



A Joint Regression Estimator for Population Mean under Double Sampling with Nonresponse and Two Auxiliary Variables

Pranjal Kaser¹, Dharmendra Kumar Gangeshwer²

¹SOS in Statistics, Pt. Ravishankar Shukla University, Raipur, Chhattisgarh

Corresponding Author Email: kaserpranjal@email.com | ORCID: <https://orcid.org/0009-0002-5976-1106>

How to Cite this Article:

Kaser, P. & Gangeshwer, D. K. (2026). A Joint Regression Estimator for Population Mean under Double Sampling with Nonresponse and Two Auxiliary Variables. International Journal of Creative and Open Research in Engineering and Management, 2(10), 1-9.
<https://doi.org/10.55041/ijcope.v2i10.021>

License:

This article is published under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

© The Author(s). Published by International Journal of Creative and Open Research in Engineering and Management.



<https://doi.org/10.55041/ijcope.v2i10.021>

Abstract-

In sample surveys, nonresponse and the high cost of collecting information on the study variable may substantially affect the efficiency of population mean estimation. Double sampling provides a useful framework in which auxiliary information can be obtained from a relatively large first-phase sample, whereas detailed information on the study variable is collected from a smaller second-phase sample. The presence of nonresponse in the second phase, however, introduces an additional source of variation and requires an appropriate follow-up procedure. The present study proposes a joint regression estimator for estimating the finite-population mean under double sampling with nonresponse using two auxiliary variables. The proposed estimator combines the Hansen–Hurwitz nonresponse-adjusted mean of the study variable with regression corrections based on the differences between first- and second-phase auxiliary means. The two auxiliary variables are incorporated jointly through their variance-covariance structure. First-order bias and mean squared error (MSE) are derived, and the optimum regression coefficients are obtained by minimizing the first-order MSE. The resulting coefficients correspond to the joint regression coefficients based on the inverse of the auxiliary covariance matrix. A second-order bias expression is also developed symbolically. An empirical illustration based on a population of 1,023 workers, with tax as the study variable, grade level as the first auxiliary variable, and gross payment as the second auxiliary variable, is presented. The proposed estimator is evaluated for follow-up factors $k=2,3,4,5$. The

calculated first-order MSEs of the proposed estimator are 41,787.54, 43,572.55, 41,815.78 and 48,922.39, respectively. These values are compared with source-reported MSEs of existing estimators. The results demonstrate the potential of joint regression adjustment for utilizing correlated auxiliary information under double sampling with nonresponse.

Keywords- Double sampling, Nonresponse, Joint regression estimator, Hansen–Hurwitz estimator, Population mean,

1. INTRODUCTION

The estimation of a population mean in a finite-population framework becomes particularly intricate when the study variable entails substantial measurement costs, is difficult to observe, or is affected by nonresponse. Under such circumstances, the incorporation of auxiliary information can substantially enhance the efficiency and precision of the resulting estimators. Double sampling, commonly designated as two-phase sampling, constitutes an effective sampling strategy in which auxiliary

characteristics are observed from a comparatively extensive first-phase sample, whereas the relatively costly study variable is measured only on a smaller second-phase subsample. In the presence of nonresponse during the second phase, an additional follow-up subsample of the nonresponding units can be selected to recover the unavailable information on the study variable.

The seminal contribution of Hansen and Hurwitz (1946) established a systematic framework for



mitigating the adverse consequences of nonresponse through the subsampling of nonrespondents. Subsequent research has produced a broad spectrum of ratio-, product-, and exponential-type estimators designed to exploit auxiliary information for more efficient estimation of the finite-population mean. Bahl and Tuteja (1991) proposed exponential ratio- and product-type estimators, thereby providing an alternative mechanism for incorporating auxiliary information into point estimation. Subsequently, Singh, Kumar and Chaudhary (2009) extended the exponential estimation framework to sampling situations characterized by nonresponse. Hazra (2015) investigated ratio-type estimation under double sampling in the presence of nonresponse, while further methodological advancements were subsequently reported by Ozel Kadilar (2016), Ünal and Kadilar (2020), Abifade (2020), Khan and Khan (2022), and Oguagbaka et al. (2024). More recently, Faweya et al. (2026) developed an improved exponential ratio-type estimator incorporating two auxiliary variables within a double-sampling framework subject to nonresponse.

Despite the demonstrated utility of exponential and ratio-type estimators, the incorporation of multiple auxiliary variables introduces an additional methodological consideration, namely, the interrelationship and covariance structure existing among the auxiliary characteristics. In situations where auxiliary variables exhibit substantial mutual association, estimating their contributions independently may fail to exploit the collective predictive information embedded within their joint covariance structure. Consequently, a methodology capable of simultaneously accounting for these interdependencies may provide a more comprehensive utilization of auxiliary information.

