Sankhyā : The Indian Journal of Statistics
2016, Volume 78-B, Part 1, pp. 119-151

Sampling from Correlated Populations: Optimal Strategies and Comparison Study

Ioulia Papageorgiou
Athens University of Economics and Business, Athens, Greece

Abstract

The problem of sampling from a population with correlated units is considered. The presence of correlation affects all stages of a survey, from the choice of the sampling scheme to the statistical inference. Ignoring or failing to identify the existing correlation can lead to incorrect inference, such as invalid standard errors, since standard sampling techniques are no longer appropriate. In this direction, several sampling methodologies have been proposed in the literature, aiming to accommodate the correlation in both the sampling and the estimation procedure. The problem can be quite difficult when the type of correlation is completely general and existing methods rely on either restricted assumptions about the population structure or limitations to practical implementation. We provide a review of currently available methodologies, drawing attention to the properties of the derived estimates, the assumptions made, the robustness of the methods under different types of correlation and the practical limitations. A question of how these methodologies compare arises because they differ on the optimality criterion they assume towards the solution. Some methodologies are even not theoretically justified, but they are commonly used as known superior in situations of correlated measurements. The comparison study is conducted on a basis of the relative efficiencies among the competing methodologies by using simulated and real data sets.

AMS (2000) *subject classification.* Primary 62D05, Secondary 62M10.
Keywords and phrases. Superpopulation, Systematic sampling, Model-based sampling, Best unbiased predictor, Optimal sampling strategy, Optimal sampling allocation, Spatial sampling.

1 Introduction

Let $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_N)'$ denote a population vector, where N is the population size and Y_i the value of the study variable, Y, for the i-th population unit. The aim of sampling is to gain information about a population parameter, say θ, based on a sample s of size n selected from the $\boldsymbol{Y}$. In general, θ is a function of $\boldsymbol{Y}$, $\theta = \theta(\boldsymbol{Y})$. Simple linear functions,

such as the mean or the population total, are typical parameters of interest. The sampling design or sampling plan, usually denoted as p or $p(s)$, which plays an important role in sampling theory, is a function that assigns a positive probability of selection to every possible sample. Altogether the probabilities sum to one. Therefore, if S is the set of all possible samples of size n drawn from the population, the function $p(s)$ defined on all samples $s \in S$ characterizes the sampling scheme. For example, in simple random sampling the probability of selection is equal among all possible samples. When the selection is without replacement $p(s) = 1/\binom{N}{n}$, since the possible samples in this case are all possible combinations of n measurements chosen from $\boldsymbol{Y}$. For a detailed study of sampling designs see Cassel et al. (1977).

For a given sample, an estimator e is used to estimate θ. The properties of the estimate e are expressed by its expected value and its mean square error (MSE) calculated in terms of the sampling design. For example the expected value of e is $E_p(e) = \sum_{s \in S} e(s)p(s)$ where $e(s)$ is the value of e for sample s.

The pair (p, e) is known as *strategy* (see for example Ramakrishnan 1975). For a given estimator e, finding the best strategy is equivalent to finding the sampling design p that minimizes the MSE of e. Estimators are usually chosen from within a class of estimators with good statistical properties, such as the class of linear unbiased estimators. In practice, however, especially with complex sampling schemes or complex population models, it is common to adopt an estimator with good practical properties, such as ease of computation. A typical example is use of the sample mean to estimate the population mean. The sample mean has good theoretical properties under certain but not under every model.

In this paper we deal with correlated measurements; and we assume the superpopulation approach that is frequently used in sampling from finite population. This approach provides a framework to introduce structure into the population or accommodate relationships among the population members. Under this approach, the population under study is a realization of a sampling procedure, a sample itself, from an infinite theoretical superpopulation. Each population unit Y_i is a random variable with mean μ_i and variance σ_i^2 and possibly covariances with other population units. More specifically, the superpopulation model is

$$Y_i = \mu_i + \varepsilon_i, \quad i = 1, 2, \ldots, N \tag{1.1}$$

where $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_N)'$ is a vector with mean zero and variance-covariance matrix $\boldsymbol{V}$. Special cases for the matrix $\boldsymbol{V}$ lead to special cases for the population vector. For example diagonal $\boldsymbol{V}$ is equivalent to uncorrelated

population units. The superpopulation model was introduced by Cochran (1977, 1953 1st ed.) and further developed by Godambe (1955), Cassel et al. (1977, Ch. 4), Tam (1984), Blight (1973), Mukerjee and Sengupta (1989, 1990) and Bolfarine and Zacks (1992) among others.

Several sampling methods proposed in the literature assume the superpopulation model and aim to provide an optimum sampling plan under various models of population structure and choices of estimator e. Most of them derive an optimal sampling strategy in a similar fashion: starting with an estimator e in a class of estimates with some good properties, the optimal sampling procedure is obtained by minimizing the MSE or some other measure of accuracy of e.

In this work we review specialized methods for deriving an optimal sampling strategy (p, e) for correlated measurements. For each method, we specify its theoretical framework, the assumptions it employs, and the optimal sampling procedure p and associated estimator e that it supports. We present exact formulas for both estimates and their measure of accuracy under the assumed population model. In addition, we characterize the properties of these estimators and discuss issues that arise during their practical implementation. From the statistical point of view, the properties of the estimators are important because they provide a basis for measuring the loss of information that occurs in practice when standard sampling techniques for independent data are employed with correlated data. The necessity of using methods specialized to correlated population will become apparent.

We also conduct a simulation study to compare several methods with respect to the efficiency of the derived estimates. The comparison is meaningful and interesting because the different methods with their different optimization criteria have different efficiencies. Moreover, an interesting aspect revealed by the comparison study is that sampling strategies arising from the various methods differ in robustness to deviations from underlying assumptions. From the practical point of view, how well an optimal strategy performs when applied to population structure slightly different from the one for which it was designed is valuable information.

A primary motivation for our work is the increasing recognition that much scientific sampling involves correlated data. For scientific areas such as quality control, spatial statistics, ecological statistics, social statistics and genetics, a sample from a population of interest plays an essential role. Although traditionally the sample has been assumed to be independent and the consequent theory has been built upon this assumption, there is increasing recognition that the assumption is often violated and that the effect of this violation can be quite significant. For example in quality control

a sample drawn from the production line is used for constructing control charts in order to monitor the line's output. Several authors, including Bagshaw and Johnson (1975), Vasilopoulos and Stamboulis (1978), Yang and Hancock (1990), Montgomery and Mastrangelo (1991) and VanBrackle and Reynolds (1997) have noted the effect that correlation can have on the control charts produced. In spatial statistics a sample in two dimensions is employed in land surveys, agricultural, ground-water monitoring, environmental statistics and socio-economic habitat surveys. Early findings confirm that failure to account for a positive correlation in the data leads to incorrect confidence intervals (Cressie, 1993, p. 14). The accommodate such correlation, the superpopulation model is extensively used in modeling geostatistical data (Cressie, 1993, ch. 1). Similarly, the time or spatial dependence may be age or generation dependence for applications in ecology and genetics, or dependence imposed from socio-economic factors in habitat surveys.

The remainder of the paper is organized as follows. In Section 2 we provide a brief description of the superpopulation model and its more general form, the regression model to provide the necessary background. The resulting estimates and their properties will be the key ingredients for the sampling strategies we deal with. Section 3 is dedicated to a review of methodologies for finding the best sampling strategies in correlated populations. The methods can be categorized into two groups: standard approaches from classical sampling theory that are frequently used in practice for correlated populations; and specialized methods developed to address correlated data. In Section 4 we evaluate the relative efficiency of each method using numerical examples based on simulated data and a real application. The paper concludes in Section 5 with a discussion on how the theory of sampling from correlated populations can be integrated into various scientific areas, providing at the same time some possible future research topics.

2 The Superpopulation Model

The superpopulation model in (1.1) can be regarded as a special case of the regression model, a model where auxiliary variables are also involved. We describe below the more general regression model from which particular superpopulation models can be derived as special cases. The presentation is brief with focus on the estimation procedure of the population parameters.

What characterizes the superpopulation model (1.1) is that the measurements are comprised of a deterministic part μ_i and a stochastic element ε_i. The superpopulation regression model results from model (1.1) if the deterministic part μ_i is modeled as a linear function of a set of auxiliary,

nonstochastic variables that may be available for the population vector. If $x_1, x_2, \ldots, x_q$ are the auxiliary variables, then the regression model is

$$Y_i = \sum_{j=1}^{q} \beta_j x_{ij} + \varepsilon_i, \quad i = 1, 2, \cdots, N, \tag{2.1}$$

where x_{ij} is the response of population member i for the auxiliary variable j and β_j are model coefficients. In matrix notation, the regression model takes the form

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{2.2}$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots \beta_q)'$, $\boldsymbol{X}$ is an $N \times q$ matrix with elements $(\boldsymbol{X})_{ij} = x_{ij}$, and the random vector $\boldsymbol{\varepsilon} = (\varepsilon_i, \varepsilon_2, \cdots, \varepsilon_N)'$ is assumed to have zero mean and covariance matrix $\boldsymbol{V}$. The regression model can accommodate various special cases of superpopulation models by assuming particular values of q and $\boldsymbol{V}$. For instance, when $q = 1$ and $\boldsymbol{X} = \mathbf{1}_N$, the unit vector of size N, the regression model (2.2) is the superpopulation model (1.1) with constant mean value for all population units. Another special case is the linear regression through the origin model, which results from (2.2) if one assumes that $q = 1$, $\boldsymbol{X} = (x_1, \ldots, x_N)'$ and $\boldsymbol{V}$ a diagonal matrix with elements $(x_1, \ldots, x_N)'$.

