Model-based Variance Estimation for Systematic Sampling

Xiaoxi Li[†] and J. D. Opsomer
Center for Survey Statistics and Methodology, and
Department of Statistics, Iowa State University, Ames, IA 50011

Abstract:
Systematic sampling is a frequently used sampling method in surveys, because of its ease of implementation and its design efficiency. An important drawback of systematic sampling, however, is that no direct estimator of the design variance is available. We describe a new estimator of the model-based expectation of the design variance, under a nonparametric model for the population. The nonparametric model is sufficiently flexible that it can be expected to hold at least approximately for many practical situations. We prove the consistency of the estimator for both the anticipated variance and the design variance under the nonparametric model, and illustrate its practical properties through a simulation experiment.
KEY WORDS: anticipated variance, nonparametric model, local polynomial regression.

1. Introduction

Systematic sampling is widely used in surveys of finite populations due to its appealing simplicity and efficiency. The method was first studied by Madow and Madow (1944), in which the expression of design-based variance for the sample mean was developed. However, it is impossible to derive an unbiased design-based estimator for this variance, because systematic sampling is equivalent to cluster sampling with only one cluster selected (Iachan 1982). Some less-than-perfect approaches for dealing with this problem exist in literature. One is to use biased variance estimators, and another one, for example, is to use an auxiliary simple random sample. For the former approach, Särndal et al. (1992) remarked that the estimator due to Yates and Grundy (1953) and Sen (1953) will overestimate the variance. A more comprehensive review can be found in Wolter (1985), where eight biased variance estimators were described and guidelines for choosing among them were given. For the latter approach, Zinger (1980) pursued an approach, defined as partially systematic sampling, that gives an unbiased variance estimator, by mixing systematic and simple random samples together.

[†]204 Snedecor Hall, Iowa State University, Ames, IA 50011; lixiaoxi@iastate.edu.

The above variance estimation methods are conditional on the design. In other words, they are design-based in the way that we treat the finite population as fixed. There also exist some model-based variance estimators where the populations are considered random realizations from a superpopulation model. For the case of a linear superpopulation, Montanari and Bartolucci (1998) proposed a model-based variance estimator, which is approximately unbiased for the *anticipated variance*, i.e. the expectation of the design-based variance for the sample mean under the superpopulation model. However, it may lack accuracy and efficiency due to a higher contribution of the bias if the systematic component of the superpopulation is significantly different from linear. A new class of unbiased estimators that includes some simple nonparametric estimators was proposed by Bartolucci and Montanari (2005) and was shown to be unbiased under linear superpopulation models.

In this article, we propose a model-based nonparametric variance estimator based on local polynomial regression. The systematic sampling framework is briefly described in Section 2, and Section 3 reviews the model-based variance results under the linear superpopulation model. In Section 4, we study the properties of the proposed local polynomial variance estimator under the nonparametric superpopulation model. Simulation results are presented in Section 5 and conclusions are drawn in Section 6.

2. Systematic sampling

Suppose the study variable Y for a finite population is $Y_1, Y_2, \cdots, Y_N$. Then the population mean is

$$\bar{Y}_N = \frac{1}{N}\sum_{j=1}^{N} Y_j,$$

which is estimated by the sample mean. Let n denote the *sample size* and $k = N/n$ denote the *sampling interval*. For simplicity, we assume k to be an integer. So the bth systematic sample ($b = 1, \cdots, k$) consists of the observations with the following labels

$$b,\ b+k,\ \ldots,\ b+(n-1)k.$$

The bth sample mean is

$$\bar{Y}_{S_b} = \frac{1}{n}\sum_{j=1}^{n} Y_{S_b j}$$

and the design-based variance for sample mean $\bar{Y}_S$, denoted by $\mathrm{Var}_p(\bar{Y}_S)$, is

$$\mathrm{Var}_p(\bar{Y}_S) = \frac{1}{k}\sum_{b=1}^{k}(\bar{Y}_{S_b} - \bar{Y}_N)^2. \qquad (1)$$

3. Variance estimation under linear models

In the model-based context, the population is considered being drawn from a superpopulation model. Let $Y_j \in \Re$ $(j = 1, 2, \cdots)$ be a set of independent and identically distributed random variables. Let $\mathbf{X}_j \in \Re^d$ $(j = 1, 2, \cdots)$ be vectors of auxiliary variables, which we consider as fixed. Suppose the linear superpopulation model, denoted by L, is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad (2)$$

where $\mathrm{E}_L(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $\mathrm{Var}_L(\boldsymbol{\varepsilon}) = \sigma_L^2\boldsymbol{\Omega}$, with E_L and Var_L denoting the expectation and variance under the model L, respectively. For simplicity, we assume $\boldsymbol{\Omega}$ to be diagonal. i.e. $\boldsymbol{\Omega} = \mathrm{diag}\{\omega_1, \omega_2, \cdots, \omega_N\}$.

In model (2), $\mathbf{Y} = (Y_1, Y_2, \cdots, Y_N)^T$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_d)^T$, and

$$\mathbf{X} = \begin{pmatrix} 1 & X_{11} & \cdots & X_{1d} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{N1} & \cdots & X_{Nd} \end{pmatrix}.$$

We can rewrite the design variance (1) as

$$\mathrm{Var}_p(\bar{Y}_S) = \frac{1}{kn^2}\mathbf{Y}^T\mathbf{N}\mathbf{Y}, \qquad (3)$$

where $\mathbf{N} = \mathbf{M}^T\mathbf{H}\mathbf{M}$, with $\mathbf{M} = \mathbf{1}_n^T \otimes \mathbf{I}_k$ and $\mathbf{H} = \mathbf{I}_k - \frac{1}{k}\mathbf{1}_k\mathbf{1}_k^T$. Here $\otimes$ is the Kronecker product and $\mathbf{1}_r$ is a column vector of 1's of length r.