The main objectives of the study are:

1. To develop a joint regression estimator for the population mean under double sampling with nonresponse.
2. To derive the first-order bias and MSE of the proposed estimator.
3. To determine the optimum joint regression coefficients.
4. To establish the minimum first-order MSE.

5. To evaluate the proposed estimator using a real-life population and compare its MSE with existing estimators.

II. SAMPLING DESIGN AND NOTATION

Consider a finite population $U = \{1, 2, \dots, N\}$. A first-phase sample s_1 of size n_1 is selected from U by simple random sampling without replacement (SRSWOR), and information on two auxiliary variables X_1 and X_2 is collected. Subsequently, a second-phase sample s_2 of size n is selected from s_1 , where $n < n_1$, and the study variable Y is observed for the selected units. Let n_r and n_{nr} denote the numbers of respondents and nonrespondents in the second-phase sample, respectively, such that $n_r + n_{nr} = n$. A follow-up subsample of the nonrespondents is selected using inverse sampling fraction k to obtain the missing information on Y . Following the Hansen–Hurwitz procedure, the nonresponse-adjusted estimator of the study-variable mean is given by $\bar{y}^* = w_r \bar{y}_r + w_{nr} \bar{y}_{nr}^*$, where $w_r = n_r/n$ and $w_{nr} = n_{nr}/n$, while $\bar{y}_{nr}^*$ represents the estimated mean of the nonrespondents obtained from the follow-up subsample. The relevant sampling fractions are defined as $f = n/n_1$ and $f' = n_1/N$, with the corresponding variance components given by $\theta = (1 - f)/n$, $\theta' = (1 - f')/n_1$, and $\theta^* = w_{nr}(k - 1)/n$. Here, W_{nr} denotes the population proportion of nonrespondents. Let $\bar{Y}$, $\bar{X}_1$, and $\bar{X}_2$ denote the finite-population means of Y , X_1 , and X_2 , respectively, and let S_Y^2 , $S_{X_1}^2$, and $S_{X_2}^2$ denote their corresponding variances, while S_{YX_1} , S_{YX_2} , and $S_{X_1X_2}$ represent the corresponding covariance terms. The covariance matrix of the two auxiliary variables is expressed as $\Sigma_X = \begin{pmatrix} S_{X_1}^2 & S_{X_1X_2} \\ S_{X_1X_2} & S_{X_2}^2 \end{pmatrix}$, with determinant $|\Sigma_X| = S_{X_1}^2 S_{X_2}^2 - S_{X_1X_2}^2$. The condition $|\Sigma_X| > 0$ guarantees a nonsingular joint covariance structure for X_1 and X_2 , thereby ensuring that the joint regression coefficients used in the proposed estimator are uniquely determined.

III. Proposed Joint Regression Estimator

The proposed estimator of the population mean is

$$\bar{y}_{PR} = \bar{y}^* + \hat{B}_x(\bar{x}' - \bar{x}^*) + \hat{B}_z(\bar{z}' - \bar{z})$$

where $\bar{y}^*$ is the Hansen–Hurwitz nonresponse-adjusted mean and $\hat{B}_x, \hat{B}_z$ are the joint regression coefficients.



For the first-order derivation, the supplied document operationalizes $\bar{x}^* = \bar{x}$.

The estimator is additive and differs structurally from exponential ratio-type estimators because it does not require an exponential transformation or an additional mixing constant.

IV. JOINT REGRESSION COEFFICIENTS

The joint regression coefficient vector is obtained from

$$\begin{pmatrix} B_x \\ B_z \end{pmatrix} = \begin{pmatrix} S_x^2 & S_{xz} \\ S_{xz} & S_z^2 \end{pmatrix}^{-1} \begin{pmatrix} S_{yx} \\ S_{yz} \end{pmatrix}.$$

Therefore,

$$B_x = \frac{S_{yx}S_z^2 - S_{yz}S_{xz}}{S_x^2S_z^2 - S_{xz}^2}$$

and

$$B_z = \frac{S_{yz}S_x^2 - S_{yx}S_{xz}}{S_x^2S_z^2 - S_{xz}^2}$$

The presence of S_{xz} in both coefficients demonstrates that the two auxiliary variables are incorporated jointly rather than independently.

V. FIRST ORDER BIAS

Let $e_y = \bar{y}^* - \bar{Y}$,

and define $\delta_x = \bar{x}' - \bar{x}$, $\delta_z = \bar{z}' - \bar{z}$.

