Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 70(3) 2016 255–264

On Estimation of Population Mean under Systematic Sampling in the Presence of a Polynomial Trend

Morteza Amini and Shahen Mohammad Faraj
College of Science, University of Tehran, Tehran, Iran

Received 11 January 2016; Revised 04 April, 2016; Accepted 28 September 2016

SUMMARY

In this paper, we discuss the conditions under which a finite population might exhibit a polynomial trend of order $k \geq 1$. The problem of estimation of a finite population mean in the presense of cubic trend is considered and a corrected estimator is obtained under a general systematic sampling scheme, which unifies the notations of linear, balanced and modified systematic sampling strategies. The performance of the estimator is evaluated using a super-population model approach. Assuming a cubic trend model, the effect of a polynomial trend of higher order is evaluated by comparing the estimators developed under parabolic and cubic trend models. The real data example of grain production in Iran is considered to compare the performance of estimators. The numerical comparisons and real data analysis suggest that estimation of the population mean under the assumption that the population exhibits a cubic trend might result to better performance of the estimators rather than under those of linear or parabolic trends.

Keywords: Empirical distribution function, Grain production, Relative efficiency, Weighted estimator.

1. INTRODUCTION

It is well-known that the sample mean under the linear systematic sampling (LSS) is more efficient, for estimation of a finite population mean, than the corresponding estimator, under simple random sampling (SRS), in the presence of linear trend (see Cochran 1977). There are attempts to improve the sampling design in populations with linear trend, including Madow (1953), Sethi (1965) and Singh *et al.* (1968). Some articles including Singh *et al.* (1968) and Yates (1948) proposed improved estimators for the population mean which coincide with the population mean under a linear trend model. Others consider the improvement of the systematic sampling for population exhibiting parabolic trend, such as Bellhouse and Rao (1975), Agrawal and Jain (1988), Bellhouse (1981), Singh *et al.* (1968), Sampath and Chandra (1991) and Sampath *et al.* (2009). Examples of populations with parabolic trend are discussed in Singh *et al.* (1968) and Bellhouse (1981).

There are two major questions about a population trend. The first question is "Under what conditions a finite population might exhibit a polynomial trend of order $k \geq 1$?" The second question is "What is the link between the knowledge of a supplementary variable which is closely related to the variable of interest and the information pertaining to lables of population units, when the list of population units are ordered by the supplementary variable? In other words, if we use an auxiliary variable to sort the list of population units, how our knowledge about the joint distribution of auxiliary variable and the variable of interest enables us to discover the population trend?". To answer these questions, let us first focus on the distributional properties of the supplementary variable. In a population with equally spaced units with equal frequencies, the ordered values of population units have an exact linear trend. This relation can be formulated by using the fact that for a population constisting

Corresponding author: Morteza Amini
E-mail address: morteza.amini@ut.ac.ir

$x_1, \ldots, x_N$, the i^{th} ordered value of $x_1, \ldots, x_N$, $x_{i:N}$, can be written as

$$x_{i:N} = F_N^{-1}\left(\frac{i}{N}\right),$$

in which

$$F_N(x) = \frac{1}{N}\sum_{i=1}^{N} I(x_i \le x)$$

denotes the empirical cumulative distribution function of $x_1, \ldots, x_N$. Thus, for example, if $x_1, \ldots, x_N$ are equally spaced with equal frequencies, which means that

$$F_N(x_{i:N}) = \frac{x_{i:N} - x_{1:N}}{x_{N:N} - x_{1:N}} = \frac{i}{N},$$

then

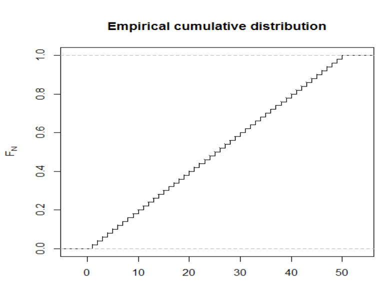
$$x_{i:N} = F_N^{-1}\left(\frac{i}{N}\right) = x_{1:N} + \frac{(x_{N:N} - x_{1:N})}{N} i,$$

that is the ordered values of the supplementary population exhibit a linear trend (see first row of Fig. 1). Now, suppose that the list of the population individuals $y_1, \ldots, y_N$ is ordered by using the auxiliary population $x_1, \ldots, x_N$, that are equally spaced with equal frequencies and the relation between pairs (x_i, y_i), $i = 1, \ldots, N$, is well approximated by a linear function, that is

$$y_i = a + bx_i + e_i,$$

where the values of $e_i = y_i - a - bx_i$, $i = 1, \ldots, N$ are close to zero, then the Y-concomitants $y_{[i:N]}$ of the i^{th} ordered value $x_{i:N}$ of $x_1, \ldots, x_N$ satisfy

$$\begin{aligned} y_{[i:N]} &= a + bx_{i:N} + e_{i:N} \\ &= a + bF_N^{-1}\left(\frac{i}{N}\right) + e_{i:N} \\ &\approx \alpha' + \beta' i. \end{aligned}$$



