# Strategies for Subsampling Nonrespondents for Economic Programs

Stephen J. Kaputa[1], Laura Bechtel, Katherine Jenny Thompson, and Daniel Whitehead

**Abstract**
Adaptive design strategies for data collection can increase the quality of response data under a reduced survey budget. In this framework, the U.S. Census Bureau is investigating nonresponse subsampling strategies, including a systematic sample of nonrespondents sorted by a measure of size, for usage in the 2017 Economic Census. Design constraints include a mandated lower bound on the Census unit response rate, along with targeted industry-specific response rates. This paper presents research on allocation procedures for subsampling nonrespondents, given a systematic subsample. We consider two approaches: (1) equal-probability sampling and (2) optimal allocation with constraints on unit response rates and sample size with the objective of selecting larger samples in industries that have initially lower response rates. Using the Annual Survey of Manufactures (ASM) sample as our original population, we present a simulation study that examines the cost, variance, relative bias, and unit response rates for the proposed allocations, assessing each procedure's sensitivity by varying the program-level sampling interval, the response mechanism, and the nonresponse adjusted estimator.

**Key words:** quadratic program, unit response rate, nonresponse adjustment, response mechanism

## 1. Introduction

Many federal programs are experiencing declining response rates along with severe reductions in survey funding. At the same time, these programs are required to maintain predetermined reliability levels and are often encouraged to collect an increased number of data items and to publish more statistics. Of course, statistics produced from a reduced sample size can be imprecise (i.e., quite variable) and can be quite sensitive to nonresponse bias. Consequently, federal agencies such as the U.S. Census Bureau are investigating adaptive collection design strategies, where the term "collection design" refers to protocol(s) for collecting current data.

With business surveys, the collection design may vary by type of unit. For the large units that are expected to contribute substantively to the survey totals, the nonresponse follow-up (NRFU) procedures become progressively more expensive in terms of cost per unit with the exception of the final contact attempt. In contrast, with the smaller units, the most expensive NRFU procedure occurs early in the data collection cycle and the follow-up procedures do not include personal contact.

---

[1] Office of Statistical Methods and Research for Economic Programs, U.S. Census Bureau, Washington, DC 20233 (Stephen.Kaputa@census.gov) This report is released to inform interested parties of research and to encourage discussion. Any views expressed on methodological or operational issues are those of the author and not necessarily those of the U.S. Census Bureau.

With an adaptive collection design, the data collection procedures can change (*adapt*) during the collection period; paradata and sample data are used to determine whether to change the current procedures (Schouten, Calinescu and Luiten, 2013). The overall budget is fixed and collection strategies are developed, but the implementation of a given strategy depends on (1) the realized sample of respondents at a point in time, (2) informative data obtained during data collection about the respondents and nonrespondents, and (3) information known in advance about the survey unit from the sampling frame. Consequently, selecting a probability sample of nonrespondents for NRFU falls under the adaptive design umbrella.

The U.S. Census Bureau is investigating nonrespondent subsampling strategies for usage in the 2017 Economic Census. The proposal under consideration is to implement a systematic sample of small single unit (SU) nonrespondents sorted by a measure of size. Design constraints include a *mandated* lower bound on the Economic Census unit response rate, along with targeted industry-specific response rates. Large single and multi unit (MU) establishments are excluded from consideration due to their high expected contribution to industry totals.

Because implementing a probability subsample of nonrespondents for NRFU represents a major procedural change from the usual full follow-up, a field test is planned for the 2015 Annual Survey of Manufactures (ASM), a PPS sample of establishments whose questionnaire is extracted from the manufactures questionnaires for the Economic Census. Using the sampled single unit (SU) establishments from the ASM, Whitehead, Kaputa, and Thompson (2013) present a simulation study that examined the cost and unit response rates at each stage of data collection under a systematic sample with a constant sampling interval, and considered the variance, relative bias, and mean squared error (MSE) of the reweighted expansion estimator under each allocation. A parallel study by Bechtel and Thompson (2013) using 2007 Economic Census data found that the targeted industry unit response rates of 70% could only be achieved in a 1-in-3 subsample if the average unit response rate in most Economic Census industries is 60% or larger *before* follow-up begins. Unfortunately, cost savings from subsampling SU nonrespondents are realized when the sample is implemented after at most one round of NRFU, and the historical check-in rates at that stage of collection are approximately 40%.

This paper extends the research presented in Whitehead, Kaputa, and Thompson (2013). This paper presents research on alternative allocation methods for subsampling nonrespondents, given a systematic subsample. We consider two approaches: (1) equal-probability (1-in-K) sampling and (2) optimal allocation that minimizes deviation between industry unit response rates or industry sampling intervals with the objective of selecting larger systematic samples in industries that have initially lower response rates. Likewise, we consider two estimators: the reweighted double expansion estimator used in the previous study and a separate ratio estimator that adjusts for unit nonresponse using a covariate that is highly correlated with both response propensity and the survey characteristic of interest. Using the ASM sample as our original population, we assess each procedure's sensitivity by varying the program-level sampling intervals, the response mechanism, and the nonresponse adjusted estimator.

## 2. Methodology

## 2.1 Survey Design and Estimation

The framework for our research is the three-stage sample design shown in Figure 1. The first stage is a stratified probability sample of $n_h$ units from stratum of size $N_h$, with a total sample size of $n$. Sampling is performed **before** data collection begins. The survey is conducted, and units either respond or do not respond. At a predetermined point in the data collection cycle, we selected a 1-in-$K_h$ systematic subsample of $nr_{h1}$ of the $nr_h$ nonrespondents: this predetermined point can be determined as a fixed calendar date or via a responsive design protocol. Only the sampled $nr_{h1}$ units receive NRFU. Of these, $nr_{h2}$ units ultimately respond: we treat them as a Bernouilli sample of nonrespondents (Särndal et al., Ch.15 and Kott, 1994). Under a missing at random (MAR) response mechanism, the sampled nonrespondents that ultimately provide response data are a random subsample.

