ASSESSING COMPLEX SAMPLE DESIGNS VIA DESIGN EFFECT DECOMPOSITIONS

Inho Park, Westat
1650 Research Boulevard, Rockville, Maryland 20850

Key Words: Stratified sampling; Clustering; Intracluster Correlation Coefficient, Nonresponse adjustment; Unequal weighting; Simulation study

1. Introduction

The design effect for survey estimates is often used as a tool for measuring sample efficiency and for survey planning. It is defined as the ratio of the variance of an estimate under the complex sample design to the variance of the same estimate that would have been obtained from a simple random sample of the same size (Kish, 1965, 8.2). Multi-stage surveys often employ complex sample designs typically involving a number of design features, such as stratification, clustering, and unequal weighting. The efficiency of the complex sample design may be evaluated for each design feature through decompositions of the design effects. A direct approach to stating the overall design effect is relatively complicated under a complex sample design with more than one complex design feature. As an indirect approach, Kish (1987) proposes to model the overall design effect for a weighted sample mean $\hat{\bar{Y}}$, $deft^2(\hat{\bar{Y}})$, as a product of two individual components $deft_W^2(\hat{\bar{Y}})$ and $deft_C^2(\hat{\bar{Y}})$ associated with unequal weighting and clustering, respectively. When the weight $w_{(g)}$ is assigned to $m_{(g)}$ cases, constituting a weighting class $g = 1,...,G$ in the sample of size $m = \sum_{g=1}^{G} m_{(g)}$, Kish's (1987) two-factor decomposition model is written as

$$deft^2(\hat{\bar{Y}}) = deft_W^2(\hat{\bar{Y}}) \cdot deft_C^2(\hat{\bar{Y}}) \qquad (1.1)$$

where

$$deft_W^2(\hat{\bar{Y}}) = \frac{m\sum_{g=1}^{G} w_{(g)}^2 m_{(g)}}{\left(\sum_{g=1}^{G} w_{(g)} m_{(g)}\right)^2}, \qquad (1.2)$$

$$deft_C^2(\hat{\bar{Y}}) = 1 + (\bar{m}_K - 1)\rho_C \qquad (1.3)$$

$\bar{m}_K = n^{-1}\sum_{i=1}^{n} m_i$ is the average of n sampled clusters of size m_i and ρ_C is the rate of homogeneity (usually denoted by roh). This approach is well accepted by many survey samplers. For example, Gabler, Haeder, and Lahiri (1999) provide a model-based justification for (1.1). Lê, Brick, and Kalton (2001) discuss the above product model for design effect decompositions and its associated component models in detail.

Component model (1.2) quantifies the increment in the variance of $\hat{\bar{Y}}$ due to unequal weighting arisen from "haphazard" or "random" sources such as nonresponse or frame problems (Kish, 1992). Gabler *et al.* (1999) obtain the same component model under a model-based framework. However, it may not be adequate for cluster sampling under the design-based approach, since its derivation ignores clustering and instead treats weighting classes as strata assuming the constant stratum variances. Expression (1.2) is, in fact, the ratio of the variance of $\hat{\bar{Y}}$ under disproportionate stratified sampling to that under the proportionate stratified sampling (Kish, 1992). In addition, it may not work well if there exists a moderate correlation between the survey characteristic and the weights (Park and Lee, 2004).

Component model (1.3) measures the increment in the variance of $\hat{\bar{Y}}$ due to clustering for multistage equal probability samples. Its form originates from the design effect for equal probability cluster sampling of equal size clusters, which is approximated by the same formula with the intracluster correlation coefficient ρ instead of ρ_C (Hansen, Hurwitz, and Madow, 1953, pp. 162-163). By viewing ρ_C as a slightly different quantity to, or a generalization of, ρ, model (1.3) is often used to study the clustering effect for multistage equal probability samples (e.g., Kish, 1965, 5.4, 1987, and Lê *et al.*, 2001). Note that ρ is defined only for a population with equal-sized clusters (Lohr, 1999, p. 140). Särndal, Swensson, and Wretman (1992, pp. 130-131) give an alternative quantity ρ_U (equivalent to the adjusted R^2 in the analysis of variance context) that is defined for populations with clusters unequal in size. See Park and Lee (2004) for detailed comparison between the two. Thus, the rather ambiguous quantity ρ_C may be better understood by comparison to the close form quantity ρ_U given by (2.11) (Sections 2.3 and 3).

When cluster sample sizes vary greatly, component model (1.3) may be no longer valid with $\bar{m}_K$ (Holt, 1980). Nonresponse adjustment to an equal probability sample leads to unequal weights. For such cases, some weighted averages of cluster sample sizes are often adopted as an effort to reflect to some extent the increment of the variance due to the cluster sample size variation. However, forcing a clustering effect under the form (1.3) with certain modifications may not be the best solution to measure its effect (Holt, 1980 and Park and Lee, 2004).

A systematic selection of clusters is often considered in multistage sampling, as it provides the effect of stratification together with the frame sort order (Hartley, 1966 and Wolter, 1985, Chapter 7). Ignoring such an implicit stratification in design effect decompositions may be misleading. Park *et al.* (2003) consider a natural extension of Kish's type production models to explicitly factor out the implicit stratification effect:

$$deft^2(\hat{\bar{Y}}) = deft_S^2(\hat{\bar{Y}}) \cdot deft_W^2(\hat{\bar{Y}}) \cdot deft_C^2(\hat{\bar{Y}}), \qquad (1.4)$$

where $deft_S^2(\hat{\bar{Y}})$ is a component model incorporating the stratification effect, $deft_W^2(\hat{\bar{Y}})$ is given by (1.2) and $deft_C^2(\hat{\bar{Y}})$ is given by either (1.3) or its variant. Under the above production form, the increase in the design effect due to unequal weighting and clustering may be compensated by the gain in precision due to efficient implicit stratification.