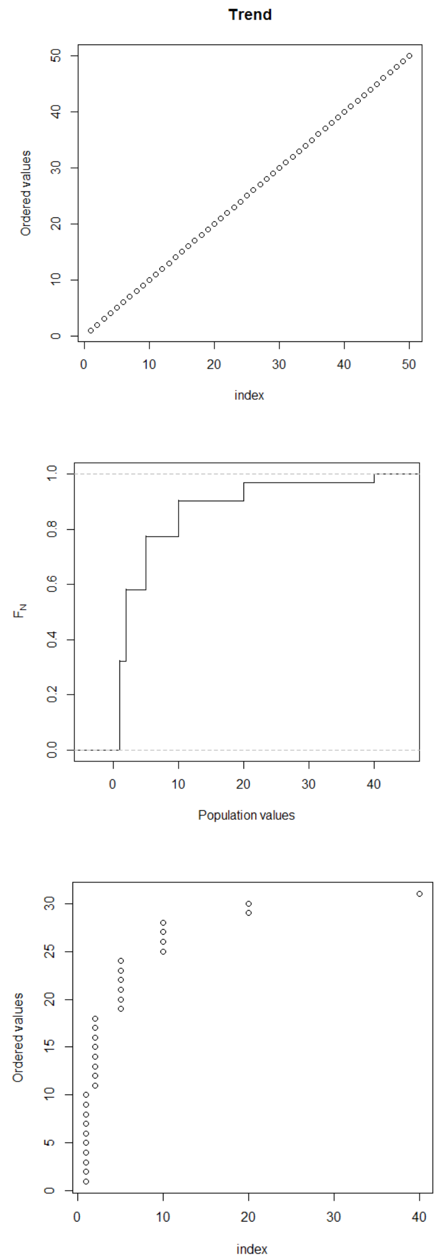


Fig. 1. Relation Between the Distribution of Y and the Population Trend

Alternatively, if the space between the values of $x_1, \ldots, x_N$ is increasing in their index and/or their frequency is decreasing, then the trend of ordered values $x_{i:N}$ is increasing and concave, which is well approximated by a parabolic function (see second row of Fig. 1). Now, suppose that F_N has two tails. Then the trend of F_N^{-1} has a return point and so the trend of the ordered values of the supplementary population can be well approximated (see Fig. 2) by a polynomial trend of order $k \geq 3$, specially by a cubic trend

$$x_{i:N} \simeq \alpha + \beta i + \gamma i^2 + \eta i^3.$$

Therefore, under the model $y_i = a + bx_i + e_i$,

$$y_{[i:N]} = a + bx_{i:N} + e_{i:N}$$
$$\approx \alpha' + \beta' i + \gamma' i^2 + \eta' i^3,$$

that is the trend of the population is parabolic.

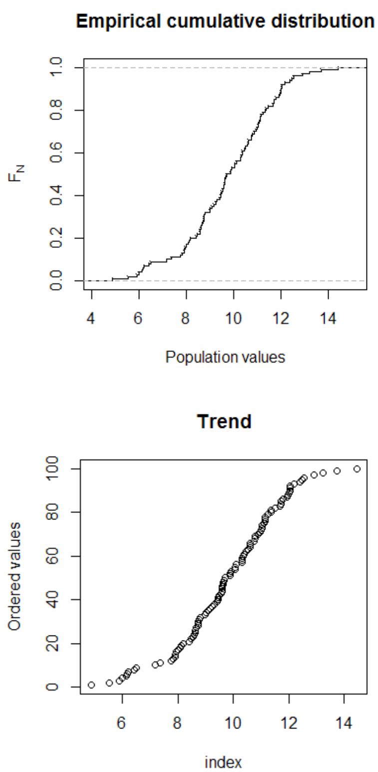


Fig. 2. The Two-Tail Distribution and the Population Trend

The higher order polynomials might also be considered as the population trend whether the relation between pairs (x_i, y_i), $i = 1, ..., N$ is well approximated by polynomials of higher order or the structure of F_N^{-1} is more complicated.

In this paper, the problem of estimation of the population mean is studied for a population with cubic trend to investigate the effect of a polynomial trend on the performannce and efficiency of the estimators under systematic sampling. In Section 2, we develope the corrected estimator of the population mean for the linear systematic sampling (LSS), the modified systematic sampling (MSS) of Singh *et al.* (1968), and for the balance systematic sampling (BSS) of Sethi (1965), under the model $y_i = \alpha + \beta i + \gamma i^2 + \eta i^3$, which coincides with the population mean. Then, using a super-population model approach, the mean square error of the corrected estimator is obtained under the model

$$y_i = \alpha + \beta i + \gamma i^2 + \eta i^3 + e_i, i = 1, \ldots, N,$$

with certain additional assumptions on the errors e_i, $i = 1, ..., N$. A comparison of three sampling schemes is performed in Section 3, with respect to the mean square errors. Furthermore, the effect of using the corrected estimator for the population with parabolic trend, under the population exhibiting a cubic trend is evaluated. Some real data set is considered in Section 4 to compare the performance of the estimators in a real life example.

2. THE CORRECTED ESTIMATOR FOR CUBIC POPULATIONS

To draw a linear systematic sample of size n from a population of size $N = nk$, a random integer $1 \leq r \leq k$ is chosen. The sample is then given by

$$y_{r+jk}, j = 0, ..., n-1$$

For the MSS case, when n is even, the sample corresponding to the random start r is

$$\left(y_{r+jk}, y_{N-r+1-jk}\right), j = 0, \ldots, \frac{n}{2} - 1.$$

The corresponding drawn sample for the BSS case, when n is even, is

$$\left(y_{r+2jk}, y_{2k-r+1+2jk}\right), j = 0, \ldots, \frac{n}{2} - 1.$$

In order to compare LSS, MSS and BSS schemes, we assume throughout the paper that n is even. For the case of n even, the notation of the three schemes can be unified as

$$\left(Y_{r+ajk}, Y_{N_{n,k,r}-ajk}\right), j = 0, \ldots, \frac{n}{2} - 1,$$

where $a = 1$, for MSS and LSS schemes, $a = 2$ for BSS, $N_{n,k,r} = (nk - r + 1)$, for MSS and BSS, and $N_{n,k,r} = r + (n-1)k$, for LSS.