The model anticipated variance is

$$\begin{aligned}\mathrm{E}_L[\mathrm{Var}_p(\bar{Y}_S)] &= \frac{1}{kn^2}\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{N}\mathbf{X}\boldsymbol{\beta} \\ &\quad + \frac{1}{kn^2}\mathrm{tr}(\mathbf{N}\boldsymbol{\Omega})\sigma_L^2. \qquad (4)\end{aligned}$$

Bartolucci and Montanari (2005) discuss an unbiased estimator for $\mathrm{E}_L[\mathrm{Var}_p(\bar{Y}_S)]$, defined as

$$\begin{aligned}\hat{\mathrm{V}}_L(\bar{Y}_S) &= \frac{1}{kn^2}\hat{\boldsymbol{\beta}}_b^T\mathbf{X}^T\mathbf{N}\mathbf{X}\hat{\boldsymbol{\beta}}_b - t\hat{\sigma}_{Lb}^2 \\ &\quad + \frac{1}{kn^2}\mathrm{tr}(\mathbf{N}\boldsymbol{\Omega})\hat{\sigma}_{Lb}^2, \qquad (5)\end{aligned}$$

where $\hat{\boldsymbol{\beta}}_b$ is the ordinary least square (OLS) estimator for $\boldsymbol{\beta}$ for the bth sample and $\hat{\sigma}_{Lb}^2$ is a model unbiased estimator for σ_L^2, which is defined as

$$\hat{\sigma}_{Lb}^2 = \frac{(\mathbf{Y}_b - \mathbf{X}\hat{\boldsymbol{\beta}}_b)^T\boldsymbol{\Omega}_b^{-1}(\mathbf{Y}_b - \mathbf{X}\hat{\boldsymbol{\beta}}_b)}{n - \mathrm{rank}(\mathbf{X})},$$

where $\boldsymbol{\Omega}_b$ is a sub matrix of $\boldsymbol{\Omega}$ corresponding to the bth sample.

In equation (5), $t\hat{\sigma}_{Lb}^2$ is a bias correction term with $t = \frac{1}{(kn)^2}\sum_{b=1}^{k}\mathrm{tr}(\mathbf{P}_b^T\mathbf{X}^T\mathbf{N}\mathbf{X}\mathbf{P}_b\boldsymbol{\Omega}_b)$ and a choice of $\mathbf{P}_b$ is $\mathbf{P}_b = (\mathbf{X}_b^T\boldsymbol{\Omega}_b^{-1}\mathbf{X}_b)^{-1}\mathbf{X}_b^T\boldsymbol{\Omega}_b^{-1}$. We assume that $(\mathbf{X}_b^T\boldsymbol{\Omega}_b^{-1}\mathbf{X}_b)^{-1}$ exists.

Li (2005) shows that, under an asymptotic framework in which $N \to \infty$ and a set of assumptions,

$$\mathrm{Var}_p(\bar{Y}_S) - \mathrm{E}_L[\mathrm{Var}_p(\bar{Y}_S)] = O_p\left(N^{-1/2}\right), \qquad (6)$$

and

$$\hat{\mathrm{V}}_L(\bar{Y}_S) - \mathrm{E}_L[\mathrm{Var}_p(\bar{Y}_S)] = O_p(n^{-1/2}) \qquad (7)$$

and hence by (6) and (7),

$$\hat{\mathrm{V}}_L(\bar{Y}_S) - \mathrm{Var}_p(\bar{Y}_S) = O_p(n^{-1/2}). \qquad (8)$$

Equation (7) suggests that $\hat{\mathrm{V}}_L(\bar{Y}_S)$ is a consistent estimator for $\mathrm{E}_L[\mathrm{Var}_p(\bar{Y}_S)]$. Equation (6) indicates that the conditional variance $\mathrm{Var}_p(\bar{Y}_S)$ converges to $\mathrm{E}_L[\mathrm{Var}_p(\bar{Y}_S)]$ in probability. So $\hat{\mathrm{V}}_L(\bar{Y}_S)$ can be used as a consistent predictor for $\mathrm{Var}_p(\bar{Y}_S)$, as (8) suggests. Details on these results are in Li (2005).

4. Variance estimators under nonparametric models

Parametric method are appropriate when we can correctly specify the superpopulation model. However, if the superpopulation model is incorrectly specified, parametric method may result in biased or inefficient estimation. We propose a consistent variance estimator under nonparametric models. We study the case where $d = 1$. Let $\mathbf{X} = (X_1, X_2, \cdots, X_N)$. The nonparametric superpopulation model, denoted by NP, is

$$\mathbf{Y} = \mathbf{m} + \boldsymbol{\varepsilon}, \qquad (9)$$

where $\mathrm{E}_{NP}(Y_j|X_j = x_j) = m(x_j)$ and $\mathrm{Var}_{NP}(\boldsymbol{\varepsilon}) = \sigma_{NP}^2\boldsymbol{\Omega}$. Let $m(\cdot)$ be a continuous and bounded function, and define $\mathbf{m} = (m(x_1), m(x_2), \cdots, m(x_N))$. We assume that the ω_j's are bounded and positive, where $j = 1, ..., N$.

Under model (9), the expected value of $\mathrm{Var}_p(\bar{Y}_S)$ is

$$\mathrm{E}_{NP}(\mathrm{Var}_p(\bar{Y}_S)) = \frac{1}{kn^2}\mathbf{m}^T\mathbf{N}\mathbf{m} + \frac{1}{kn^2}\mathrm{tr}(\mathbf{N}\boldsymbol{\Omega})\sigma^2_{NP}. \quad (10)$$

As an estimator for $\mathrm{E}_{NP}(\mathrm{Var}_p(\bar{Y}_S))$, we propose

$$\hat{\mathrm{V}}_{NP}(\bar{Y}_S) = \frac{1}{kn^2}(\hat{\mathbf{m}}_b^T\mathbf{N}\hat{\mathbf{m}}_b) + \frac{1}{kn^2}\mathrm{tr}(\mathbf{N}\boldsymbol{\Omega})\hat{\sigma}^2_{NPb}. \quad (11)$$

