A Note on the Effect of Auxiliary Information on the Variance of Cluster Sampling

Nina Hagesæther[1] *and Li-Chun Zhang*[1]

A model-based synthesis is presented for a number of well-known results concerning the design effect of cluster sampling. This helps to yield a simpler perspective, making the design-based expressions more transparent. In particular, it is shown that the use of auxiliary information removes the extra variance that is due to the variation in the cluster sizes. Moreover, it can reduce the loss of efficiency to the extent it reduces the conditional intra-cluster correlation given the covariates. The design effect then depends on the remaining conditional intra-cluster correlation and an average cluster size.

Key words: Cluster sampling; design effect; cluster size; intra-cluster correlation; general regression estimator.

1. Introduction

Clustering of elements of interest exists in many natural populations. In cluster sampling one generally needs to balance between the likely loss of efficiency as compared to sampling of elements and the potential administrative and operational advantages that are important in practice. For example, cluster sampling of household-like units is employed in the Norwegian Labor Force Survey (NLFS), mainly due to cost and operational concerns. In contrast the Swedish LFS is based on direct sampling of persons. Strictly speaking, one may consider the design effect (DEFF) in this case to be the ratio between the variance of the Horvitz-Thompson (HT) estimator of a population total under cluster sampling and that under simple random sampling (SRS) of elements. But the comparison may be affected by several factors, including the sampling design, the choice of estimator and the use of auxiliary information that is available. We refer to Park and Lee (2004) for a study of the design effects for the weighted total estimator and the weighted mean estimator under a purely design-based perspective.

In this article we synthesize a number of results that scatter in the literature (see e.g., Cochran 1977, Chapter 9A; Kish 1987; Särndal et al. 1992, Chapters 4 and 8). The idea is to regard the finite population as drawn from an infinite super-population that has certain properties, and compare the model expectations of the various design variances, also known as the anticipated variances (Isaki and Fuller 1982). The model-based evaluation helps to yield a simpler perspective, making the design-based

expressions more transparent and providing conditions under which similar results can be expected. Gabler et al. (1999) used the same approach to justify Kish's formula (Kish 1987) for two-stage sampling and multiple weighting classes. More generally, this is an example of the combined approach to inference that involves both the sampling design and the super-population model (see Pfeffermann 1993).

We shall distinguish between two situations of auxiliary information. In the first one, to be referred to as the minimal case, the only information available is the cluster sizes in the sample and the number of elements in the population. In the second one, to be referred to as the general case, nontrivial auxiliary information is available at cluster or element level. SRS of clusters in the minimal case is considered in Section 2, and the general case in Section 3. In Section 4 we illustrate the theoretical results using the NLFS data. In Section 5 we consider cluster sampling with varying sampling probabilities.

It is shown that the use of auxiliary information in the general case affects the DEFF of cluster sampling in two respects: (a) it can remove the extra variance that is due to the variation in the cluster sizes, and (b) it can reduce the loss of efficiency to the extent it reduces the conditional intra-cluster correlation given the covariates. The average cluster size remains the other deciding factor. In household surveys, or other sampling situations where the cluster sizes are small, the variance increase by cluster sampling will be small unless the intra-cluster correlation is very high. But there will be a great loss of efficiency if the cluster sizes are large, unless the conditional intra-cluster correlation is basically zero.

2.   Decomposition of DEFF in the Minimal Case

We start with the minimal case of auxiliary information. Denote by (ki) element i from cluster k. Assume the constant intra-cluster correlation model

$$y_{ki} = \mu_y + \varepsilon_{ki}$$

where y_{ki} is the variable of interest and μ_y is its model expectation. The residuals are such that $E(\varepsilon_{ki}) = 0$, and $V(\varepsilon_{ki}) = \sigma_y^2$, and $Cov(\varepsilon_{ki}, \varepsilon_{kj}) = \rho_y \sigma_y^2$ for two elements in the same cluster and $Cov(\varepsilon_{ki}, \varepsilon_{gj}) = 0$ for two elements from different clusters, and ρ_y is the intra-cluster correlation coefficient. All the expectations are with respect to the model. We have then, at the cluster level, $y_k = N_k \mu_y + \varepsilon_k$ and $\varepsilon_k = \sum_{i=1}^{N_k} \varepsilon_{ki}$ where N_k is the size of the cluster. The above model is often used to account for the cluster effect (see e.g., Gabler et al. 1999).