Under the above unified notation, we consider the weighted estimator of the population mean, $\bar{y}_N$, as

$$\bar{y}_{(r)w}^{Cub} = \frac{1}{n}\left[w_1 y_r + w_2 y_{r+a(\frac{n}{2}-1)k} + w_3 y_{N_{n,k,r}-a(\frac{n}{2}-1)k} \right.$$

$$\left. + w_4 y_{N_{n,k,r}} + \sum_{j=1}^{\frac{n}{2}-2} [y_{r+ajk} + y_{N_{n,k,r}-ajk}] \right]. \quad (1)$$

The weights w_i, $i = 1, ..., 4$ are determined so that $\bar{y}_{(r)w} = \bar{y}_N$, under the model

$$y_i = \alpha + \beta i + \gamma i^2 + \eta i^3, i = 1, 2, ..., N. \quad (2)$$

Using the fact that under the model (2), we have

$$\bar{y}_N = \alpha + \beta\frac{(N+1)}{2} + \gamma\frac{(N+1)(2N+1)}{6} + \eta\frac{N(N+1)^2}{4},$$

the following system of equations is obtained:

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ r & \mu(r) & \chi(r) & N_{n,k,r} \\ r^2 & \mu^2(r) & \chi^2(r) & N_{n,k,r}^2 \\ r^3 & \mu^3(r) & \chi^3(r) & N_{n,k,r}^3 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix} = \begin{bmatrix} 4 \\ g_1(N) \\ g_2(N) \\ g_3(N) \end{bmatrix}$$

where

$$\mu(r) = \left(r + a\left(\frac{n}{2} - 1\right)k \right),$$

$$\chi(r) = \left(N_{n,k,r} - a\left(\frac{n}{2} - 1\right)k \right),$$

$$g_1(N) = \frac{n(N+1)}{2} - \left(\frac{n}{2} - 2\right)\left(r + N_{n,k,r}\right),$$

$$g_2(N) = \frac{n(N+1)(2N+1)}{6} - \sum_{j=1}^{\frac{n}{2}-2}\left[\left(r + ajk\right)^2 + \left(N_{n,k,r} - ajk\right)^2\right],$$

and

$$g_3(N) = \frac{nN(N+1)^2}{4} - \sum_{j=1}^{\frac{n}{2}-2}\left[\left(r + ajk\right)^3 + \left(N_{n,k,r} - ajk\right)^3\right].$$

The solution of the above system of equations is

$$w_4 = \frac{g_3(N,r)}{\delta_3(r)}; \quad (3)$$

$$w_3 = \frac{(\delta_3(r)g_2(N,r)) - (\delta_2(r)g_3(N,r))}{\delta_3(r)\chi_2(r)}; \quad (4)$$

$$w_2 = \frac{\begin{matrix}[g_1(N,r)\delta_3(r)\chi_2(r) - g_2(N,r)\delta_3(r)\chi_1(r) \\ +g_3(N,r)\delta_2(r)\chi_1(r) - g_3(N,r)\delta_1(r)\chi_2(r)]\end{matrix}}{\delta_3(r)\chi_2(r)\mu_1(r)} \quad (5)$$

and

$$w_1 = 4 - w_2 - w_3 - w_4, \quad (6)$$

where

$$\mu_1(r) = \mu(r) - r,$$

$$\chi_1(r) = \chi(r) - r,$$

$$\delta_1(r) = N_{n,k,r} - r,$$

$$\chi_2(r) = (\chi^2(r) - r^2) - (\mu(r) + r)\chi_1(r),$$

$$\delta_2(r) = (N_{n,k,r}^2 - r^2) - (\mu(r) + r)\delta_1(r),$$

$$\delta_3(r) = (N_{n,k,r}^3 - r^3) - (\mu^2(r) + r\mu(r) + r^2)$$

$$\delta_1(r) - (\chi(r) + \mu(r) + r)\delta_2(r),$$

$$g_1(N,r) = g_1(N) - 4r,$$

$$g_2(N,r) = g_2(N) - 4r^2 - (\mu(r) + r)g_1(N,r)$$

and

$$g_3(N,r) = g_3(N) - 4r^3 - (\mu^2(r) + r\mu(r) + r^2)$$

$$g_1(N,r) - (\chi(r) + \mu(r) + r)g_2(N,r).$$

Using the weights given in (3)-(6), the mean square error of the estimator in (1) is zero under the model (2). In the sequel, the mean square error of the estimator in (1) is obtained using a super-population model approach. In this approach, the mean square error is calculated by averaging with respect to probability distribution function of y_1, ..., y_N. To model the joint distribution of y_1, ..., y_N, we utilize the approach used in Sampath *et al.* (2009) and Agrawal and Jain (1988), by considering the model

$$y_i = \alpha + \beta i + \gamma i^2 + \eta i^3 + e_i,\ i = 1, \ldots, N, \qquad (7)$$

with $E(e_i) = 0$ and $Cov(e_i, e_j) = 0$, $i \neq j = 1, ..., N$. To include the homoscedastic and heteroscedastic errors in the model, we further assume that $Var(e_i) = \sigma^2 i^g$, where g is a real number. The homoscedasticity is then determined by the case $g = 0$. It is easy to see that

