# ASYMPTOTICALLY OPTIMAL BILINEAR AND QR ESTIMATORS FOR SMALL AREA PARAMETERS.

Wietse Dol and Ton Steerneman, University of Groningen
Wietse Dol, Department of Econometrics, University of Groningen,
P.O.Box 800, 9700 AV Groningen, The Nertherlands

KEYWORDS: Superpopulation model, replication scheme, bilinear estimators, QR estimators small area parameters.

## 1. Introduction

The superpopulation approach is often used to estimate certain population parameters more accurately than is possible by applying the classical sampling theory. By postulating a superpopulation model it is for instance possible to make inferences about subpopulations (also called small areas) using the whole sample instead of only the sampled elements that are in the subpopulation. If such a superpopulation model applies, more accurate results can be obtained then by classical sampling theory. This is especially the case if we are dealing with a small subpopulation and a sample that contains only a very small number of individuals from this subpopulation. The accuracy of the superpopulation approach, however, depends on the validity of the model used.

This paper considers bilinear estimators, i.e. estimators that are linear with regard to both the model and the sampling design. The mean-squared error is used as a yardstick. In the class of bilinear estimators, optimal estimators are derived under several possible restrictions like (asymptotic) design–, model–, and model/design–unbiasedness. When the superpopulation model parameters are known this can be done easily. If the parameters are unknown and have to be estimated, it is more difficult to find optimal estimators. The QR estimator is an intuitive extension of the bilinear estimator when we have to estimate the superpopulation model parameters. To compare QR estimators we will use an asymptotic method similar to the one described by Brewer (1979).

Quite frequently asymptotic design–unbiased estimators are preferred, but we will see that the ordinary least squares (OLS) estimator is better than asymptotic design–unbiased estimators and that the OLS estimator is an optimal QR estimator.

We will use the superpopulation approach to estimate a (sub)population parameter $T=c'y$, with $y$ an $N$x1 vector of characteristics under interest and $c$ some $N$x1 vector. If we are e.g. interested in the population mean we will use $c=N^{-1}\mathbb{1}$ with $\mathbb{1}$ being the $N$x1 vector of ones. It is postulated that the actual value $y_i$ of the characteristic under interest belonging to the $i^{th}$ population element is a realization of the random variable $Y_i$ $(i=1,...,N)$. The random vector $Y=(Y_1,...,Y_N)'$ satisfies the following linear model:

$$Y=X\beta+\varepsilon, \qquad \varepsilon\sim(0,\ \sigma^2 I), \qquad (1.1)$$

where $X$ is an $N$x$k$ matrix of known auxiliary variables, $\beta$ a $k$x1 vector of parameters and $\varepsilon$ an $N$x1 vector of uncorrelated identically distributed random disturbances.

In section 2 we will give some definitions and notations needed in the subsequent sections. In section 3 we will give optimal bilinear estimators when model parameters are known. Section 4 is devoted to the QR estimators when the model parameters are unknown. For proofs and more examples we refer to Dol and Steerneman (1989). Section 5 contains a summary of the results and some remarks.

## 2. Some Definitions and Notations

Consider a population that consists of $N$ elements, labelled 1,...,$N$. The labels are supposed to uniquely determine the population units. The number of elements $N$ is called the **population size.** The **characteristic under interest** of population unit $i$ $(i=1,...,N)$ is denoted by $Y_i$. The $Y_i$'s are not known. By drawing a sample of size $n$ we get to know some of the $Y_i$'s. For all elements $i$ there are $k$ auxiliary variables with scores $x_{i1},...,$ $x_{ik}$ on population unit $i$ which are known to us.

It is very convenient to write formulas in matrix notation. We will use the following matrices:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & & & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Nk} \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix},$$

$$Y = (Y_1, \ldots, Y_N),$$

$$\Pi = \operatorname{diag}(\pi_1, \pi_2, \ldots, \pi_N),$$

where $\Pi$ denotes the diagonal matrix of the **first order inclusion probabilities**, e.g. $\pi_i$ is the probability that element $i$ is sampled. The **sample matrix**

$$P = \operatorname{diag}(p_1, p_2, \ldots, p_N)$$

$$\text{with} \quad p_i = \begin{cases} 1 & \text{if } i \in \text{sample } s \\ 0 & \text{if } i \notin \text{sample } s \end{cases}$$

denotes which population elements are sampled. In the sequel we frequently identify a sample with the matrix $P$. Because we do not know $Y$ but only a sample of $Y$ we are looking at the model