Here $\hat{\sigma}^2_{NPb}$ is defined as

$$\hat{\sigma}^2_{NPb} = \frac{(\mathbf{Y}_b - \hat{\mathbf{m}}_b)^T\boldsymbol{\Omega}_b^{-1}(\mathbf{Y}_b - \hat{\mathbf{m}}_b)}{n}, \quad (12)$$

and $\hat{\mathbf{m}}_b = (\hat{m}(x_1), \cdots, \hat{m}(x_n))$ where $\hat{m}(x_j)$ is the local polynomial regression estimator

$$\hat{m}(x_j) = \mathbf{e}_1^T(\mathbf{X}_{bj}^T\mathbf{W}_{bj}\mathbf{X}_{bj})^{-1}\mathbf{X}_{bj}^T\mathbf{W}_{bj}\mathbf{Y}_b,$$

where $\mathbf{e}_1$ is the $(q+1)\times 1$ vector having 1 in the first entry and all other entries 0, and

$$\mathbf{X}_{bj} = \begin{pmatrix} 1 & (x_1 - x_j) & \cdots & (x_1 - x_j)^q \\ \vdots & \vdots & \vdots & \vdots \\ 1 & (x_n - x_j) & \cdots & (x_n - x_j)^q \end{pmatrix}, \quad (13)$$

$$\mathbf{W}_{bj} = \mathrm{diag}\left\{K\left(\frac{x_i - x_j}{h}\right)\frac{1}{h}, \quad i = 1, \cdots n\right\}, \quad (14)$$

where h denotes the bandwidth, q denotes the degree of local polynomial regression and $K\left(\frac{x_i - x_j}{h}\right)$ is the kernel function. We refer to Wand and Jones (1995) for more information on the local polynomial regression estimator. Note that we are not including the bias correction term $t\hat{\sigma}^2_{Lb}$ used in equation (11) because that term is asymptotically negligible.

To study the convergence properties of $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$, we make the following assumptions.

A 1 *The errors ε_j's are independent with mean zero, variance $\omega_j\sigma^2_{NP}$ and compact support, uniformly for all N.*

A 2 *For each N, we consider the x_j's as fixed with respect to the superpopulation model NP. The x_j's are independent and identically distributed with $F(x) = \int_{-\infty}^x f(t)dt$, where $f(\cdot)$ is a density function with compact support $[a_x, b_x]$ and $f(x) > 0$ for all $x \in [a_x, b_x]$.*

A 3 *As $N \to \infty$, $nN^{-1} \to p \in (0,1), h \to 0$ and $Nh^2/(\log\log N) \to \infty$.*

A 4 *The $(q+1)th$ derivative of the function $m(\cdot)$ exists and is bounded on $[a_x, b_x]$.*

Theorem 1 *Under assumption A1 - A4,*

$$Var_p(\bar{Y}_S) - E_{NP}(Var_p(\bar{Y}_S)) = O_p(N^{-1/2}), \quad (15)$$

$$\hat{V}_{NP}(\bar{Y}_S) - E_{NP}(Var_p(\bar{Y}_S)) = O_p\left(h^{q+1}\right) + O_p\left(\frac{1}{\sqrt{nh}}\right), \quad (16)$$

and

$$\hat{V}_{NP}(\bar{Y}_S) - Var_p(\bar{Y}_S) = O_p\left(h^{q+1}\right) + O_p\left(\frac{1}{\sqrt{nh}}\right). \quad (17)$$

Theorem1 shows that $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ is a consistent estimator for $\mathrm{E}_{NP}(\mathrm{Var}_p(\bar{Y}_S))$ and a consistent predictor for $\mathrm{Var}_p(\bar{Y}_S)$ under the nonparametric model NP. We provide an outline of proof for (16) in Appendix. The proof for (15) and (17) can be found in Li (2005).

5. Simulation Study

To further investigate the statistical properties of the above variance estimators and predictors, we perform a simulation study. For simplicity, we assume that the errors are independently and normally distributed with homogeneous variances. Two superpopulation models are examined: the linear model

$$y_j = 5 + 2x_j + \varepsilon_j, \quad (18)$$

where $j = 1, \ldots, N$ and $\varepsilon_j \sim N(0, \sigma_1^2)$, and the quadratic model

$$y_j = 5 + 2x_j - 2x_j^2 + \varepsilon_j, \quad (19)$$

where $j = 1, \ldots, N$ and $\varepsilon_j \sim N(0, \sigma_2^2)$.

Let R_1^2 and R_2^2 denote the coefficient of determination for model (18) and (19), respectively. The coefficient of determination, also known as R-square, is the fraction of variation in the response that is explained by the model. So bigger R-square means bigger predictive power of the model. We investigated two levels of σ_1^2 and two levels of σ_2^2, which correspondingly determined two levels of R_1^2 and two levels of R_2^2. Specifically, we have four different cases: (1)$R_1^2 \approx 0.75$; (2)$R_1^2 \approx 0.25$; (3)$R_2^2 \approx 0.75$; (4)$R_2^2 \approx 0.25$.

We compare the performance of $\hat{\mathrm{V}}_L(\bar{Y}_S)$ and $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ with the variance estimator for simple random sampling (SI) design, i.e.

$$\hat{\mathrm{V}}_{SI}(\bar{Y}_S) = \frac{1-f}{n}S^2_{YS}$$

where $f = n/N$ and $S^2_{YS} = \frac{1}{n-1}\sum_S (Y_k - \bar{Y}_S)^2$.

We generate populations of size $N = 2,000$, and we choose three different systematic sample sizes $n = 500, 100$ and 10, with corresponding sampling intervals $k = 4, 20$ and 200, respectively. To draw a systematic sample, we first sort the data by x, from the smallest to the largest, then we randomly choose an observation from the first k observations, say the bth one. Then our sample consists of the observations with the following subscripts: $b,\ b+k,\ ...,\ b+(n-1)k$. For each sample, we calculate the corresponding $\mathrm{Var}_p(\bar{Y}_S)$, $\mathrm{E}_L[\mathrm{Var}_p(\bar{Y}_S)]$, $\hat{\mathrm{V}}_L(\bar{Y}_S)$, $\mathrm{E}_{NP}[\mathrm{Var}_p(\bar{Y}_S)]$ and $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ as defined in (3), (4), (5), (10) and (11) respectively. For $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$, we calculate it using two bandwidth values: $h = 0.50$ and $h = 0.25$. Each simulation setting is repeated $B = 10,000$ times.