$$\overline{y}_{(r)w}^{\text{Cub}} = \frac{1}{n}\left[w_1 e_r + w_2 e_{r+a(\frac{n}{2}-1)k} + w_3 e_{N_{n,k,r}-a(\frac{n}{2}-1)k} \right.$$

$$\left. + w_4 e_{N_{n,k,r}} + \sum_{j=1}^{\frac{n}{2}-2} (e_{r+ajk} + e_{N_{n,k,r}-ajk}) \right] - \frac{1}{N}\sum_{i=1}^{N} e_i,$$

and therefore

$$\frac{1}{k}\sum_{r=1}^{k} E(\overline{y}_{(r)w}^{Cub} - \overline{y}_N)^2 = \frac{1}{k}\sum_{r=1}^{k}\frac{1}{n^2}$$

$$\left[w_1^2 E\left(e_r^2\right) + w_2^2 E\left(e^2_{r+a(\frac{n}{2}-1)k}\right) + w_3^2 E\left(e^2_{N_{n,k,r}-a(\frac{n}{2}-1)k)}\right) \right.$$

$$\left. + w_4^2 E\left(e^2_{N_{n,k,r}}\right) + \sum_{j=1}^{\frac{n}{2}-2} [E\left(e^2_{r+ajk}\right) + E\left(e^2_{N_{n,k,r}-ajk}\right) \right]$$

$$+ \frac{1}{N^2}\sum_{i=1}^{N} E\left(e_i^2\right) - 2\left(\frac{1}{nN}\right)\left[w_1 E(e_r^2) + w_2 E\left(e^2_{r+a(\frac{n}{2}-1)k}\right) \right.$$

$$+ w_3 E\left(e^2_{N_{n,k,r}-a\left(\frac{n}{2}-1\right)k}\right) + w_4 E\left(e^2_{N_{n,k,r}}\right)$$

$$\left. + \sum_{j=1}^{\frac{n}{2}-2}\left[E\left(e^2_{r+ajk}\right) + E\left(e^2_{N_{n,k,r}-ajk}\right) \right] \right].$$

Consequently,

$$\frac{1}{k}\sum_{r=1}^{k} E\left(\overline{y}_{(r)w}^{Cub} - \overline{y}_N\right)^2$$

$$= \frac{\sigma^2}{k}\sum_{r=1}^{k}\left[W_1 r^g + W_2\left(r + a\left(\frac{n}{2}-1\right)k\right)^g \right.$$

$$+ W_3\left(N_{n,k,r} - a\left(\frac{n}{2}-1\right)k\right)^g + W_4(N_{n,k,r})^g + \frac{N-2n}{Nn^2}$$

$$\left. \sum_{j=1}^{\frac{n}{2}-2}\left[\left(r + ajk\right)^g + \left(N_{n,k,r} - ajk\right)^g \right] + \frac{1}{N^2}\sum_{i=1}^{N} i^g \right], \qquad (8)$$

where $W_i = \left(\left(\frac{w_i}{n}\right)^2 - 2\left(\frac{w_i}{nN}\right)\right), i = 1, \ldots, 4.$

3. COMPARISON

We start with comparing the mean square errors of the estimators under LSS, BSS and MSS sampling strategies. Then, for a population with cubic trend, the estimator developed under the assumption of cubic trend is compared with that developed under the assumption of parabolic trend to measure the effect of polynomial trends of higher order on the performance of the corrected estimators.

In order to compare the precision of LSS, BSS and MSS methods under the model

$$y_i = \alpha + \beta i + \gamma i^2 + \eta i^3 + e_i,,\ i = 1, \ldots, N,$$

the relative efficiencies

$$RE_1 = \frac{\frac{1}{k}\sum_{r=1}^{k} E(\overline{y}_{(r)w}^{\text{Cub}.LSS} - \overline{y}_N)^2}{\frac{1}{k}\sum_{r=1}^{k} E(\overline{y}_{(r)w}^{\text{Cub}.MSS} - \overline{y}_N)^2}$$

and

$$RE_2 = \frac{\frac{1}{k}\sum_{r=1}^{k} E(\overline{y}_{(r)w}^{\text{Cub}.LSS} - \overline{y}_N)^2}{\frac{1}{k}\sum_{r=1}^{k} E(\overline{y}_{(r)w}^{\text{Cub}.BSS} - \overline{y}_N)^2},$$

are computed and given in Table 1 for different values of g, n and k, where $\overline{y}_{(r)w}^{\text{Cub}.Type}$, *Type* = L*SS*, *MSS* and B*SS*, is the estimator in (1) with corresponding values of a and $N_{n,k,r}$ for L*SS*, *MSS*

Table 1. Relative efficiencies of *MSS* and *BSS* with respect to *LSS*.