Often the parameters of interest are $\boldsymbol{\beta}$, the large scale variation parameters in model (2.1), and for this purpose we assume the small variation parameters associated with the errors as known (Bolfarine and Zacks, 1992, ch. 2). In other words, the covariance matrix $\boldsymbol{V}$ is assumed known. When $\boldsymbol{V}$ is diagonal the superpopulation model generates uncorrelated measurements whereas if $\boldsymbol{V}$ is a general positive-definite matrix the model refers to an autocorrelated superpopulation model. Let s be a sample of size n selected from the superpopulation described by (2.2) and suppose the population parameter of interest is the population total $T = \sum_{i=1}^{N} Y_i$. Let also $(\boldsymbol{Y}_s, \boldsymbol{Y}_r)'$, $(\boldsymbol{X}_s, \boldsymbol{X}_r)'$ denote the partitions corresponding to the sampled and non-sampled part of $\boldsymbol{Y}$ and $\boldsymbol{X}$ respectively. If $\boldsymbol{V}_s$, $\boldsymbol{V}_{sr}$, $\boldsymbol{V}_{rs}$ and $\boldsymbol{V}_r$ are also the corresponding partitions of matrix $\boldsymbol{V}$, the best linear unbiased estimator for T is

$$\hat{T}_{BLU} = \mathbf{1}_s'\boldsymbol{Y}_s + \mathbf{1}_r'[\boldsymbol{X}_r\hat{\boldsymbol{\beta}}_s + \boldsymbol{V}_{rs}\boldsymbol{V}_s^{-1}(\boldsymbol{Y}_s - \boldsymbol{X}_s\hat{\boldsymbol{\beta}}_s)] \tag{2.3}$$

with variance

$$\begin{aligned}\mathrm{Var}(\hat{T}_{BLU}) = {} & \mathbf{1}_r'\boldsymbol{V}_r\mathbf{1}_r' - \mathbf{1}_r'\boldsymbol{V}_{rs}\boldsymbol{V}_s^{-1}\boldsymbol{V}_{sr}\mathbf{1}_r \\ & + \mathbf{1}_r'(\boldsymbol{X}_r - \boldsymbol{V}_{rs}\boldsymbol{V}_s^{-1}\boldsymbol{X}_s)(\boldsymbol{X}_s'\boldsymbol{V}_s^{-1}\boldsymbol{X}_s)^{-1}(\boldsymbol{X}_r - \boldsymbol{V}_{rs}\boldsymbol{V}_s^{-1}\boldsymbol{X}_s)'\mathbf{1}_r.\end{aligned} \tag{2.4}$$

The estimator $\hat{\boldsymbol{\beta}}_s$ in (2.3) is the best unbiased estimator for the regression model coefficients $\boldsymbol{\beta}$ provided by the weighted least-squares estimator of $\boldsymbol{\beta}$,

$$\hat{\boldsymbol{\beta}}_s = (\boldsymbol{X}_s \boldsymbol{V}_s^{-1} \boldsymbol{X}_s)^{-1} \boldsymbol{X}_s' \boldsymbol{V}_s^{-1} \boldsymbol{Y}_s \tag{2.5}$$

(see for example Särndal, 1982) for which $\mathrm{Var}(\hat{\boldsymbol{\beta}}_s) = (\boldsymbol{X}_s \boldsymbol{V}_s^{-1} \boldsymbol{X}_s)^{-1}$. The estimator $\hat{\boldsymbol{\beta}}_s$ coincides with the maximum likelihood estimator of $\boldsymbol{\beta}$ if normality for the population units holds. The ordinary least squares estimator $\hat{\boldsymbol{\beta}}_{ols} = (\boldsymbol{X}_s \boldsymbol{X}_s)^{-1} \boldsymbol{X}_s' \boldsymbol{Y}_s$ is inefficient when $\boldsymbol{V}$ is not diagonal (Cressie, 1993, p. 21). Results on the best unbiased estimator are also available from the literature and is based on a complete predictive sufficient statistic. For a detailed study of both estimates we refer to Bolfarine and Zacks (1992, §2.1 and §2.2).

An alternative estimator of T is $T^\star = Nn^{-1} \mathbf{1}_s' \boldsymbol{Y}_s$ which results from the simple sample mean as an estimate of the corresponding population mean. The estimate $T^\star$ is frequently used in practice due to its simple expression and its optimality under more restricted models. More particularly, $T^\star$ is the ordinary least squares estimate of μ for model (1.1) under the special case of $\mu_i = \mu$ for every $i, i = 1, 2, \ldots, N$ and diagonal matrix $\boldsymbol{V}$. The estimate $T^\star$ is often used in practice for correlated data.

When the matrix $\boldsymbol{V}$ is not known in advance, but is instead depended on an unknown small scale parameter $\boldsymbol{\phi}$, i.e. $\boldsymbol{V} = \boldsymbol{V}(\boldsymbol{\phi})$, the estimated weighted least-squares estimator of $\boldsymbol{\beta}_s$, is used in expression (2.3). This estimate denoted by $\hat{\boldsymbol{\beta}}_{e,s}$ is derived from the corresponding expression of $\hat{\boldsymbol{\beta}}_s = (\boldsymbol{X}_s \boldsymbol{V}_s^{-1}(\hat{\boldsymbol{\phi}}) \boldsymbol{X}_s)^{-1} \boldsymbol{X}_s' \boldsymbol{V}_s^{-1}(\hat{\boldsymbol{\phi}}) \boldsymbol{Y}_s$ provided an estimate of $\boldsymbol{\phi}$. Two possible procedures for estimating $\boldsymbol{\phi}$ are often employed in practice (Cressie, 1993, p. 22). One makes use of the maximum likelihood estimator of $\boldsymbol{\phi}$, derived simultaneously with the estimator of $\boldsymbol{\beta}$, when the distribution is known and the second non-parametric and more general method is an iterative procedure. In this approach, an initial estimate of $\boldsymbol{\beta}$ that does not depend on $\boldsymbol{V}$, say $\tilde{\boldsymbol{\beta}}$, is used and the residuals $Y - X\tilde{\boldsymbol{\beta}}$ are constructed. Using those residuals as proxy for error $\boldsymbol{\varepsilon}$ in (2.2) an estimate of $\boldsymbol{\phi}$ can be derived and by plugging in this estimate in $\hat{\boldsymbol{\beta}}_{e,s}$ an improved estimate of $\boldsymbol{\beta}_s$ is obtained. Repetition of the same procedure will lead to final improved estimates for both parameters $\boldsymbol{\phi}$ and $\boldsymbol{\beta}$. When the random process that generates the data is a time series process, identification of the model and estimation of its parameters with time series methodologies will provide with the estimate of $\boldsymbol{V}$, $\boldsymbol{V}(\hat{\boldsymbol{\phi}})$. A preliminary sample is needed for such cases in order to first identify the process mechanism.

3 Review of Methodologies for Correlated Measurements

In this section we present a brief description of the existing methodologies for autocorrelated populations and consider the problem of finding the best pair of sampling scheme and exact estimator. The problem can be very complex depending on the assumed population model. The resulting optimum sampling procedure itself also depends on the population structure. We initially proceed with some broad results and we focus later in this section on the specific methodologies under study.

The choice of optimal sampling procedures for correlated populations has been considered by several authors including Blight (1973), Papageorgiou and Karakostas (1998) and Mukerjee and Sengupta (1989, 1990). Blight (1973) assumes that the finite population is generated by a simple linear Markov process introducing correlation $\rho(h) = \lambda^h$ among population units that are separated by lag h, $h = 1, 2, ..., N-1$. The optimal choice of the sampling units is determined by employing the sample mean $\bar{Y}_s$ as the estimator of the population mean and the conditional variance $\mathrm{Var}(\bar{Y}_s | Y_j, j \in s)$ as an optimization criterion. The effect of the autocorrelation $\rho(h) = \lambda^h$ to the resulting optimal sampling scheme when λ is positive or negative is studied.

Papageorgiou and Karakostas (1998) generalized Blight's results to any decreasing, positive and convex autocorrelation function. This class of autocorrelation functions is very wide (Bellhouse, 1984) including the linear and exponential functions (Cochran, 1977, p. 221). The feature of decreasing correlation with lag is common in practice. Papageorgiou and Karakostas (2001) provide the optimal sampling strategy based on the best unbiased estimator of the population mean. The optimum sampling design combines centrally located systematic (Blight, 1973; Madow, 1953) and end correction (Yates, 1948).

Tam (1984) also considers the linear regression model assuming that the autocorrelation between two units Y_i and Y_j depends on the corresponding values of the auxiliary variables. A study on the optimal sampling strategy under this model is provided. Bartlett (1986) presents an application of Tam's optimal results. Mukerjee and Sengupta (1990) study the properties of the optimal sampling strategy for a population model that can be seen as a special case of the regression model studied in Tam (1984). A more general model is assumed in Mukerjee and Sengupta (1989) and a methodology to obtain the optimal sampling design with respect to the assumed autocorrelation is proposed. A more analytical description of this result is provided in the next paragraph. Royall (1970) has studied the linear regression through the origin superpopulation model with respect to the best

strategy. For certain variance assumptions that involve variable x and for various ratio type estimates the best sampling scheme is determined. For all examined cases the sampling scheme is purposive, i.e. nonrandom.

A more detailed description of the methodologies considered for the study follows. The criterion of selecting those techniques is the broad assumptions about the correlation that they rely on.

3.1. Strategy 1: Mukerjee and Sengupta (1989). Mukerjee and Sengupta (1989) assume a general superpopulation model and proposed a method for deriving the best pair of estimator and sampling scheme. The model is described mathematically as

$$E(Y_i) = \mu_i \quad \& \quad E[(Y_i - \mu_i)(Y_j - \mu_j)] = v_{ij}, \quad i, j = 1, \ldots, N, \tag{3.1}$$

where the population mean values are general and the covariance matrix $\boldsymbol{V}$ with elements v_{ij} can be any non-diagonal semi-positive square matrix. The optimal strategy $(p^\star, e^\star)$ is derived by minimizing the mean square error (MSE) of the population total estimate within the class of all linear unbiased estimators of the population total. Although quite general, the result of the optimal strategy is rather theoretical. Analytically, the process of finding $(p^\star, e^\star)$ consists of the following steps:

1. Construct an $N \times N$ matrix $\boldsymbol{\Phi}$, with elements $(\boldsymbol{\Phi})_{ij} = \phi_{ij}$ given by

$$\phi_{ij} = \sum_{s \supset ij} v_s^{ij} p(s), \quad i, j = 1, \ldots, N, \tag{3.2}$$

   where v_s^{ij} are the elements of the $N \times N$ inverse matrix, $\boldsymbol{V}_s^{-1}$, of the variance-covariance matrix of the population. The notation $s \supset ij$ denotes that the summation is over all possible samples s that include both units i and j from the population while $p(s)$ is the probability of selection for sample s. These probabilities are the unknown quantities in this method. When the probabilities $p(s)$ are determined the optimal sample will correspond to the largest probability. Initially the matrix $\boldsymbol{\Phi}$ is a function of these unknown probabilities.

2. Define an $N \times 1$ vector $\boldsymbol{\lambda}$ as

$$\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_N)' = \boldsymbol{\Phi}^{-1} \mathbf{1_N} \tag{3.3}$$

   where $\mathbf{1}$ is the $N \times 1$ vector of ones and let $\boldsymbol{\lambda}_s$ be the $n \times 1$ subvector of $\boldsymbol{\lambda}$ that retains only the elements λ_i with $i \in s$.

3. We also assume the class of linear unbiased estimators for the population total, e, with the form

$$e = \alpha_s + \sum_{i \in s} b_{si} Y_i, \tag{3.4}$$

with α_s and b_{si} real constants, satisfying

$$E_p(\alpha_s) = 0 \quad \& \quad \sum_{s \supset i} b_{si} p(s) = 1 \tag{3.5}$$

where E_p denotes expectation with respect to the sampling design and $p(s)$ is the probability of selection for the sample s.

4. Minimize the quantity $\mathbf{1}'\mathbf{\Phi}^{-1}\mathbf{1}$ with respect to the probabilities $p(s)$, $s \in P_n$. The sampling design for which the minimum is achieved is the optimum, $p^\star$, in our notation.