Denote by U the population of elements and denote by s the sample of elements. Denote by U_0 and s_0 the population and sample of clusters. Let N be the size of U, and let M be the size of U_0. Let n and m be, respectively, the number of elements and clusters in the sample. Take first simple random sampling (SRS) of elements. Denote by $\hat{Y}_E = (N/n) \sum_{(ki) \in s} y_{ki}$ the Horvitz-Thompson (HT) estimator. We assume that $n = m\bar{N}$, where $\bar{N} = \sum_{k=1}^{M} N_k / M = N/M$ is the population average cluster size, which is comparable to cluster sampling with m sample clusters. Take next SRS of clusters. Denote by $\hat{Y}_R = (N/n) \sum_{k \in s_0} y_k$ the ratio-to-size estimator (Cochran 1977, Section 9A.1), which looks just like $\hat{Y}_E$. The difference is that n is fixed for $\hat{Y}_E$ but variable for $\hat{Y}_R$; and inclusion of an element implies inclusion of the whole cluster in the case of $\hat{Y}_R$ but not in the case of $\hat{Y}_E$. Finally, denote by $\hat{Y}_C = (M/m) \sum_{k \in s_0} y_k$ the HT estimator given SRS of clusters.

Denote by V_D the design-based sampling variance, which is readily given by standard formulae; see e.g., Cochran 1977, Equation (2.13) for $V_D(\hat{Y}_E)$, Equation (9A.2) for $V_D(\hat{Y}_C)$, and Equation (9A.3) for $V_D(\hat{Y}_R)$. Denote by $\vec{N} = (N_1, N_2, \ldots, N_M)^T$ the vector of cluster sizes in the population. It is straightforward to verify that, conditional on $\vec{N}$, the anticipated variances under the intra-cluster correlation model are given by

$$E\{V_D(\hat{Y}_E)|\vec{N}\} = \frac{M}{m}\frac{M-m}{M-1/\bar{N}}E\left\{\sum_{(ki)\in U}(\varepsilon_{ki}-\bar{\varepsilon})^2|\vec{N}\right\}$$

$$= \frac{M}{m}\frac{M-m}{M-1/\bar{N}}N\sigma_y^2\left\{1-\frac{1}{N^2}\sum_{k\in U_0}v_k(y)/\sigma_y^2\right\} \tag{1}$$

$$E\{V_D(\hat{Y}_R)|\vec{N}\} = \frac{M}{m}\frac{M-m}{M-1}E\left\{\sum_{k\in U_0}(\varepsilon_k-N_k\bar{\varepsilon})^2|\vec{N}\right\}$$

$$= \frac{M}{m}\frac{M-m}{M-1}\left\{\sum_{k\in U_0}v_k(y)\right\}$$

$$\times\left\{1-\frac{2}{N}\sum_{k\in U_0}N_kv_k(y)/\left(\sum_{k\in U_0}v_k(y)\right)+\frac{1}{N^2}\sum_{k\in U_0}N_k^2\right\} \tag{2}$$

$$E\{V_D(\hat{Y}_C)|\vec{N}\} = \frac{M}{m}\frac{M-m}{M-1}E\left\{\sum_{k\in U_0}[(N_k-\bar{N})\mu_y+(\varepsilon_k-\bar{N}\bar{\varepsilon})]^2|\vec{N}\right\}$$

$$= \frac{M}{m}\frac{M-m}{M-1}\left\{\sum_{k\in U_0}v_k(y)\right\}$$

$$\times\left\{1+\mu_y^2\sum_{k\in U_0}(N_k-\bar{N})^2/\sum_{k\in U_0}v_k(y)-\frac{1}{M}\right\} \tag{3}$$

where $\bar{\varepsilon}=\sum_{(ki)\in U}\varepsilon_{ki}/N=\sum_{k\in U_0}\varepsilon_k/N$ and $v_k(y)=V(y_k|N_k)=N_k\sigma_y^2+N_k(N_k-1)\rho_y\sigma_y^2$.