		g = 0		*g = 1*		*g = 2*		*g = 3*	
k	*n*	RE_1	RE_2	RE_1	RE_2	RE_1	RE_2	RE_1	RE_2
5	12	1.0679	1.0338	1.0679	1.0338	1.0538	1.0024	1.0397	0.9719
	16	1.0471	1.0233	1.0471	1.0233	1.0369	0.9985	1.0266	0.9743
	20	1.0360	1.0178	1.0360	1.0178	1.0279	0.9973	1.0198	0.9773
	24	1.0291	1.0143	1.0291	1.0143	1.0225	0.9970	1.0158	0.9800
	28	1.0244	1.0120	1.0244	1.0120	1.0188	0.9970	1.0131	0.9822
	32	1.0210	1.0104	1.0210	1.0104	1.0161	0.9971	1.0112	0.9840
	36	1.0184	1.0091	1.0184	1.0091	1.0141	0.9972	1.0097	0.9855
8	12	1.0642	0.9822	1.0642	0.9822	1.0508	0.9427	1.0373	0.9051
	16	1.0445	0.9871	1.0445	0.9871	1.0347	0.9546	1.0250	0.9235
	20	1.0339	0.9899	1.0339	0.9899	1.0263	0.9625	1.0186	0.9362
	24	1.0274	0.9917	1.0274	0.9917	1.0211	0.9682	1.0148	0.9454
	28	1.0230	0.9930	1.0230	0.9930	1.0176	0.9724	1.0123	0.9523
	32	1.0198	0.9939	1.0198	0.9939	1.0151	0.9756	1.0105	0.9577
	36	1.0173	0.9947	1.0173	0.9947	1.0132	0.9781	1.0091	0.9620
10	12	1.0629	0.9523	1.0629	0.9523	1.0498	0.9082	1.0365	0.8668
	16	1.0436	0.9657	1.0436	0.9657	1.0340	0.9286	1.0244	0.8935
	20	1.0332	0.9733	1.0332	0.9733	1.0257	0.9416	1.0182	0.9115
	24	1.0268	0.9781	1.0268	0.9781	1.0207	0.9507	1.0145	0.9244
	28	1.0225	0.9815	1.0225	0.9815	1.0173	0.9573	1.0120	0.9341
	32	1.0193	0.9840	1.0193	0.9840	1.0148	0.9624	1.0103	0.9415
	36	1.0170	0.9859	1.0170	0.9859	1.0130	0.9664	1.0089	0.9475
12	12	1.0621	0.9248	1.0621	0.9248	1.0490	0.8766	1.0360	0.8321
	16	1.0429	0.9456	1.0429	0.9456	1.0335	0.9043	1.0241	0.8658
	20	1.0327	0.9575	1.0327	0.9575	1.0254	0.9219	1.0180	0.8884
	24	1.0264	0.9652	1.0264	0.9652	1.0204	0.9341	1.0143	0.9046
	28	1.0221	0.9705	1.0221	0.9705	1.0170	0.9430	1.0118	0.9167
	32	1.0190	0.9744	1.0190	0.9744	1.0146	0.9497	1.0101	0.9261
	36	1.0167	0.9774	1.0167	0.9774	1.0128	0.9551	1.0088	0.9336

and B*SS* schemes, respectively. As it can be seen from Table 1,

(1) The estimator $\overline{y}_{(r)w}^{\text{Cub}.MSS}$ is more efficient than $\overline{y}_{(r)w}^{\text{Cub}.LSS}$ and $\overline{y}_{(r)w}^{\text{Cub}.BSS}$, for all values of n =12(4)36, k = 5, 8, 10, 12 and g = 0, 1, 2, 3.

(2) The values of RE_2 suggest that there exist integers k^* and n^*, both non-increasing in g, such that for $k \geq k^*$ and $n \geq n^*$ $\overline{y}_{(r)w}^{\text{Cub}.LSS}$ is more efficient than $\overline{y}_{(r)w}^{\text{Cub}.BSS}$.

(3) It seems that the values of relative efficiencies tend to 1 as n grows larger.

To evaluate the effect of a polynomial trend of higher order on the performance of the estimators, suppose that the population exhibits a cubic trend

$$y_i = \alpha + \beta i + \gamma i^2 + \eta i^3 + e_i,\ i = 1, \ldots, N,$$

with $E(e_i) = 0$ and $Cov(e_i, e_j) = 0$, $i \neq j = 1, \ldots, N$ and $Var(e_i) = \sigma^2 i^g$, where g is a real number. and that we use the weighted estimator

$$\overline{y}_{(r)w}^{\text{Par}} = \frac{1}{n}\left[\lambda_1 Y_r + \lambda_2\left(Y_{r+a\left(\frac{n}{2}-1\right)k} + Y_{N_{n,k,r}-a\left(\frac{n}{2}-1\right)k}\right) + \lambda_3 Y_{N_{n,k,r}} + \sum_{j=1}^{\frac{n}{2}-2}\left[Y_{r+ajk} + Y_{N_{n,k,r}-ajk}\right]\right], \quad (9)$$

where λ_i, i = 1, 2, 3 are determined such that estimator in (9) coincide with the population mean under the assumption that $y_i = \alpha + \beta i + \gamma i^2$, $i = 1, \ldots, N$.

Remark 1: It is worth noting that the estimator in (9) is different from the estimator proposed in Sampath *et al*. (2009) for the case of BSS in the way that in Sampath et al. (2009) l2 is the weight of the last drawn pair $\left(y_{r+(n-1)k}, y_{nk-r+1}\right)$, while in (9) it is the weight of the middle pair $\left(y_{r+2(\frac{n}{2}-1)k}, y_{nk-r+1-2(\frac{n}{2}-1)k}\right)$.