This paper briefly reviews Kish's type design effect decompositions in the literature and attempts to empirically evaluate two decomposition models (1.1) and (1.4). Section 2 defines these models and their variants in detail under a simple situation involving the aforementioned three design features. Also, Section 2 discusses the intracluster correlation coefficient and its generalizations including the rate of homogeneity. Section 3 describes a simulation study. Section 4 concludes the paper with a brief discussion.

2. Kish's Type Design Effect Decomposition Models

2.1 Two-Factor Models

Consider that a finite population of M elements is partitioned into N clusters of size M_i, denoted by $U=\{(i,j):$ $i=1,...,N; j=1,...,M_i\}$, where $M=\sum_{i=1}^{N} M_i$. Suppose that an equal-probability sample S' of n clusters is selected in two stages from U and weight adjustments for nonresponse result in the unequal weights w_{ij} assigned to each of m_i respondents from cluster $i=1,...,n.$ Let $S=\{(i,j): i=1,...,n; \quad j=1,...,m_i\}$ denote the set of respondents, which is a subset of the full sample S'. Furthermore, assume that there are G unique weights in S, each cluster i with $m_{i(g)}$ elements having the g th unique weight $w_{(g)}$ for $g=1,...,G$. Call the set of those $m_{(g)}=\sum_{i=1}^{n} m_{i(g)}$ elements the g th weighting class. Note that $m_i=\sum_{g=1}^{G} m_{i(g)}$. The population mean $\bar{Y}=Y/M$ is then estimated by the weighted mean

$$\hat{\bar{Y}}=\sum_{(i,j)\in S} w_{ij} y_{ij} \Big/ \sum_{(i,j)\in S} w_{ij}\,, \qquad (2.1)$$

where $Y=\sum_{(ij)\in U} y_{ij}$.

Kish's (1987) decomposition model for $\hat{\bar{Y}}$ is given in (1.1) assuming: D1) the G weighting classes are random (or haphazard) with respect to y with the common variance, and D2) S' is an equal-probability sample and the variation among m_i in S is not significant. Note that component model (1.2) is often written as

$$deft_W^2(\hat{\bar{Y}})=1+cv_w^2=\frac{m\sum_{k=1}^{m} w_k^2}{\left(\sum_{k=1}^{m} w_k\right)^2}=\frac{\sum_{k=1}^{m} w_k^2 \Big/ m}{\left(\sum_{k=1}^{m} w_k \Big/ m\right)^2}$$

by treating as if each element $k\in S$ has its own weight (or weighting class of size one), where k is a single index for the pair of indices (i,j) (Kish, 1992).

Gabler *et al*. (1999) justify (1.1) assuming a model-based analogue to condition D1 along with D3) y_{ij} is a realization from the one-way random effects model

$$\operatorname{cov}(y_{ij}, y_{i'j'})=\begin{cases}\sigma^2 & i=i', j=j',\\ \rho_\xi\sigma^2 & i=i', j\neq j',\\ 0 & i\neq i',\end{cases} \qquad (2.2)$$

and has its simple random sampling version with nonzero $\operatorname{cov}(y_{ij}, y_{i'j'})=\sigma^2$ only for $i=i'$ and $j=j'$. They derive

$$deft^2(\hat{\bar{Y}})=deft_W^2(\hat{\bar{Y}})\cdot deft_{G_1}^2(\hat{\bar{Y}})\le deft_W^2(\hat{\bar{Y}})\cdot deft_{G_2}^2(\hat{\bar{Y}}) \qquad (2.3)$$

where $deft_{G_t}^2(\bar{y})=1+(\bar{m}_{G_t}-1)\rho_\xi$ for $t=1,2$,

$$\bar{m}_{G_1}=\frac{\sum_{i=1}^{n}\left(\sum_{g=1}^{G} w_{(g)} m_{i(g)}\right)^2}{\sum_{g=1}^{G} w_{(g)}^2 m_{(g)}} \quad\text{and}\quad \bar{m}_{G_2}=\frac{\sum_{i=1}^{n} m_i \sum_{g=1}^{G} w_{(g)}^2 m_{i(g)}}{\sum_{i=1}^{n}\sum_{g=1}^{G} w_{(g)}^2 m_{i(g)}}.$$

Note that $deft_W^2(\hat{\bar{Y}})\cdot deft_{G_2}^2(\hat{\bar{Y}})$ is an upper bound of $deft^2(\hat{\bar{Y}})$ and $\bar{m}_{G_2}$ is a weighted average of cluster sample sizes with weights $\sum_{g=1}^{G} w_{(g)}^2 m_{i(g)}$. For the special case of $m_i\equiv m_0$ for all i, it becomes $\bar{m}_{G2}=m_0$ and the upper bound becomes (1.1) except for ρ_ξ. Note that ρ_ξ is a model parameter different from the other quantities, ρ, ρ_U and ρ_C. See Valliant, Dorfman, and Royall (2000, Chapter 8) for more discussion of ρ_ξ.

