Determining the Optimal Subsampling Rate for Refusal Conversion in RDD Surveys

Inho Park[1,a]

[a]Economic Statistics Department, The Bank of Korea

Abstract

Under recent dramatic declines in response rates, various procedures have been considered among survey practitioners to reduce nonresponse in order to avoid its potential impairment to the inference. In the random digit dialing telephone surveys, substantial efforts are often required to obtain the initial contact for the screener interview. To reduce a burden with higher data collection costs, refusal conversion can be administered only to a random portion of the sample, reducing nonresponse (bias) with an expense of sample variability increment due to the associated weight adjustment. In this paper, we provide ways to determine the optimal subsampling rate using a linear cost model. Our approach for refusal subsampling is to predetermine a random portion from the full sample and to apply refusal conversion efforts if needed only to the subsample.

Keywords: Screener interview, cost model, design effect, weight adjustment.

1. Introduction

Under recent dramatic declines in response rates (*e.g.*, Brick *et al.*, 2005), various procedures have been considered among survey practitioners to reduce nonresponse in order to avoid its potential impairment to the inference, that is, nonresponse bias. In the random digit dialing(RDD) surveys, each sampled number is contacted for a screener interview followed by a main (or extended) interview. Since the former component (*i.e.*, the screener) often requires substantial efforts for an initial contact, refusal conversion (or a serious of multiple callbacks) is very costly and time-consuming. Thus, it can be administered only to a random portion of the sample, reducing data collection costs but improving (weighted) response rates. In other words, subsampling refusals would lead to bias reduction but also to design inefficiency due to the associated weight adjustment. Thus, at subsampling, there is a trade-off between bias reduction and variance increase. Related discussion can be found in Hansen and Hurwitz (1946), Deming (1953), Särndal *et al.* (1992), Elliott *et al.* (2000), Brick *et al.* (2002), Harter *et al.* (2007) and references cited therein.

In this paper, we provide ways to determine the optimal subsampling rate for refusal conversion under a linear cost model. Our approach for refusal conversion is to predetermine before survey a random portion from the full sample for refusal conversion if needed. It is different from those of, for example, Hansen and Hurwitz (1946) and Harter *et al.* (2007), where the entire sample is released and a subsample is selected from the initial refusals among all the cases from the entire sample for refusal conversion. In Section 2, we briefly describe the screening interview process under subsampling for refusal conversion. In Section 3, we demonstrate how to obtain the optimal refusal subsampling rate with some artificial examples. In Section 4, we discuss limitations and further considerations to our approach.

---

[1] Statistician, Economic Statistics Department, The Bank of Korea, Namdaemun-Ro 106, Jung-Gu, Seoul 100-794, Korea.
E-mail: ipark@bok.or.kr

2. Interview Process under Refusal Subsampling

For the illustrations in this paper we consider an unstratified single-stage simple random sampling in RDD surveys. With the initial calling effort for the screener interview, the full sample s of n numbers selected from the frame U can be classified into four non-overlapping sets as follows:

s_{iC} = set of eligible sample numbers completing the screener after the initial calling effort,

s_{iR} = set of eligible sample numbers refusing the screener after the initial calling effort,

s_{IE} = set of ineligible sample numbers,

s_{NW} = set of nonresidential and nonworking sample numbers,

where $s_R = s_{iC} \cup s_{iR} \cup s_{IE} = s \setminus s_{NW}$ denotes the set of residential sample numbers. Administering a certain amount of refusal conversion efforts, initially refused numbers can be converted eventually to the screener completes (*i.e.*, the final screener completes), that is, $s_{iR} = s_{vC} \cup s_{fR}$, where the two non-overlapping sets are defined as follows:

s_{vC} = set of eligible sample numbers completing the screener with conversion efforts,

s_{fR} = set of eligible sample numbers eventually refusing the screener with conversion efforts.

Based on the above decomposition, we can define seven rates as follows:

$r_{iC} = \sharp s_{iC}/\sharp s$ = initial screener completion rate,

$r_{iR} = \sharp s_{iR}/\sharp s$ = initial screener refusal rate,

$r_{IE} = \sharp s_{IE}/\sharp s$ = ineligibility rate,

$r_R = \sharp s_R/\sharp s$ = residency rate,

$r_{NW} = \sharp s_{NW}/\sharp s$ = non-residency and non-working rate,

$r_{vC} = \sharp s_{vC}/\sharp s_{iR}$ = screener refusal conversion rate,

$r_{fR} = \sharp s_{fR}/\sharp s_{iR}$ = screener refusal unconversion rate,

where $\sharp A$ represents the number of units in the set A. Note that $r_{iC} + r_{iR} + r_{IE} + r_{NW} = 1$, $r_{iC} + r_{iR} + r_{IE} = r_R$, $r_R + r_W = 1$ and $r_{vC} + r_{fR} = r_{iR}$.