5. The best estimator, $e^\star$, which minimizes the variance among the class of linear unbiased estimators, is the estimator given by (3.4) for which

$$\alpha_s = \sum_{i=1}^{N} \mu_i - \sum_{i \in s} b_{si} \mu_i \quad \text{and} \quad \boldsymbol{b}_s = \boldsymbol{b}_s^\star \quad \text{where} \quad \boldsymbol{b}_s^\star = \boldsymbol{V}_s^{-1} \boldsymbol{\lambda}_s \tag{3.6}$$

Matrix $\mathbf{\Phi}$ is calculated for the optimum probabilities of the sampling design $p^\star$.

In summary, the solution is the result of a minimization problem based on the inverse of the population variance-covariance matrix. This minimization can be difficult in practice, involving the inverse of an $N \times N$ matrix where the population size N is potentially large and it may not always be possible to obtain the minimum. The number of unknown parameters can also be high, for example $\binom{N}{n}$ when sampling without replacement. The practical consequences of these issues are illustrated in Section 3.4. We shall refer to this model and optimum strategy as "Strategy 1" in the remainder of the paper.

We provide a simple numerical example below to illustrate the implementation of this method and provide a tutorial of this and subsequent strategies. We assume a population of size $N = 5$ with variance-covariance matrix $(\boldsymbol{V})_{ij} = v_{ij}$ given by:

$$v_{ij} = \begin{cases} \sigma^2, & \text{if } i = j \\ 0.4\sigma^2, & \text{if } |i-j| = 1 \\ 0.25\sigma^2, & \text{if } |i-j| = 2 \\ 0, & \text{otherwise.} \end{cases} \tag{3.7}$$

The true population variance σ^2 is in general assumed unknown. Suppose also that the population vector $\boldsymbol{Y}$ is normally distributed with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{V}$ and moreover $\boldsymbol{\mu} = \mu\mathbf{1}$, $\mu = 2$ and $\sigma = 1$ for this example. If $\boldsymbol{Y} = (2.29,\ 0.89,\ 2.28,\ 3.34,\ 2.13)$ is the population vector with true population mean $\theta = 2.19$, we consider the problem of estimating $\theta = T/n$ based on a sample of size $n = 2$.

Strategy 1 begins by constructing $\boldsymbol{\Phi}$ in equation (3.2). Mukerjee and Sengupta (1989) provide an equivalent but more convenient way to compute $\boldsymbol{\Phi}$ via

$$\boldsymbol{\Phi} = \sum_{s \in P_n} p(s)T(s), \tag{3.8}$$

where P_n is the collection of all possible samples of size n and $T(s)$ is an $N \times N$ matrix with zero elements except from lines and columns with indices that correspond to the selected for the sample s units. The non-zero part of $T(s)$ is taken to be the inverse of the partition matrix $\boldsymbol{V}_s$, the partition of $\boldsymbol{V}$ corresponding to the sample s. For example for the numerical example where $n = 2$, matrix $T(\{1, 2\})$ is given by

$$T(\{1,2\}) = \begin{bmatrix} V_{\{1,2\}}^{-1} & 0_{2\times 3} \\ 0_{3\times 2} & 0_{3\times 3} \end{bmatrix} = \begin{bmatrix} 1.19 & -0.48 & 0 & 0 & 0 \\ -0.48 & 1.19 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\text{where} \quad \begin{bmatrix} 1.19 & -0.48 \\ -0.48 & 1.19 \end{bmatrix} = \begin{bmatrix} 1 & 0.4 \\ 0.4 & 1 \end{bmatrix}^{-1}$$

while

$$\mathrm{T}(\{3,5\}) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.07 & 0 & -0.27 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.27 & 0 & 1.07 \end{bmatrix} \quad \text{with} \quad V_{\{3,5\}}^{-1} = \begin{bmatrix} 1 & 0.25 \\ 0.25 & 1 \end{bmatrix}^{-1}$$

The collection of all possible samples of size $n = 2$ is $P_2 = (\{Y_1, Y_2\}, \{Y_1, Y_3\}, \cdots, \{Y_4, Y_5\})$, consisting of 10 samples and if $q_1, q_2, \ldots, q_{10}$ are the corresponding selection probabilities, then matrix $\boldsymbol{\Phi}$ is calculated in terms of the unknown q_i from equation (3.8) with the sum running over the 10 possible samples. In general, the procedure proceeds by first producing all possible $\binom{N}{n}$ T-matrices, creating matrix $\boldsymbol{\Phi}$ and then minimizing the quantity $\mathbf{1}'\boldsymbol{\Phi}^{-1}\mathbf{1}$ with respect to the selection probabilities. As the authors point out

in their paper, it is important to reduce the number of unknowns for the minimization problem. For instance, the structure of the population variance matrix (3.7) implies that samples $\{Y_1, Y_2\}, \{Y_2, Y_3\}, \{Y_3, Y_4\}$, and $\{Y_4, Y_5\}$ will all have the same probability q. This is characterized as an intuitive consideration in the paper justified from the fact that the corresponding T matrices have identical non-zero elements. They differ on their locations, but the operation $\mathbf{1}'\mathbf{\Phi}^{-1}\mathbf{1}$ will sum the elements of matrix $\mathbf{\Phi}$ and location will not matter for the minimization stage. For the same reason, samples $\{Y_1, Y_3\}, \{Y_2, Y_4\}, \{Y_3, Y_5\}$ have the same inclusion probability and so on. Moreover, the condition that holds over all samples is $\sum_i q_i = 1$ and the number of unknown selection probabilities is reduced to 2 in this case.

The minimization problem is therefore relatively straightforward in this example and the resulting optimum probabilities, $p^\star$, for all possible samples are:

$$p^\star = \{q_1, q_2, \ldots, q_{10}\} = \{0,\ 0.176,\ 0.156,\ 0.156,\ 0,\ 0.176,\ 0.156,\ 0,\ 0.176,\ 0\}. \tag{3.9}$$

The maximum probability is thus 0.176, corresponding to samples $\{Y_1, Y_3\}$, $\{Y_2, Y_4\}$ and $\{Y_3, Y_5\}$. The collection of those three samples is the optimum sampling design p_0 with equal probabilities, while the optimum estimate $e^\star$ is calculated from formula (3.4) multiplied by N^{-1} to account for population mean. It is

$$e^\star(s) = N^{-1}\left\{\sum_i^N \mu_i + \sum_{i\in s} \boldsymbol{b}^\star_{s,i}(y_i - \mu_i)\right\}, \tag{3.10}$$

with $s \in p_0$ and $\boldsymbol{b}^\star = \boldsymbol{V}_s^{-1}\boldsymbol{\lambda}_s$ with $\boldsymbol{\lambda} = \mathbf{\Phi}^{-1}1$ and $\boldsymbol{\lambda}_s$ its partition that corresponds to s. Note that matrix $\mathbf{\Phi}$ is known at this point and $\boldsymbol{\lambda}$ is easily calculated. Namely, $\boldsymbol{\lambda}^\star = (2.295, 3.355, 3.229, 3.355, 2.295)$ and hence the estimates $e^\star$ that correspond to the optimal samples $s \in p_0$ are:

$$e^\star(\{Y_1, Y_3\}) = 2.31, \quad e^\star(\{Y_2, Y_4\}) = 2.15, \quad e^\star(\{Y_3, Y_5\}) = 2.24. \tag{3.11}$$

3.2. Strategy 2: Papageorgiou and Karakostas (2001). Papageorgiou and Karakostas (2001) building on the work of Papageorgiou and Karakostas (1998) and Blight (1973) considered a superpopulation model given by

$$E(Y_i) = \mu \quad \& \quad E[(Y_i - \mu)(Y_j - \mu)] = \sigma^2\rho(|i-j|), \quad i, j = 1, \ldots, N \tag{3.12}$$

where the superpopulation parameters μ and σ^2 are unknown parameters and $\rho(\cdot)$ is a function with values depending on the lag of the population units. According to this model the mean value is the same for every population unit and the correlation between two units depends solely on their distance and not on their specific position. As a result, the variance-covariance

matrix $\boldsymbol{V}$ of the population vector is a non-diagonal matrix with the property that the elements of the $k-$th diagonal are equal.

The assumed model can be seen as a special case of the regression model (2.2) with $q = 1$, $\boldsymbol{X} = \mathbf{1}_N$ a vector of units and $\beta = \mu$. The quantity of interest remains the population mean or total as above. Under the assumption that $\rho(\cdot)$ is a positive decreasing and convex function, such that $\rho(i) + \rho(i+2) \geq 2\rho(i+1)$ for example, the optimum strategy derived by Papageorgiou and Karakostas (2001) is optimum in terms of minimum variance among all unbiased estimators for the population mean. Note that the optimality of the strategy is not restricted to the class of linear type estimators. The best pair (p, e) for estimating the population mean is:

$$e = N^{-1} E\left(\sum_{i=1}^{N} Y_i | \boldsymbol{Y}_s\right) \tag{3.13}$$

where $\boldsymbol{Y}_s$ is the collection of population units for the selected sample s, and p is the sampling design characterized by the property of "almost equally spaced" sampling units. This property implies that the sampling units are separated with equal distance, say h, or $h+1$. In other words the optimal sampling design p derived under this model proposes a uniform (or almost uniform) coverage of the population. The best pair for the population total is $(p, e_T = Ne)$. Under the assumption of normality for $\boldsymbol{Y}$, the estimator e_T takes the form

$$e_T = \mathbf{1}_s' \boldsymbol{Y}_s + \mathbf{1}_r' [\mathbf{1}_r \hat{\mu} + \boldsymbol{V}_{rs} \boldsymbol{V}_s^{-1} (\boldsymbol{Y}_s - \mathbf{1}_s \hat{\mu})] \tag{3.14}$$

which coincides with the best linear estimator (2.3) when $\hat{\mu}$ is the weighted least-squares estimator $\hat{\mu}_s$ of μ. Note that under the same normality assumption $\hat{\mu}_s$ is the maximum likelihood estimate of μ given by (2.5) which moreover under model (3.12) takes the form $\hat{\mu}_s = (\mathbf{1}_s \boldsymbol{V}_s^{-1} \mathbf{1}_s)^{-1} \mathbf{1}_s^{-1} \boldsymbol{V}_s^{-1} \boldsymbol{Y}_s$. It is therefore straightforward under this assumption to calculate the estimate and its exact variance from (2.3) and (2.4) respectively that become (3.14) and

$$\begin{aligned} \mathrm{Var}(\hat{T}_{BLU}) = {} & \mathbf{1}_r' \boldsymbol{V}_r \mathbf{1}_r - \mathbf{1}_r' \boldsymbol{V}_{rs} \boldsymbol{V}_s^{-1} \boldsymbol{V}_{sr} \mathbf{1}_r + \\ & \mathbf{1}_r' (\mathbf{1}_r - \boldsymbol{V}_{rs} \boldsymbol{V}_s^{-1} \mathbf{1}_s)(\mathbf{1}_s' \boldsymbol{V}_s^{-1} \mathbf{1}_s)^{-1} (\mathbf{1}_r - \boldsymbol{V}_{rs} \boldsymbol{V}_s^{-1} \mathbf{1}_s)' \mathbf{1}_r. \end{aligned} \tag{3.15}$$

respectively.