$PY = PX\beta + P\varepsilon.$

The samples we shall consider are so-called **Fixed Effective Sample size $n$ samples** ($FES(n)$ samples): only samples with $n$ different population elements have positive probability of being selected. So e.g. simple random sampling with replacement is excluded. We also only consider samples with positive first order inclusion probability for all population elements ($\pi_i > 0$, $i = 1, \ldots, N$). This means that every population element has a positive probability of being sampled. Another restriction is that we only consider **noninformative** (or **exogenous**) designs. Noninformative sampling designs are designs where the inclusion probabilities do not depend on the characteristics under interest $Y_i$. The reason why we use noninformative sampling designs is that the random mechanisms due to the model and the sampling design are then independent.

There are two stochastic components: model stochastics and sampling design stochastics. We use $E_p$ to denote the conditional expectation with respect to sampling given $Y$, that is

$$E_p(\hat{T}) = E(\hat{T} \mid Y)$$

and $E_\varepsilon$ to denote the conditional expectation with respect to the disturbances given the sample $P$, that is

$$E_\varepsilon(\hat{T}) = E(\hat{T} \mid P).$$

For example we have $E_p(P) = \Pi$ and $E(Y) = X\beta$.

Because we have two stochastic components, we have several forms of unbiasedness:

Definition 2.6. An estimator $\hat{T}$ is called **$\varepsilon$-unbiased** or **model-unbiased** if and only if $E_\varepsilon(\hat{T}-T) = 0$ for all samples $s$ with $p(s)>0$. An estimator $\hat{T}$ is called **$p$-unbiased** or **design-unbiased** if and only if $E_p(\hat{T}-T) = 0$ almost everywhere. An estimator $\hat{T}$ is called **$p\varepsilon$-unbiased** or **unbiased** if and only if $E(\hat{T}-T) = E_\varepsilon E_p(\hat{T}-T) = E_p E_\varepsilon(\hat{T}-T) = 0$.

In order to find an optimal estimator for the population parameter $T$ we have to take into account both the sampling design and the superpopulation model as well. The sampling design has to be considered, because the way the characteristics under interest are gathered, is important for choosing an estimator. The superpopulation model gives us extra structure and this should be employed when constructing an estimator. To compare estimators we will use the mean-squared error ($MSE$). The mean-squared error of an estimator $\hat{T}$ for $T$ is defined as

$$MSE(\hat{T}) = E_\varepsilon E_p(\hat{T} - T)^2.$$

It would be nice to find estimators $\hat{T}$ that are optimal in the sense that they have the smallest possible (asymptotic) mean-squared error. Often there is no estimator with uniformly smallest mean-squared error, so we limit ourselves to smaller classes of estimators. It turns out that we can find such optimal estimators in a restricted class, e.g. the class of bilinear estimators.

## 3. Bilinear estimators

In this section we will consider the class of bilinear estimators. These estimators are, given the vector $Y$, linear in the matrix $P$ and, given $P$, linear in $PY$. Since we only get to know the sampled elements $PY$, instead of the vector $Y$, we can only use $PY$ in the estimator $\hat{T}$. The general form of a bilinear estimator is

$$\hat{T} = c'APY + c'BPa + c'b, \qquad (3.1)$$

where $A$, $B$ are $N{\times}N$ matrices and $a$, $b$ are $N{\times}1$ vectors. We have to choose $A$, $B$, $a$, and $b$ in such a way that the mean-squared error (risk function) is as small as possible for each $FES(n)$ sampling design. This results in the following theorem:

Theorem 3.1. Consider the model $Y = X\beta + \varepsilon$ with $\varepsilon \sim (0, \sigma^2\mathrm{I})$ and $\beta$, $\sigma^2$ are known. Given the class of bilinear estimators (3.1), the estimator

$$\hat{T}_o = c'PY + c'(I-P)X\beta$$

is the unique optimal bilinear estimator for $T = c'Y$ with regard to $MSE$ and

$$MSE(\hat{T}_o) = \sigma^2 c'(I - \Pi)c.$$

Because $\hat{T}_o$ is an $\varepsilon$-unbiased estimator, $\hat{T}_o$ is also the unique optimal $\varepsilon$-unbiased bilinear estimator.