Kish (1987) observes component model (1.3) to be fairly robust when samples deviate moderately from condition D2. However, a considerable variation among cluster sample sizes may increase the variance. Holt (1980) considers a design effect for $\bar{y}_H=\sum_{i=1}^{n} w_i' \bar{y}_i$ under condition D3, and also D4) w_i' is a constant weight for cluster i, where $\bar{y}_i=m_i^{-1}\sum_{j=1}^{m_i} y_{ij}$ is the simple sample mean for cluster i. Note that $\bar{y}_H$ is equivalent to (2.1) for a special case of $S=S'$, i.e., full response. With unequal sample sizes m_i, Holt (1980) derives $deft_H^2(\bar{y}_H)=$ $\sum_{i=1}^{n}(w_i'^2/m_i)\{1+\rho_\xi(m_i-1)\}$, which is further simplified to

$$deft_H^2(\bar{y}_H)=1+\rho_\xi(\bar{m}_H-1) \qquad (2.4)$$

for equal-probability samples with $w_i' = m_i / m$ and

$$\overline{m}_H = \sum_{i=1}^{n} m_i^2 \Big/ \sum_{i=1}^{n} m_i .$$

Note that $\overline{m}_H$ is another weighted average of cluster sample sizes with weights m_i and $\overline{m}_H \geq \overline{m}_K$ by the Cauchy-Schwarz inequality. Taking the above into consideration, Lê *et al*. (2001) propose the following decomposition model:

$$deft^2(\hat{\bar{Y}}) = deft_W^2(\hat{\bar{Y}}) \cdot deft_H^2(\hat{\bar{Y}}) \qquad (2.5)$$

The subscript(s) G_1, G_2 and H in (2.3) and (2.5) are used to articulate the fact that their component model for clustering is of the form (1.3) but is derived under different setups by Gabler *et al*. and Holt, respectively.

2.2 Three-Factor Models

Suppose that D5) the first stage sampling is systematic selection with unequal probabilities p_i from the list of clusters sorted in a specific order. We can obtain variance of $\hat{\bar{Y}}$ from its Taylor approximation given as $\hat{\bar{Y}} - \bar{Y} \doteq \hat{Z} = \sum_{i=1}^{n} \sum_{j=1}^{m_i} w_{ij} z_{ij}$, where $z_{ij} = M^{-1}(y_{ij} - \bar{Y})$. Letting $w_i = (np_i)^{-1}$ denote the sampling weights for cluster i, $\hat{Z}$ can be written as $\hat{Z} = \sum_{i=1}^{n} w_i \hat{Z}_i$, where $\hat{Z}_i = \sum_{j=1}^{m_i} w_i^{-1} w_{ij} z_{ij}$ is an unbiased estimator of the cluster total $Z_i = \sum_{j=1}^{M_i} z_{ij}$. The implicit stratification effect may be reflected in variance estimation by pairing selected clusters in the listed order. See, for example, Wolter (1985, pp. 286-288), Westat (2001, Appendix-D), and Megill and Gomez (1987). We may consider a variance estimator:

$$v_{syspps}(\hat{Z}) = n^{-2} \sum_{h=1}^{n/2} (p_{2h-1}^{-1} \hat{Z}_{2h-1} - p_{2h}^{-1} \hat{Z}_{2h})^2 \qquad (2.6)$$

assuming a two-cluster-per-stratum sampling design. Comparing to the probability-proportional-to-size with-replacement (*ppswr*) sampling variance given as $v_{ppswr}(\hat{Z}) = \{n(n-1)\}^{-1} \sum_{i=1}^{n} (p_i^{-1} \hat{Z}_i - \hat{Z})^2$, we may propose a component model for the implicit stratification effect applying in decomposition (1.4):

$$deft_S^2(\hat{\bar{Y}}) = v_{syspps}(\hat{Z}) \Big/ v_{ppswr}(\hat{Z}) \qquad (2.7)$$

We may also consider another component model given by replacing $v_{syspps}(\hat{Z})$ in (2.7) by $v_{overlap}(\hat{Z}) = \{2n(n-1)\}^{-1} \sum_{i=2}^{n} (p_i^{-1} \hat{Z}_i - p_{i-1}^{-1} \hat{Z}_{i-1})^2$, which allows overlapping differences (see, e.g., Wolter, 1985, pp. 287-288). Hartley (1966) measures the efficiency of the implicit stratification by comparing the true variance to the unstratified sampling variance under the one-stage element sampling. Park *et al*. (2003) discuss when $deft_S^2(\hat{\bar{Y}})$ can be smaller than one, that is, efficient implicit stratification.

Under both decomposition models (1.1) and (1.4), we will write the component model for clustering in the form $deft_C^2(\hat{\bar{Y}}) = 1 + (\overline{m}_{C_1} - 1)\rho_{C_1}$ to address the fact that the overall design effect is decomposed into either two or three factors with a choice of (weighted) average sample sizes $\overline{m}_{C_1}$, where ρ_{C_1} is its associated estimate of ρ_C.