In the asymptotic case, where N and n are large the above sampling design collapses to a special case of systematic sampling design from the

classical sampling theory. The special case is the centrally located systematic scheme. According to the systematic sampling scheme, the sampling units for every possible sample s are equally spaced by distance $h = N/n$, called the step of the systematic. When this ratio is integer every possible sample has exactly n units. There are h such samples in total and they all have equal selection probabilities, i.e. $p(s) = \frac{1}{h}$. The selection of a systematic sample is equivalent with the selection of a random number, say r, within the range $\{1, 2, \ldots, h\}$. When r is selected the corresponding sample is $s = (Y_r, Y_{r+h}, \ldots, Y_{r+(n-1)h})$. The resulting systematic samples do not share any common population unit and therefore the inclusion probability π_i for each Y_i belonging to a sample is also $\frac{1}{h}$. If $N \neq nh$ a slight complication and need for modification arises, but the effect of it is negligible (Yates, 1948). If the start is chosen such as $2r = N + 1 - (n-1)h$ then the systematic is called systematic centrally located (Blight, 1973). This is a systematic sample with equally spaced population units selected with moreover equal the two end spacings. The estimate of the population mean under the systematic sampling scheme is according to the classical sampling theory the sample mean $e_{sy} = n^{-1} \sum Y_s$. The sample mean does not take into account any correlation among population units if it exists. For a detailed study of systematic sampling design we refer to Cochran (1977, chapter 8).

For this asymptotic case and under the normality assumption for $\boldsymbol{Y}$ the variance of the best unbiased estimator (3.13) is given by

$$\begin{aligned} Var(e) &= \frac{\sigma^2}{N^2}\left[V(i) + V(j) + (n-1)\left\{(h+1) + 2\sum_{k=1}^{h}(h+1-k)\rho(k)\right.\right. \\ &\qquad \left.\left. - \frac{2(\sum_{k=0}^{h}\rho(k))^2}{(1+\rho(h))}\right\}\right] \end{aligned} \tag{3.16}$$

with

$$V(m) = \frac{1}{N^2}\left[m + 2\sum_{k=1}^{m-1}(m-k)\rho(k) - \left(\sum_{k=0}^{m-1}\rho(k)\right)^2\right]$$

The quantities $i-1$ and $j-1$ are the two end spacings and h is the step in the systematic design. Expressions (3.14) and (3.16) can be easily programmed and calculated for given values of the function $\rho(\cdot)$. Estimator e_T given by (3.14), or equivalently $e = e_T/N$, both depend on the correlation. Any present correlation for the population under study will therefore reflect to the derived estimate. In contrast to Strategy 1, this approach does not

depend on the population size and does not require an optimization problem involving the population covariance matrix. The computational effort here is therefore minimal. We shall refer to this model as "Strategy 2" in the remainder of this paper.

For illustration, for the same small numerical example of Section 3.1.2, the correlation values in the notation of Strategy 2 are

$$\rho(0) = 1, \rho(1) = 0.4, \rho(2) = 0.25 \text{ and } \rho(k) = 0 \text{ for } k \geq 3, \tag{3.17}$$

derived from matrix $\boldsymbol{V}$ given by (3.7). The sampling design p with centrally located and equally spaced sampled units consists of samples $\{Y_2, Y_4\}$ and $\{Y_1, Y_5\}$ in this example. From equation (3.14) we obtain the estimate for the population mean $\theta = e_T/N = 2.19$. That is

$$e(\{Y_2, Y_4\}) = \hat{\bar{Y}} = 2.01, \quad e(\{Y_1, Y_5\}) = \hat{\bar{Y}} = 2.20\,. \tag{3.18}$$

3.3. Methodologies Suggested by Chao (2004). Chao (2004) also adopted the superpopulation approach but developed sampling methodologies for correlated measurements borrowing ideas from principal component analysis. More specifically, the superpopulation model (1.1) with mean vector $\boldsymbol{\mu}$, and known variance-covariance matrix $\boldsymbol{V}$ is assumed. The matrix $\boldsymbol{V}$ can have any structure but is assumed known in order to be able to calculate its eigenvalues. The idea is to choose those population units pointed from the n most important components to consist the sample, assuming that these components contain most of the information from the population. Chao (2004) proposed two algorithms, called *Design I* and *Design II* with the second a slight modification of the former, both based on the spectral decomposition of matrix $\boldsymbol{V}$.

3.3.1. Strategy 3: Design I by Chao (2004). Let $\boldsymbol{e}_i = (e_{i1}, \ldots, e_{iN})$ be the eigenvector of matrix $\boldsymbol{V}$ corresponding to the eigenvalue λ_i, $i = 1, \ldots, N$, with $\lambda_1 > \lambda_2 > \cdots > \lambda_N$. The algorithm chooses the sample with indices $s = \{i_1, i_2, \ldots, i_n\}$ sequentially in n steps as follows:

Step 1: The population unit that contributes the most in eigenvector $\boldsymbol{e}_1$ is selected in the sample. That is

$$i_1 = j, |e_{1j}| = \max_{1 \leq t \leq N} |e_{1t}|$$

Step k: For $2 \leq k \leq n$, of the remaining population units, the unit corresponding to the largest component of eigenvector $\boldsymbol{e}_k$ is selected. That is

$$i_k = j, |e_{kj}| = \max_{t, t \notin s} |e_{kt}|$$

We shall refer to this procedure as "Strategy 3" for the rest of the paper.

For illustration, for the same small example considered earlier, the first step is to calculate the eigenvectors of matrix $\boldsymbol{V}$ given by (3.7). The 5×5 matrix that contains the eigenvectors of $\boldsymbol{V}$ as columns is

$$\begin{array}{rrrrr} 0.3351 & -0.5697 & 0.6171 & 0.0831 & 0.4188 \\ 0.4818 & -0.4188 & -0.3167 & 0.4093 & -0.5697 \\ 0.5578 & 0.0000 & -0.1943 & -0.8069 & 0.0000 \\ 0.4818 & 0.4188 & -0.3167 & 0.4093 & 0.5697 \\ 0.3351 & 0.5697 & 0.6171 & 0.0831 & -0.4188 \end{array} \tag{3.19}$$

The largest in absolute value components in the first and second eigenvectors are respectively, $\{3\}$ and $\{1\}$ or $\{5\}$ resulting in the samples $\{Y_3, Y_1\}$ and $\{Y_3, Y_5\}$ respectively. Employing the sample mean as the estimator of the population mean the corresponding estimates are

$$e_1 = \hat{\bar{Y}}_{de1} = 2.29, \qquad e_2 = \hat{\bar{Y}}_{de1} = 2.20$$

Since the optimality is derived independently with the estimate, the best linear estimate $\hat{T}_{BLU}$ from (2.3) can also be used and the new estimates for the same samples are

$$T_1 = \hat{\bar{Y}}_{de1} = 2.28 \quad \text{and} \quad T_2 = \hat{\bar{Y}}_{de1} = 2.21$$

respectively.

3.3.2. Strategy 4: Design II by Chao (2004). Here a modification of the previous algorithm is followed in an attempt to incorporate the sign as well as the magnitude of the components of the eigenvectors in the selection procedure. The steps are:

Step 1: In the trivial case of $n = 1$ the single selected unit is chosen by following the first step of the Design I.
If $n > 1$, let $s' = \{j_1, j_2, \ldots, j_m\}$ be an initial sub-sample of size $m < N$ consisting of the first m largest in magnitude components of the first eigenvector, namely

$$|e_{1j_1}| \geq |e_{1j_2}| \geq \cdots \geq |e_{1j_m}| \geq \cdots \geq |e_{1j_N}|.$$

Chao (2004) suggests that m can be specified before the survey according to the population size, N. The next steps are repetitions of the following general step starting with $k = 2$.

General step k: At each subsequent step, the population units l_1 and l_2 are selected such that

- l_1 and l_2 are not in the sample s already
- $|e_{kl_1}| = \max_i |e_{ki}|$, and
- $|e_{kl_2}| = \max_{j:e_{kj}\cdot e_{kl_1}<0} |e_{kj}|$.

The purpose of this step is to choose the population units corresponding to the two leading in magnitude but opposite in sign elements of the $k-$th principal component. The following rule locates these two new indexes in the sample:

$$\begin{aligned} i_{2(k-1)} &= l_1, i_{2k-1} = l_2, \text{ if } n \geq 2k-1 \\ i_{2(k-1)} &= l_1, \text{ if } n = 2(k-1). \end{aligned}$$

The step is repeated until $k = (n+1)/2$ if n is odd and $k = n/2+1$ if n is even. Sample s will reach the sampling size in the meanwhile and will consist of $s = (i_2, i_3, \ldots, i_n)$.

A final adjustment in the sample s is needed concerning the selection of the first sampling unit. The new sample unit will be chosen from the initial sub-sample s' and the selection criterion is the minimum multiple correlation coefficient between itself and the remaining $n-1$ sample units already in the sample.

We shall refer to this procedure as "Strategy 4" for the rest of the paper.

Using the same numerical example for Strategy 4 towards the optimal sample and estimate, we initially need to choose m. If, say $m = 3$, the temporary sub-sample s' defined from the principal in magnitude components of the first eigenvector is $s' = \{j_3, j_2, j_4\}$.

The sample size is small and only one step of the algorithm is required. In this second and final step the index $i_2 = l_1$ is obtained based on the second eigenvector and is $i_2 = 5$, or equally $i_2 = 1$, meaning that two equivalent samples can result with Y_5 or Y_1 selected respectively. The final step in order to complete the sample(s) is to select the first unit with criterion the minimum multiple correlation coefficient (simple correlation in this small example) with the units already in the sample. For the first sample with unit Y_5, Y_2 from s' fulfills this criterion. Similarly for the second possible sample with Y_1 at the second place, unit Y_4 is selected to complete the sample. Altogether the two samples suggested from design II are

$\{Y_2, Y_5\}$ and $\{Y_1, Y_4\}$. Assuming first the sample mean as the estimator, the corresponding calculated estimates are

$$e_1 = \hat{\bar{Y}}_{de2} = 1.51 \quad \text{and} \quad e_2 = \hat{\bar{Y}}_{de2} = 2.81$$

while if $\hat{T}_{BLU}$ estimate is assumed the corresponding calculated estimates of population mean $\theta = 2.19$, are

$$e_1 = \hat{\bar{Y}}_{de2} = 1.54 \quad \text{and} \quad e_2 = \hat{\bar{Y}}_{de2} = 2.88 \, .$$

For all strategies studied in this section matrix $\boldsymbol{V}$ or equivalently the autocorrelation function is assumed to be known. If not, an estimate of it can be used instead by using techniques discussed in Section 2.