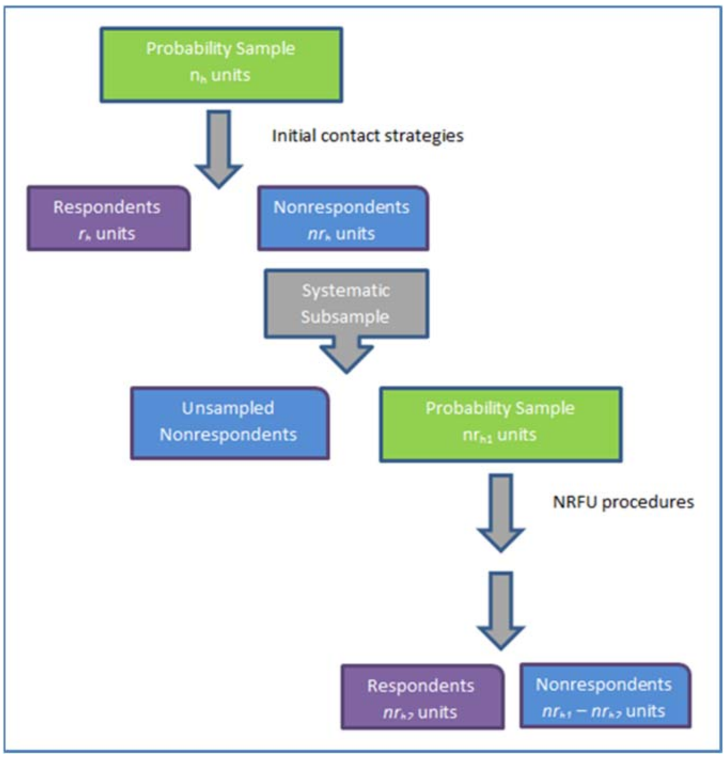


**Figure 1:** Nonrespondent subsample from probability sample, selected during data collections. Unsampled nonrespondents do not receive NRFU

Our objective is to estimate $\hat{Y}$, the population total of characteristic $y$, from the realized sample of respondents. Let

$S_{hi}$ = 1 if unit $i$ in stratum $h$ was in original sample; 0 otherwise
$\theta_{hi}$ = the probability of sampling unit $i$ in stratum $h$ into the original sample ($w_{hi} = 1/\theta_{hi}$)
$R_{hi}$ = 1 if unit $i$ in stratum $h$ provided a response before subsampling time $t$ (value for $y$); 0 otherwise
$I_{hi}$ = 1 if unit $i$ in stratum $h$ was selected for NRFU (i.e., was a subsampled nonrespondent); 0 otherwise
$J_{hi}$ = 1 if unit $i$ in stratum $h$ responds, given selection into nonrespondent subsample; 0 otherwise
$f_{hi}$ = adjustment factor for nonrespondent subsampling and unit nonresponse after NRFU

$y_{hi}$ = value of characteristic $y$ for unit $i$ in stratum $h$, available only for respondents
$x_{hi}$ = value of characteristic $x$ for unit $i$ in stratum $h$, available for all sampled units

Then $\hat{Y} = \sum_h \sum_i w_{hi}\, y_{hi} S_{hi} R_{hi} + \sum_h \sum_i w_{hi}\; f_{hi} y_{hi}\, S_{hi}(1 - R_{hi}) I_{hi} J_{hi} = \hat{Y}_R + \hat{Y}_{NR}$. We consider two different estimators of $\hat{Y}_{NR}$, each implementing a variant of the recommended reweighting procedure described in Brick (2013):

Double Expansion: $$\hat{Y}^{W}_{NR} = \sum_h \sum_{i \in h} w_{hi} K_h \left( \frac{nr_{h1}}{nr_{h2}} \right) y_{hi} S_{hi} (1 - R_{hi}) I_{hi} J_{hi} \qquad (2.1)$$

Ratio Estimator:

$$\hat{Y}^{R}_{NR} = \sum_h \sum_{i \in h} w_{hi} K_h \left( \frac{nr_{h1}}{nr_{h2}} \right) \left( \frac{\sum_{i \in nr_{h1}} w_{hi} K_h x_{hi}}{\sum_{i \in nr_{h2}} w_{hi} K_h \left( \frac{nr_{h1}}{nr_{h2}} \right) x_{hi}} \right) y_{hi} S_{hi} (1 - R_{hi}) I_{hi} J_{hi} \qquad (2.2)$$

In a given stratum, a minimum requirement for variance estimation is that $n_{rh2} \geq 2$. Ultimately, a careful sample design can often accommodate this requirement, but this is often not true during the early stages of NRFU collection. The double expansion estimator (2.1) is used for strata with $nr_{h2} \geq 2$ , strata with $nr_{h2} < 2$ are estimated using a reweighted expansion estimator ($\hat{Y}^{W}_{h}$) that includes the initial responders. With the ratio estimator ($\hat{Y}^{R}_{NR}$), we used the separate ratio estimation procedure (2.2) for the subsampled nonrespondents when $nr_{h2} \geq 2$. Otherwise (when $nr_{h2} <2$), we used a combined ratio estimator ($\hat{Y}^{R}_{h}$) that includes initial responders for that stratum. Where,

$$\hat{Y}^{W}_{h} = \sum_{i \in h} w_{hi} \left( \frac{n_h}{r_h} \right) y_{hi} S_{hi} R_{hi} \qquad (2.3)$$

$$\hat{Y}^{R}_{h} = \sum_{i \in h} w_{hi} \left( \frac{n_h}{r_h} \right) \left( \frac{\sum_{i \in r_1 + nr_1} w_{hi} x_{hi}}{\sum_{i \in r_1} w_{hi} \left( \frac{n_h}{r_h} \right) x_{hi}} \right) y_{hi} S_{hi} R_{hi} \qquad (2.4)$$

In the case study described below in Section 3, the separate ratio estimate is used to approximate the current procedure estimates (no nonrespondent subsampling, i.e., $K_h$=1).