For each of $\hat{\mathrm{V}}_L(\bar{Y}_S)$, $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ and $\hat{\mathrm{V}}_{SI}(\bar{Y}_S)$, we calculate the relative bias (RB), mean squared error (MSE) and mean squared prediction error (MSPE), which are defined as follows:

$$\begin{aligned}
\mathrm{RB} &= \frac{\mathrm{E}(\hat{\mathrm{V}}(\bar{Y}_S)) - \mathrm{E}_L[\mathrm{Var}_p(\bar{Y}_S)]}{\mathrm{E}_L[\mathrm{Var}_p(\bar{Y}_S)]}, \\
\mathrm{MSE} &= \mathrm{E}(\hat{\mathrm{V}}(\bar{Y}_S) - \mathrm{E}_L[\mathrm{Var}_p(\bar{Y}_S)])^2, \\
\text{and}\quad \mathrm{MSPE} &= \mathrm{E}(\hat{\mathrm{V}}(\bar{Y}_S) - \mathrm{Var}_p(\bar{Y}_S))^2.
\end{aligned}$$

where $\hat{\mathrm{V}}(\bar{Y}_S)$ denotes one of $\hat{\mathrm{V}}_L(\bar{Y}_S)$, $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ and $\hat{\mathrm{V}}_{SI}(\bar{Y}_S)$.

For all four cases, $\hat{\mathrm{V}}_L(\bar{Y}_S)$ assumes linear trend for the superpopulation model. So it is expected that for case 3 and 4, which generate populations from quadratic model (19) , $\hat{\mathrm{V}}_L(\bar{Y}_S)$ will have poor variance estimation results.

Table 1 reports the relative bias for $\hat{\mathrm{V}}_L(\bar{Y}_S)$, $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ (evaluated at two bandwidth values) and $\hat{\mathrm{V}}_{SI}(\bar{Y}_S)$. We can see that for case 3 and 4, $\hat{\mathrm{V}}_L(\bar{Y}_S)$ is a significantly biased estimator. And its bias is of similar magnitude to $\hat{\mathrm{V}}_{SI}(\bar{Y}_S)$ in corresponding cases. For case 1 and 2, in which the populations have linear trends, the relative biases of $\hat{\mathrm{V}}_L(\bar{Y}_S)$ are generally small, and tend to get smaller when sample size gets larger. The relative biases of $\hat{\mathrm{V}}_{SI}(\bar{Y}_S)$ in case 1 and 2 are similar to those in case 3 and 4, respectively, and behave similarly poorly. When we use the model based variance estimator $\hat{\mathrm{V}}_L(\bar{Y}_S)$, assuming the wrong model can be a serious problem.

The results in Table 1 also suggest that $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ performs better than $\hat{\mathrm{V}}_L(\bar{Y}_S)$ in almost all cases, especially when the superpopulation model is quadratic. This is because $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ does not require a linear specification of the model, as it only requires smoothness of the superpopulation mean function. We also see that its relative biases are generally small, except for case 3 and $n = 10$, where the relative bias values are -21.49% and -18.99% for $h = 0.50$ and $h = 0.25$, respectively. This is mostly likely due to the extremely small sample size for local polynomial regression. Also as far as the relative bias is concerned, there seems to be no difference between these two bandwidth values.

Table 2 reports the ratios of MSE and MSPE between $\hat{\mathrm{V}}_L(\bar{Y}_S)$ and $\hat{\mathrm{V}}_{SI}(\bar{Y}_S)$, and the ratios of MSE and MSPE between $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ and $\hat{\mathrm{V}}_{SI}(\bar{Y}_S)$, evaluated at two different bandwidth values. MSE measures the variability of an estimator and MSPE measures the variability of a predictor. Smaller MSE and MSPE are desired. We see that when the superpopulation model is linear, i.e. case 1 and 2, $\hat{\mathrm{V}}_L(\bar{Y}_S)$ performs better than $\hat{\mathrm{V}}_{SI}(\bar{Y}_S)$, because those ratios are less than one. And when the linear superpopulation is more precise, i.e. $R^2 \approx 0.75$, the advantage of using $\hat{\mathrm{V}}_L(\bar{Y}_S)$ is more obvious. However, when it comes to case 3 and 4, where we incorrectly assume the superpopulation model, $\hat{\mathrm{V}}_{SI}(\bar{Y}_S)$ performs slightly better than $\hat{\mathrm{V}}_L(\bar{Y}_S)$. For $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$, the ratios are less than one in all categories, suggesting that $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ is a less variable estimator and a less variable predictor than $\hat{\mathrm{V}}_{SI}(\bar{Y}_S)$.

6. Conclusions

From the above study, we conclude that $\hat{\mathrm{V}}_L(\bar{Y}_S)$ and $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ are consistent estimators for $\mathrm{E}_L[\mathrm{Var}_p(\bar{Y}_S)]$ and consistent predictors for $\mathrm{Var}_p(\bar{Y}_S)$, under the assumed models (2) and (9), respectively. The linear estimator $\hat{\mathrm{V}}_L(\bar{Y}_S)$ only performs well if the model is correctly specified. The nonparametric model (9) is less restrictive than the linear regression model (2), and the corresponding variance estimator $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ performs well in almost all cases. So in practice, we recommend the nonparametric estimator $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$.