3.4. Some General Comments. Some first general comments can be drawn from the description and illustration of the methods. The superpopulation model adopted by all four strategies is quite general with respect to the underlying correlation structure, a characteristic that makes all studied techniques competitive. However, the methods differ in many other aspects, such as the basis on which the derivations rely on, the theoretical or mathematical background, the computational effort and ease of application, and the properties of the derived estimators.

Comparing them at the level of the theoretical justification of the results, Strategies 1 and 2 rely on a mathematical background, whereas Strategies 3 and 4 are based on the logic that keeping the most important components seems related with optimality. The mathematical basis of Strategies 1 and 2 allows the methods to address the problem of finding the sampling design and estimate simultaneously as a pair. For Strategies 3 and 4, the criterion of sample selection is not defined in terms of a measure of accuracy of the estimator and therefore the estimate can be chosen independently. Moreover, the mathematical basis of Strategies 1 and 2 provide exact formulas for the variance calculation of the considered estimate whereas the same calculation for Strategies 3 and 4 can only be made through simulations.

On the other hand, with respect to complexity and ease of application, Strategies 3 and 4 are straightforward, very fast and simple in structure. They only require the spectral decomposition of an $N \times N$ matrix and the proposed algorithms for both strategies use steps that need simple calculations. At the same time, despite the lack of mathematical background, intuitively it seems reasonable to expect Strategies 3 and 4 to perform better than simple random sampling.

Between Strategies 1 and 2, Stategy 2 does not depend on the population size N and the estimation process does not involve a numerical optimization problem. Therefore the computational effort and time is significatively

smaller than in Strategy 1. In practice, the computational requirements of Strategy 1 are so high that allow the implementation only for a small population and sample size. We demonstrate the practical limitations of Strategy 1 below by using a larger example than that considered earlier in this section. Strategy 1 on the other hand, if applicable, provides an informative output with the complete vector of probabilities over all possible samples.

Let us assume the same correlation structure for the population, namely a correlation up to distance two among the population units and assume $N = 10$ and $n = 4$, that is the population and sample sizes are doubled with respect to the first example. For the population values $\boldsymbol{Y} = (2.81,\ 2.55,\ 4.69,\ 4.26,\ 3.32,\ 3.56,\ 2.63,\ 2.39,\ 2.99,\ 4.39)$ simulated from a normal distribution with mean 4 and variance-covariance matrix $\boldsymbol{V}$, the population mean is $\theta = \bar{Y} = 3.36$.

The number of all possible samples is 210 and therefore 210 unknown probabilities of inclusion subject to the restriction that they sum to one. It is interesting to note that the number of the unknowns increases in binomial coefficient order with respect to n and soon becomes unmanageable. However, due to the simple structure of matrix $\boldsymbol{V}$, the number of the unknowns can be reduced to 27 in this example, subject to the condition

$$\begin{aligned} 7q_1 \ &+\ 6q_2 + 15q_3 + 6q_4 + 5q_5 + 10q_6 + 15q_7 + 10q_8 + 10q_9 + 6q_{10} + 5q_{11} \\ &+\ 10q_{12} + 5q_{13} + 4q_{14} + 6q_{15} + 10q_{16} + 6q_{17} + 4q_{18} + 15q_{19} + 10q_{20} \\ &+\ 10q_{21} + 10q_{22} + 6q_{23} + 4q_{24} + 10q_{25} + 4q_{26} + q_{27} = 1, \end{aligned}$$

where the coefficients of the q_i's denote the sample multiplicities corresponding to a same partition of matrix $\boldsymbol{V}$.

Matrix $\boldsymbol{\Phi}$ is calculated from (3.8) and the quantity $\mathbf{1}'\boldsymbol{\Phi}^{-1}\mathbf{1}$ is next to be minimized with respect to the unknown inclusion probabilities. However, it is not possible to derive the inverse of $\boldsymbol{\Phi}$ symbolically in this case, and the minimization of $\mathbf{1}'\boldsymbol{\Phi}^{-1}\mathbf{1}$ fails even when $n = 2$ or 3 where a smaller number of parameters is involved. In general, as the population size increases, the dimension of matrix $\boldsymbol{\Phi}$ increases and as the sample size increases the number of unknowns also increases. The relation is not linear and the problem becomes infeasible even for small problems that do not correspond to realistic situations.

For completeness of the example, we provide the optimal sampling results according to the remaining strategies. According to Strategy 2, the

optimum samples satisfying the property of centrally located and equally spaced sampled units, are

$$s_1 = \{Y_1, Y_4, Y_7, Y_{10}\}, \quad s_2 = \{Y_2, Y_4, Y_6, Y_8\},$$
$$s_3 = \{Y_3, Y_5, Y_7, Y_9\} \text{ and } s_4 = \{Y_4, Y_5, Y_6, Y_7\}.$$

The sampling scheme defined as $p = \{s_1, s_2, s_3, s_4\}$ that corresponds equal probability of selection to all four samples is the optimum given the structure $\boldsymbol{V}$ of the population. Calculating the estimate from (3.13) for all possible samples, the estimates of $\theta = 3.36$ are

$$e(s_1) = 3.51, \quad e(s_2) = 3.10, \quad e(s_3) = 3.47 \text{ and } e(s_4) = 3.44.$$

respectively.

For Strategies 3 and 4, the first 4 eigenvalues of the matrix $\boldsymbol{V}$ and their leading components are the primary ingredients of the corresponding algorithms. These are easily derived and the suggested optimal samples according to Strategy 3 and 4 are

$s_1 = \{Y_1, Y_3, Y_4, Y_5\}$, $s_2 = \{Y_1, Y_3, Y_4, Y_6\}$, $s_3 = \{Y_1, Y_4, Y_5, Y_8\}$, $s_4 = \{Y_1, Y_4, Y_6, Y_8\}$,
$s_5 = \{Y_3, Y_4, Y_5, Y_{10}\}$, $s_6 = \{Y_3, Y_4, Y_6, Y_{10}\}$, $s_7 = \{Y_3, Y_5, Y_7, Y_{10}\}$, $s_8 = \{Y_3, Y_6, Y_7, Y_{10}\}$,
$s_9 = \{Y_1, Y_3, Y_5, Y_7\}$, $s_{10} = \{Y_1, Y_3, Y_6, Y_7\}$, $s_{11} = \{Y_4, Y_5, Y_8, Y_{10}\}$, $s_{12} = \{Y_4, Y_6, Y_8, Y_{10}\}$,
$s_{13} = \{Y_5, Y_7, Y_8, Y_{10}\}$, $s_{14} = \{Y_6, Y_7, Y_8, Y_{10}\}$, $s_{15} = \{Y_1, Y_5, Y_7, Y_8\}$, $s_{16} = \{Y_1, Y_6, Y_7, Y_8\}$

4 Comparison Study

The last section provided a comparison of different sampling strategies dedicated to correlated measurements with respect to their theoretical framework, the assumptions they employ and the derived optimal sampling procedure p. It is however important to compare the methodologies from the statistical point of view and evaluate them with respect to their estimation performance which is the principal aim of each sampling strategy. The comparison is even more important for those strategies that they do not rely on a mathematical derivation. Their efficiency and the logic behind the method can be evaluated in practice and tested whether it is related with optimality or not.

A comparison study can also reveal information whether the relative performance of the strategies depends on the underlying correlation structure. Another issue that will be examined is the robustness of the sampling strategies and to what extend the optimality properties hold when some

deviations from the assumed population structure occur in reality. This is essential because the exact population structure is never known in practice but rather estimated from practical considerations or pilot studies. The sample size n is also considered as a factor and examined through the comparison study whether it differentiates the relative performance or the degree of superiority among the sampling strategies.

The insight gained from a comparison study can be useful as a guidance to practitioners to two main directions: (i) the choice of the most appropriate strategy for the correlation type they are dealing with and (ii) having a measure of how safe the choice would be if the actual correlation is slightly different.

4.1. Simulation Design. We describe in this paragraph the simulation design followed for all the experiments included later in the current section. Simulation is imposed from the fact that the calculation of the mean squared error (MSE) for the estimates proposed from strategies 3 and 4 by Chao (2004) is not straightforward. The comparison among the competitive sampling schemes is based on the relative efficiency of each method with respect to simple random sampling (srs). The srs was chosen because it is population model free, and although not optimal in the presence of correlation, it is often a safe choice in practice. Moreover, the use of srs allows us to measure the loss of efficiency if the existing correlation is not taken into account.

For each experiment we assume that the population vector follows a normal distribution with mean value μ, the parameter of interest, and a variance covariance matrix $\boldsymbol{V}$ that differs from one experiment to another. The reasoning behind this is to cover a range of different population structures and explore the relative efficiency of the sampling methods. Given the matrix $\boldsymbol{V}$, we consider a population vector of size $N = 100$ generated from the superpopulation $N(\mu \mathbf{1}_N, \boldsymbol{V})$ and assume as the sampling problem to be the estimation of the population total based on a sample of size n. We allow n to range from $n = 2$ to $n = 25$ and for each pair of $(N = 100, n)$ we determine the sampling designs suggested by each competitive sampling method by using matrix $\boldsymbol{V}$. The calculation of the MSEs for the estimates that each sampling design suggests is made by simulation. More precisely, for each pair of $(N = 100, n)$ a population vector is generated 15000 times and let θ_j be the population total for the j iteration of population generation. If in turn, $\hat{\theta}_j^d$ is the estimate of population parameter based on the sample size

n and proposed by the design d, the corresponding mean square error for design d is calculated by

$$E_d(\theta - \hat{\theta})^2 = \frac{1}{K}\sum_{j=1}^{K}(\theta_j - \hat{\theta}_j^d)^2 \tag{4.1}$$

The formula used for the calculation of $\hat{\theta}_j^d$ for each sampling design is the best unbiased estimate provided by Strategy 2, given by formula (3.14), in order to be able to compare the results. It is already mentioned that under the normality assumption this estimate coincides with the best linear estimate given by (2.3). The estimate $T^\star = Nn^{-1}\sum_{i\in s} Y_i$, is also considered for the comparison study. Although only optimal under a restricted model, the estimate $T^\star$ is frequently used in practice due to its simple expression, as explained in Section 2.

The above procedure of generating 15,000 population vectors and calculating the MSEs of each sampling design is repeated for each sample size n from 2 to 25. Similarly, the complete simulation design as described is repeated for each population occasion with variance covariance structure $\boldsymbol{V}$ to vary from one experiment to another. The relative efficiency is calculated from the MSE obtained with design d through divided with the corresponding obtained with srs. Relative efficiency values less than one indicate better performance. The programming and all the relevant calculations are made in Matlab.