Then $\bar{Y}_{PR} - \bar{Y} = e_y + B_x\delta_x + B_z\delta_z$.

Consequently,

$$\begin{aligned} Bias(\bar{Y}_{PR}) &= E(\bar{Y}_{PR} - \bar{Y}) \\ &= E(e_y) + B_xE(\delta_x) + B_zE(\delta_z) \\ &= 0. \end{aligned}$$

Thus, $Bias_1(\bar{Y}_{PR}) = 0$

This result is conditional on the unbiasedness of the Hansen–Hurwitz nonresponse-adjusted mean and on treating the regression coefficients as fixed population coefficients.

VI. First-Order Mean Squared Error

The relevant first-order moment expressions are

$$\begin{aligned} Var(e_y) &= \bar{Y}^2(\theta C_y^2 + \theta^* C_{y2}^2), \\ Var(\delta_x) &= (\theta - \theta')S_x^2, \\ Var(\delta_z) &= (\theta - \theta')S_z^2, \end{aligned}$$

$$Cov(\delta_x, \delta_z) = (\theta - \theta')S_{xz},$$

$$\begin{aligned} And \quad Cov(e_y, \delta_x) &= -(\theta - \theta')S_{yx}, \\ Cov(e_y, \delta_z) &= -(\theta - \theta')S_{yz}. \end{aligned}$$

Therefore,

$$\begin{aligned} MSE(B_x, B_z) &= \bar{Y}^2(\theta C_y^2 + \theta^* C_{y2}^2) \\ &\quad + (\theta - \theta')[B_x^2S_x^2 + B_z^2S_z^2 + 2B_xB_zS_{xz} \\ &\quad - 2B_xS_{yx} - 2B_zS_{yz}]. \end{aligned}$$

$$MSE(B_x, B_z) = V_0 + (\theta - \theta')(\beta^T \Sigma_a \beta - 2\beta^T c)$$

Where,

$$\begin{aligned} V_0 &= \bar{Y}^2(\theta C_y^2 + \theta^* C_{y2}^2), \\ \beta &= \begin{pmatrix} B_x \\ B_z \end{pmatrix}, \quad c = \begin{pmatrix} S_{yx} \\ S_{yz} \end{pmatrix}. \end{aligned}$$

This quadratic MSE expression is the principal theoretical result of the proposed estimator.

VII. OPTIMUM REGRESSION COEFFICIENTS

Differentiating the MSE with respect to B_x and B_z ,

$$\frac{\partial MSE}{\partial B_x} = 0, \quad \frac{\partial MSE}{\partial B_z} = 0.$$

This produces the normal equations

$$\begin{aligned} S_x^2B_x + S_{xz}B_z &= S_{yx}, \\ S_{xz}B_x + S_z^2B_z &= S_{yz}. \end{aligned}$$

In matrix form,

$$\Sigma_a \beta = c.$$

Hence,

$$\beta^{opt} = \Sigma_a^{-1}c$$

The Hessian matrix is

$$H = 2(\theta - \theta') \begin{pmatrix} S_x^2 & S_{xz} \\ S_{xz} & S_z^2 \end{pmatrix}.$$

When

$$\theta > \theta'$$

and

$$\Delta = S_x^2S_z^2 - S_{xz}^2 > 0,$$



the Hessian is positive definite, confirming that the stationary point corresponds to a minimum of the first-order MSE.

VIII. MINIMUM MSE

Substituting the optimum coefficients into the MSE expression gives

$$MSE_{\min} = V_0 - (\theta - \theta')c^T \Sigma_a^{-1}c$$

Or
$$MSE_{\min} = \bar{Y}^2(\theta C_y^2 + \theta^* C_{y2}^2) - (\theta - \theta')c^T \Sigma_a^{-1}c.$$

Therefore, the improvement relative to the Hansen–Hurwitz nonresponse-adjusted baseline is represented by

$$MSE(\bar{y}^*) - MSE_{\min} = (\theta - \theta')c^T \Sigma_a^{-1}c.$$

The magnitude of the reduction depends on the joint covariance between the study variable and the two auxiliary variables.

IX. Empirical Study

The empirical illustration uses the real-life population consists of 1,023 workers from Akure South Local Government, Nigeria. The study variable is tax, while grade level and gross payment are used as the two auxiliary variables.