7. Appendix

Outline of proof of equation (16):

$$\begin{aligned}
&\hat{\mathrm{V}}_{NP}(\bar{Y}_S) - \mathrm{E}_{NP}(\mathrm{Var}_p(\bar{Y}_S)) \\
&\qquad = \frac{1}{kn^2}(\hat{\mathbf{m}}_b^T\mathbf{N}\hat{\mathbf{m}}_b - \mathbf{m}^T\mathbf{N}\mathbf{m}) \\
&\qquad\quad + \frac{1}{kn^2}\mathrm{tr}(\mathbf{N}\boldsymbol{\Omega})(\hat{\sigma}^2_{NPb} - \sigma^2_{NP}) \\
&\qquad \equiv (*) + (**). \qquad (20)
\end{aligned}$$

We can write $(*)$ as

$$\begin{aligned}
(*) &= \frac{1}{kn^2}(\hat{\mathbf{m}}_b - \mathbf{m})^T\mathbf{N}(\hat{\mathbf{m}}_b - \mathbf{m}) \\
&\quad + \frac{1}{kn^2}\mathbf{m}^T\mathbf{N}(\hat{\mathbf{m}}_b - \mathbf{m})
\end{aligned}$$

Relative Bias (%)		Linear 1: $R_1^2 \approx 0.75$	Linear 2: $R_1^2 \approx 0.25$	Quadratic 3: $R_2^2 \approx 0.75$	Quadratic 4: $R_2^2 \approx 0.25$
$\hat{\mathrm{V}}_L(\bar{Y}_S)$	$n = 500$	0.00	-0.06	318.22	31.77
	$n = 100$	-0.18	-0.04	323.68	32.67
	$n = 10$	1.49	1.19	376.10	38.11
$\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$, $h = 0.5$	$n = 500$	0.00	0.00	-0.74	-0.07
	$n = 100$	0.01	0.01	-3.01	-0.30
	$n = 10$	1.21	1.54	-21.49	-1.29
$\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$, $h = 0.25$	$n = 500$	0.00	0.00	-0.74	-0.07
	$n = 100$	0.03	0.03	-2.99	-0.28
	$n = 10$	3.71	4.77	-18.99	1.86
$\hat{\mathrm{V}}_{SI}(\bar{Y}_S)$	$n = 500$	330.24	33.20	317.84	31.73
	$n = 100$	321.65	32.96	320.82	32.39
	$n = 10$	250.86	32.67	349.15	34.91

Table 1: Simulated relative bias for $\hat{\mathrm{V}}_L(\bar{Y}_S)$, $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ (at two bandwidth values) and $\hat{\mathrm{V}}_{SI}(\bar{Y}_S)$ for four populations and three systematic sample sizes (in percent).

		Linear 1: $R_1^2 \approx 0.75$	Linear 2: $R_1^2 \approx 0.25$	Quadratic 3: $R_2^2 \approx 0.75$	Quadratic 4: $R_2^2 \approx 0.25$
$\frac{\mathrm{MSE}(\hat{\mathrm{V}}_L(\bar{Y}_S))}{\mathrm{MSE}(\hat{\mathrm{V}}_{SI}(\bar{Y}_S))}$	$n = 500$	0.00	0.03	1.00	1.00
	$n = 100$	0.00	0.14	1.02	1.02
	$n = 10$	0.02	0.50	1.16	1.17
$\frac{\mathrm{MSPE}(\hat{\mathrm{V}}_L(\bar{Y}_S))}{\mathrm{MSPE}(\hat{\mathrm{V}}_{SI}(\bar{Y}_S))}$	$n = 500$	0.06	0.85	1.00	1.00
	$n = 100$	0.01	0.51	1.02	1.01
	$n = 10$	0.02	0.51	1.16	1.16
$\frac{\mathrm{MSE}(\hat{\mathrm{V}}_{NP}(\bar{Y}_S), h=0.50)}{\mathrm{MSE}(\hat{\mathrm{V}}_{SI}(\bar{Y}_S))}$	$n = 500$	0.00	0.00	0.00	0.00
	$n = 100$	0.00	0.00	0.00	0.00
	$n = 10$	0.00	0.01	0.01	0.00
$\frac{\mathrm{MSPE}(\hat{\mathrm{V}}_{NP}(\bar{Y}_S), h=0.50)}{\mathrm{MSPE}(\hat{\mathrm{V}}_{SI}(\bar{Y}_S))}$	$n = 500$	0.79	0.68	0.82	0.69
	$n = 100$	0.01	0.42	0.01	0.43
	$n = 10$	0.00	0.03	0.00	0.02
$\frac{\mathrm{MSE}(\hat{\mathrm{V}}_{NP}(\bar{Y}_S), h=0.25)}{\mathrm{MSE}(\hat{\mathrm{V}}_{SI}(\bar{Y}_S))}$	$n = 500$	0.00	0.00	0.00	0.00
	$n = 100$	0.00	0.00	0.00	0.00
	$n = 10$	0.00	0.03	0.00	0.01
$\frac{\mathrm{MSPE}(\hat{\mathrm{V}}_{NP}(\bar{Y}_S), h=0.25)}{\mathrm{MSPE}(\hat{\mathrm{V}}_{SI}(\bar{Y}_S))}$	$n = 500$	0.06	0.85	0.06	0.85
	$n = 100$	0.01	0.42	0.01	0.43
	$n = 10$	0.01	0.05	0.00	0.04

Table 2: Ratios of MSE and MSPE between $\hat{\mathrm{V}}_L(\bar{Y}_S)$ and $\hat{\mathrm{V}}_{SI}(\bar{Y}_S)$, between $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ and $\hat{\mathrm{V}}_{SI}(\bar{Y}_S)$ at two bandwidth values.

$$
\begin{aligned}
& +\frac{1}{kn^2}(\hat{\mathbf{m}}_b-\mathbf{m})^T\mathbf{N}\mathbf{m} \\
\equiv\ & (A)+2(B).
\end{aligned}
$$