Asymptotically, provided $m, M\to\infty$, and $m/M\to r$ for $0<r<1$, while N_k remains bounded, we have

$$\frac{E\{V_D(\hat{Y}_R)|\vec{N}\}}{E\{V_D(\hat{Y}_E)|\vec{N}\}} = \{1+\rho_y(\bar{N}^*-1)\}\cdot\left\{1+O\left(\frac{1}{M}\right)\right\} = \{1+\gamma\}\cdot\left\{1+O\left(\frac{1}{M}\right)\right\}$$

$$\frac{E\{V_D(\hat{Y}_C)|\vec{N}\}}{E\{V_D(\hat{Y}_R)|\vec{N}\}} = \left\{1+\frac{\mu_y^2\sum_{k\in U_0}(N_k-\bar{N})^2}{\sum_{k\in U_0}v_k(y)}\right\}\cdot\left\{1+O\left(\frac{1}{M}\right)\right\} = \{1+\lambda\}\cdot\left\{1+O\left(\frac{1}{M}\right)\right\}$$

$$\frac{E\{V_D(\hat{Y}_C)|\vec{N}\}}{E\{V_D(\hat{Y}_E)|\vec{N}\}} = \{1+\lambda\}\cdot\{1+\gamma\}\cdot\left\{1+O\left(\frac{1}{M}\right)\right\}$$

where $\bar{N}^* = \sum_{k \in U_0} N_k^2/N$ is a weighted average cluster size. We may call $E\{V_D(\hat{Y}_C)|\vec{N}\}/E\{V_D(\hat{Y}_E)|\vec{N}\}$ the anticipated DEFF. Asymptotically, it is neatly decomposed into a product of the anticipated variance ratio between $\hat{Y}_C$ and $\hat{Y}_R$ and that between $\hat{Y}_R$ and $\hat{Y}_E$. The term λ is approximately given by the ratio $V\{E(y_k|N_k)\}/E\{V(y_k|N_k)\}$, where $V\{E(y_k|N_k)\} = \mu_y^2 V(N_k)$ and $E\{V(y_k|N_k)\} = E(\nu_k)$, which are the two components of the unconditional variance $V(y_k)$ under some joint model of (y_k, N_k). It seems difficult to lay down general conditions of the joint distribution of (y_k, N_k) that affect the term λ monotonically, apart from some special situations. For instance, for a given population of clusters, λ increases with $|\mu_y|$ provided the covariance parameters (σ_y^2, ρ_y) remain the same; and it increases with μ_y^2/σ_y^2 provided the intra-cluster correlation ρ_y remains the same. The cluster size variability is important. Indeed, in the extreme case of $\sigma_y^2 = 0$, i.e., $y_{ki} \equiv \mu_y$, we have $V_D(\hat{Y}_R) = 0$, but $V_D(\hat{Y}_C) > 0$, which is entirely due to the variation in the cluster sizes. The situation is greatly simplified through the use of $\hat{Y}_R$. The anticipated variance ratio, i.e., $1 + \gamma$, no longer depends on the cluster size variability or the parameters (μ_y^2, σ_y^2), but only on the intra-cluster correlation ρ_y and the average cluster size $\bar{N}^*$.

3.   Variance Decomposition for GREG Estimators

Consider now the general case of auxiliary information. A GREG estimator can be written as $\sum_{(ki) \in s} a_{ki} g_{ki} y_{ki}$ where $a_{ki} = 1/\pi_{ki}$ is the initial HT weight and π_{ki} is the inclusion probability of $(ki) \in U$. Suppose first that the GREG estimator is calculated at the cluster level under the model

$$y_k = x_k^T \beta + e_k$$

where x_k is the auxiliary vector for cluster k and β contains the linear regression coefficients, and $E(e_k) = 0$, and $V(e_k) = \sigma_k^2$ which is specified up to a proportional constant, and $Cov(e_k, e_g) = 0$. By nontrivial x_k we mean to exclude the case of $x_k \equiv 1$. Denote by $\hat{Y}_{Cgreg}$ the GREG estimator, where $a_{ki} = a_k = M/m$ and $g_{ki} = g_k = 1 + \left\{ X - (M/m) \sum_{k \in s_0} x_k \right\}^T \left(\sum_{k \in s_0} x_k x_k^T/\sigma_k^2 \right)^{-1} \left(\sum_{k \in s_0} x_k/\sigma_k^2 \right)$. Denote by $V_D(\hat{Y}_{Cgreg})$ the approximate sampling variance, which is available from Result 8.4.1 in Särndal et al. (1992). Given SRS of clusters, it can be obtained from $V_D(\hat{Y}_C)$ by replacing y_k with e_k.