However, for the sake of comparability with the results of Section 2, we use (9) as the corrected estimator for the parabolic population. Hence, for the case of BSS, the weights *li*, $i = 1, 2, 3$ in (9) are different from the weights obtained in Section 3 of Sampath et al. (2009). Indeed, for the estimator in (9) to coincide with the population mean, under the model $y_i = \alpha + \beta i + \gamma i^2, i = 1, \ldots, N$, we obtain

$$\lambda_3 = \frac{g_2'(N,r)}{\chi_2'(r)}; \tag{10}$$

$$\lambda_2 = \frac{g_1'(N,r)}{\chi_1'(r)} - \frac{g_2'(N,r)}{\chi_2'(r)}; \tag{11}$$

$$\lambda_1 = 4 - 2\lambda_2 - \lambda_3, \tag{12}$$

where

$$\mu'(r) = r + N_{n,k,r},$$

$$\mu_2'(r) = (r + a(n/2-1)k)^2 + (N_{n,k,r} - a(n/2-1)k)^2,$$

$$\chi_1'(r) = N_{n,k,r} - r$$

$$\delta_1'(r) = n(N+1)/2 - (n/2-2)\mu'(r),$$

$$\delta_2'(r) = n(N+1)(2N+1)/6 - \sum_{j=1}^{n/2-2}\left[\left(r + ajk\right)^2 + \left(N_{n,k,r} - ajk\right)^2\right],$$

$$\eta'(r) = \frac{-\mu_2'(r) + 2r^2}{chi_1'(r)},$$

$$\chi_2'(r) = N_{n,k,r}^2 + r^2 - \mu_2'(r),$$

$$g_1'(N,r) = \delta_1'(r) - 4r$$

and

$$g_2'(N,r) = \delta_2'(r) - 4r^2 + \eta'(r)g_1'(N,r).$$

From the computation results, the following result arises. Unfortunately, the proof of this result is very tedious and untractable and we could not handle it.

Table 2: The values of $\sigma\sqrt{\dfrac{\frac{1}{k}\sum_{r=1}^{k} A_r^2}{E_{\mathrm{Par}}(\overline{y}_{(r)w}^{\mathrm{Par}.LSS} - \overline{y}_N)^2}}$ for different values of *n, k* and *g*.

k	*n*	*g = 0*	*g = 1*	*g = 2*
5	12	3004.317	543.996	85.361
	20	10945.243	1540.209	187.739
	28	25544.682	3042.327	313.824
	36	48045.137	5050.393	459.788
8	12	5781.699	830.204	103.168
	20	21109.608	2352.785	226.965
	28	49313.805	4649.350	379.441
	36	92800.231	7719.961	555.964
10	12	7952.366	1022.395	113.704
	20	29056.145	2898.378	250.168
	28	67899.205	728.310	418.249
	36	127797.595	9512.268	612.841
12	12	10345.345	1214.999	123.398
	20	37817.989	3445.124	271.515
	28	88392.925	6809.537	453.951
	36	166389.990	11308.324	665.164

Conjecture 1: The estimator in (9) coincide with the estimator in (1) for the MSS and the BSS strategies and for $n \geq 6$($N = nk$, n is even), in the way that

$w_3 = \lambda_3$ and $w_2 = w_1 = \lambda_2$.

For the LSS case, it is easy to see that

$$\frac{\frac{1}{k}\sum_{r=1}^{k} E_{\mathrm{Cub}}(\overline{y}_{(r)w}^{\mathrm{Cub}.LSS} - \overline{y}_N)^2}{\frac{1}{k}\sum_{r=1}^{k} E_{\mathrm{Cub}}(\overline{y}_{(r)w}^{\mathrm{Par}.LSS} - \overline{y}_N)^2} = \frac{1}{\dfrac{\frac{1}{k}\sum_{r=1}^{k} E_{\mathrm{Par}}(\overline{y}_{(r)w}^{\mathrm{Par}.LSS} - \overline{y}_N)^2}{\frac{1}{k}\sum_{r=1}^{k} E_{\mathrm{Cub}}(\overline{y}_{(r)w}^{\mathrm{Cub}.LSS} - \overline{y}_N)^2} + \eta^2 \dfrac{\frac{1}{k}\sum_{r=1}^{k} A_r^2}{\frac{1}{k}\sum_{r=1}^{k} E_{\mathrm{Cub}}(\overline{y}_{(r)w}^{\mathrm{Cub}.LSS} - \overline{y}_N)^2}},$$

where $\overline{y}_{(r)w}^{Cub.LSS}$ and $\overline{y}_{(r)w}^{\mathrm{Par}.LSS}$ are given in (1) and (9) with $a = 1$ and $N_{n,k,r} = r + (n-1)k$, respectively, E_{Cub} and E_{Par} denote the expectation under the model (7) and the model

$$y_i = \alpha + \beta i + \gamma i^2 + e_i, i = 1, \ldots, N,$$

respectively, and

$$A_r = \lambda_1 r^3 + \lambda_2 \left[\left(r + \left(\frac{n}{2} - 1\right)k\right)^3\right.$$

$$+\left(r+(n-1)k-\left(\frac{n}{2}-1\right)k\right)^3\Bigg]+\lambda_3(r+(n-1)k)^3$$

$$+\sum_{j=1}^{\frac{n}{2}-2}[(r+jk)^3)+(r+(n-1)k-jk)^3]-\frac{N(N+1)^2}{4}.$$

Therefore

$$\frac{E_{Cub}(\bar{y}_{(r)w}^{Cub.LSS}-\bar{y}_N)^2}{E_{Cub}(\bar{y}_{(r)w}^{Par.LSS}-\bar{y}_N)^2}<1,$$

for the values of η satisfying

$$|\eta|>\sqrt{\frac{\frac{1}{k}\sum_{r=1}^{k}A_r^2}{E_{\mathrm{Par}}\left(\bar{y}_{(r)w}^{\mathrm{Par}.LSS}-\bar{y}_N\right)^2}}.$$

Table 2 presents the values of $\sigma\sqrt{\frac{\frac{1}{k}\sum_{r=1}^{k}A_r^2}{E_{\mathrm{Par}}(\bar{y}_{(r)w}^{\mathrm{Par}.LSS}-\bar{y}_N)^2}}$ for different values of *n, k* and *g*. It should be kept in mind that as the lower bound of |η| increases, the effect of the cubic trend decreases. So, as it can be seen from Table 2, the effect of the cubic trend decreases as *n* and *k* increases. Also, under hetroscedastic error models the cubic trend is more effective.