Note that for Strategy 2, the true MSE of the estimate can be calculated in exact form by Eq. (3.15). However, the simulation is used for all samplings methods for comparison reasons. One last comment is that Strategy 1 was not included in the numerical experiments because the population and sample sizes considered are prohibitive for its implementation.

4.2. Convex Type of Correlation. As a first category we assume a population model generated by normal distribution and autocorrelation that belongs in the class of convex, positive and decaying to zero functions. Three possible scenarios for populations with mild, moderate and strong correlation among its units are considered and use autoregressive model of order one, AR(1), to generate the specific cases. Analytically the three assumed models for this paragraph are (i) $\rho(h) = 0.2^h$, (ii) $\rho(h) = 0.4^h$ and (iii) $\rho(h) = 0.7^h$ plotted in Fig. 1d. The general population parameters μ and σ^2 are assumed to be zero and one respectively.

Following the simulation plan as described earlier, the relative efficiencies of each sampling design with respect to srs for models (i), (ii) and (iii) are calculated and plotted in Fig. 1a, b, c respectively. As expected, Strategy 2 outperforms the remaining strategies, since it is the optimal scheme for positive, convex and decreasing functions of $\rho(h)$. Moreover the gain in efficiency depends on the degree of the correlation with greater gain achieved at higher correlation. For strong correlation (model iii) the reduction of the corresponding to srs MSE is around 60 percent.

Between Strategies 3 and 4, Strategy 4 is superior to Strategy 3 while the relative efficiency of both strategies is generally smaller than one indicating better performance compared to srs. In general, the performance of the competing strategies is consistent with respect to the sample size with their differences being apparent after sample size 5.

4.3. Positive, decreasing, non-convex correlation. For the second set of simulations we assume three examples of autocorrelation functions that

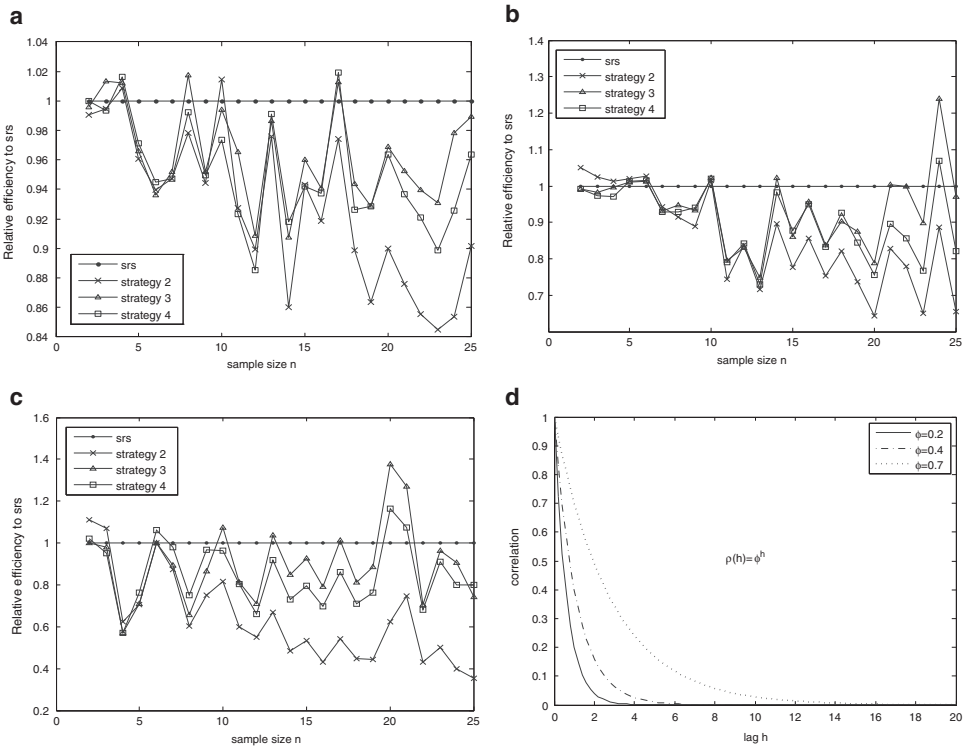

Figure 1: Relative efficiencies for strategies 2, 3 and 4 and model $\rho(h) = \phi^h$. **a** ϕ=0.2, **b** ϕ=0.4, **c** ϕ=0.7

are positive and decaying to zero with lag as above, but are no longer convex. Such types of correlation are frequently met in applications of spatial statistics and quality control. The considered autocorrelation function is a gaussian shape correlation function (Cressie, 1993)

$$\rho(h) = \exp(-h^2/c^2)$$

where the parameter c controls the strength of the correlation between two units at distance h. Three values of c are taken (i) c=1.5, (ii) c=3.5 and (iii) c=4.5 leading to three autocorrelation functions plotted in Fig. 2d. All other assumptions about the distribution and the general parameters remain the same.

Figure 2a, b, c plots the relative efficiencies to srs for population cases that correspond to (i), (ii) and (iii) correlation functions respectively and

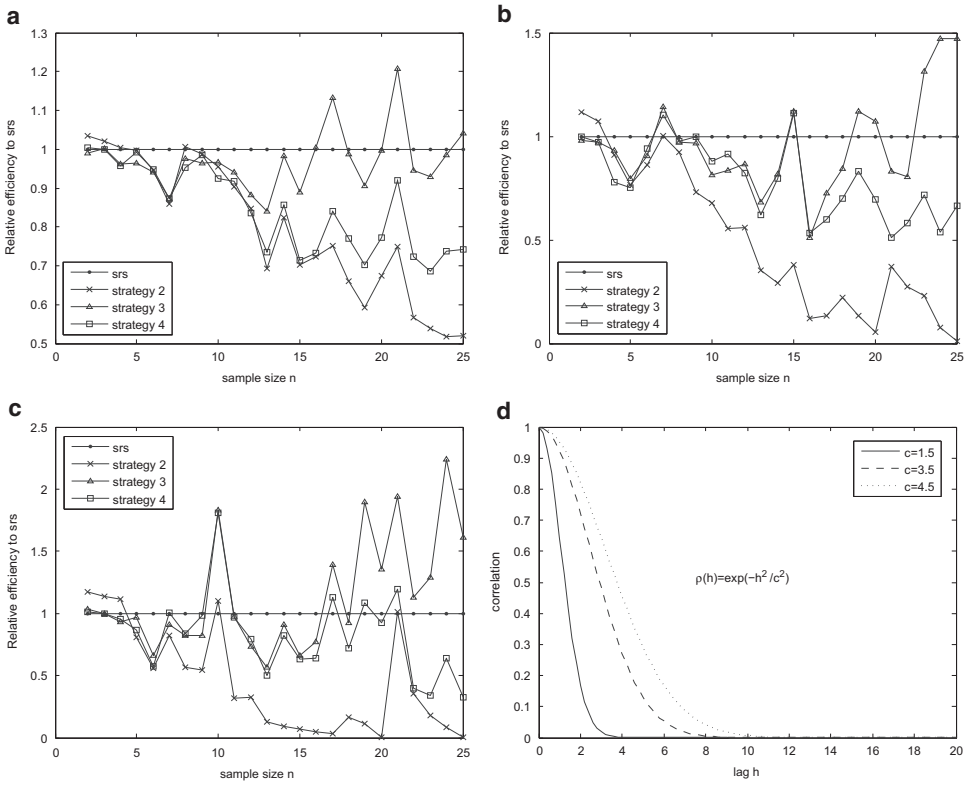

Figure 2: Relative efficiencies for strategies 2,3 and 4 and model $\rho(h) = \exp(-h^2/c^2)$. **a** c=1.5, **b**=3.5, **c**=4.5

shows that the centrally located systematic design is robust in efficiency under a non convex, but positive and decreasing autocorrelation function. The gain in efficiency is greater for this case of correlation than the gain in the convex case (Fig. 1). This finding can be explained from the stronger correlation among neighbor units that $\rho(h) = \exp(-h^2/c^2)$ imposes compared to the convex function $\rho(h) = \phi^h$.

The between strategies comparison is consistent with Fig. 1, with Strategy 4 performing better than Strategy 3 and srs. However Strategy 3 is very often less favorable compared to srs, especially when the correlation is higher ($c = 3.5, 4.5$), and Strategy 4 although superior to srs is starting to become problematic when c takes larger values. To gain further insight we provide in Fig. 3 the relative efficiencies of Strategies 2 and 4 when $c = 5$. We can see that the efficiency of Strategy 2 is improved because of the stronger correlation, while Strategy 4 is comparable if not worse than srs. This last occasion with $c = 5$ and a strong correlation among population units may not be a common population scenario, on the contrary it might be considered as an extreme one, but the result from the theoretical point of

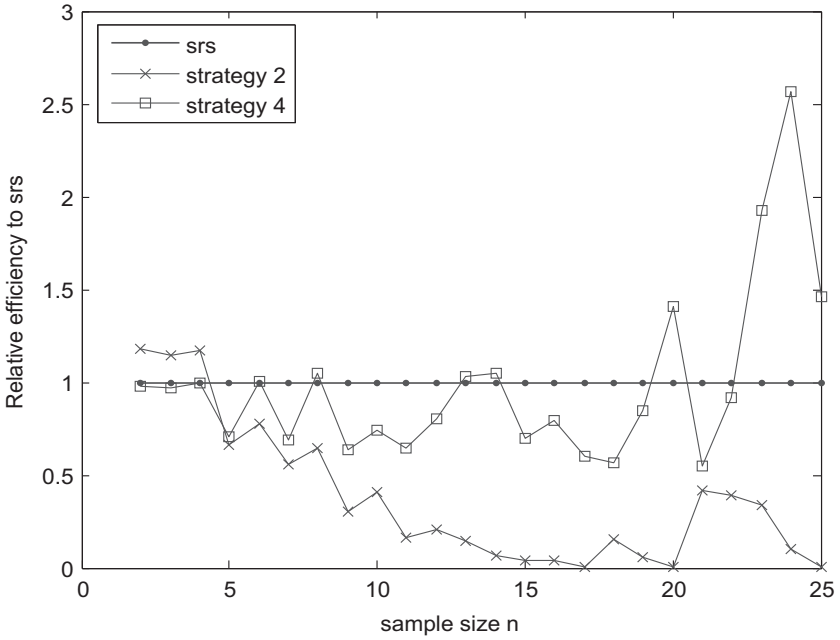

Figure 3: Relative efficiencies for strategies 2 and 4 and model $\rho(h) = \exp(-h^2/5^2)$

view is interesting providing useful insight about the efficiency of strategies 2 and 4.

4.4. More General Types of Correlation. In the third set of simulations we relaxed the properties of convexity, positiveness and decreasing monotonicity in the correlation function in order to examine the performance of the methods under more general types of correlation. It is reasonable to expect Strategies 3 and 4 to outperform both srs and systematic in this case as they do not make any assumption about the underlying correlation structure. For the systematic, which is proven to be optimal within a certain class of correlation functions, it is a matter of how robust the result is outside this class.