Next, suppose that the GREG estimator is calculated at the element level under the model

$$y_{ki} = x_{ki}^T \beta + e_{ki}$$

where x_{ki} is the auxiliary vector for element (ki) and β contains the regression coefficients, and $E(e_{ki}) = 0$, and $V_M(e_{ki}) = \sigma_{ki}^2$ which is specified up to a proportional constant, and $Cov(e_{ki}, e_{kj}) = Cov(e_{ki}, e_{gj}) = 0$, i.e., zero intra-cluster correlation, which is common in practice. We assume that the models at the cluster and the element level are connected in that $x_k = \sum_i x_{ki}$. Denote by $\hat{Y}_{Rgreg}$ the GREG estimator, where $a_{ki} = a_k = M/m$ and $g_{ki} = 1 + \left\{ X - (M/m) \sum_{(ki) \in s} x_{ki} \right\}^T \left(\sum_{(ki) \in s} x_{ki} x_{ki}^T/\sigma_{ki}^2 \right)^{-1} \left(\sum_{(ki) \in s} x_{ki}/\sigma_{ki}^2 \right)$. The approximate sampling variance, denoted by $V_D(\hat{Y}_{Rgreg})$, can be obtained from $V_D(\hat{Y}_R)$ simply by replacing $y_k = \sum_i y_{ki}$ with $e_k = \sum_i e_{ki}$ (see Särndal et al. 1992, Result 8.9.1).

Finally, denote by $\hat{Y}_{Egreg}$ the GREG estimator under the same element-level model given SRS of elements, where $g_{ki} = 1 + \left\{ X - (N/n) \sum_{(ki) \in s} x_{ki} \right\}^T \left(\sum_{(ki) \in s} x_{ki} x_{ki}^T / \sigma_{ki}^2 \right)^{-1}$ $\left(\sum_{(ki) \in s} x_{ki} / \sigma_{ki}^2 \right)$ and $a_{ki} = a_k = N/n$. Denote by $V_D(\hat{Y}_{Egreg})$ the approximate sampling variance, which can be obtained from $V_D(\hat{Y}_E)$ simply by replacing y_{ki} with e_{ki} (see Särndal et al. 1992, Result 6.6.1).

For the anticipated variances of these GREG estimators we use the following model

$$y_{ki} = x_{ki}^T \beta + e_{ki}$$

where $E(e_{ki}) = 0$, and e_{ki} has the same intra-cluster covariance structure as that of ε_{ki} in Section 2, but the parameters are now given by (σ_e^2, ρ_e). The anticipated variances can be obtained from the corresponding formulae (1)–(3) above by replacing $v_k(y)$ with $v_k(e) = N_k \sigma_e^2 + N_k(N_k - 1)\rho_e \sigma_e^2$, and $\left(\mu_y, \sigma_y^2, \rho_y \right)$ with $(0, \sigma_e^2, \rho_e)$. Under the same asymptotic setting as in Section 2, we have

$$\frac{E\{V_D(\hat{Y}_{Cgreg})|\vec{N}\}}{E\{V_D(\hat{Y}_{Egreg})|\vec{N}\}} = \frac{E\{V_D(\hat{Y}_{Rgreg})|\vec{N}\}}{E\{V_D(\hat{Y}_{Egreg})|\vec{N}\}} \cdot \left\{ 1 + O\left(\frac{1}{M}\right) \right\}$$

$$= \{1 + \rho_e(\bar{N}^* - 1)\} \cdot \left\{ 1 + O\left(\frac{1}{M}\right) \right\}$$

Thus, while the anticipated DEFF $E\{V_D(\hat{Y}_C)|\vec{N}\}/E\{V_D(\hat{Y}_E)|\vec{N}\}$ remains the same, the use of nontrivial auxiliary information in the general case can reduce its impact in two respects. (a) $\hat{Y}_{Cgreg}$ no longer has an extra variance term compared to $\hat{Y}_{Rgreg}$. In comparison, $\hat{Y}_C$ can be regarded as a GREG estimator in the minimal case with $x_k \equiv 1$, and $\hat{Y}_R$ as a GREG starting with $\hat{Y}_C$ and using $x_{ki} \equiv 1$ at the element level. (b) The anticipated variance ratio between $\hat{Y}_{Rgreg}$ and $\hat{Y}_{Egreg}$ is reduced as compared to that between $\hat{Y}_R$ and $\hat{Y}_E$ to the extent that ρ_e is smaller than ρ_y. Notice also that, provided σ_e^2 is smaller than σ_y^2, a GREG estimator in the general case has a smaller anticipated variance than the corresponding GREG estimator in the minimal case, such as $\hat{Y}_{Rgreg}$ compared to $\hat{Y}_R$.