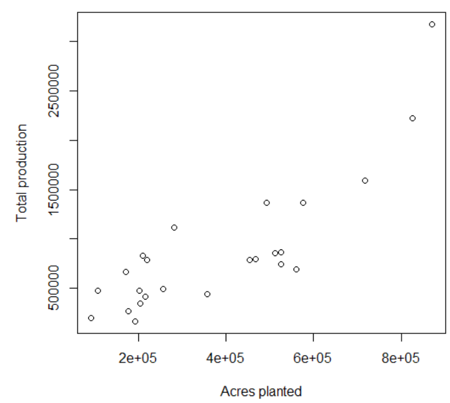
4. A REAL DATA EXAMPLE

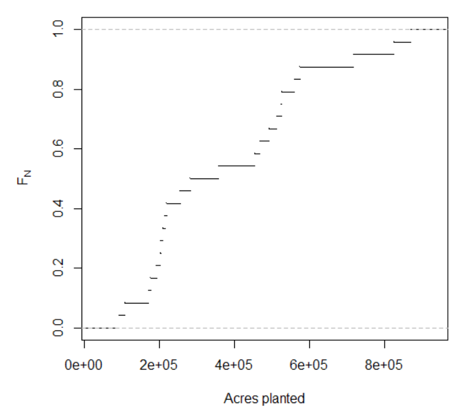
We use the real data set of the grain production in 24 non dry states of Iran for the cropping year 2004/05. The acres planted and total production of grain are given in Table 3 for all 24 non dry states in 2004/05 cropping season. Suppose that the variable of interest is the total production and the aim is to estimate the population mean. Assume further that the list of states is ordered by the acres planted variable.

Fig. 3 (top) shows the scatter plot of total production versus acres planted. The relation between two variables might be modeled by a linear model and the variance of the errors of the linear model is increasing in acres planted. The empirical distribution function of the auxiliary variable is shown in Fig. 3 (top-second). The distribution of the acres planted variable has two tails. Also, the plot of total production in the ordered list versus the index of the list units is shown in Fig. 3 (bottom).

Table 3. The acres planted and total production of grain for 24 non dry states of iran in 2004/05 cropping season

State	Acres Planted (Acre)	Total Production (Tons)
Fars	870730	3176283
Khuzestan	825003	2223741
Khorasan Razavi	717043	1591215
Kermanshah	574852	1364382
Golestan	491980	1362293
Mazandaran	281919	1118081
East Azarbayejan	525111	867414
Hamedan	512719	852208
Isfahan	209101	832006
Ardabil	467854	793168
West Azarbayejan	454437	788573
Gilan	219579	784868
Lorestan	525829	740919
Kordestan	559556	686907
Kerman	170874	663903
Markazi	254924	486280
Kazvin	202424	473079
Tehran	107436	469668
Zanjan	357013	439254
North Khorasan	214775	409167
Ealam	203223	341997
Kohkiluye & Boyerahmad	176088	268816
Chaharmahal & Bakhtyari	91539	199007
Bushehr	191808	163964





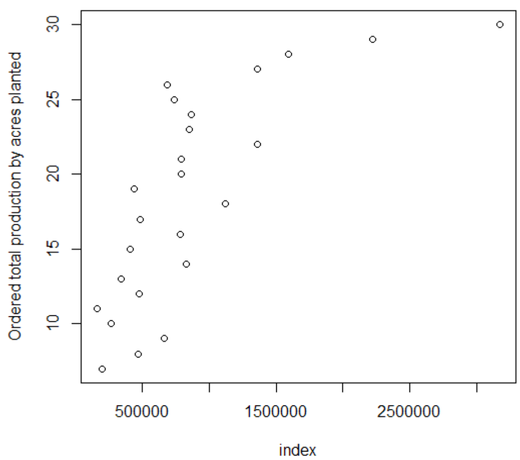


Fig. 3. The scatter plot (top), the empirical distribution function of auxiliary variable (top-second) and the trend of ordered total production by acres planted (bottom) for grain data set

As, it is claimed in Conjecture 1 the estimator $\overline{y}_{(r)w}^{Cub.type}$ coincides with $\overline{y}_{(r)w}^{Par.type}$ for *type* = *MSS* and *BSS* for n = 6, 8 and 12. In order to compare the estimators under different sampling strategies and different trend models the following quantities are computed

$$RE_1 = \frac{\frac{1}{k}\sum_{r=1}^{k}(\overline{y}_{(r)w}^{Par.LSS} - \overline{y}_N)^2}{\frac{1}{k}\sum_{r=1}^{k}(\overline{y}_{(r)w}^{Cub.LSS} - \overline{y}_N)^2},$$

$$RE_2 = \frac{\frac{1}{k}\sum_{r=1}^{k}(\overline{y}_{(r)w}^{Lin.LSS} - \overline{y}_N)^2}{\frac{1}{k}\sum_{r=1}^{k}(\overline{y}_{(r)w}^{Par.LSS} - \overline{y}_N)^2},$$