To obtain variance estimates for the double expansion and ratio estimators, we used the linearization methods outlined in Binder et al. (2000), adopted for a three-stage sample, with the nonrespondent subsample representing a two-phase sample. For strata without subsampling or $nr_{h2} < 2$, we use the linearization methods for a two-phase sample. Expressions for the variance estimates used on our case study are available upon demand; we omit them as they are specifically developed for a first stage Poisson sample with a stratified systematic subsample of nonrespondents that have a missing at random (MAR) response mechanism.

## 2.2 Allocation Strategies

The objective of nonrespondent subsampling is to obtain a set of respondents that are a random subsample (i.e., are a representative subsample). When all cases are subjected to

NRFU, respondent contact strategies focus on improving response rates, and analysts may focus primarily on obtaining responses from soft refusal cases that have similar characteristics. With a probability sample, the targeted cases represent a cross-section of the refusal population. By 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 and by using weighting or imputation methods (Brick, 2013).

With a business survey that collects little or no demographic information, we often have very little information on the nonrespondents to use for the subsample design. We have frame information, such as industry and unit size (e.g., total payroll, total value of shipments) and we have response status. By sorting the nonrespondents within strata by unit size and selecting a systematic sample, we hope to obtain a subsample that better resembles the originally designed sample (Lohr, 2009).

Here, we consider two allocation approaches for our systematic sample: (1) equal-probability sampling; and (2) optimal allocation with constraints on unit response rates and sample size with the objective of selecting larger samples in industries that have initially lower response rates. Equal probability sampling is easy to implement and should have the lowest sampling variance among these allocations. However, since the same proportion of nonrespondents are sampled in each stratum, it can produce biased estimated totals especially when there are “hard to reach” populations in selected strata.

Our optimal allocation methods addresses the latter concern by concentrating NRFU efforts in strata that have low response rates, attempting to select sufficient cases to achieve the performance benchmarks. This strategy is designed to reduce the effect of nonresponse bias on the totals. However, it can lead to increased variances, as the subsampling intervals will differ. To minimize the additional sampling variance caused by differing sampling intervals, the strata sampling intervals should be close to a single constant sampling interval (K). To control costs, the optimal allocation should not select more units from NRFU than the budgeted 1-in-*K* subsample for the program. We recognize that there other criteria that could be considered, such as minimizing the nonresponse sampling variance (Haziza and Beaumont, 2011). However, our application context requires that target response be achieved or nearly achieved as mandated.

We formulate optimal allocation as a quadratic program and consider two different objective functions. The first quadratic program minimizes the squared deviation in stratum unit response rates from the *target unit response rate* (URR) and is subject to the same linear constraints on the sample size of nonrespondents. The second quadratic program minimizes the squared deviation in stratum sampling intervals from the *target constant interval* (*K)* subject to linear constraints on the unit response rates in each stratum and on the number of sampled nonrespondents. Hereafter, we refer to the allocations obtained from these quadratic programs as Min-URR and Min-K, respectively.

Constraints (1) through (3) in Table 1 are used in both quadratic programs. Constraint (4) is included in the Min-K allocation to ensure that the realized subsample attains the required target unit response rates as feasible. In reality, the only constraint submitted to the quadratic program is the restriction on the strata response rates. This constraint ensures that the optimization solution is not $K_h = K$ for all strata *h*. There are two limiting scenarios that have to be addressed before the optimization. First, strata that have achieved the target unit response rate *before* subsampling must be removed from the optimization problem ($K_h = \infty$). Second, if a target unit response rate cannot be achieved given the number of nonrespondents prior to subsampling for an assumed nonrespondent conversion rate, $q_h$, then all units in the strata are selected for NRFU ($K_h = 1$).

**Table 1:** Optimal Allocation Quadratic Programs

| | | Min-URR | Min-K | Purpose |
|---|---|---|---|---|
| Objective Function | | $\min \sum_h (URR_h - URR)^2$ <br> $= \min \sum_h \left( \frac{r_h + q_h nr_{h1}}{n_h} - URR \right)^2$ | $\min \sum_h (K_h - K)^2$ <br> $= \min \sum_h \left( \frac{nr_h}{nr_{h1}} - K \right)^2$ | |
| Constraints | 1 | $K \le \frac{\sum_h nr_h}{\sum_h nr_{h1}}$ | | subsample size cannot exceed overall 1-in-K size |
| | 2 | $\frac{nr_h}{nr_{h1}} = K_h \ge 1$ | | subsample cannot exceed number of nonrespondents |
| | 3 | $nr_{h1} \ge 0$ | | Non-negativity constraint |
| | 4 | | $\frac{r_h + q_h nr_h}{n_h} < URR \quad K_h = 1$ <br> $\frac{r_h}{n_h} \ge URR \quad K_h = \infty$ <br> $otherwise \quad URR_h \ge URR$ | Ensures that $K_h$ ≠K in all stratum and that all strata URR achieve target. |

Both quadratic programs are primarily deterministic. However, at the allocation stage, we have to estimate the number of respondents in the subsample. For this, we set $r_{h2} = (q_h)\times(nr_{h1})$ where $q_h$ can be estimated from historic data or can be implemented as a constant value whose sensitivity can be tested.

Using sample data containing respondents and nonrespondents, along with values for $q_h$, we use the SAS® PROC NLP[2] to solve the quadratic programs. The realized allocations are not integer values, but the real valued intervals can be used in SAS® PROC SURVEYSELECT[2] to select a stratified systematic subsample of nonrespondents.

## 3. Case Study

In this section, we present the results of a simulation study that evaluates the considered allocation procedures and NRFU subsample designs on sample data from the ASM. We compare four different allocation strategies: *Full follow-up* (no subsampling), *Constant - K* (an across-the-board systematic sample with sampling interval *K*), *Min-K (*optimal allocation targeting uniform sampling intervals), and *Min-URR* (optimal allocation targeting uniform response rates).