4.   An Illustration Based on NLFS

We now illustrate the above results using the NLFS data from the first quarter of 2005. The NLFS is based on a single-stage cluster sample. The clusters are families in the Central Population Register, which is a kind of household unit. The population size is slightly above 3.3 million people, and there are 21,525 persons in our sample. To illustrate the general case of auxiliary information we use register variables age (12 groups), sex and a register-based employment status (i.e., "employed" or "not employed"), which can be linked to the sample through a unique Personal Identification Number.

The estimated standard errors (SEs) of all the six estimators considered in Sections 2 and 3 are given in Table 1. The formulae for variance estimators are available from the above-cited sources for the corresponding sampling variances. For the GREG estimators we set $\sigma_k^2 \propto 1$ and $\sigma_{ki}^2 \propto 1$, as is common in practice. We have also carried out the calculation with $\sigma_k^2 \propto N_k$ for $\hat{Y}_{Cgreg}$, and the resulting estimates are very close to the ones reported in Table 1. The estimates on the "Person" line are calculated as if the sample were selected using SRS of elements. This is plausible provided $n/m \approx N/M$, which holds in

Table 1. Estimated standard errors for total employment and unemployment in NLFS (2005, 1st Quarter). Minimal auxiliary information: number of elements in population and sample cluster sizes. Register auxiliary information: age, sex and register employment status

Sampling unit	Minimal auxiliary information			Register auxiliary information		
	Estimator	Employment	Unemployment	Estimator	Employment	Unemployment
Cluster	$\hat{Y}_C$	15,791	3,958	$\hat{Y}_{Cgreg}$	7,094	4,192
Cluster	$\hat{Y}_R$	10,859	3,931	$\hat{Y}_{Rgreg}$	7,071	4,152
Person	$\hat{Y}_E$	10,341	3,856	$\hat{Y}_{Egreg}$	6,948	4,077

a large sample situation as the NLFS, and provides the basis for assessing the design effect of SRS of clusters in practice.

Take first the minimal case. We note that cluster sampling has a large DEFF of over 200% for total employment. But almost all the DEFF can be removed by switching to the ratio-to-size estimator, reducing the SE from 15,791 to 10,859 as compared to 10,341 based on SRS of elements. For the NLFS we have $\bar{N}^* \approx 1.6$, and $\rho_y \approx 0.2$ for the binary variable "employed" or "not employed" (see Rafat 2002), such that $1 + \gamma = 1 + \rho_y(\bar{N}^* - 1) \approx 1.12$; whereas $\hat{V}_D(\hat{Y}_R)/\hat{V}_D(\hat{Y}_E) = (10,859/10,341)^2 = 1.10$ in Table 1. The intra-cluster correlation model seems to be a good approximation in the NLFS situation. Meanwhile, the differences are small for total unemployment, and the DEFF is only 105%. The difference between $\hat{Y}_C$ and $\hat{Y}_R$ is small because the mean per element for the binary unemployment variable is only about 0.03, compared to about 0.7 for the binary employment variable. The difference between $\hat{Y}_E$ and $\hat{Y}_R$ is small because the intra-cluster correlation is low. Indeed, judging from the results in Table 1, we have $\rho_y \approx \{\hat{V}_D(\hat{Y}_R)/\hat{V}_D(\hat{Y}_E) - 1\}/(\bar{N}^* - 1) = 0.065$ for the unemployment variable.