Note that $\mathbf{m}^T\mathbf{N}(\hat{\mathbf{m}}_b-\mathbf{m}) = (\hat{\mathbf{m}}_b-\mathbf{m})^T\mathbf{N}\mathbf{m}$ because they are both scalars. By the definition of matrix $\mathbf{N}$, we can write (A) as

$$
\begin{aligned}
(A) &= \frac{1}{kn^2}(\hat{\mathbf{m}}_b-\mathbf{m})^T\mathbf{N}(\hat{\mathbf{m}}_b-\mathbf{m}) \\
&= \frac{1}{k}\sum_{b=1}^{k}\left[\frac{1}{n}\sum_{j\in s_b}(\hat{m}(x_j)-m(x_j))\right. \\
&\quad \left. -\frac{1}{N}\sum_{j\in U}(\hat{m}(x_j)-m(x_j))\right]^2 \\
&= \frac{1}{k}\sum_{b=1}^{k}\left\{(a1)+(a2)+(a3)\right\},
\end{aligned}
$$

where

$$
\begin{aligned}
(a1) &= \frac{1}{n^2}[\sum_{j\in s_b}(\hat{m}(x_j)-m(x_j))]^2, \\
(a2) &= \frac{1}{N^2}[\sum_{j\in U}(\hat{m}(x_j)-m(x_j))]^2, \\
\text{and}\quad (a3) &= -\frac{2}{nN}\sum_{j\in s_b}(\hat{m}(x_j)-m(x_j)) \\
&\quad \cdot\sum_{j\in U}(\hat{m}(x_j)-m(x_j)).
\end{aligned}
$$

Note that

$$
\begin{aligned}
\hat{m}(x_j)-m(x_j) &= \mathbf{s}_{bj}\mathbf{Y}_b - m(x_j) \\
&= \mathbf{s}_{bj}(\mathbf{m}_b+\boldsymbol{\varepsilon}_b) - m(x_j) \\
&= b_b(x_j)+\mathbf{s}_{bj}\boldsymbol{\varepsilon}_b.
\end{aligned}
$$

Here $\mathbf{s}_{bj}$ is the smoother matrix and $\mathbf{s}_{bj} = \mathbf{e}_1^T(\mathbf{X}_{bj}^T\mathbf{W}_{bj}\mathbf{X}_{bj})^{-1}\mathbf{X}_{bj}^T\mathbf{W}_{bj}$, where $\mathbf{X}_{bj}$ and $\mathbf{W}_{bj}$ are defined as in (13) and (14), respectively. For simplicity, we will use K_{ij} to denote $K\left(\frac{x_i-x_j}{h}\right)$ in future notation.

Now expand the parentheses in $(a1)$, we have

$$
\begin{aligned}
\mathrm{E}(a1) &= \frac{1}{n^2}\sum_{j\in s_b}b_b^2(x_j)+\frac{1}{n^2}\mathrm{E}\left(\sum_{j\in s_b}\mathbf{s}_{bj}\boldsymbol{\varepsilon}_b\boldsymbol{\varepsilon}_b^T\mathbf{s}_{bj}^T\right) \\
&\quad +\frac{1}{n^2}\sum_{j\in s_b}\sum_{l\in s_b, j\neq l}b_b(x_j)b_b(x_l) \\
&\quad +\frac{1}{n^2}\mathrm{E}\left(\sum_{j\in s_b}\sum_{l\in s_b, j\neq l}\mathbf{s}_{bj}\boldsymbol{\varepsilon}_b\boldsymbol{\varepsilon}_b^T\mathbf{s}_{bl}^T\right). \qquad (21)
\end{aligned}
$$

The right-hand side of (21) contains four terms. We will calculate them one by one.

(i) First let us investigate $\frac{1}{n^2}\sum_{j\in s_b}b_b^2(x_j)$. We will use a technique similar to Ruppert and Wand (1994). Let $\mathbf{m}_b = (m(x_1), m(x_2), \cdots, m(x_n))$. By Taylor's theorem,

$$
\mathbf{m}_b = \mathbf{X}_{bj}\begin{pmatrix} m(x_j) \\ \mathbf{D}_m(x_j)\end{pmatrix} + \mathbf{R}_m(x_j),
$$

where $\mathbf{D}_m(x_j) = (m'(x_j), \frac{1}{2}m''(x_j), \cdots, \frac{1}{q!}m^{(q)}(x_j))$ and $\mathbf{R}_m(x_j)$ is a vector of Taylor series remainder terms. So

$$
\begin{aligned}
b_b(x_j) &= \mathbf{s}_{bj}\mathbf{m}_b - m(x_j) = \mathbf{s}_{bj}\mathbf{R}_m(x_j) \\
&= \mathbf{e}_1^T(\mathbf{X}_{bj}^T\mathbf{W}_{bj}\mathbf{X}_{bj})^{-1}\mathbf{X}_{bj}^T\mathbf{W}_{bj} \\
&\quad \cdot\begin{pmatrix} \frac{1}{(q+1)!}m^{(q+1)}(x_{j1}^*)(x_1-x_j)^{q+1} \\ \vdots \\ \frac{1}{(q+1)!}m^{(q+1)}(x_{jn}^*)(x_n-x_j)^{q+1}\end{pmatrix} \\
&= \mathbf{e}_1^T\begin{pmatrix} z_{11j} & \cdots & z_{1(q+1)j} \\ \vdots & \vdots & \vdots \\ z_{(q+1)1j} & \cdots & z_{(q+1)(q+1)j}\end{pmatrix}^{-1} \\
&\quad \cdot\begin{pmatrix} t_{1j} \\ \vdots \\ t_{(q+1)j}\end{pmatrix} \\
&\equiv \mathbf{e}_1^T\mathbf{Z}_j^{*-1}\mathbf{T},
\end{aligned}
$$

where

$$
z_{stj} = \frac{1}{nh}\sum_{i=1}^{n}K_{ij}(x_i-x_j)^{s+t-2},
$$

$$
t_{kj} = \frac{1}{nh}\sum_{i=1}^{n}K_{ij}(x_i-x_j)^{k+q}\frac{m^{(q+1)}(x_{ji}^*)}{(q+1)!},
$$

with $s,t = 1,2,\cdots q+1$, $k = 1,2,\cdots q+1$ and x_{ji}^* is some point between x_i and x_j.