Remark: We see that there is an optimal choice for $A$, $B$, $a$, and $b$ that does not depend on the vector $c$. The subpopulation elements that are in the sample are directly used ($PY$) and those not in the sample are estimated by the best model estimator $(I-P)X\beta$.

Using $\hat{T}_o$ the following corollary will give us the optimal sampling design and hence the optimal strategy.

Corollary 3.1. Let the population elements be ordered such that $c_1^2 \geq c_2^2 \geq \ldots \geq c_N^2$. An optimal sampling design, using the optimal bilinear estimator $\hat{T}_o$ is such that:

$$\pi_i = \begin{cases} 1 & \text{for } i = 1, \ldots, n, \\ 0 & \text{for } i = n+1, \ldots, N. \end{cases}$$

In the classical sampling theory $p$-unbiasedness is a desirable property of an estimator. Suppose we limit ourselves to the class of bilinear $p$-unbiased estimators then the following theorem holds:

Theorem 3.2. Consider the model $Y = X\beta + \varepsilon$ with $\varepsilon \sim (0, \sigma^2 I)$ and $\beta$, $\sigma^2$ are known, then the estimator

$$\hat{T}_p = c'\Pi^{-1}PY + c'(I - \Pi^{-1}P)X\beta \qquad (3.2)$$

is the unique optimal bilinear $p$-unbiased estimator of $T = c'Y$ in the class of bilinear $p$-unbiased estimators, with

$$MSE(\hat{T}_p) = \sigma^2 c'(\Pi^{-1} - I)c.$$

Remark: Note that $\hat{T}_p$ defined in (3.2) is also $\varepsilon$-unbiased. So both $\hat{T}_p$ and $\hat{T}_o$ are $p\varepsilon$-unbiased estimators but $MSE(\hat{T}_o)$ is smaller than $MSE(\hat{T}_p)$. The optimal $p$-unbiased bilinear estimator has the same $MSE$ as the optimal ($\varepsilon$-unbiased) bilinear estimator if and only if $\pi_i = 1$ for $i = 1, \ldots, N$. This means that we take the whole population as a sample. The difference between $MSE(\hat{T}_p)$ and $MSE(\hat{T}_o)$ will in general be large. We conclude that if a superpopulation model is adequate, we should use that model to estimate $T$ and not trying to estimate $T$ using an $p$-unbiased estimator as in classical sampling theory.

For the moment suppose we no longer limit ourselves to $\pi_i > 0$ for all $i$, but use $\pi_i \geq 0$ instead. We like to derive the optimal sampling design that minimizes the $MSE$ of the optimal $p$-unbiased bilinear estimator for subpopulation parameter $T = c'Y$. Choosing an optimal sampling design when one is only interested in a subpopulation parameter will result in a sampling design that depends on the vector $c$.

Corollary 3.2. An optimal sampling design, using the optimal $p$-unbiased bilinear estimator $\hat{T}_p$ is such that:

$$\pi_i = n \; |c_i| \Big/ \sum_{j=1}^{N} |c_j|, \qquad i = 1, \ldots, N.$$

## 4. QR estimators

In section 3 we have derived the optimal bilinear, $p$-unbiased, and $\varepsilon$-unbiased bilinear estimators. They all appeared to have the same form

$$\hat{T} = c'RPY + c'(I - RP)X\beta \qquad (4.1)$$

with $R = I$ if we look at the optimal ($\varepsilon$-unbiased) bilinear estimators and $R = \Pi^{-1}$ if we look at the optimal $p$-unbiased bilinear estimators. In this section we will consider the case that we do not know the superpopulation model parameters. In order to estimate the population parameter $T$ we have to estimate the model parameters. This means that we have to estimate $\beta$ by some estimator $\hat{\beta}$. In the literature several estimators are proposed. Two frequently applied estimators are the OLS and the GLS estimators. If we estimate $\beta$ conditionally given the sample, that is given the matrix $P$, we estimate $\beta$ by the OLS method:

$$\hat{\beta}_{ols} = (X'PX)^{-1}X'PY. \qquad (4.2)$$

We can calculate (4.2) only if $(X'PX)^{-1}$ has rank $k$ for all matrices $P$ and therefore we will assume that the rank of $PX$ is $k$, with sampling design probability one. If we estimate $\beta$ unconditionally, then we want to estimate the model

$$PY = PX\beta + P\varepsilon,$$

with

$$E(P\varepsilon) = 0 \text{ and } E(P\varepsilon\varepsilon'P) = E_p(PE_\varepsilon(\varepsilon\varepsilon')P) = \sigma^2\Pi,$$

and GLS results in

$$\hat{\beta}_{reg} = (X'P\Pi^{-1}PX)^{-1}X'P\Pi^{-1}PY = (X'\Pi^{-1}PX)^{-1}X'\Pi^{-1}PY.$$

In the literature the GLS estimator is also known as the regression estimator.

