diagnostic statistics

5. For each allocation, repeat Step 4 until either ten rounds of follow-up have been conducted or the total budget has been expended.

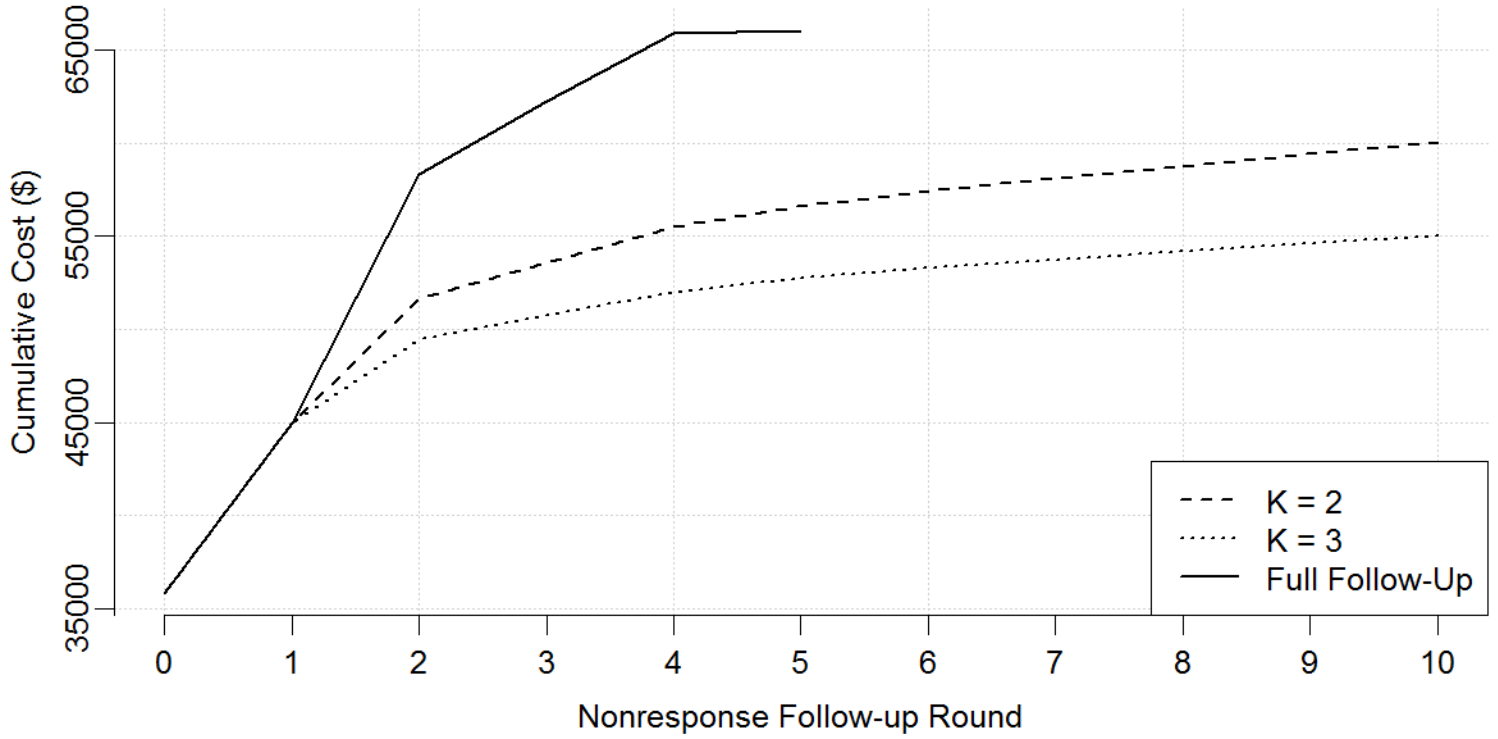
1-in-3 Subsampling Rate

- Selecting the smallest subsample provides the largest cost savings for all allocations

- But, yields estimates that are not of good quality as measured by
  - Response Rates
  - Variance
  - MSE