[2] The data analysis for this paper was generated using SAS software. Copyright, SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

### 3.1 Background on the Annual Survey of Manufactures (ASM)

The Annual Survey of Manufactures (ASM) is an establishment survey designed to produce "sample estimates of statistics for all manufacturing establishments with one or more paid employee(s)" (http://www.census.gov/manufacturing/asm/). The ASM is a Pareto-PPS sample of approximately 50,000 establishments selected from a universe of 328,500 manufactures. Approximately 20,000 establishments are included with certainty (probability =1), and the remaining establishments are selected with probability proportional to a composite measure of size (≈ 30,000 establishments). Selected units are in the sample for the four years between censuses. Although the ASM uses a Pareto sample, the publication variance estimates use the Poisson sampling variance formula.

The ASM estimates totals with a difference estimator (Särndal et al., 1992), with the difference estimates computed at the establishment level. The ASM imputes the complete record of unit nonrespondents. For additional information concerning the sample design of the ASM, see the ASM website (http://www.census.gov/manufacturing/asm/).

With the ASM and the Economic Census, implementing a probability subsample of nonrespondents for follow-up represents a major procedural change. The ASM NRFU procedures are very similar to those implemented in the Economic Census, focusing on obtaining respondent data from the largest or most difficult to impute cases. With the ASM, the largest units are included with certainty and have the highest priority for phone follow-up. Similarly, because a given company can comprise several establishments, multi-unit (MU) establishments can be designated for phone follow-up, as company data may need to be allocated to the establishment level. As with the Economic Census, all the remaining nonresponding cases receive some form of reminder, but the noncertainty single unit (SU) establishments are very unlikely to receive personal phone follow-up.

While a field test of the new sampling and data collection procedures is highly desirable before implementing a large-scale change in the 2017 Economic Census, the ASM has reliability requirements. Consequently, only the noncertainty SU population is considered for nonrespondent subsampling. In the simulation study described below, we consider a single item (total value of shipments) and do not use the difference estimator implemented in the ASM. Similarly, we do not implement the composite ratio estimation procedures used in the ASM and restrict our ratio estimator to a single covariate. The noncertainty SU cases account for approximately five-percent of the expected value of total shipments and eleven-percent of the total variance of this estimate. For these reasons, the results presented in this paper are not directly applicable to the ASM.

### 3.2 Simulation Study

Our simulation study compares the statistical properties of total shipment estimates obtained from the different nonrespondent subsampling designs over repeated samples, using two different estimators. Our sampling frame of nonrespondents is derived from the fully imputed 2011 ASM sample and is limited to the SU noncertainty units: we use the complete ASM sample for optimal allocation but present simulation results only for the subsampled domain.

To simulate NRFU for the ASM SU noncertainty population, we removed the MU and SU certainty cases from the ASM sample data. The first NRFU attempt is very effective, so the considered subsample of nonrespondents would occur before the second NRFU attempt. All SU noncertainty units are therefore assumed to have received a questionnaire and one reminder letter if necessary. We use the following procedure, repeating steps 1-5 independently 5,000 times.

1. Using the response probabilities derived from historic data, randomly induce nonresponse in the complete dataset using a MAR response mechanism.
2. Select a stratified systematic sample (ordered by weight) using the nonrespondent strata subsampling rates for a given allocation strategy, yielding seven independent samples per replicate.
3. Simulate unit response to each round of NRFU. The response propensities used for each NRFU contact phase are available upon demand. After assigning response status to each unit, compute comparison statistics.
4. For each allocation, repeat Step 4 until either ten rounds of follow-up have been conducted or the total budget has been expended.

Strata-level response propensities are estimated for each round of NRFU contact. These statistics use historic ASM paradata; this same file is used to obtain optimal allocations (Fink and Lineback, 2013). Unfortunately, we only use this data to estimate response propensities through the fourth round of NRFU. For subsequent rounds of NRFU, we estimated response propensities heuristically under the conditions advocated in Olson and Groves (2012)[3], with a minimum allowable response propensity of 0.02. Mail and phone response propensities were provided by subject matter experts (0.59 and 0.41 respectively), as were the approximate costs of mail out ($2.75/form, $0.75/letter), mail response cost ($0.90), and phone response cost ($5.60). The overall budget is estimated from the response propensity model and cost information given four rounds of complete NRFU.

We obtained average cost, response rates, quality response rates, total shipments estimates and variance estimates (double expansion and ratio estimates) from the 5,000 samples. After, we computed the relative bias and mean squared error of each estimate to evaluate the statistical properties of the estimates obtained with each allocation method. The relative bias for each total at NRFU phase $t$ for a given sampling interval, allocation method and estimator is

$$B(T)^{e}_{Kmt} = 100 * \left[ \frac{\sum_{s}^{5000} \hat{T}^{e}_{Kmts}}{5000} \Big/ T \right] - 1$$

The mean squared error at NRFU phase $t$ for a given sampling interval, allocation method and estimator is

$$MSE(T)^{e}_{Kmt} = \left[ \sum_{s}^{5000} \left( \hat{T}^{e}_{Kmts} - T \right)^{2} \right] \Big/ 5000$$

Where $\hat{T}^{e}_{kmts}$ is the estimated total for estimator $e$ (double expansion or ratio), overall subsampling interval $K$, and allocation method $m$ (Constant-K, Min-K, Min-URR) at NRFU phase $t$ in sample $s$, and $T$ is the population total.

[3] The authors postulate that the response propensities change over the collection cycle, especially as data collection protocols are modified. With the ASM, the reminder letters become more stringent at each NRFU contact phase. They also demonstrate that response propensities decline over the collection phase when a stable data collection protocol is used, as reflected in our heuristically obtained response propensities.