$$RE_3 = \frac{\frac{1}{k}\sum_{r=1}^{k}(\overline{y}_{(r)w}^{Lin.MSS} - \overline{y}_N)^2}{\frac{1}{k}\sum_{r=1}^{k}(\overline{y}_{(r)w}^{Par.MSS} - \overline{y}_N)^2},$$

$$RE_4 = \frac{\frac{1}{k}\sum_{r=1}^{k}(\overline{y}_{(r)w}^{Lin.BSS} - \overline{y}_N)^2}{\frac{1}{k}\sum_{r=1}^{k}(\overline{y}_{(r)w}^{Par.BSS} - \overline{y}_N)^2},$$

$$RE_5 = \frac{\frac{1}{k}\sum_{r=1}^{k}(\overline{y}_{(r)w}^{Cub.LSS} - \overline{y}_N)^2}{\frac{1}{k}\sum_{r=1}^{k}(\overline{y}_{(r)w}^{Cub.MSS} - \overline{y}_N)^2}, \quad \text{and}$$

$$RE_6 = \frac{\frac{1}{k}\sum_{r=1}^{k}(\overline{y}_{(r)w}^{Cub.LSS} - \overline{y}_N)^2}{\frac{1}{k}\sum_{r=1}^{k}(\overline{y}_{(r)w}^{Cub.BSS} - \overline{y}_N)^2},$$

where the superscripts Cub, Par and Lin stand for the cubic, parabolic and linear trend models, respectively, and the superscripts LSS, MSS and BSS stand for the corresponding sampling strategies. The relative efficiencies RE_i, i = 1, ..., 6 are given in Table 4, for n = 6, 8 and 12. As one can see from Table 4, the corrected LSS estimators under cubic trend model is more efficient than that under parabolic trend model for n = 6, 8 and 12. The corrected LSS estimator under parabolic trend model is more efficient than that under linear trend model for n = 6, 8 and 12, while, for the MSS case, this holds only for n = 6 and 8 and for the BSS case it holds only for n = 6 and 12. Furthermore, under cubic trend model, the corrected LSS estimators is less efficient than the corrected MSS estimator for n = 6, 8 and 12, while it is more efficient than the corrected BSS estimator for n = 6 and 8.

Table 4. Relative efficiencies of cubic estimator relative to the parabolic estimator for different sampling schemes and values of n for the grain data set

n	RE_1	RE_2	RE_3	RE_4	RE_5	RE_6
6	1.743	3.248	4.732	2.221	6.769	0.755
8	3.634	3.000	3.479	0.819	1.856	0.940
12	1.488	2.315	0.377	3.533	10.793	34.244

5. CONCLUDING REMARKS

As a result of numerical comparisons and real data analysis, we suggest using the estimator $\overline{y}_{(r)w}^{Cub.MSS}$ rather than $\overline{y}_{(r)w}^{Cub.LSS}$ and $\overline{y}_{(r)w}^{Cub.BSS}$. Furthermore, the numerical comparisons and real data analysis suggest that estimation of the population mean under the assumption that the population exhibits a cubic trend might result to better performance of the estimators rather than under those of linear or parabolic trends.

It is well-known that under the ordinary systematic sampling the second order inclusion probabilities are zero for several pairs of units which makes variance estimation difficult. To overcome this problem, Tukey (1950) and later on Gautschi (1957) suggested using multiple random starts for systematic sampling. The problem of estimation of population mean under systematic sampling with multiple random starts in the presence of a cubic trend remains as a future work.

ACKNOWLEDGEMENT

The authors would like to thank an anonymous referee for his/her comments and suggestions on an earlier version of this manuscript which resulted in this improved version.

REFERENCES

Agrawal, M.C. and Jain, N. (1988). Comparison of some sampling strategies in the presence of trend. *J. Ind. Soc. Agril. Statist.*, **40**,191-201.

Bellhouse, D.R. (1981). Spatial sampling in the presence of a trend. *J. Statist. Plann. Infer.*, **5**, 365-375.

Bellhouse, D.R. and Rao, J.N.K. (1975). Systematic sampling in the presence of trend. *Biometrika*, **62**, 694-697.

Cochran, W.G. (1977). *Sampling Techniques*. Wiley Eastern Limited, New Delhi.

Gautschi, W. (1957). Some remarks on systematic sampling. *Ann. Math. Statist.*, **28**, 385-394.

Madow, W.G. (1953). On the theory of systematic sampling III. *Ann. Math. Statist.*, **24**, 101-106.

Sampath, S. and Chandra, S.K. (1991). Estimation of population mean in the presence of parabolic trend. *J. Ind. Soc. Agril. Statist.*, **43**, 321-328.

Sampath S., Varalakshmi, V. and Geetha, R. (2009). Estimation under systematic sampling schemes for parabolic populations. *J. App. Statist.*, **36**, 1281-1292.

Sethi, V.K. (1965). On optimum pairing of units. *Sankhya B*, **27**, 315-320.

Singh, P., Jindal, K.K. and Garg, J.N. (1968). On modified systematic sampling. *Biometrika*, **55**, 541-546.

Tukey, J.W. (1950). Some sampling simplified. *J.. Amer. Statist. Assoc.*, **45**, 501-519.

Yates, F. (1948). Systematic sampling. *Philos. Trans. R. Soc. Lond. Ser. A*, **241**, 345-377.

Iran Ministry of Agriculture, Summary Statistics of 2004/05 Cropping Year, Volume 1, 2005, available at: http://www.maj.ir/portal/File/ShowFile.aspx?ID=437e1a37-525b-453f-b4af-185c42758ef5.