2.3 Within Cluster Homogeneity Measures

Given the actual design effect $deft^2(\hat{\bar{Y}})$, the rate of homogeneity ρ_C can be estimated from the aforementioned models. For example, it follows from (1.1) that

$$\hat{\rho}_{K|\bullet W} = (\overline{m}_K - 1)^{-1} [d_{C|\bullet W}^2(\hat{\bar{Y}}) - 1] \qquad (2.8)$$

and $\hat{\rho}_{G_1|\bullet W}$, $\hat{\rho}_{G_2|\bullet W}$, $\hat{\rho}_{H|\bullet W}$ are estimated similarly with (2.3) and (2.4), where $d_{C|\bullet W}^2(\hat{\bar{Y}}) = deft^2(\hat{\bar{Y}}) / deft_W^2(\hat{\bar{Y}})$. Note that additional subscripts in (2.8) denote explicitly how the clustering component and the rate of homogeneity are estimated. For example, $C \mid \bullet W$ represents that the clustering component is estimated from the two-factor production model (1.1) ignoring stratification effect and $G_1 \mid \bullet W$ indicates that $\overline{m}_{G_1}$ from (2.3) is used for estimating ρ_C. Also, it follows from (1.4) and (2.7) that

$$\hat{\rho}_{C_1|SW} = (\overline{m}_{C_1} - 1)^{-1} [d_{C|SW}^2(\hat{\bar{Y}}) - 1] \qquad (2.9)$$

where $\overline{m}_{C_1}$ can take any of the four average cluster sizes in Section 2.1 and $d_{C|SW}^2(\hat{\bar{Y}}) = deft^2(\hat{\bar{Y}}) / [deft_W^2(\hat{\bar{Y}}) deft_S^2(\hat{\bar{Y}})]$.

The rate of homogeneity defined in the form $1 + (\overline{m}^* - 1)\rho^*$ is intended to be a generalization of the intracluster correlation coefficient defined by

$$\rho = \frac{2 \sum_{i=1}^{N} \sum_{j>j'}^{M_0} (y_{ij} - Y)(y_{ij'} - \bar{Y})}{(M_0 - 1) \sum_{i=1}^{N} \sum_{j=1}^{M_0} (y_{ij} - \bar{Y})^2} \qquad (2.10)$$

to a broad-spectrum of clustered sampling with the hope of encapsulating all of the related sources of deviation related to clustering (Kish, 1992). Note that ρ is originally defined for the population of equal-sized clusters and can be generalized as

$$\rho_U = 1 - S_{yW}^2 \big/ S_y^2 \qquad (2.11)$$

for unequal-sized cluster populations, where $S_{yW}^2 = (M-N)^{-1} \sum_{i=1}^{N} \sum_{j=1}^{M_i} (y_{ij} - \bar{Y}_i)^2$ is the pooled within-cluster variance, $\bar{Y}_i = \sum_{j=1}^{M_i} y_{ij} / M_i$ and $S_y^2 = (M-1)^{-1} \sum_{i=1}^{N} \sum_{j=1}^{M_i} (y_{ij} - \bar{Y})^2$ is the population variance. See Särndal et al. (1992, p. 130), Lohr (1999, 5.2.2) and Park and Lee (2004) for more discussions. Analytic comparison is possible between the rate of homogeneity and ρ_U for simple sample designs. For example, for simple random cluster sampling (Särndal *et al.*, 1992, p. 317), $Deft^2(\hat{\bar{Y}}) \doteq 1 + (\bar{M} - 1)\rho_U - N^{-1} + M^{-1} S_y^{-2} \sum_{i=1}^{N} M_i (M_i - \bar{M})(\bar{Y}_i - \bar{Y})^2$, where $\bar{M} = M / N$. Thus, we can show that $\rho_C \doteq \rho_U + \{(\bar{M} - 1) M S_y^2\}^{-1} \sum_{i=1}^{N} (M_i / \bar{M})(M_i - \bar{M})(\bar{Y}_i - \bar{Y})^2$, which becomes $\rho_C \doteq \rho_U$ for the special case of $M_i \equiv M_0$ for all i. For two-stage equal-probability sampling with $m_i \equiv m_0$ for all i, we can show $\rho_C \doteq a\rho_U + b$, where $a = 1 - 1/\bar{M}$ and $b = (m_0 M)^{-1} \sum_{i=1}^{N} (S_{yi} / S_y)^2 + \bar{M}^{-1} - m_0^{-1} - M^{-1}$, where $S_{yi}^2 = (M_i - 1)^{-1} \sum_{j=1}^{M_i} (y_{ij} - \bar{Y}_i)^2$. Note that $a \approx 1$ and $b \approx 0$ for many cases and the above two sample designs are all equal-probability selection method (*epsem*).

3. Simulation Study

To compare decomposition models discussed in Sections 1 and 2, a simulation study was conducted with regard to two aspects among others: random weighting (D1) and significant stratification effect (D5). The former was evaluated through two kinds of response classes (ignorable vs. nonignorable) created for the population. The latter was compared with two sets of study characteristics having different relationships to a frame sort order (without vs. with significant stratification effect). The one-way random effects model (cf. D3) was utilized to generate the two sets of study characteristics by differing cluster means (Section 3.2). In addition, two-stage equal-probability sampling with equal-size cluster samples was considered (D2 and D4) and variations in both sizes of cluster respondents and their final weights (cf. D3) were assumed to arise from nonresponse (and its weight adjustments). Section 3.1 discusses how the population was created. Section 3.2 describes how both overall and various design effect components were estimated. Section 3.3 summarizes the simulation results.