Optimal allocations were obtained once using the historic 2011 ASM data with the quadratic programs. Recall that the target response rate is for the *entire* ASM program, not for the domain under consideration. Consequently, the SU certainty and MU units are included in the input data, but their response rates are held constant. We used 2010 ASM data to model their final response rates. We used three-digit industry as NRFU sampling strata to ensure that each stratum contained a sufficient nonrespondent sample to obtain a feasible solution.

Both quadratic programs require an estimated probability of responding to NRFU ($q_h$). In the absence of any historic paradata, we used a constant value of $q$=0.50 for all strata. For a given $K$, an estimated initial response probability, and an assumed respondent conversion probability, we obtained maximum achievable unit response rates for the ASM of 0.764 ($K$=2) and 0.758 ($K$=3). A target unit response rate of 100% was used for the Min-URR quadratic program. An optimal solution targeting 100% response cannot be obtained with the Min-K quadratic program because of the additional strata-level URR constraint, the maximum achievable unit response rate was used as the target response rate. If the solution is not feasible, we use heuristics to obtain the maximum target response rate that will result in a feasible solution. Table 2 presents the optimal allocations obtained from both quadratic programs.

**Table 2:** Stratum Level Sampling Rates for Optimal Allocations for an overall target value of $K$=2 and 3 with an assumed value of $q$=0.5

| Strata | Target $K$=2 | | Target $K$=3 | |
|---|---|---|---|---|
| | Min-K | Min-URR | Min-K | Min-URR |
| **1** | 0 | 0 | 0 | 0 |
| **2** | 0 | 1 | 0 | 1 |
| **3** | 2 | 1 | 0 | 1 |
| **4** | 1 | 1 | 1.23418 | 1 |
| **5** | 1 | 1 | 1 | 1 |
| **6** | 1 | 1 | 1 | 1 |
| **7** | 0 | 1 | 0 | 3.02273 |
| **8** | 0 | 1 | 0 | 1 |
| **9** | 1.71602 | 1 | 3.08191 | 1 |
| **10** | 0 | 4.18239 | 0 | 1.56987 |
| **11** | 0 | 0 | 0 | 0 |
| **12** | 0 | 1.27897 | 0 | 0 |
| **13** | 0 | 0 | 0 | 0 |
| **14** | 0 | 1 | 0 | 1 |
| **15** | 2 | 0 | 3.22485 | 0 |
| **16** | 1.89359 | 0 | 3.08026 | 0 |
| **17** | 1 | 1 | 2.44343 | 1 |
| **18** | 0 | 1 | 0 | 1 |
| **19** | 0 | 1 | 0 | 1 |
| **20** | 1.42041 | 1 | 2.2165 | 1 |
| **21** | 1 | 1 | 1 | 0 |

For $K$ = 3, only 2 out of 21 cells have exactly the same allocations from the two quadratic programs. When $K$=2, there is slightly more agreement between the allocations. With the Min-K quadratic program, strata that had already achieved the target response rate are removed ($K_h = 0$) and the strata that cannot achieve the target response rate receive full follow-up ($K_h = 1$). Interestingly, often the same strata targeted for no NRFU with the Min-K quadratic program are designed for full follow-up with the Min-URR quadratic program and vice-versa. In general, the same strata are sampled within the same quadratic program, even with the differing overall sampling intervals.

## 3.3. Results

Table 3 presents summary statistics on cost, response rates, and relative bias for each allocation method at the completion of the fourth and tenth phases of NRFU (where applicable). Double expansion estimates are denoted DE. Hereafter, statistics obtained under the current procedure are labeled as “Full”; statistics obtained with an across-the-board allocation are labeled as “Constant-K”; and those obtained with the two optimal allocations are labeled as “Min-K” and “Min-URR,” respectively. All of these statistics are computed from the noncertainty SU cases and do not reflect the subsampling effects on the entire ASM sample.

**Table 3:** Summary of Cost, Response Rate, and Relative Bias of the Estimate

| K | Method | NRFU Round | Cost | Response Rate | DE RBE | Ratio RBE |
|---|---|---|---|---|---|---|
| 1 | Full | 4 | 65,930 | 78.5 | -0.022 | 0.042 |
| 2 | Constant-K | 4 | 55,517 | 67.9 | 0.022 | 0.147 |
| 2 | Constant-K | 10 | 60,035 | 70.8 | 0.022 | 0.092 |
| 2 | Min-K | 4 | 54,923 | 67.1 | 0.001 | 0.063 |
| 2 | Min-K | 10 | 59,302 | 69.8 | 0.004 | 0.053 |
| 2 | Min-URR | 4 | 55,412 | 67.4 | -0.014 | 0.053 |
| 2 | Min-URR | 10 | 60,132 | 70.2 | -0.011 | 0.030 |
| 3 | Constant-K | 4 | 52,013 | 64.3 | 0.000 | 0.222 |
| 3 | Constant-K | 10 | 55,027 | 66.2 | 0.010 | 0.136 |
| 3 | Min-K | 4 | 51,959 | 64.1 | 0.034 | 0.087 |
| 3 | Min-K | 10 | 55,075 | 65.9 | 0.018 | 0.071 |
| 3 | Min-URR | 4 | 51,911 | 63.8 | -0.004 | 0.057 |
| 3 | Min-URR | 10 | 55,126 | 65.7 | -0.012 | 0.034 |

Clearly, selecting a subsample of nonrespondents decreases the budget, with the reduction increasing as the overall subsampling interval increases. The cost reductions are not great: for $K=2$, the maximum budget reduction is about 9 percent, whereas for $K$=3, the maximum budget reduction is approximately 21-percent.

However, the reduced budget with the smaller subsample does not outweigh the detrimental effects on quality. First, it is impossible to approach the complete NRFU unit response rate of 78.5 percent with a $K$=3 overall subsample, even with the additional rounds of NRFU. The decrease in response rate for the larger sampling interval (K=3) is a concern. Turning to the relative bias of the estimates, the DE estimator yields unbiased estimates, whereas the ratio estimator yields slightly bias estimates, even with the additional rounds of NRFU.