4.4.1. Decreasing, Non-positive, Non-convex Correlation. The population values are generated from an ARMA(p, q) model with parameters p and q selected such that the autocorrelation function of the generated data has the form of a decreasing but non positive or convex function shown in Fig. 4b. The efficiency comparison (Fig. 4a) indicates that the centrally systematic is still preferable in this case and this is probably because the negative correlations are not very strong. Strategies 3 and 4 seem to have a similar behavior. Compared to srs they both outperform for sample sizes less than fifteen and become comparable after this point.

4.4.2. Non Decreasing Correlation. For this general case we use both real and simulated data. The real data set, from Box et al. (1994) represents annual mink fur sales of the Hudson's Bay Company between 1850-1911. The autocorrelation function and relative efficiencies are plotted in Fig. 5a and b respectively. For this type of autocorrelation neither of the strategies studied

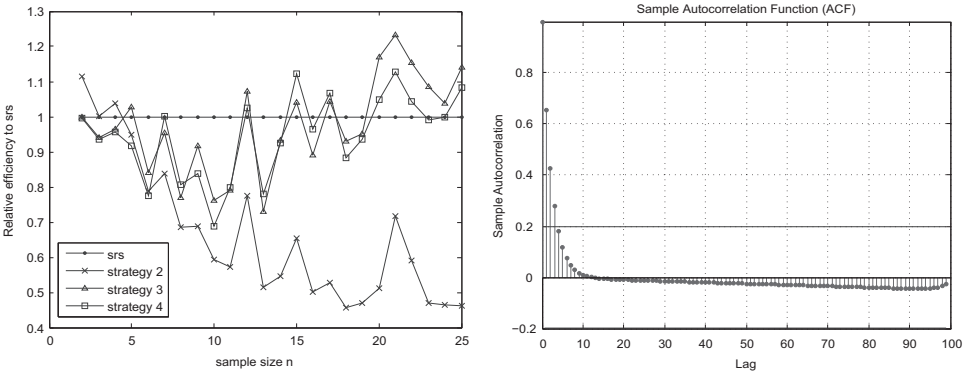

Figure 4: Relative efficiencies for strategies 2, 3 and 4 for ARMA model with ACF in (**b**)

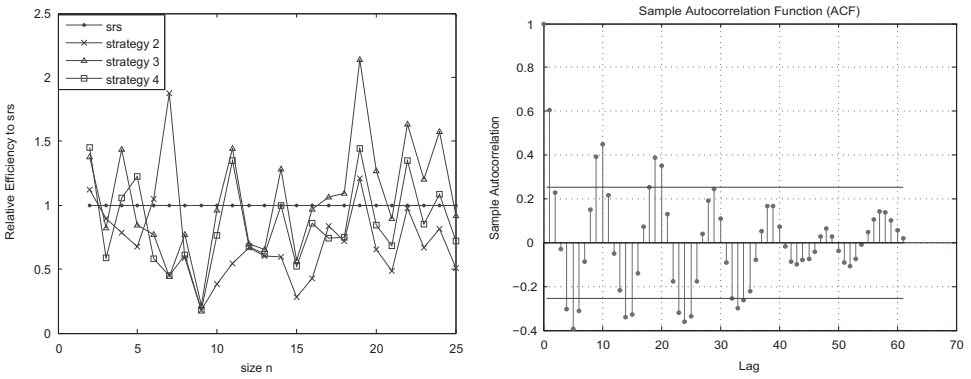

Figure 5: Relative efficiencies for strategies 2, 3 and 4 for the mink data set. ACF in (**b**)

seem to provide reliable results. The efficiency of each methodology depends on the sample size and there is no consistently superior method with respect to efficiency in this case. Strategy 2 is no longer robust, which confirms the known from the literature result that there is no uniform optimal sampling strategy (Godambe, 1955). The optimal sampling design depends closely on the particular type of correlation.

Similar experiments with simulated data from a general autocorrelation function result to the same conclusion that there is no clear optimality among the strategies under study. For example, if measurements are generated according to an ARMA$(-0.6, 0.2)$ model with autocorrelation function plotted in Fig. 6b the same mixed results with respect to efficiency are observed (Fig. 6a). However this is a type of correlation that is close to Blight's (1973) studied case with AR model and negative parameter where the sign of correlation interchanges from positive to negative. For this special case, Blight has proved that the sample which locates the units into two groups one placed at the beginning and the other at the end of the population is the optimal. The optimality criterion he has adopted is the minimization of $Var(\bar{Y}|Y_j, j \in s)$. Although the model and optimality criterion considered by Blight (1973) are not exactly the same here, Blight's sample as shown in Fig. 6a is superior to any other strategy. This concludes that although general strategies which work relatively well under the presence of autocorrelation and can be superior over the srs, the gain of a strategy that makes full use of the exact correlation is unrivaled.

For all examined cases above, the comparison between the centrally systematic and the known systematic from the classical theory showed not

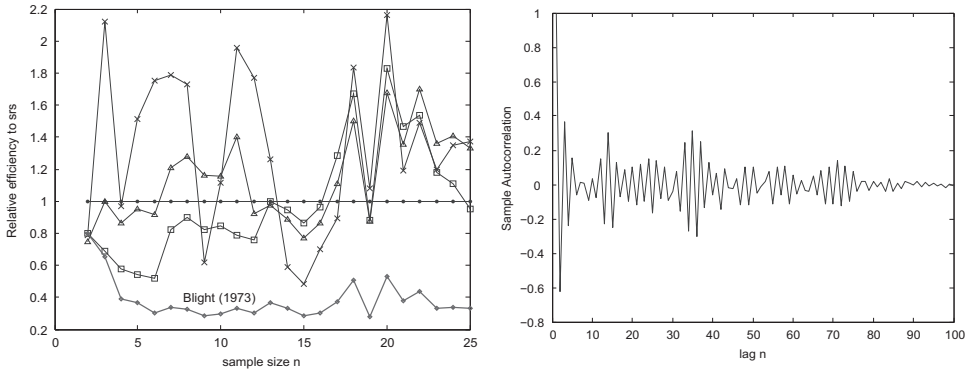

Figure 6: Relative efficiencies for an autocorrelation function with sign that alternates

significant difference. The centrally located systematic has a slight advantage over the standard systematic with a random start. The exact variance of the centrally located systematic sample as calculated from (3.15) is very close to the simulated value obtained from (4.1). A common characteristic in plotting relative efficiencies for all examined population models is a jagged feature as the sample size increases. This is due to the MSE calculation through simulation by (4.1) for the srs design and the large deviation that this design exhibits for proposed estimate over all possible samples. The number of possible samples with srs is $\binom{N}{n}$ and as n increases this number is excessively larger than 15000 iterations. For all other designs the number of possible samples is small and the number of iterations is sufficient.

The least squares estimate of the population mean, $T^{\star}$ has also been considered for the comparison study. The calculated relative efficiencies adopting this estimate are quite similar with those obtained for the best unbiased estimate. The only difference is that the gain in efficiency for all studied designs is smaller compared to the same efficiency calculated with the best unbiased estimate.

5. Discussion

Several methodologies have been proposed in the literature that are specialized for sampling from a correlated population. The plurality of the methodologies results from the lack of a universal technique that holds for any type of correlation among the population units. A methodology that could combine generality in assumptions and mathematical derivation for the

results is a demanding task due to the required mathematical optimization and the dimensionality of the problem.

Some of the existing methodologies rely on advanced theoretical derivations and are computationally demanding, while others are based on a rather intuitive principle. The same characteristics that define a methodology impose their own limitations. Strategy 1 (Mukerjee and Sengupta, 1989) is quite general in assumptions about the population model, but involves a difficult and computationally demanding minimization problem which becomes impractical for even small population and sample sizes.

Strategy 2 (Papageorgiou and Karakostas, 2001), which is proven optimal within a more restricted but still wide class of autocorrelation functions has several advantages including ease of implementation, good practical performance and robustness when relaxing the assumptions about the correlation and the assumed estimates or the minimization criterions. Moreover, it does not require the knowledge of the variance-covariance matrix of the population for its implementation.

Strategies 3 and 4 (Chao, 2004) do not rely on a theoretical result of optimality and do not involve a minimization procedure. The selection criterion in sampling units is based on the principal eigenvalues of the population covariance matrix. Both strategies are straightforward in implementation and do not impose any restrictions on the population parameters or the type of correlation. It is however necessary to know or estimate the variance-covariance matrix of the population vector before the implementation. On the other hand Strategies 3 and 4 are not stable in performance in terms of relative efficiency compared to srs. This is also reported in the original paper of Chao (2004). The comparison study illustrates that their performances is not stable even for the class of decreasing autocorrelation functions (Fig. 4). They do have some logic and some times this produces better results, but this is not for a general autocorrelation function, as they are designed for. Moreover their efficiency depends on the sample size.

The conducted comparison study confirms that the problem of sampling from correlated population has no unified answer with respect to the suggested sampling mechanism. The optimum sampling procedure varies per population occasion and depends considerably on the specific type of correlation. On the other hand, once the optimal sampling designs are obtained under a specific type of correlation they seem to retain optimality under different estimates or different optimization criterions.

It is apparent from the comparison study that estimation and efficiency can be improved when the correlation structure of the population is fully integrated into the sampling procedure. The stronger the existing

correlation among population units the more significant the gain in efficiency is, a realization that makes the use of srs inappropriate, if not unacceptable for some cases. Systematic sampling is a popular sampling scheme with advantages when the correlation is in general, positive and decreasing. Systematic scheme performs well even when these assumptions are relaxed but good performance is not guaranteed in any general case.

The presence of correlation among population units affects both sampling selection procedure and statistical inference on the population parameter considered to be estimated. If variance covariance matrix $\boldsymbol{V}$, e.g. in formula (2.4) is incorrectly assumed as diagonal the MSE calculation will be misleading. The non optimal choice of sampling design may prevent the researcher from achieving best accuracy or receive a representative sample at a relative small sample scale, but the correct calculation of accuracy is essential and affects all consequent steps of analysis. The necessity of testing for dependence and estimating as accurately as possible the type of correlation among the population units is therefore apparent.

The appropriate use of the specialized sampling strategies and the lesson gained from their study can lead to a significant contribution for a number of research fields dealing with correlated populations. In statistical process control and in particular in construction of control and warning charts, the violation of the independence assumption and its significant effect has long been known in the literature. The problem can be addressed with two different approaches. The first and most popular in the literature approach is to model the process measurements according to a time series model and apply the traditional control charts on the residuals of the model which reserve the independent and identical distributed property (see among others Alwan and Roberts, 1988; Harris and Ross, 1991; Mastrangelo and Montgomery, 1995; Apley and Lee, 2003; Lu and Reynolds, 1999a, b and Glasbey, 2001). Following this approach the method can account for the process' serial correlation of theoretically any time series model as long as it is identified and fitted in the data. The assumptions for the traditional charts are fulfilled and no need for specialized to correlated measurements sampling theory is required. However, the approach has been questioned mainly because of the lack in consistency of detecting shifts from the mean. It seems that control charts based on the residuals work best when the level of correlation is high while in moderate or low correlation levels they are less effective in mean changes detection (VanBrackle and Reynolds, 1997). The second more direct approach is to adjust the constructed control chart levels by taking account of the present correlation between the sampled units. Little work has been done in this direction and only for quite special types of

correlation. We mention for example Yang and Hancock (1990) who assume constant degree of correlation. The specialized sampling strategies can be used for selecting the sample and for constructing the control limits using the present correlation in both estimates and their standard errors. In that way the error rates used for evaluation of a control chart can be calculated correctly.