3.1 Population

An artificial finite population with N=500 clusters was created with its cluster sizes M_i generated from $round[50*(1+Z)]$, where Z are random numbers from $p(z) = 1 + \exp(-z)$ for $z > 0$ (i.e., standard exponential distribution). Two study characteristics were generated by differing μ_i from the one-way random effects model:

$$y_{ij} = \mu_i + a_i + e_{ij}, \qquad (3.1)$$

where μ_i is the cluster mean, a_i and e_{ij} are two independent Gaussian random errors with zero means and variances of $\rho_i \sigma_i^2$ and $(1-\rho_i)\sigma_i^2$, respectively, for $i = 1, \ldots, N$, and $j = 1, \ldots, M_i$. Two sets of μ_i were considered:

A) (*no stratification effect*) a common $\mu_i = 106$ for all clusters; and

B) (*stratification effect*) μ_i decreasing in M_i with a mean $100 + 500 / M_i$ and a variance $80^2 M_i^{2(1-.000036)}$.

Note that the former does not yield a significant stratification effect but the latter does. The scheme for the latter was borrowed from the idea of Lee, Rancourt, and Särndal (1994). Also, the other two parameters were set to constants, $\sigma_i^2 = 10$ and $\rho_i = 0.05$ for all clusters for simplicity. In addition, nonresponse was assumed to arise with response probabilities $p_g = 0.4 + 0.1(g-1)$ for class $g = 1, \ldots, 6$ that are determined in two ways:

I) (*ignorable nonresponse*) random response classes; and

N) (*nonignorable nonresponse*) nonrandom response classes determined from quintiles of x_{ij} that were generated by adding a Gaussian random error of zero mean and variance of one to y_{ij}.

3.2 Estimation of Design Effect Components

A total of $R = 10{,}000$ equal-probability samples were selected in two stages from the finite population generated in Section 3.1. The first-stage selection was systematic probability-proportional-to-size (*syspps*) sampling of $n = 50$ clusters with selection probability $p_i = M_i / M$ from the frame sorted by size. The second-stage selection was without-replacement simple random sampling (*srswor*) of size $m_i' = 12$ for all $i = 1, \ldots, n$. Nonresponse resulted in m_i (≤ 12) respondents in cluster i, and its customary weight adjustment (*e.g.*, Valliant, 2003) led to positive unequal weights w_{ij} that were assigned only to a total of $m = \sum_{i=1}^{n} m_i$ respondents. Note that six response classes mentioned in Section 3.1 produced up to six different weights $w_{(g)}$ for $m_{i(g)}$ respondents within cluster i. With full response, the final weights w_{ij} should not have been altered from the design weights $w_{ij}' = M / m'$ that are constant for all $m' = \sum_{i=1}^{n} m'_i = 600$ sampled elements. The overall design effect was computed as $deft^2(\hat{\bar{Y}}) = v_{syspps}(\hat{Z}) / v_{srswr}(\hat{\bar{Y}})$, where $v_{syspps}(\hat{Z})$ is given in (2.6) and $v_{srswr}(\hat{\bar{Y}}) = (m\sum_S w_{ij})^{-1}\sum_S w_{ij}(y_{ij} - \hat{\bar{Y}})^2$ is a customary formula for a variance estimate under SRSWR used in packages for complex survey data analysis (e.g., Westat, B-26). Table 1 lists overall design effect, its components under both two-factor and three-factor decompositions and the associated rates of homogeneity. Note that only the component estimates due to clustering depend upon the choice of a decomposition model.

Table 1. Overall Design Effect, Component Models and Rate of Homogeneity

		Two-factor decomposition	Three-factor decomposition
Overall Design Effect	$deft(\hat{\bar{Y}})$	d_O^2	-
Component Models	$deft_S^2(\hat{\bar{Y}})$	-	d_S^2
	$deft_C^2(\hat{\bar{Y}})$	$d_{C\|\bullet W}^2$	$d_{C\|SW}^2$
	$deft_W^2(\hat{\bar{Y}})$	d_W^2	d_W^2
Rate of homogeneity	ρ_K	$\rho_{K\|\bullet W}$	$\rho_{K\|SW}$
	ρ_{G_1}	$\rho_{G_1\|\bullet W}$	$\rho_{G_1\|SW}$
	ρ_{G_2}	$\rho_{G_2\|\bullet W}$	$\rho_{G_2\|SW}$
	ρ_H	$\rho_{H\|\bullet W}$	$\rho_{H\|SW}$

We selected three other samples for each of the R main samples independently under different sampling designs in order to obtain simulation-based true values of design effect components. The first auxiliary sample, denoted by PPSWOR, was selected similar to the main sample except for a random order in sorting the frame (i.e., no stratification). The other two auxiliary samples, denoted by SRSWR1 and SRSWR2, were selected by with-replacement simple random sampling of sizes m' and m, respectively. The former was for referencing the full sample S' and the latter was for respondents only S. Table 2 shows sample means in the form (2.1) with appropriate samples (full vs. respondents only) and weights (base weights vs. nonresponse adjusted weights).

Table 2. $\hat{\bar{Y}}$ from four samples

	Main sample	Auxiliary samples		
		PPSWOR	SRSWR1	SRSWR2
S	$\hat{\bar{Y}}_{SCW}$	$\hat{\bar{Y}}_{CW}$	-	$\hat{\bar{Y}}_{srs}$
S'	$\hat{\bar{Y}}'_{SC}$	$\hat{\bar{Y}}'_C$	$\hat{\bar{Y}}'_{srs}$	-

Note: S and S' denote complete only and full samples, respectively.