Under assumptions A2 and A3, by Lemma 2 (ii) of Breidt and Opsomer (2000), for a certain point x_j, there are at least $q+1$ points in the interval $[x_j-h, x_j+h]$. So $\mathbf{Z}_j^*$ is invertible.

Lemma 1 *Assume that the kernel function K_{ij} is bounded above, then*

$$
\frac{1}{nh}\sum_{i=1}^{n}K_{ij}(x_i-x_j)^r = O(h^r),
$$

where $r = 0,1,2,\cdots$.

The proof of Lemma 1 is provided by Li (2005). Thus, suppose assumption A4 holds, by Lemma 1, we have $z_{stj} = O(h^{s+t-2})$ and $t_{kj} = O(h^{k+q})$.

So

$$b_b(x_j) = \mathbf{e}_1^T \begin{pmatrix} O(1) & \cdots & O(h^q) \\ \vdots & \vdots & \vdots \\ O(h^q) & \cdots & O(h^{2q}) \end{pmatrix}^{-1} \cdot \begin{pmatrix} O(h^{q+1}) \\ \vdots \\ O(h^{2q+1}) \end{pmatrix}.$$

Note that the order of a matrix is the same as its inverse. So

$$b_b(x_j) = O(h^{q+1}) + \cdots + O(h^{3q+1}) = O(h^{q+1}), \tag{22}$$

and thus

$$\begin{aligned} \frac{1}{n^2}\sum_{j\in s_b} b^2(x_j) &= \frac{1}{n^2}\sum_{j\in s_b} O(h^{2q+2}) \\ &= O\left(\frac{h^{2q+2}}{n}\right). \end{aligned} \tag{23}$$

(ii) Secondly, let us compute $\frac{1}{n^2}\mathrm{E}\left(\sum_{j\in s_b} \mathbf{s}_{bj}\boldsymbol{\varepsilon}_b\boldsymbol{\varepsilon}_b^T\mathbf{s}_{bj}^T\right)$ in (21).

$$\begin{aligned} \frac{1}{n^2}\mathrm{E}\left(\sum_{j\in s_b} \mathbf{s}_{bj}\boldsymbol{\varepsilon}_b\boldsymbol{\varepsilon}_b^T\mathbf{s}_{bj}^T\right) &= \frac{1}{n^2}\sum_{j\in s_b} \mathbf{s}_{bj}\boldsymbol{\Omega}_b\mathbf{s}_{bj}^T \\ = &\ \frac{1}{n^2}\sum_{j\in s_b} \mathbf{e}_1^T(\mathbf{X}_{bj}^T\mathbf{W}_{bj}\mathbf{X}_{bj})^{-1}\mathbf{X}_{bj}^T\mathbf{W}_{bj}\boldsymbol{\Omega}_b\mathbf{W}_{bj}^T\mathbf{X}_{bj} \\ &\cdot(\mathbf{X}_{bj}^T\mathbf{W}_{bj}\mathbf{X}_{bj})^{-1}\mathbf{e}_1 \\ \equiv &\ \frac{1}{n^2}\sum_{j\in s_b} \mathbf{e}_1^T\mathbf{Z}_j^{*-1}\mathbf{C}_j^*\mathbf{Z}_j^{*-1}\mathbf{e}_1, \end{aligned}$$

where $\boldsymbol{\Omega}_b$ is the variance-covariance matrix of model (9) and $\boldsymbol{\Omega}_b = \mathrm{diag}(\omega_1, \omega_2, \cdots, \omega_n)$, and

$$\mathbf{C}_j^* = \begin{pmatrix} c_{11j} & \cdots & c_{1(q+1)j} \\ \vdots & \vdots & \vdots \\ c_{(q+1)1j} & \cdots & c_{(q+1)(q+1)j} \end{pmatrix}$$

with

$$\begin{aligned} c_{stj} &= \frac{1}{n^2h^2}\sum_{i=1}^n K_{ij}^2\omega_i(x_i - x_j)^{s+t-2} \\ &= O\left(\frac{h^{s+t-3}}{n}\right) \quad \text{by Lemma 1.} \end{aligned}$$

Li (2005) shows that

$$\mathbf{e}_1^T\mathbf{Z}_j^{*-1}\mathbf{C}_j^*\mathbf{Z}_j^{*-1}\mathbf{e}_1 = O\left(\frac{1}{nh}\right),$$

and thus

$$\frac{1}{n^2}\mathrm{E}\left(\sum_{j\in s_b} \mathbf{s}_{bj}\boldsymbol{\varepsilon}_b\boldsymbol{\varepsilon}_b^T\mathbf{s}_{bj}^T\right) = O\left(\frac{1}{n^2h}\right). \tag{24}$$

(iii) Thirdly we will calculate $\frac{1}{n^2}\sum_{j\in s_b}\sum_{l\in s_b, j\neq l} b_b(x_j)b_b(x_l)$ in (21). Using the result in (22), we get

$$\frac{1}{n^2}\sum_{j\in s_b}\sum_{l\in s_b, j\neq l} b_b(x_j)b_b(x_l) = O\left(h^{2q+2}\right). \tag{25}$$

(iv) The last term on the right-hand side of (21) is $\frac{1}{n^2}\mathrm{E}\left(\sum_{j\in s_b}\sum_{l\in s_b, j\neq l} \mathbf{s}_{bj}\boldsymbol{\varepsilon}_b\boldsymbol{\varepsilon}_b^T\mathbf{s}_{bl}^T\right)$, and Li (2005) shows that

$$\frac{1}{n^2}\mathrm{E}\left(\sum_{j\in s_b}\sum_{l\in s_b, j\neq l} \mathbf{s}_{bj}\boldsymbol{\varepsilon}_b\boldsymbol{\varepsilon}_b^T\mathbf{s}_{bl}^T\right) = O\left(\frac{1}{n}\right). \tag{26}$$

Assumption A3 implies that $nh \to \infty$, and by (23), (24), (25) and (26),

$$\mathrm{E}(a1) = O\left(h^{2q+2}\right) + O\left(\frac{1}{n}\right). \tag{27}$$