The use of register auxiliary information reduces the sampling variance for total employment as compared to the minimal case. It practically removes the difference between $\hat{Y}_{Cgreg}$ and $\hat{Y}_{Rgreg}$. The difference between $\hat{Y}_{Rgreg}$ and $\hat{Y}_{Egreg}$ is also reduced because, while $\bar{N}^*$ is the same, the conditional intra-cluster correlation is smaller, i.e., the results suggest $\rho_e \approx \{\hat{V}_D(\hat{Y}_{Rgreg})/\hat{V}_D(\hat{Y}_{Egreg}) - 1\}/(\bar{N}^* - 1) = 0.060$ as compared to $\rho_y \approx 0.2$. As for the total unemployment, the register auxiliary information here is not efficient; and the conditional intra-cluster correlation is approximately $\rho_e \approx 0.062$, which is basically the same as $\rho_y \approx 0.065$. Indeed, the estimated variances of the GREG estimators are slightly increased using the register information, which is a price one pays for using good auxiliary information for employment that at the same time is ineffective for unemployment.

5. On Varying Sampling Probability

Varying sampling probability can be achieved by stratification or some probability proportional-to-size (PPS) scheme. For simplicity we consider PPS cluster sampling *with* replacement, where the sampling probability is given by $p_k = N_k/N$ for $k \in U_0$. Denote by $\hat{Y}_P = (N/m) \sum_{k \in s_0} (y_k/N_k)$ the HT estimator. Under the intra-cluster correlation model in the minimal case (Section 2), we have

$$E\{V_D(\hat{Y}_P)|\vec{N}\} = \frac{N^2}{m} E\left\{\sum_{k \in U_0} p_k(\bar{\varepsilon}_k - \bar{\varepsilon})^2 |\vec{N}\right\}$$

$$= \frac{M}{m} N\sigma_y^2 \left\{1 + \rho_y(\bar{N} - 1) - \frac{1}{M \cdot N} \sum_{k \in U_0} v_k(y)/\sigma_y^2\right\}$$

where $\bar{\varepsilon}_k = \varepsilon_k/N_k$ and $V_D(\hat{Y}_P)$ is given by Cochran (1977, Equation (9A.6)). Thus, asymptotically and disregarding the finite population correction, we obtain the anticipated DEFF of PPS cluster sampling as

$$E\{V_D(\hat{Y}_P)|\vec{N}\}/E\{V_D(\hat{Y}_E)|\vec{N}\} = \{1 + \rho_y(\bar{N} - 1)\} \cdot \{1 + O(1/M)\}$$

which is very similar to the corresponding ratio between $\hat{Y}_R$ and $\hat{Y}_E$ in Section 2. This is intuitive because it is the same information of cluster size that is being used in both cases: in PPS sampling it is used before the sample is selected, whereas for the ratio-to-size estimator it is used afterwards.

Now, consider the general case of nontrivial auxiliary information as in Section 3, and the GREG estimator applied to PPS sampling, denoted by $\hat{Y}_{Pgreg}$. The "assisting" model can either be the cluster- or element-level model above. The approximate sampling variance, denoted by $V_D(\hat{Y}_{Pgreg})$, is given by $V_D(\hat{Y}_P)$ on replacing y_k with e_k and $\bar{Y}$ with $\bar{e}$. For the anticipated variance we assume the constant intra-cluster correlation model (Section 3) for e_{ki} with parameters $\left(\sigma_e^2, \rho_e\right)$. Again, the extra auxiliary information has two effects: (i) it reduces the anticipated variances provided $\sigma_e^2 < \sigma_y^2$, and (ii) it reduces the difference between the sampling variances under PPS of clusters and SRS of elements, i.e., provided $\rho_e < \rho_y$, $E\{V_D(\hat{Y}_{Pgreg})\}/E\{V_D(\hat{Y}_{Egreg})\}$ is closer to unity compared to $E\{V_D(\hat{Y}_P)\}/E\{V_D(\hat{Y}_E)\}$.

6. References

Cochran, W.G. (1977). Sampling Techniques, (3rd ed.). New York: Wiley.

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, 105–106.

Isaki, C. and Fuller, W. (1982). Survey Design Under the Regression Superpopulation Model. Journal of the American Statistical Association, 77, 89–96.

Kish, L. (1987). Weighting in Deft[2]. The Survey Statistician, June.

Park, I. and Lee, H. (2004). Design Effects for the Weighted Mean and Total Estimators Under Complex Survey Sampling. Survey Methodology, 30, 183–193.

Pfeffermann, D. (1993). The Role of Sampling Weights When Modeling Survey Data. International Statistical Review, 61, 317–337.

Rafat, D. (2002). Analyse av sammenheng mellom ektefellers sysselsetting i en familie. Statistics Norway, Notater 2002/35. [In Norwegian].

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