Table 3 provides how the true values for the overall design effect and its components were estimated based on $\hat{\bar{Y}}$ from $R = 10{,}000$ sets of four samples.

Table 3. Simulation-based true values for design effect and its components

True value	Simulation-based estimation of the true value
Overall design effect	$D_O^{sim}(\hat{\bar{Y}}) = V_{sim}(\hat{\bar{Y}}_{SCW}) \big/ V_{sim}(\hat{\bar{Y}}_{srs})$
Stratification effect	$D_S^{sim}(\hat{\bar{Y}}) = V_{sim}(\hat{\bar{Y}}_{SCW}) \big/ V_{sim}(\hat{\bar{Y}}_{CW})$
Clustering effect	$D_C^{sim}(\hat{\bar{Y}}) = V_{sim}(\hat{\bar{Y}}'_C) \big/ V_{sim}(\hat{\bar{Y}}'_{srs})$
Weighting effect	$D_W^{sim}(\hat{\bar{Y}}) = V_{sim}(\hat{\bar{Y}}_{SCW}) \big/ V_{sim}(\hat{\bar{Y}}'_{SC})$

Note: V_{sim} represents a variance of R estimates.

3.3 Summary of Simulation Results

Table 4 and Figure 1 summarize the simulation results for design effect decompositions under two- and three-factor models (1.1) and (1.4) for four situations described in Section 3.1. That is, (a) *AI*: no stratification effect with ignorable nonresponse, (b) *AN*: no stratification effect with nonignorable nonresponse, (c) *BI*: stratification effect with ignorable nonresponse, (d) *BN*: stratification effect with nonignorable nonresponse. Table 4 gives nine quantities including true value (True), average (Ave), bias (Bias), root mean square error (Rmse), minimum (MIN), maximum (MAX), and quartiles (Q1, Q2, Q3) of $R = 10{,}000$ sets of the estimated overall design effect and its components. Figure 1 shows boxplots of the distributions of deviations of estimated design effect and its components from their corresponding true value.

For the overall design effect, the bias is fairly small for the two situations with ignorable nonresponse: (a) and (c), but it is relatively large for the others with nonignorable nonresponse: (b) and (d). For the weighting component d_W^2, both Bias and Rmse are very small and similar in their absolute magnitude. Its bias is negative for those with ignorable nonresponse but it is positive for those with nonignorable nonresponse. The stratification component d_S^2 is negatively biased but its magnitude is relatively small for all situations. Such a negative bias may have arisen from the fact that the numerator $v_{syspps}(\hat{Z})$ does not reflect the without-replacement sampling of the clusters, slightly overstating the variance. Both clustering components $d_{C|SW}^2$ and $d_{C|\bullet W}^2$ are negatively biased for all situations and the magnitude of the bias is fairly large for those situations with significant stratification effects. $d_{C|\bullet W}^2$ is considerably more biased than $d_{C|SW}^2$. Variation in d_O^2 is mostly attributed to $d_{C|\bullet W}^2$, while two components d_S^2 and $d_{C|SW}^2$ share that in d_O^2 evenly to a certain degree.

Under situation (a), to which Kish's decomposition model (1.1) may best fit among the four considered, all components are almost unbiased with obvious underestimation of the weighting component. Under situation (b), both decomposition models (1.1) and (1.4) tend to work well with obvious overestimation of both overall and weighting component. For two situations (c) and (d) with significant stratification effect, Kish's decomposition model (1.1) gives a fair amount of negative bias for the clustering effect. That is, failure to account for the stratification effect tends to produce significant underestimation of the clustering effect.

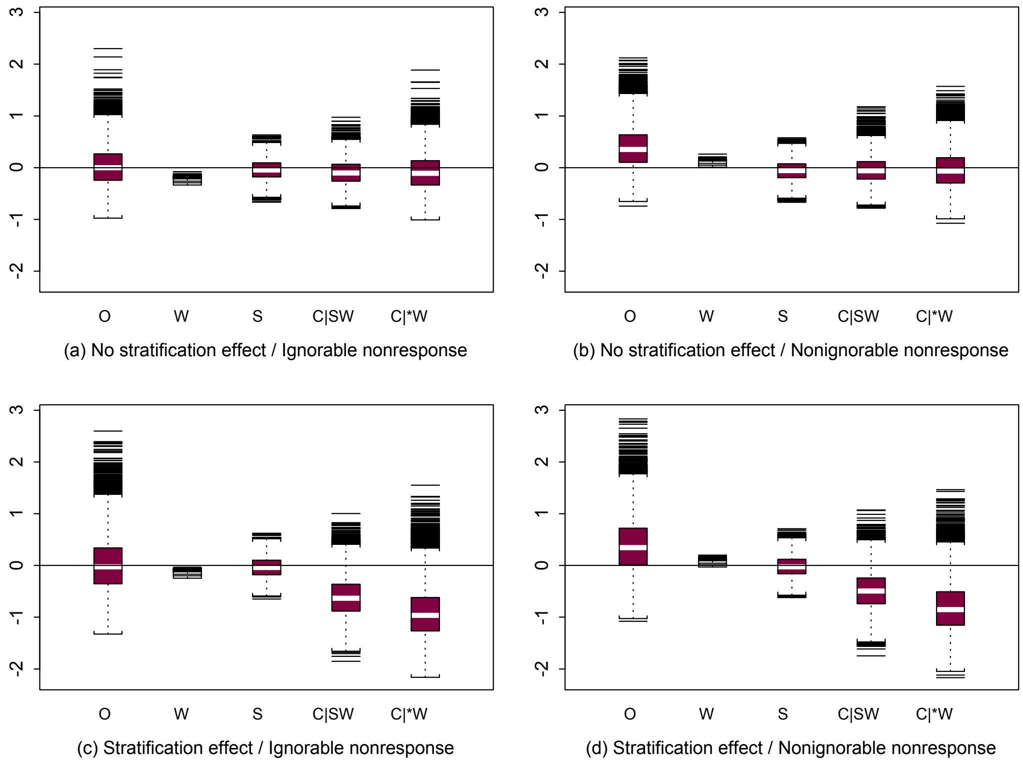
Figure 2 shows the distribution of $R = 10{,}000$ sets of estimated rates of homogeneity using various averages of cluster sample sizes from two- and three-factor decomposition models. The horizontal reference line in each plot denotes ρ_U from (2.12) for each situation. As seen, rates of homogeneity are mostly close to ρ_U for two situations (a) and (b) without significant stratification effect, while they are all considerably smaller than ρ_U for the other two situations (c) and (d) with significant stratification effect. However, their magnitude is not fluctuating across different averages of cluster sample sizes as Kish (1987) observes.