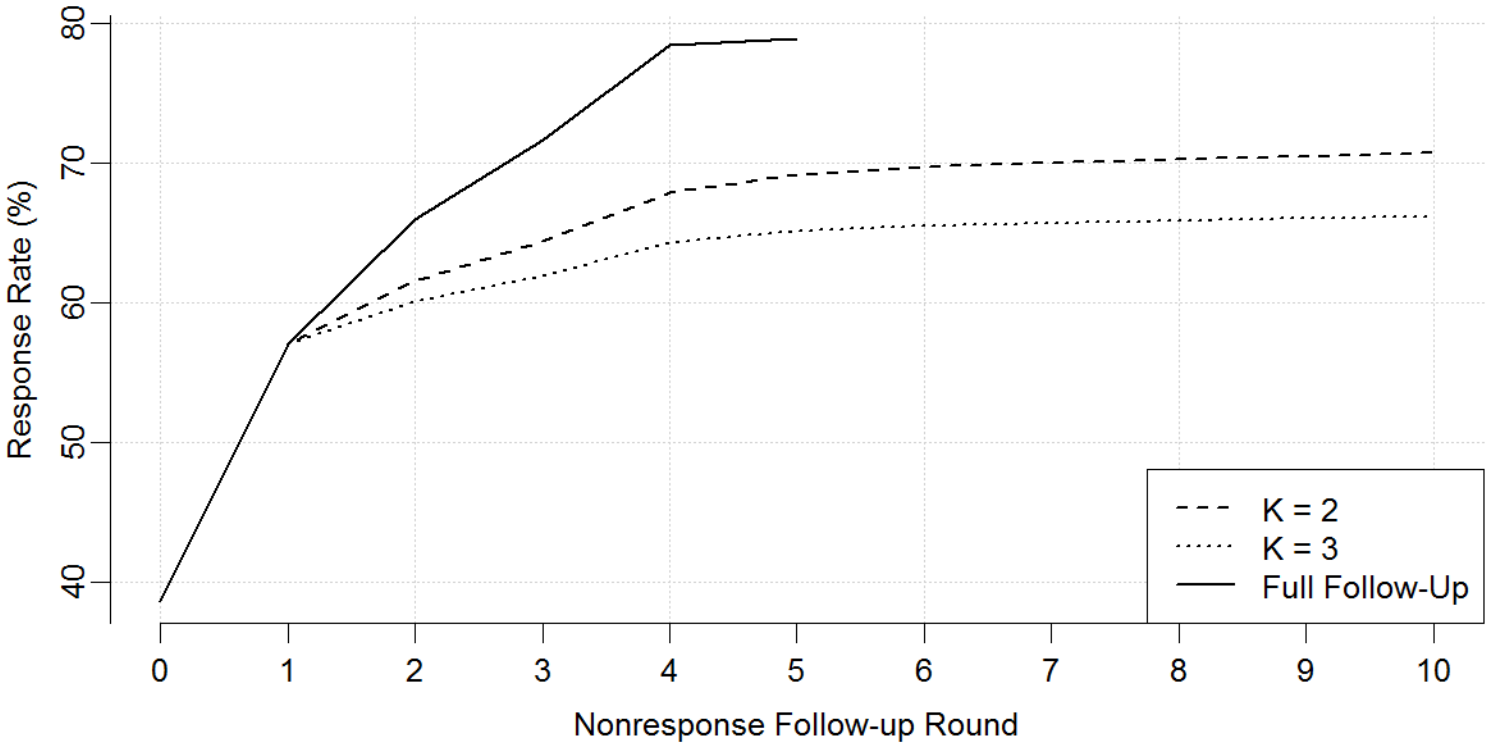
Cumulative Cost

Across the board systematic sample of nonrespondents



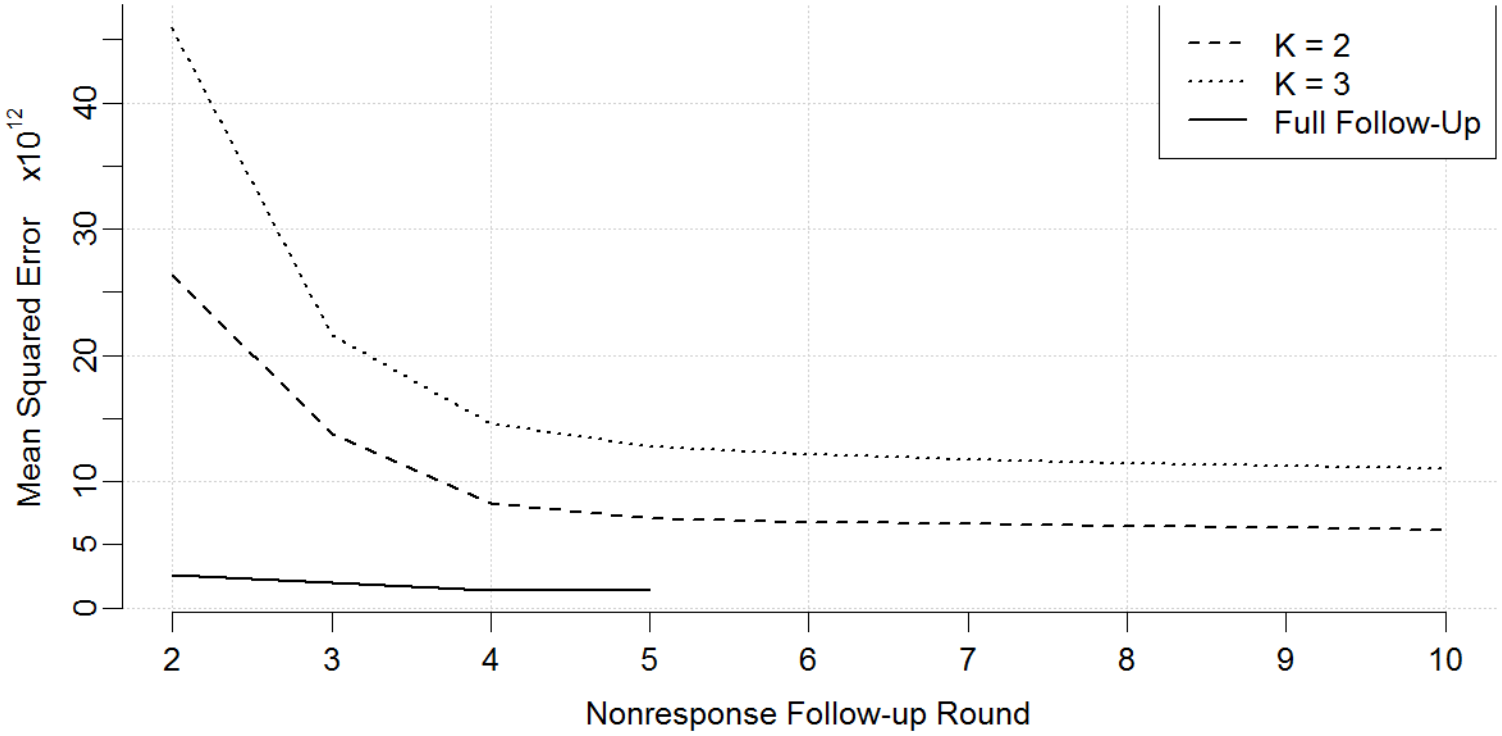
Response Rates

Across the board systematic sample of nonrespondents



Mean Squared Error