Assume that our approach is to select a random portion at a rate of $0 < f \leq 1$ from the full sample prior to the data collection and to apply refusal conversion to the initial refusals only from the selected portion (*i.e.*, the subsample). Let s_1 denote the random portion designated for refusal conversion and s_2 the remainder of the full sample, where the subsampling rate is defined as $f = \sharp s_1/\sharp s$. To compensate for the associated under representation due to the refusal subsampling, a weight of $w_2 = 1/f$ is assumed to be assigned only to the cases of s_1 that are refusals after the initial calling effort (*i.e.*, $s_{iR} \cap s_1$) and a weight of $w_1 = 1$ for the rest. The loss of the (potential) final respondents due to the exclusion from refusal conversion will be compensated by increasing the initial full sample s of size n at a rate of $\alpha(f) \geq 1$. Then $n\alpha(f)$ is the size of the full sample increased to compensate for the loss due to refusal subsampling at the rate of f and $\alpha(1) \equiv 1$ denotes the rate when no exclusion (or

full subsampling, *i.e.*, $f = 1$) is made. Furthermore, the number of total screener completes $n_{CS}(f)$ can be written as

$$n_{CS}(f) = n\alpha(f)r_R(r_{iC} + fr_{iR}r_{vC}) = n_{CS,1}(f) + n_{CS,2}(f),$$

where $n_{CS,1}(f) = n\alpha(f)r_Rr_{iC}$ and $n_{CS,2}(f) = n\alpha(f)r_Rfr_{iR}r_{vC}$. Use of the weight adjustment for refusal subsampling would lead to some design inefficiency expressible by so-called the effective sample size:

$$n_e(f) = \frac{n_{CS}(f)}{d(f)} = n\alpha(f)r_R\frac{(r_{iC} + r_{iR}r_{vC})^2}{r_{iC} + r_{iR}r_{vC}/f},$$

where the design effect (Kish, 1992) is given as

$$d(f) = \frac{n_{CS}(f)\left[n_{CS,1}(f)w_1^2 + n_{CS,2}(f)w_2^2\right]}{\left[n_{CS,1}(f)w_1 + n_{CS,2}(f)w_2\right]^2}.$$

If the precision requirement for refusal subsampling is to achieve the resultant effective sample size equivalent to the one with no exclusion ($f = 1$), that is, $n_e(f) = n_{CS}(1) = nr_R(r_{iC} + r_{iR}r_{vC})$, (*e.g.*, Brick *et al.*, 2005), then

$$\alpha(f) = \frac{r_{iC} + r_{iR}r_{vC}/f}{r_{iC} + r_{iR}r_{vC}},$$

which is one at $f = 1$ as required.

3. Optimal Refusal Subsampling Rate

Suppose that cost components with no exclusion for refusal conversion (that may be figured out based on previous or reference survey(s)) can be written based on a simple linear model as follows:

$$c(1) = nr_R\{c_{iC}r_{iC} + c_{iV}r_{iR}r_{vC} + c_{iW}r_{iR}(1 - r_{vC}) + c_{OS}[1 - (r_{iC} + r_{iR})]\} + n(1 - r_R)c_{NW},$$

where

$c_{iC} = $ cost of carrying out the screener with a cooperative respondent with no refusal,

$c_{iV} = $ cost of carrying out the screener with a cooperative respondent who was initially refused but is converted as cooperative with the refusal conversion efforts,

$c_{iW} = $ cost of finalizing the screener with either other non-respondent who was initially refused or that of other types using the refusal conversion efforts,

$c_{OS} = $ cost of finalizing a non-human contact,

$c_{NW} = $ cost of finalizing a non-residential (including non-working) number.