4. Discussion

In this paper, we reviewed Kish's type design effect decomposition models in the literature, focusing on the assumptions on which their component models are derived (or postulated). We empirically evaluated two- and three-factor component models through a simulation study. Given a significant stratification effect, Kish's type models tend to underestimate both weighting and clustering effects, the latter with a considerable magnitude. With moderate variations in sizes of respondents within clusters, substitution by a weighted average of cluster sample sizes in the form (1.3) may not make a significant difference. Under the presence of a significant stratification effect, estimated rates of homogeneity deviate much from the closed form generalization of ρ_U to the population intracluster correlation coefficient ρ and vary greatly by models. Thus, their interpretation may not be straightforward. In summary, the simulation results show that Kish's type design effect decomposition models require caution in their interpretation since they may not work properly for many of the practical situations deviating from the one under which Kish's model is originally considered.

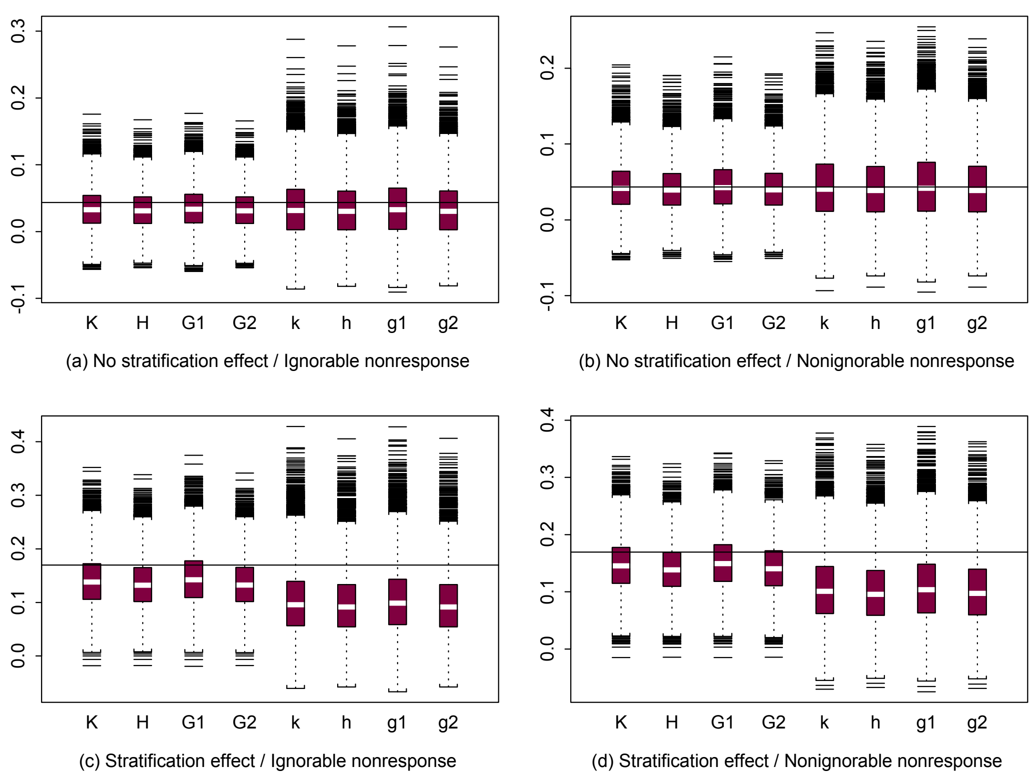
Table 4. Comparisons of design effects and their component models