Across the board systematic subsample with the combined ratio estimator

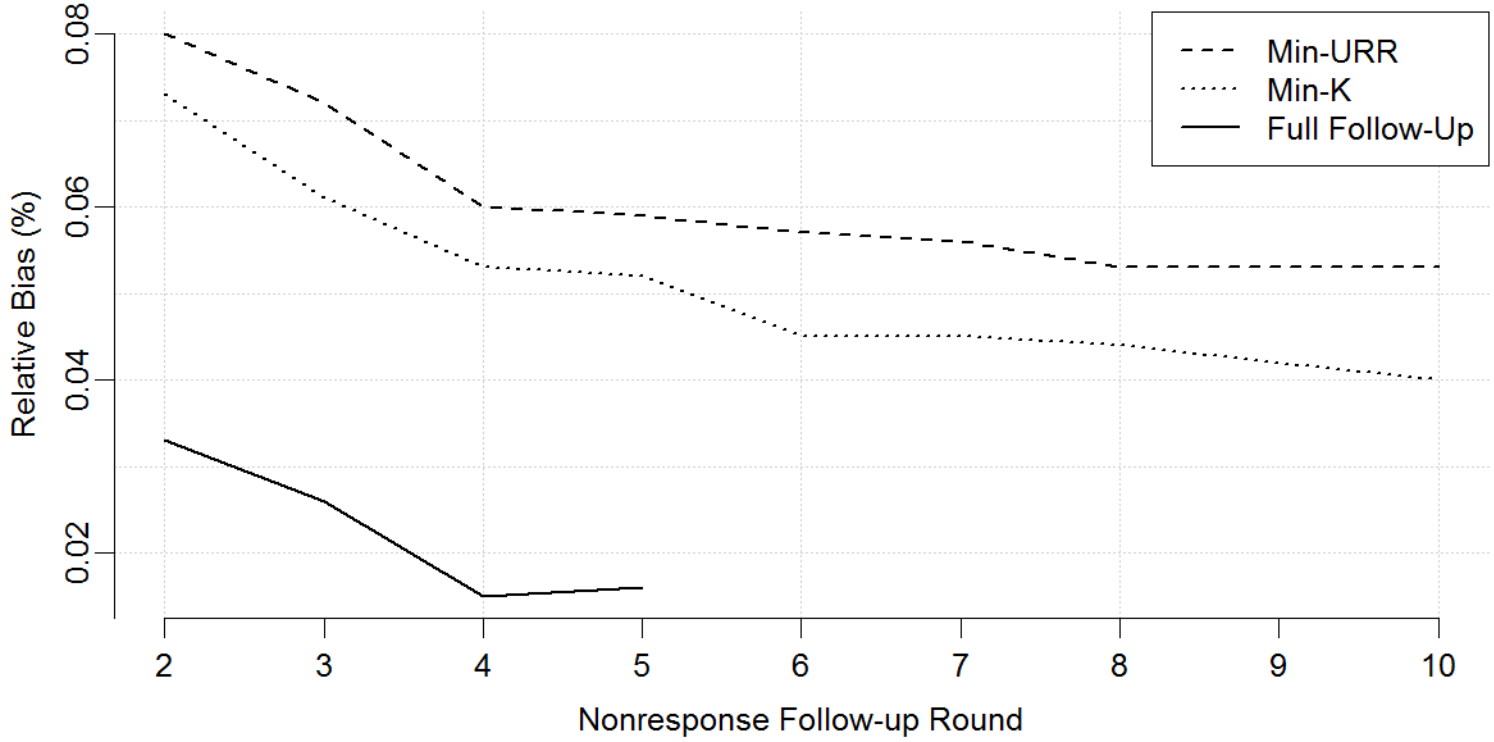


1-in-2 Subsampling Rate

- No method or estimator approaches the quality of full NRFU
- The optimal allocation methods with the combined ratio estimator performed the best
- The estimated conversion rate and constant conversion rates of 0.30 and 0.50 all performed well
  - A constant rate of 0.70 never performed well
  - We recommend using the estimated conversion rate