Under refusal subsampling, it can be changed to

$$c(f) = n\alpha(f)r_R\{c_{iC}r_{iC} + f[c_{iV}r_{iR}r_{vC} + c_{iW}r_{iR}(1 - r_{vC})] + c_{OS}[1 - (r_{iC} + r_{iR})]\} + n\alpha(f)(1 - r_R)c_{NW}.$$

Then the relative cost (*i.e.*, cost gain or loss) due to refusal subsampling can be written by taking a ratio of the two costs $c(f)$ and $c(1)$, that is, $\phi(f) = c(f)/c(1)$, which is rewritten as
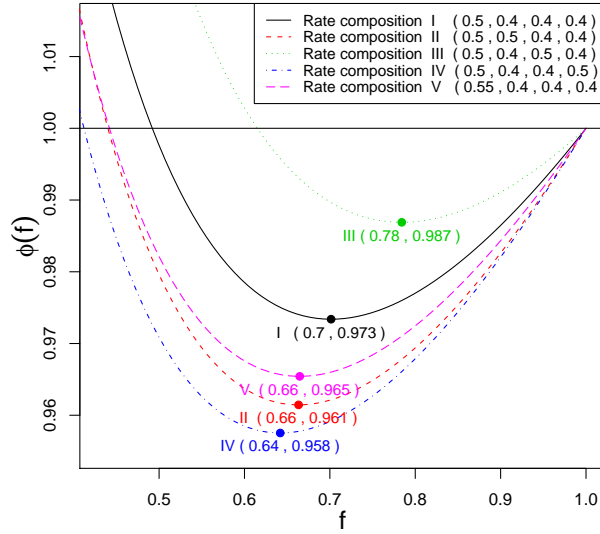
$$\phi(f) = \alpha(f)\beta(f), \tag{3.1}$$

Figure 1: *Comparisons of relative costs $\phi(f)$ for varying rate compositions*

where

$$\alpha(f) = \alpha_1 + \frac{\alpha_2}{f},$$

$$\beta(f) = \beta_1 + \beta_2 f,$$

$$\alpha_1 = \frac{r_{iC}}{r_{iC} + r_{iR}r_{vC}},$$

$$\alpha_2 = 1 - \alpha_1,$$

$$\beta_1 = c(1)^{-1}\Big(r_R\{c_{iC}r_{iC} + c_{OS}[1 - (r_{iC} + r_{iR})]\} + (1 - r_R)c_{NW}\Big),$$

$$\beta_2 = 1 - \beta_1.$$

Following a common strategy adopted in the optimal sample allocation theory (*e.g.*, Särndal *et al.*, 1992), the optimal subsampling rate may be obtained by maximizing the relative cost for a specified precision requirement. Routine calculation (see Appendix) shows that the relative cost is maximized at

$$f^* = \sqrt{\frac{\alpha_2\beta_1}{\alpha_1\beta_2}} \qquad (3.2)$$

with $0 < f^* \leq 1$ and $0 < \phi(f^*) \leq 1$ as required.

Figure 1 compares relative costs of refusal subsampling with varying rate compositions ($r_{iC}, r_{iR}, r_{vC}, r_R$) at a fixed cost structure ($c_{iC}, c_{iV}, c_{iW}, c_{OS}, c_{NW}$) = (1.0, 2.0, 2.0, 1.3, 0.4). The curves of the relative costs $\phi(f)$ for all of the five rate compositions (I through V) are all convex with their inflection point indicating their respective optimal rate for refusal subsampling. For comparisons, consider a line of composition I as a reference. We see, from composition I, that a subsampling rate of $f = .70$ might lead to a relative cost of .973, *i.e.*, 2.7% gain in the survey cost. For a higher initial refusal rate r_{iR} (composition II), cost gains are much greater with optimum attained at smaller f. For a higher conversion rate r_{vC} (composition III), optimum gain in cost is attained at a greater f but with
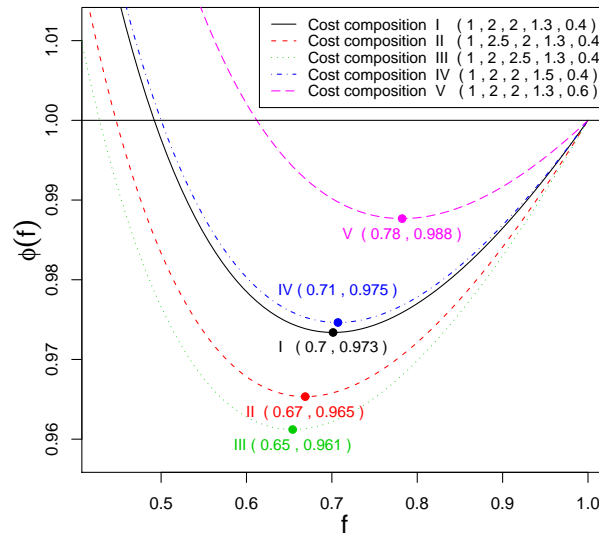
Figure 2: *Comparisons of relative costs $\phi(f)$ for varying cost compositions*

smaller relative gain. For a higher residential rate r_R (composition IV), cost gains are much larger with retaining the lowest, that is, the optimal rate. Finally, for a smaller initial screener completion rate r_{iC} (composition V), cost gains are much smaller and the optimal rate is higher.