Table 4 presents average cost, average sampling variance and MSE. For clarity, the values of the statistics are presented for the current procedure only (Full), whereas the others are presented as a ratio to the corresponding measure from the Full procedure. For example, the ratio estimate from the Constant-K allocation method with K=2 has an MSE that is 3.47 times larger than the full follow-up MSE after four rounds of NRFU. Note that that the purpose of Table 4 is to assess the differences between each allocation method *within* estimator; it should not be used to compare corresponding results *between* double expansion and ratio estimators.

**Table 4:** Average Cost, Sampling Variance and MSE with respect to Full NRFU

| K | Method | Time | Cost | DE Variance | Ratio Variance | DE MSE | Ratio MSE |
|---|---|---|---|---|---|---|---|
| 1 | Full | 4 | 65,930 | 1.414E+14 | 5.491E+13 | 4.165E+12 | 2.321E+12 |
| 2 | Constant | 4 | 0.84 | 2.82 | 4.12 | 2.81 | 3.47 |
| 2 | Constant | 10 | 0.91 | 2.02 | 3.34 | 1.90 | 2.55 |
| 2 | Min-k | 4 | 0.83 | 2.81 | 5.80 | 1.88 | 2.06 |
| 2 | Min-k | 10 | 0.90 | 2.47 | 5.51 | 1.61 | 1.80 |
| 2 | Min-URR | 4 | 0.84 | 3.12 | 7.62 | 1.47 | 1.40 |
| 2 | Min-URR | 10 | 0.91 | 3.03 | 7.53 | 1.24 | 1.17 |
| 3 | Constant | 4 | 0.79 | 4.68 | 7.38 | 4.73 | 6.24 |
| 3 | Constant | 10 | 0.83 | 3.46 | 6.11 | 3.42 | 4.82 |
| 3 | Min-k | 4 | 0.79 | 3.83 | 7.54 | 2.52 | 2.87 |
| 3 | Min-k | 10 | 0.84 | 3.27 | 7.06 | 2.10 | 2.52 |
| 3 | Min-URR | 4 | 0.79 | 3.44 | 8.57 | 1.72 | 1.62 |
| 3 | Min-URR | 10 | 0.84 | 3.38 | 8.52 | 1.53 | 1.45 |

The varying sampling intervals obtained from the optimal allocations increased sampling variances in comparison to those obtained with the Constant-K allocations. With the ratio estimator, regardless of the overall sampling interval (*K=2* or *K*=3), the Constant-K allocation produces the smallest sampling variance. Although the sampling variance increases for the Min-K allocations over the corresponding Constant-K allocations, the increase is still considerably less than with the Min-URR allocations with the ratio estimator. The double expansion estimator does not always follow the same pattern.

More important, reducing the overall sample size by increasing the sampling interval greatly affects the reliability of both estimators. The large sampling variance obtained with any of the *K*=3 allocations is reflected in substantively larger MSEs. Within allocation procedure, the smallest MSE's are obtained using the min-URR procedure. Recall that this allocation method oversamples strata with the lowest response rates, which are under-represented. This strategy reduced the nonresponse bias, leading in turn to a lower MSE.

Ultimately, selecting the smallest subsample (*K*=3) provides the largest cost savings, but yields estimates that are not of good quality as measured by response rates, variance, and MSE. In short, the additional cost savings do not outweigh the accompanying reductions in quality. For these reasons, we eliminated the usage of a 1-in-3 subsample. Hereafter, we restrict analyses to the allocations obtained with *K*=2.

Figure 2 plots average sampling variance obtained from each allocation (Constant-K, Min-K, and Min-URR) against NRFU contact attempt for the double expansion and ratio estimators, respectively. Average sampling variances obtained with the current NRFU procedure (*K*=1) are also provided as a baseline for comparison. Both figures are on the same scale. Regardless of allocation procedure, the ratio estimator variance estimates are more precise than their double expansion counterparts are, although the difference in precision between corresponding allocations decreases with the number of NRFU contact attempts. Recall that the conditional probability of response is very close to zero after the fifth NRFU contact attempt. As a result, the average variances level off after the fourth or fifth NRFU contact attempt, regardless of allocation procedure or estimator. The same patterns displayed in Table 4 reappear, with the Constant-K method producing the smallest variances and the Min-URR method producing the largest variances.

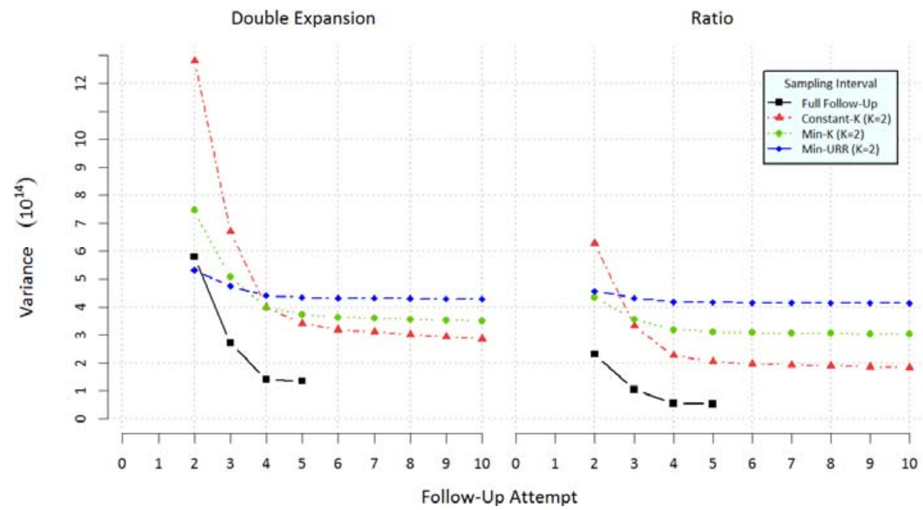


**Figure 2:** Double Expansion/Ratio Average Variance vs NRFU Contact Attempt

Figure 3 plots the MSE obtained from each allocation (Constant-K, Min-K, and Min-URR) against NRFU contact attempt for the double expansion and ratio estimators, respectively. The improvements in overall accuracy with the ratio estimator over the double expansion estimator are quite clear, with the increased bias obtained by the ratio estimator offset by the much-reduced variance, regardless of allocation. Again, the MSEs level off near the fourth or fifth NRFU contact attempt.