	Design effect component	Ignorable Nonresponse									Nonignorable Nonresponse								
		True	Ave	Bias	Rmse	Min	Q1	Q2	Q3	Max	True	Ave	Bias	Rmse	Min	Q1	Q2	Q3	Max
No stratification effect	d_O^2	1.35	1.39	0.03	0.38	0.38	1.11	1.36	1.62	3.65	1.08	1.47	0.39	0.56	0.33	1.18	1.43	1.71	3.20
	d_W^2	1.36	1.09	-0.27	0.27	1.03	1.07	1.08	1.10	1.28	1.02	1.09	0.07	0.07	1.03	1.07	1.09	1.10	1.28
	d_S^2	1.05	1.01	-0.04	0.20	0.39	0.87	1.01	1.14	1.68	1.07	1.01	-0.06	0.20	0.40	0.88	1.01	1.14	1.65
	$d_{C\mid SW}^2$	1.36	1.27	-0.09	0.26	0.57	1.10	1.25	1.42	2.33	1.38	1.34	-0.04	0.26	0.60	1.16	1.32	1.50	2.56
	$d_{C\mid \bullet W}^2$	1.36	1.28	-0.08	0.36	0.35	1.02	1.25	1.49	3.25	1.38	1.35	-0.03	0.37	0.31	1.09	1.32	1.57	2.95
Significant stratification effect	d_O^2	1.92	1.95	0.03	0.54	0.60	1.57	1.89	2.26	4.52	1.59	1.98	0.39	0.66	0.51	1.61	1.93	2.31	4.42
	d_W^2	1.27	1.09	-0.18	0.19	1.02	1.07	1.08	1.10	1.24	1.05	1.08	0.03	0.04	1.02	1.07	1.08	1.10	1.25
	d_S^2	0.90	0.86	-0.04	0.20	0.25	0.72	0.86	1.00	1.52	0.88	0.86	-0.02	0.20	0.25	0.71	0.85	0.99	1.58
	$d_{C\mid SW}^2$	2.71	2.10	-0.61	0.72	0.86	1.82	2.07	2.34	3.71	2.63	2.15	-0.49	0.61	0.89	1.89	2.14	2.39	3.70
	$d_{C\mid \bullet W}^2$	2.71	1.80	-0.91	1.04	0.54	1.44	1.74	2.08	4.26	2.63	1.83	-0.81	0.94	0.46	1.48	1.78	2.13	4.10



Note: Symbols O, W, S, C|SW and C|*W represent overall design effect, weighting, stratification, and two clustering components, respectively.

Figure 1. Comparisons of deviation of design effect and their component models from their corresponding true value, under varying simulation conditions



Note: Upper cases represent rates of homogeneity with various averages of cluster sample sizes due to Kish, Holt and Gabler *et al*., respectively, under the three-factor decomposition models. Lower cases correspond to those under the two-factor decomposition models.

Figure 2. Comparisons of Rates of Homogeneity under varying Simulation Conditions

5. References

Cochran, W.G. (1977). *Sampling Techniques.* New York: John Wiley & Sons.

Gabler, S., Haeder, S., and Lahiri, P. (1999). A Model Based Justification of Kish's Formula for Design Effects for Weighting and Clustering. *Survey Methodology*, **25**, pp.105-106.

Hansen, M.H., Hurwitz, W.N. and Madow, W.G. (1953). *Sample Survey Methods and Theory.* Vol. II, New York: John Wiley & Sons, Inc.

Hartley, H.O. (1966). Systematic Sampling With Unequal Probability and Without Replacement. *Journal of American Statistical Association*, pp. 739-748.

Holt, D. (1980). Discussion of the Paper by Verma, Scott, and O'Muircheartaigh. *Journal of the Royal Statistical Society A*, 143, Part 4, pp. 468-469.

Kish, L. (1965). *Survey Sampling*, New York: John Wiley & Sons.

Kish, L. (1987). Weighting in *Deft*^{2}. *The Survey Statistician*, June, pp. 26-30.

Kish, L. (1992). Weighting for Unequal p_i . *Journal of Official Statistics.* 8, pp. 183-200.

Lê, T., Brick, J.M., and Kalton, G. (2001). Decomposing Design Effects. *Proceedings of the American Statistical Association,* Section on Survey Research Methods [CD-ROM], Alexandra VA.

Lee, H., Rancourt, E., and Särndal, C.E. (1994). Experiments with Variance Estimation from Survey Data with Imputed Values. 10, pp. 231-243.

Lohr, S.L. (1999). *Sampling: Design and Analysis*. Pacific Grove, CA:Brooks/Cole.

Megill, D.J. and Gomez, E.E. (1987). Measuring the Efficiency of Implicit Stratification in Multistage Sample Surveys, *Proceedings of the American Statistical Association,* Section on Survey Research Methods, pp. 166-171, Alexandra VA.

Park, I., and Lee, H. (2004). Design Effects for the Weighted Mean and Total Estimators Under Complex Survey Sampling. *Survey Methodology*, 30, in press.

Park, I., Winglee, M., Clark, J., Rust, K., Sedlak, A., and Morganstein, D. (2003). Design Effects and Survey Planning. *Proceedings of the American Statistical Association,* Section on Survey Research Methods [CD-ROM], Alexandra VA.

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

Valliant, R. (2003). The Effect of Multiple Weighting Steps on Variance Estimation. *Journal of Official Statistics,* 20, pp. 1-18.

Valliant, R., Dorfman, A.H., and Royall, R.M. (2000). *Finite Population Sampling and Inference: A Prediction Approach*. New York: Wiley.

Westat (2001). *WesVar 4.0 User's Guide.* Rockville, MD: Westat. Inc.

Wolter, K. (1985). *Introduction to Variance Estimation.* New York: Springer-Verlag.