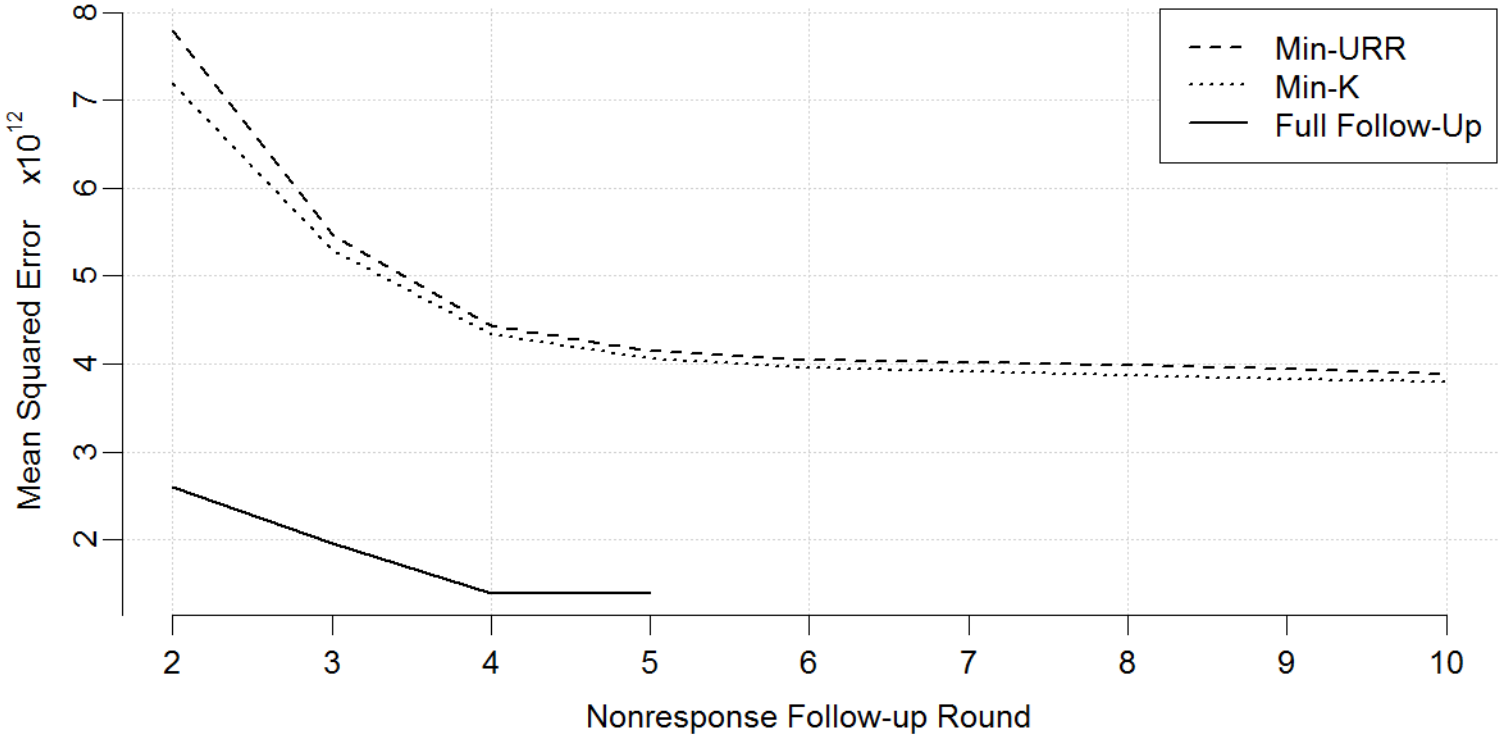
Relative Bias of the Estimate

Combined ratio estimator with an estimated conversion rate



Mean Squared Error

Combined ratio estimator with an estimated conversion rate



Conclusion

- Emphasis on obtaining responses from the larger units at the cost of the lower unit response in turn creates a bias in the estimates
  - Imputed or adjusted values for smaller units resemble the large unit values

- Subsampling nonrespondents with optimal allocation methods increases the potential of obtaining a representative sample by targeting the low responding areas

Conclusion

- But, subsampling nonrespondents without changing the data collection procedure may have minimal tangible benefits besides cost reduction

- Subsampling nonrespondents paired with a new contact strategy for these “hard to reach” establishments would create a truly adaptive approach

Thank You

Stephen J. Kaputa

Statistical Methods and Sample Design Staff
Economic Statistical Methods Division
U.S. Census Bureau, Washington, DC 20233

Stephen.Kaputa@census.gov