Similarly, we can calculate $\mathrm{E}(a2)$ and $\mathrm{E}(a3)$. Under A3, $nh \to \infty$, so

$$\mathrm{E}(a2) = O\left(h^{2q+2}\right) + O\left(\frac{1}{N}\right) \tag{28}$$

$$\text{and} \quad \mathrm{E}(a3) = O\left(h^{2q+2}\right) + O\left(\frac{1}{n}\right). \tag{29}$$

Also note that $(A) > 0$, so $|(A)| = A$. Thus, by (27), (28) and (29),

$$\begin{aligned} \mathrm{E}|(A)| = \mathrm{E}(A) &= \frac{1}{k}\sum_{b=1}^k \{\mathrm{E}(a1) + \mathrm{E}(a2) + \mathrm{E}(a3)\} \\ &= O\left(h^{2q+2}\right) + O\left(\frac{1}{n}\right), \end{aligned}$$

which implies

$$(A) = O_p\left(h^{2q+2}\right) + O_p\left(\frac{1}{n}\right).$$

Next, using a similar technique to that of (A),

$$(B) = O_p\left(h^{q+1}\right) + O_p\left(\frac{1}{\sqrt{n}}\right).$$

Thus,

$$
\begin{aligned}
(*) &= (A) + 2(B) \\
&= O_p\left(h^{q+1}\right) + O_p\left(\frac{1}{\sqrt{n}}\right). \qquad (30)
\end{aligned}
$$

Now let us calculate $(**) = \frac{1}{kn^2}\mathrm{tr}(\mathbf{N\Omega})(\hat{\sigma}^2_{NPb} - \sigma^2_{NP})$ in (20), where $\hat{\sigma}^2_{NPb}$ is defined as in (12). So

$$
\begin{aligned}
&\hat{\sigma}^2_{NPb} - \sigma^2_{NP} \\
&= \frac{1}{n}\sum_{j=1}^{n}\frac{(Y_j - m(x_j))^2}{\omega_j} - \sigma^2_{NP} \\
&\quad + \frac{1}{n}\sum_{j=1}^{n}\frac{(\hat{m}(x_j) - m(x_j))^2}{\omega_j} \\
&\quad + \frac{2}{n}\sum_{j=1}^{n}\frac{(Y_j - m(x_j))(\hat{m}(x_j) - m(x_j))}{\omega_j}.
\end{aligned}
$$

Li (2005) shows that

$$
\frac{1}{n}\sum_{j=1}^{n}\frac{(Y_j - m(x_j))^2}{\omega_j} - \sigma^2_{NP} = O_p\left(\frac{1}{\sqrt{n}}\right), \qquad (31)
$$

$$
\begin{aligned}
&\frac{1}{n}\sum_{j=1}^{n}\frac{(\hat{m}(x_j) - m(x_j))^2}{\omega_j} \\
&= O_p\left(h^{2q+2}\right) + O_p\left(\frac{1}{nh}\right), \qquad (32)
\end{aligned}
$$

and

$$
\begin{aligned}
&\frac{2}{n}\sum_{j=1}^{n}\frac{(Y_j - m(x_j))(\hat{m}(x_j) - m(x_j))}{\omega_j} \\
&= O_p\left(h^{q+1}\right) + O_p\left(\frac{1}{\sqrt{nh}}\right). \qquad (33)
\end{aligned}
$$

Since $\frac{1}{kn^2}\mathrm{tr}(\mathbf{N\Omega}) = O(1)$, and by (31), (32) and (33), we have

$$
\begin{aligned}
&\frac{1}{kn^2}\mathrm{tr}(\mathbf{N\Omega})(\hat{\sigma}^2_{NPb} - \sigma^2_{NP}) \\
&= O_p\left(h^{q+1}\right) + O_p\left(\frac{1}{\sqrt{nh}}\right). \qquad (34)
\end{aligned}
$$

Therefore by (30) and (34),

$$
\begin{aligned}
&\hat{\mathrm{V}}(\bar{Y}_S) - \mathrm{E}_{NP}(\mathrm{Var}_p(\bar{Y}_S)) \\
&= O_p\left(h^{q+1}\right) + O_p\left(\frac{1}{\sqrt{nh}}\right).
\end{aligned}
$$

References

Bartolucci, F. and G. Montanari (2005). A new class of unbiased estimators of the variance of the systematic sample mean. *Journal of Statistical Planning and Inference*. To appear.

Breidt, F. J. and J. D. Opsomer (2000). Local polynomial regresssion estimators in survey sampling. *The Annals of Statistics 28*(4), 1026–1053.

Iachan, R. (1982). Systematic sampling: A critical review. *International Statistical Review 50*, 293–303.

Li, X. (2005). *Model-based variance estimation for systematic sampling*. Ph. D. thesis. Iowa State University. Draft.

Madow, W. G. and L. G. Madow (1944). On the theory of systematic sampling, I. *Annals of Mathematical Statistics 15*, 1–24.

Montanari, G. and F. Bartolucci (1998). On estimating the variance of the systematic sample mean. *Journal of the Italian Statistical Society 7*, 185–196.

Ruppert, D. and M. P. Wand (1994). Multivariate locally weighted least squares regression. *The Annals of Statistics 22*, 1346–1370.

Särndal, C.-E., B. Swensson, and J. H. Wretman (1992). *Model assisted survey sampling.* Springer-Verlag Inc.

Sen, A. R. (1953). On the estimate of the variance in sampling with varying probabilities. *Journal of the Indian Society of Agricultural Statistics 5*, 119–127.

Wand, M.-P. and M.-C. Jones (1995). *Kernel smoothing.* Chapman & Hall Ltd.

Wolter, K. M. (1985). *Introduction to variance estimation.* Springer-Verlag Inc.

Yates, F. and P. M. Grundy (1953). Selection without replacement from within strata with probability proportional to size. *Journal of the Royal Statistical Society. Series B 15*, 253–261.

Zinger, A. (1980). Variance estimation in partially systematic sampling. *Journal of the American Statistical Association 75*, 206–211.