Geostatistical data in spatial statistics very often exhibit a small-scale variation, typically a strong correlation between data at neighboring locations (Watson, 1972). If the population mean is for example the parameter of interest, failure to realize the presence of positive correlation in the data leads to very narrow confidence interval (Cressie, 1993, p. 14). The superpopulation model is therefore extensively used in modeling geostatistical data in order to accommodate this present correlation (Cressie, 1993, ch. 1). In this context, let $\boldsymbol{s} \in \mathbb{R}^d$ be the data location in $d-$dimensional Euclidean space and $\boldsymbol{Y}(\boldsymbol{s})$ the measured data, that is moreover assumed random, at location $\boldsymbol{s}$. Assuming that $\boldsymbol{s}$ takes values over an index set $D \subset \mathbb{R}^d$, the superpopulation model results as a realization $\{\boldsymbol{y}(\boldsymbol{s}) : \boldsymbol{s} \in D\}$ from the multivariate random field $\{\boldsymbol{Y}(\boldsymbol{s}) : \boldsymbol{s} \in D\}$.

The superpopulation regression model is defined in accordance with model (2.2) with the only difference that the population index $i = 1, 2, \ldots, N$ is in this context the locations, say $(\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_N)$ that cover the area under study. When $\boldsymbol{s} \in D \subset \mathbb{R}$ and $\boldsymbol{V}$ is not a diagonal matrix, the superpopulation describes a population with serial correlation. A common hypothesis for superpopulation model describing geostatistical data is the property of intrinsic stationarity. Intrinsic stationarity means that the degree of correlation among two population units depends only on their between distance and not on the specific location. The same assumption is made under model (3.12) and strategy 2 studied in this paper. When the autocorrelation function ρ also has the property of being decreasing, positive and convex the use of optimal sampling results provided by strategy 2 is straightforward.

Another interesting and recently developed application of sampling from correlated populations is composite marginal likelihoods (CML). CML finds application in Clustered or longitudinal data (Neuhaus and Kalbfleisch, 1998), in survival analysis, time series, spatial data (Lande, 1991; Bjørnstad et al., 1999) and image analysis among others. CML are pseudolikelihoods constructed by compounding marginal densities as a substitute of the ordinary likelihood when this second is time consuming or impractical to compute due to its dimensionality and complexity (see for instance Cox and Reid, 2004 and Varin, 2008). CML focuses on low dimension marginals such as univariate or bivariate densities. Univariate CML works under the assumption

of independence while bivariate leading to pairwise marginal likelihood (PML) is a more standard approach. The efficiency of the estimates derived from the PML instead of the ordinary likelihood can be improved by either weighting the different components or selecting a sample from them (Varin, 2008). The key ingredient for approaches is the correlation between measurements and how this can be incorporated in either the weights or the selection procedure. Glasbey (2001) has proposed the quasi likelihood of pairs based on adjacent pairs and in the same direction Hjort and Varin (2008) proposed the inclusion of pairs at lagged distances not exceeding a distance m. The resulting MPL, called pairwise likelihood of order m is proved by simulations to be more efficient than the complete PML. It is natural to expect an even improved scheme of selection to result if an appropriate sampling strategy scheme is applied.

Sampling from correlated populations is a complex problem and requires special methodologies in both sampling and statistical inference stages. The merits from making complete use of the existing correlation and the appropriate techniques can be significant while ignoring the correlation can be misleading towards the calculation of essential quantities such as estimates and their measure of accuracy. A unified answer to the problem of the optimal sampling scheme is not possible and the need of a very general in assumptions methodology, feasible in practice, that can answer the problem providing the correlation matrix of the population still remains.

Acknowledgments. The author would like to thank the Associate Editor and the anonymous referee for their comments on an earlier version of the manuscript. The specialized in the topic comments and suggestions of the Referee, but also in the structure and appearance of the paper led to a significant improvement of this article.

References

ALWAN, L.C. and ROBERTS, H.V. (1988). Time series modeling for statistical process control. *J. Bus. Econom. Statist.* **6**, 87–95.

APLEY, D.W. and LEE, H.C. (2003). Design of exponentially weighted moving average control charts for Autocorrelated processes with model uncertainty. *Technometrics* **45**, 3, 187–198.

BAGSHAW, M. and JOHNSON, R.A. (1975). The Effect of Serial Correlation on the Performance of CUSUM Tests II. *Technometrics* **17**, 1, 73–80.

BARTLETT, R.F. (1986). Sampling a finite population in the presence of trend and correlation: Estimation of total 305-day lactation production in cattle. *Canad. J. Statist.* **14**, 3, 201–210.

BELLHOUSE, D.R. (1984). A review of optimal designs in survey sampling. *Canad. J. Statist.* **12**, 1, 53–65.

BJØRNSTAD, O.N., IMS, R.A. and LAMBIN, X. (1999). Spatial population dynamics: Analysing patterns and processes of population synchrony. *Trends Ecol. Evol.* **11**, 427–431.

BOLFARINE, H. and ZACKS, S. (1992). *Prediction Theory for Finite Population*. Springer, New York.

BOX, G.E.P., JENKINS, G.M. and REINSEL, G.C. (1994). *Time series analysis: forecasting and control*. Prentice-Hall, New Jersey.

BLIGHT, B.J.N. (1973). Sampling from an autocorrelated finite population. *Biometrika* **60**, 375–385.

CASSEL, C.M., SÄRNDAL, C.E. and WRETMAN, J.H. (1977). *Foundations of Inference in Survey Sampling*. Wiley, New York.

COCHRAN, W.G. (1977). *Sampling techniques*, 3rd edition. Wiley, New York.

COX, D.R. and REID, N. (2004). A note on Pseudolikelihood constructed from marginal densities. *Biometrika* **91**, 729–737.

CHAO, C-T. (2004). Selection of sampling units under a correlated population based on the eigensystem of the population covariance matrix. *Environmetrics* **15**, 757–775.

CRESSIE, N.A. (1993). *Statistics for Spatial Data, Revised version*. Wiley, New York.

GLASBEY, C. (2001). Non-linear autoregressive time series with multivariate Gaussian mixtures as marginal distributions. *Appl. Stat.* **50**, 2, 143–154.

GODAMBE, V.P. (1955). A unified theory of sampling from finite populations. *J. R. Statist. Soc. B* **17**, 269–78.

HARRIS, T.J. and ROSS, W.H. (1991). Statistical Process Control Procedures for Correlated Observations. *Can. J. Chem. Eng.* **69**, 48–57.

HJORT, N.L. and VARIN, C. (2008). ML, PL and QL for Markov chain models. *Scand. J. Stat.* **35**, 64–82.

LANDE, R. (1991). Isolation by distance in a quantitative trait. *Genetics* **128**, 443–453.

LU, C.-W. and REYNOLDS, M.R. (1999a). EWMA control charts for monitoring the mean of autocorrelated processes. *J. Qual. Technol.* **31**, 2, 166–188.

LU, C.-W. and REYNOLDS, M.R. (1999b). Control charts for monitoring the mean and variance of autocorrelated processes. *J. Qual. Technol.* **31**, 3, 259–274.

MADOW, W. G. (1953). On the theory of the systematic sampling III. Comparison of centered and random start systematic sampling. *Ann. Math. Statist.* **24**, 101–106.

MASTRANGELO, C.M. and MONTGOMERY, D.C. (1995). SPC with correlated observations for the chemical and process industries. *Qual. Reliab. Eng. Int.* **11**, 79–89.

MONTGOMERY, D.C. and MASTRANGELO, C.M. (1991). Some Statistical Process Control methods for Autocorrelated data. *J. Qual. Technol.* **23**, 179–193.

MUKERJEE, R. and SENGUPTA, S. (1989). Optimal estimation of finite population total under a general correlated model. *Biometrika* **76**, 4, 789–94.

MUKERJEE, R. and SENGUPTA, S. (1990). Optimal estimation of finite mean in the presence of linear trend. *Biometrika* **77**, 3, 625–630.

NEUHAUS, J.M. and KALBFLEISCH, J.D. (1998). Between- and Within-Cluster Covariate effects in the Analysis of Clustered Data. *Biometrics* **54**, 2, 638–645.

PAPAGEORGIOU, I. and KARAKOSTAS, K.X. (2001). Model-complete strategies for sampling from convex autocorrelated finite populations. *J. Stat. Plan. Infer.* **99**, 71–89.

PAPAGEORGIOU, I. and KARAKOSTAS, K.X. (1998). On optimal sampling designs for autocorrelated finite populations. *Biometrika* **85**, 2, 482–486.

RAMAKRISHNAN, M.K. (1975). Choice of an optimum sampling strategy. *I. Ann. Statist.* **3**, 669–679.

ROYALL, R.M. (1970). On finite population sampling theory under certain linear regression models. *Biometrika* **57**, 377–387.

SÄRNDAL, C.E. (1982). Implications of survey design for generalised regression estimation of linear functions. *J. Statist. Plann. Inference* **19**, 155–170.

TAM, S.M. (1984). Optimal estimation in survey sampling under a regression superpopulation model. *Biometrika* **71**, 3, 645–647.

VANBRACKLE, L.N. III and REYNOLDS, M.R. (1997). EWMA and Cusum Control Charts in the presence of correlation. *Comm. Statist. Simulation Comput.* **26**, 3, 979–1008.

VARIN, C. (2008). On composite marginal likelihoods. *Adv. Stat. Anal.* **92**, 1, 1–28.

VASILOPOULOS, A.V. and STAMBOULIS, A.P. (1978). Modification of Control Chart Limits in the Presence of Data Correlation. *J. Qual. Technol.* **10**, 20–30.

WATSON, G.S. (1972). Trent-surface analysis and spatial correlation. *Geol. Soc. Am. Spec. Pap.* **146**, 39–46.

YANG, K. and HANCOCK, W.M. (1990). Statistical quality control for correlated samples. *Int. J. Prod. Res.* **28**, 3, 595–608.

YATES, F. (1948). Systematic sampling. *Phil. Trans. R. Soc A* **241**, 345–377.

Ioulia Papageorgiou
Department of Statistics, Athens
University of Economics and Business,
Athens, Greece
E-mail: ioulia@aueb.gr

Paper received: 28 May 2013.