Both estimators are $\varepsilon$-unbiased but are in general not $p$-unbiased and they can be considered as weighted least squares estimators:

$$\hat{\beta} = (X'PQPX)^{-1}X'PQPY, \qquad (4.3)$$

with $\hat{\beta} = \hat{\beta}_{ols}$ if $Q = I$ and $\hat{\beta} = \hat{\beta}_{reg}$ if $Q = \Pi^{-1}$. The question is: 'How to choose $Q$, do we take $Q = I$, $Q = \Pi^{-1}$ or should $Q$ be something else ?'

By replacing $\beta$ in (4.1) by $\hat{\beta}$ given in (4.3) we obtain the so called QR estimator. If $\pi_{ij} = 0$, the probability that $i$ and $j$ are sampled both, then $(PQP)_{ij} = p_i p_j q_{ij} = 0$ with probability one. So, without loss of generality we may assume that $q_{ij} = 0$ for all pairs $(i,j)$ with $\pi_{ij} = 0$. As in (4.2) we have to assume that $X'PQPX$ has rank $k$ with design probability one and therefore we will assume that $Q$ has full rank $N$ and that $PX$ has rank $k$ with design probability one.

Definition 4.1. An estimator for the population parameter $T = c'Y$ is called a QR estimator if it is of the form

$$\hat{T}_{QR} = c'RPY + c'(I - RP)X\hat{\beta}_{QR},$$

where

$$\hat{\beta}_{QR} = (X\ PQPX)^{-1}X'PQPY.$$

The matrices $R$ and $Q$ are $N \times N$ matrices and $Q$ is of full rank. The matrix $Q$ is symmetric with $q_{ij} = 0$ for all pairs $(i,j)$ with $\pi_{ij} = 0$.

Our aim is to find matrices $R$ and $Q$ that minimize the asymptotic $MSE$ ($AMSE$) in the class of the QR estimators. A problem in calculating the $MSE$ of a QR estimator is that we no longer have simple expressions when taking the design expectation. That is why we will consider the asymptotic distribution of $\hat{T} - T$, if the expiriment is replicated. For asymptotics we will use a method slightly different from the one introduced by Brewer (1979), known as the **replica method**. Our replication scheme considers the asymptotics as if the experiment is repeated infinely many times. At "time" $t \in \{1,2,3,...\}$ a new vector $Y_t$ is generated and a new sample $P_t$ is chosen according to the fixed sampling design and we aggregate the $t$ vectors $Y_t$ and $P_t$ to get the population parameter

$$T_t = \sum_{h=1}^{t} c'Y_h.$$

In the situation of replication we estimate $T_t$ by

$$\hat{T}_t = \sum_{h=1}^{t} [c'RP_hY_h + c'(I - RP_h)X\hat{\beta}_t],$$

where

$$\hat{\beta}_t = (\sum_{h=1}^{t} X'P_hQP_hX)^{-1} \sum_{h=1}^{t} X'P_hQP_hY.$$

For examples of the replica approach see e.g. Ten Cate (1986) and Rao (1984). In Brewer's replica approach the same vector $Y$ is used at all times $t$. We, however, use the superpopulation model approach in full and use a different $Y_t$ for each time $t$.

The remainder of this section will be devoted to the derivation of the asymptotic distribution of a QR estimator in case of replication. The following theorem shows that for $Q = R = I$ the asymptotic optimal QR estimator $\hat{T}_{ols}$ under the replication scheme is obtained. We will see that it is not the unique optimal asymptotic QR estimator. If we substitute $\hat{\beta}_{ols}$ for $\hat{\beta}$ in the general expression for $\hat{T}$, then it can be shown that the choice $R=I$ leads to the unique optimal QR estimator in the subclass of QR estimators based on the optimal $\hat{\beta}_{ols}$.