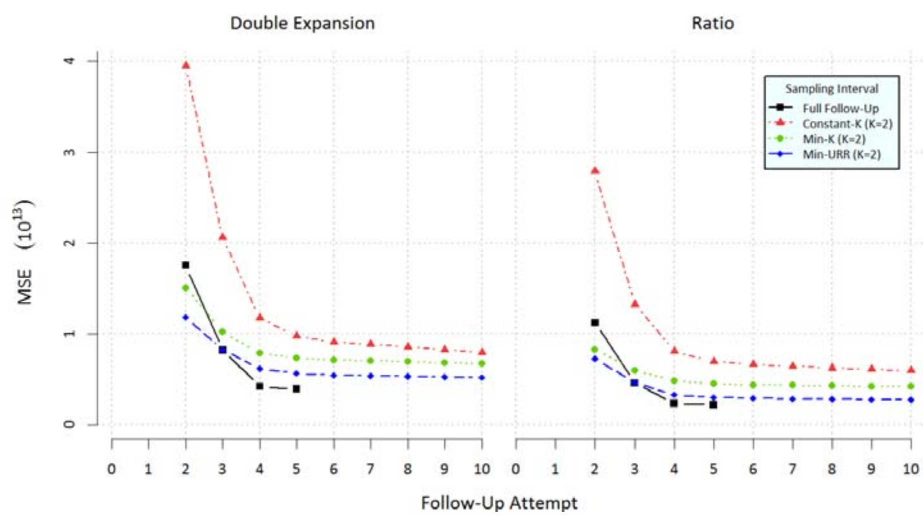


**Figure 3:** Double Expansion/Ratio MSE vs NRFU Contact Attempt

Not surprisingly, the current ASM method has the lowest MSE, regardless of estimators. Interestingly, the estimates obtained with the Min-K allocations approach this minimum MSE after five NRFU contact attempts, and the corresponding ratio estimator MSEs are even closer. This provides evidence that the oversampling strategy used for the Min-URR allocation is in fact reducing the nonresponse bias effects in the estimator. These allocations select a larger proportion of sampling in low responding areas. This yields similar response rates across sampling strata, indicative of a representative sample (Wagner, 2012 and Schouten et al., 2009). The reduced variances obtained from this allocation do not offset the increased bias. Selecting a 1-in-2 subsample across all NRFU strata (Constant-K) obtains the least representative subsample of the three considered allocations.

As always, the Full NRFU results are the benchmark. Ideally, the subsampled estimates should achieve close to the same values at the conclusion of the NRFU contacts. The ratio estimates from the Constant-K allocation subsamples have the smallest sampling variance after four NRFU contact attempts, but are likewise the most biased. These MSEs do not approach the benchmark levels. In contrast, although the Min-URR allocation subsamples produce ratio estimates with the largest variances, their MSEs are comparable to the benchmark, even before all contact attempts have been exhausted.

### 3.4. Discussion

Our case study demonstrates that selecting a systematic subsample of SU noncertainty nonrespondents in the ASM should not have a detrimental effect on their total estimates for this domain. The additional stage of sampling increases the sampling variance, but the level of the variance is reduced by the judicious choice of a ratio adjustment procedure. That said, we recommend limiting the overall subsampling interval to be no larger than two.

Of the three considered allocation methods, the Min-URR allocation was the most effective in achieving close-to-benchmark response rates and achieve reliable estimates in terms of MSE; the larger sampling variances caused by the varying strata sampling intervals is generally offset by the reduced nonresponse bias. Although both optimal allocation are producing improved subsamples over the across-board-allocations, the two objective functions attain very different allocations, each having very different benefits. Ideally, we would like to blend the two approaches together into a single optimization problem. Possible approaches include additional constraints on the stratum level sampling rates or removing the restriction on stratum level response rates (min-K only).

If reducing cost is the overall goal, then we note that additional NRFU contact attempts beyond the fifth contact did not improve the bias, sampling variance, or MSE of the subsampled estimates. Furthermore, ratio estimates obtained from the Min-URR subsamples ($K$=2) had comparable MSEs to those obtained from the Full NRFU (no subsampling) procedure. Thus, dropping the final fourth to fifth NRFU contact attempts would save approximately an additional $4,500, rendering this allocation more comparable in cost savings to a 1-in-3 subsample without making similar sacrifices in quality. Of course, if the achieved cost reduction for a 1-in-2 subsample with up to ten NRFU contact attempts is acceptable, the funds allocated to these final contact attempts might be better expended earlier in the collection cycles, using other contact strategies.

It is still too early to make any recommendations for the 2017 Economic Census, but we are currently planning field tests for the 2014 and 2015 ASM to further explore improvements to this adaptive design process.

## 4. Conclusion

In general, the NRFU procedures for economic programs conducted by the U.S. Census Bureau follow a calendar schedule. Since economic populations are highly skewed and the statistics of interest are totals, a large fraction of the NRFU budget is allocated to the larger units. Given that the NRFU procedures rely on obtaining response data from the larger units, the response rates from smaller units tend to be much lower. It is quite likely that the realized respondent set is neither “balanced…which means (the selected sample has) the same or almost the same characteristics as the whole population” for selected items (Särndal, 2011) nor “representative… with respect to the sample if the response propensities $\rho_i$ are the same for all units in the population” (Schouten et al., 2009). The emphasis on obtaining responses from the larger units at the cost of small unit response in turn creates a bias in the estimates (Thompson and Washington, 2013).