Figure 2 compares relative costs of subsampling with varying cost components ($c_{iC}, c_{iV}, c_{iW}, c_{OS}, c_{NW}$) at a fixed rate structure ($r_{iC}, r_{iR}, r_{vC}, r_R$) = $(.5, .4, .5, .4)$. As before, the relative cost ($\phi(f)$) curves are all convex and consider a line of composition I as a reference one. For either larger c_{iV} (composition II) or c_{iW} (composition III), cost gains are larger with optimum at a smaller f. For either larger c_{OS} or c_{NW}, cost gains get smaller, thus less efficient with a higher optimum refusal subsampling rate.

4. Discussion

Our approach can be further modified to reflect more of the practical aspects, for example:

1) additional IE and NW numbers may be determined during refusal conversion;

2) cost/rate composition may vary across surveys and in time;

3) stratification, unequal probability selection and other complex design features may involve in the sample design;

4) alternative precision requirements may be considered to better incorporate two-phase nature of refusal subsampling (within the predetermined portion).

Nonetheless, our findings in Sections 2 and 3 show important aspects related to survey (cost) efficiency with refusal subsampling. The optimal strategy for refusal subsampling, that is, survey cost savings can be pursued assuming a linear cost model based on the equal effective sample size constraint. For a given cost structure, cost gains are larger for higher rates of initial refusal, residency or initial completion. For a given rate structure, cost gains are larger for higher costs for refusal conversion. In summary, one can expect efficiency gains in the survey cost by adopting refusal subsampling

methodology, when data collection is challenging and costly especially for converting initial screener nonrespondents into the screener completes (that may be described with higher initial refusal and residential rates and much more expensive refusal conversion). Albeit its limitations, our approach may provide a reasonable guideline for determining a reasonable subsampling rate.

Appendix: Proof of (3.2)

From (3.1), the relative cost can be written as $\phi(f) = \alpha_1\beta_1 + \alpha_2\beta_2 + \alpha_1\beta_2 f + \alpha_2\beta_1/f$. Then, the first and second derivatives with respect to f are given as $\partial\phi(f)/\partial f = \alpha_1\beta_2 - \alpha_2\beta_1 f^{-2}$ and $\partial^2\phi(f)/\partial f^2 = 2\alpha_2\beta_1 f^{-3}$, since $0 < \alpha_i, \beta_i < 1$ for all $i = 1, 2$. Solving $\partial\phi(f)/\partial f = 0$ for f, we get $f^* = (\alpha_1^{-1}\alpha_2\beta_1\beta_2^{-1})^{1/2} \in (0, 1]$. Furthermore, using Cauchy-Schwartz inequality, we have

$$\phi(f^*) = \alpha_1\beta_1 + \alpha_2\beta_2 + 2\sqrt{\alpha_1\alpha_2\beta_1\beta_2} = \left(\sqrt{\alpha_1 + \alpha_2} + \sqrt{\beta_1 + \beta_2}\right)^2 \le (\alpha_1 + \alpha_2)(\beta_1 + \beta_2) = 1,$$

which completes the proof.

References

Brick, J. M., Judkins, D., Montaquila, J. and Morganstein, D. (2002). Two-phase list-assisted RDD sampling, *Journal of Official Statistics,* **18**, 203–215.

Brick, J. M., Montaquila, J., Hagedorn, M. C., Roth, S. B. and Chapman, C. (2005). Implications for RDD design from an incentive experiment, *Journal of Official Statistics,* **21**, 571–589.

Deming, W. E. (1953). On a probability mechanism to attain an economic balance between the resultant error of response and the bias of nonresponse, *Journal of the American Statistical Association,* **48**, 743–772.

Elliott, M. R., Little, R. J. A. and Lewitzky, S. (2000). Subsampling callbacks to improve survey efficiency, *Journal of the American Statistical Association,* **95**, 730–738.

Hansen, M. H. and Hurwitz, W. N. (1946). The problem of non-response in sample surveys, *Journal of the American Statistical Association,* **41**, 517–529.

Harter, R. M., Mach, T., Wolken, J. and Chapline, J. (2007). Determining subsampling rates for nonrespondents, presented at the third international conference of establishment surveys, Montreal, Canada.

Kish, L. (1992). Weighting for unequal p_i, *Journal of Official Statistics,* **8**, 183–200.

Särndal, C. E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*, Springer-Verlag, New York.