Theorem 4.3. A) If the population parameter $T=c'Y$ is estimated by the OLS estimator

$$\hat{T}_{ols}=c'PY+c'(I-P)X\hat{\beta}_{ols},$$

$$\hat{\beta}_{ols}=(X'PX)^{-1}X'PY,$$

then under the replication scheme

$$\mathcal{L}\ \ t^{-½}(\hat{T}_{ols,t}-T_t)\xrightarrow{\mathcal{L}}$$

$$\mathrm{N}(\ 0,\ \sigma^2c'(I-\Pi)c+\sigma^2c'(I-\Pi)X(X'\Pi X)^{-1}X'(I-\Pi)c),$$

and it is an asymptotic optimal estimator in the class of QR estimators of definition 4.1.

B) If the population parameter $T=c'Y$ is estimated by an R estimator

$$\hat{T}_R=c'RPY+c'(I-RP)X\hat{\beta}_{ols},$$

and $R$ is an $N$x$N$ matrix, then under the replication scheme $\hat{T}_{ols}$ is the unique optimal R estimator.

C) The QR estimator $\hat{T}_{QR}$ is asymptotic optimal in the class of QR estimators if and only if it can be written in the form

$$\hat{T}_{QR}=c'PY+c'(I-P)X\hat{\beta}_{QR}+c'(\Pi-I)X(X'\Pi X)^{-1}X'P(Y-X\hat{\beta}_{QR}),$$

where the matrix $Q$ satisfies

$$SPQPX=PSQSX$$

for all independent sample design matrices $P$ and $S$ that have positive probability.

Theorem 4.3 shows that the OLS procedure is optimal in an asymptotic sense. This is not true for the regression estimator

$$\hat{T}_{reg}=c'\Pi^{-1}PY+c'(\mathrm{I}-\Pi^{-1}P)X\hat{\beta}_{reg},$$

$$\hat{\beta}_{reg}=(X'\Pi^{-1}PX)^{-1}X'\Pi^{-1}PY,$$

which is a QR estimator with $Q=R=\Pi^{-1}$. It can be shown that its *AMSE* is equal to

$$AMSE(\hat{T}_{reg})=\sigma^2c'(\Pi^{-1}-I)c$$

which is strictly larger than the *AMSE* of the OLS estimator in case that $\pi_i<1$ for all $i$. The theorem shows that the OLS procedure is in general not unique in having the smallest *AMSE*. In fact any QR estimator with $R=I$ and $Q$ is diagonal gives an optimal estimator. So, one could take $Q=\Pi^{-1}$. The theorem gives an optimal QR estimator based on $\hat{\beta}_{reg}$:

$$c'PY+c'(I-P)X\hat{\beta}_{reg}+c'(\Pi-I)X(X'\Pi X)^{-1}X'P(Y-X\hat{\beta}_{reg}).$$

The correction term can be interpreted as being due to the fact that the residuals contain additional information in comparison to the information already contained in $PY$ and $X\hat{\beta}_{reg}$. If OLS is used the correction term is equal to zero. This suggests that OLS is more appropriate in the sense that no correction based on residuals is needed.

The OLS procedure is optimal in an asymptotic sense. The choice for $Q$ and $R$ in this case does not depend on the particular value of $c$. What happens is that in fact the vector of population values $Y$ is estimated. The elements in the sample are known ($PY$) and the elements not sampled are estimated (by $(I-P)X\hat{\beta}_{ols}$). The OLS procedure estimates $Y$ by

$$\hat{Y}_{ols}=PY+(I-P)X\hat{\beta}_{ols}.$$

A particular choice for $c$ leads to

$$\hat{T}_{ols}=c'\hat{Y}_{ols}$$

as an estimator for $T=c'Y$. In practice various values of $c$ will be of interest. All that will be needed is in fact $\hat{Y}_{ols}$, because an estimator for some population parameter $c'Y$ can almost immediately be obtained.

The reason why the regression estimator is so popular is that many sampling designers prefer asymptotic design-unbiased estimators, although their *AMSE* can be relatively large in the superpopulation context.