By limiting the target subpopulation for nonrespondent subsampling to the smaller units, we can potentially reduce this unmeasurable bias. Using a probability sample allows us to measure the sampling variance component. Our optimal allocation method increases the potential of obtaining a balanced and representative sample by targeting the low responding areas that usually would not receive any special treatment. We acknowledge that the increased variability in design weights and reduction in response rates are less than desirable effects caused by subsampling

Certainly, without probability subsampling, the contention that the realized respondent set of small businesses remains a probability sample is debatable. Several discussions over the summary report of the AAPOR Task Force on Non-probability sampling (Baker et al., 2013) specifically question whether “a probability sample with less than full coverage and high nonresponse should still be considered a probability sample.” That question is certainty relevant in our studied context, where sampled smaller units truly “opt in” to respond. Selecting a probability subsample of nonrespondents and instructing survey analysts to limit NRFU contact to these cases attempts to avoid this phenomenon. In addition, with a probability subsample, one can use accepted measures of quality such as sampling error or response rates (unit or quantity) for evaluation.

All of the results presented for our case study assume that the existing NRFU contact strategies are used with the subsampled designs. However, subsampling nonrespondents without changing the data collection procedure may have minimal tangible benefits besides cost reduction. The reverse is also true: for example, Kirgis and Lepkowski (2013) present improved response data results for targeted small domains obtained with probability samples and revised contact strategies. Business surveys can draw on a wealth of cognitive research on data collection strategies for large companies: for example, see Chapters 2, 3, 6, 7, and 8 in Snijkers et al (2013). In contrast, the small SU establishments receive very little personal contact (if any) and there is limited cognitive research on preferable contact strategies to draw upon. Additional cognitive research for small establishments could yield better contact strategies, especially for the hard to reach small establishments. Subsampling nonrespondents paired with a new contact strategy for these ”hard to reach” establishments would create a truly adaptive approach for all units, not just the larger ones.

## Acknowledgements

The authors thank Eric Fink, Xijian Liu, Edward Watkins III, and Hannah Thaw for their review and comments and Barry Schouten for his useful suggestions on the optimization problems. We also thank Michelle Vile Karlsson and Michael Zabelsky for providing ASM cost data for our study.

## References

Baker, R., Brick, J.M., Bates, N., Battaglia, M., Couper, M., Dever, J., Gile, K., and Tourangeau, R. (2013). Summary Report of the AAPOR Task Force on Non-Probability Sampling – Report and Rejoinder. *Journal of Survey Statistics and Methodology*, **1**, pp. 90-137.

Bechtel, L. and Thompson, K.J. (2013). Optimizing Unit Nonresponse Adjustment Procedures After Subsampling Nonrespondents In The Economic Census. *Proceedings of the Federal Committee on Statistical Methods Research Conference*, www.fcsm.gov.

Binder, D. , Babyak, c., Brodeur, M., Hidiroglou, M., and Wisner, J. (2000). Variance Estimation for Two-Phase Stratified Sampling. *The Canadian Journal of Statistics*, **28**, pp. 751-764.

Brick, J.M. (2013) Unit Nonresponse and Weighting Adjustments: A Critical Review. *Journal of Official Statistics*, **29**, pp. 329-353.

Haziza, D. and Beaumont, J.F. (2011). A Theoretical Framework for Adaptive Collection Designs. *Proceedings of the International Total Survey Error Workshop.*

Fink, E. and Lineback, J.F. (2013). Using Paradata to Understand Business Survey Reporting Patterns. *Proceedings of the Federal Committee on Statistical Methods Research Conference*, www.fcsm.gov.

Kott, P. (1994). A Note on Handling Nonresponse in Sample Surveys. Journal of the American Statistical Association, 89, pp. 693 –696.

Kirgis, N. and Lepkowski, J. (2013). Design and Management Strategies for Paradata-Driven Responsive Design: Illustrations for the 2006-2010 National Survey of Family Growth. Improving Surveys With Paradata (Frauke Kreuter, Ed). Hoboken, NJ: John Wiley & Sons, Inc.

Lohr. S.L. (2009). Sampling: Design and Analysis. Boston: Brooks/Cole.

Olson, K. and Groves, R.M. (2012). An Examination of Within-Person Variation in Response Propensity Over the Data Collection Field Period. *Journal of Official Statistics*, **28**, pp. 29-51.

Särndal, C.E. (2011). The 2010 Morris Hansen Lecture: Dealing with Survey Nonresponse in Data Collection, in Estimation. *Journal of Official Statistics*, ***27***, pp. 1-21.

Schouten, B., Cobben, F., and Bethlehem, J. (2009). Indicators for The Representativeness Of Survey Response. *Survey Methodology*, **35**, pp. 101-113.

Schouten, B., Calinescu, M., and Luiten, A. (2013). Optimizing Quality of Response Through Adaptive Survey Designs. *Survey Methodology*, **39(2)**, pp. 29-58.

Snijkers, G. Haraldsen, G., Jones, J.,and Willimack, D. K. (2013). Designing and Conducting Business Surveys. Hoboken, NJ: John Wiley & Sons, Inc.

Thompson, K.J. and Oliver, B. (2012). Response Rates in Business Surveys: Going Beyond the Usual Performance Measure. *Journal of Official Statistics*, **27**, pp. 221-237.

Thompson K. J., and Washington K. T. (2013). Challenges in the Treatment of Unit Nonresponse for Selected Business Surveys: A Case Study. *Survey Methods: Insights from the Field.* Retrieved from http://surveyinsights.org/?p=2991.

Wagner, J. (2012). Research Synthesis: A Comparison of Alternative Indicators for the Risk of Nonresponse Bias. *Public Opinion Quarterly*, 76 (3), pp. 555-575.

Whitehead, D. , Kaputa, S., and Thompson, K.J. (2013). Determining Allocation Requirements For Subsampling Nonrespondents From the Annual Survey of Manufactures. *Proceedings of the Federal Committee on Statistical Methods Research Conference*, www.fcsm.gov.