## 5. Summary and remarks

In this article we have presented optimal bilinear as well as optimal QR estimators. Knowing the superpopulation model parameters it is easy to derive optimal bilinear estimators. Bilinear estimators are linear in the model as well as in the sampling design. Section 3 gives several optimal estimators under several restrictions. We seldomly know the superpopulation parameters, so we have to estimate the vector of model parameters $\beta$ in order to estimate the population parameter $T$. Estimating $\beta$ by $\hat{\beta} = (X'PQPX)^{-1}X'PQPY$ results in a QR estimator $\hat{T}$ that is no longer linear in the sampling design. We use the QR estimator as an intuitive extension of a bilinear estimator. The optimal bilinear estimator is given by

$$\hat{T}_o = c'PY + c'(I-P)X\beta \qquad (5.1)$$

and when we do not know $\beta$, we simply estimate $\beta$ by $\hat{\beta}_{ols}$ and (5.1) becomes

$$\hat{T}_{ols} = c'PY + c'(I-P)X\hat{\beta}_{ols}.$$

When we apply the optimal $p$-unbiased bilinear estimator

$$\hat{T}_p = c'\Pi^{-1}PY + c'(I-\Pi^{-1}P)X\beta,$$

a popular estimator in case $\beta$ is unknown becomes

$$\hat{T}_{reg} = c'\Pi^{-1}PY + c'(I-\Pi^{-1}P)X\hat{\beta}, \qquad (5.2)$$

where $\beta$ is estimated by $\hat{\beta}$.

From theorem 4.3. we see that the OLS estimator is an optimal QR estimator. Wright (1983) has shown that (5.2) is not exactly $p$-unbiased but an asymptotic design-unbiased (ADU) QR estimator. The regression estimator is not an optimal QR estimator.

The literature mainly concentrates on the ADU estimators (see e.g. Brewer (1979), Särndal (1980), Wright (1983), Hansen, Madow, and Tepping (1983) and Brewer, Hanif, and Tam (1988)). The advantage of an ADU QR estimator is that it remains asymptotic design-unbiased whatever model is used. This means that if we specify a wrong model the ADU estimators remain asymptotic design-unbiased. When we use a model-unbiased estimator and have specified a wrong model it will result in a model-bias. However not only the bias is important, the *AMSE* is the risk function we use. Using a wrong model, the OLS estimator (optimal model-unbiased QR estimator) as well as the regression estimator (asymptotic design-unbiased QR estimator) will have a larger *AMSE* as expected. The question is which estimator will have a larger *AMSE*.

The optimal OLS estimator has several nice properties when looking at small areas: the optimal solution is independent of the vector $c$. This means that the OLS estimator is optimal, independent of the small area under interest. We do not have to know the whole $(N\times k)$ matrix $X$ but only the $(k\text{x}1)$ vectors $X_i$ of the sampled elements and the small area totals of the auxiliary variables $X_i$ $(=c'PX_i)$.

## ACKNOWLEDGEMENTS

The authors thank Tom Wansbeek for his useful suggestions and comments. The work of the first author was supported by the Netherlands Organisation for Scientific Research (NWO).

## REFERENCES

BREWER,K.R.W.(1979),"A Class of Robust Sampling Designs for Large-Scale Surveys"; *Journal of the American Statistical Association*, **74**, pp.911-915.

BREWER,K.R.W., HANIF,M. and TAM,S.M.(1988), "How nearly can Model-based Prediction and Design-Based Estimation be Reconciled?"; *Journal of the American Statistical Association*, **83**, pp.128-132.

DOL,W. and STEERNEMAN,A.G.M.(1989),"Asymptotically optimal Bilinear and QR estimators for Small Area parameters", *IER Research memorandum*, Groningen, The Netherlands.

HANSEN,M.H., MADOW,W.G., and TEPPING,B.J.(1983),"An Evaluation of Model- Dependent and Probability-Sampling Inference in Sample Surveys"; *Journal of the American Statistical Association*, **78**, pp.776-793.

RAO,T.J.(1984),"On Brewer's Class of Robust Sampling Designs for Large-Scale Surveys"; *Metrika*, **31**, pp.319-322.

SÄRNDAL,C.E.(1980),"On $\Pi$-inverse Weighting vs. Best Linear Unbiased Weighting in Probability Sampling"; *Biometrika*, **67**, pp.639-650.

TEN CATE,A.(1986),"Analysis using Survey Data, with Endogenous Design"; *Survey Methodology*, **12**, pp.121-136.

WRIGHT,R.L.(1983),"Finite Population Sampling with Multivariate Auxiliary Information"; *Journal of the American Statistical Association*, **78**, pp.879-884.