The first-phase sample size is

$$n' = 600,$$

and the second-phase sample size is

$$n = 450.$$

A 5% nonresponse setting is considered, with follow-up factors

$$k = 2,3,4,5.$$

Table 1. Empirical Setting

Quantity	Value
Population size N	1,023
Study variable y	Tax
Auxiliary variable x	Grade level
Auxiliary variable z	Gross payment
First-phase sample n'	600
Second-phase sample n	450

Nonresponse	5%
Follow-up factors	2, 3, 4, 5
$\bar{Y}$	5,528.6505
$\bar{X}$	9.03
$\bar{Z}$	87,114.2835
S_y	7,176.68099
S_x	3.259
S_z	63,589.33128
S_{yx}	15,958.438
S_{yz}	450,727,541.1
S_{xz}	150,959.333

Table 1 presents the empirical setting considered for evaluating the proposed estimator. The study is based on a population of 1,023 workers, where **Tax** is taken as the study variable and **Grade Level** and **Gross Payment** are used as two auxiliary variables. A first-phase sample of 600 units and a second-phase sample of 450 units are considered. A 5% nonresponse rate is assumed, with follow-up factors of 2, 3, 4, and 5. This setting provides the basis for examining the performance of the proposed estimator under double sampling with nonresponse and two auxiliary variables.

Table 2. MSE comparison of estimators

k	Prop osed Y_{PR}	Y_{FA1}	Y_{Set}	Y_H	Y_{KK}	Y_{Oget}	Y_{HH}
2	41,7 87.5 4	54,6 01.3 5	61,2 83.4 4	72,6 48.3 1	58,4 63.7 2	61,7 02.4 8	75,9 16.4 8
3	43,5 72.5 5	56,4 82.1 9	63,1 52.7 6	74,9 02.5 6	60,8 94.3 6	63,5 94.2 1	78,3 04.1 2
4	41,8 15.7 8	58,7 74.6 1	65,3 08.4 1	77,3 81.9 4	63,8 11.4 9	66,1 04.7 3	80,8 92.7 7
5	48,9 22.3 9	61,9 02.4 4	67,8 19.2 2	79,9 42.8 7	67,2 95.1 8	68,3 37.9 0	83,7 46.5 5

Table 2 presents the comparison of the Mean Squared Error (MSE) of the proposed estimator with the



existing estimators for different follow-up factors. The MSE of the proposed estimator is 41,787.54, 43,572.55, 41,815.78, and 48,922.39 for follow-up factors 2, 3, 4, and 5, respectively. The proposed estimator shows lower MSE values than the corresponding comparison estimators reported in the table. These results indicate that the joint incorporation of the two auxiliary variables along with the Hansen–Hurwitz nonresponse adjustment contributes to reducing the estimation error.

Table 3. Percentage Relative Efficiency

k	Y_{FA1}	Y_{Set}	Y_H	Y_{KK}	Y_{Oget}	Y_{HH}
2	130.6 6	146.6 5	173.8 5	139.9 1	147.6 6	181.6 7
3	129.6 3	144.9 4	171.9 0	139.7 5	145.9 5	179.7 1
4	140.5 6	156.1 8	185.0 5	152.6 0	158.0 9	193.4 5
5	126.5 3	138.6 3	163.4 1	137.5 5	139.6 9	171.1 8

Table 3 presents the Percentage Relative Efficiency (PRE) results for the proposed estimator under different follow-up factors. The reported PRE values vary across the follow-up settings, reflecting the relative efficiency of the proposed estimator compared with the corresponding estimators. The results demonstrate the contribution of jointly incorporating the two auxiliary variables in the estimation procedure. The joint regression structure utilizes the covariance relationship between the auxiliary variables and thereby incorporates their combined information into the estimation of the population mean.

X. DISCUSSION

The theoretical results show that the proposed estimator integrates three important components: double sampling, Hansen–Hurwitz nonresponse adjustment, and joint utilization of two auxiliary variables. The first-phase sample supplies auxiliary information at relatively large sample size, while the second phase supplies information on the study variable.

The regression correction operates through the phase differences

$$\bar{x}' - \bar{x}$$

and

$$\bar{z}' - \bar{z}.$$

Because these differences are correlated with the study-variable estimation error, the regression adjustment can reduce the first-order MSE. The cross-covariance S_{xz} is explicitly incorporated into the optimum coefficients and therefore allows the joint contribution of the auxiliary variables to be represented.

The nonresponse component remains important. The term

$$\theta^* = \frac{W_2(k-1)}{n'}$$

increases with k , indicating greater uncertainty associated with a smaller follow-up subsample. Thus, although the regression correction can reduce the total MSE, it does not eliminate the statistical consequences of nonresponse.

The empirical results show that the proposed estimator produces MSE values of approximately 41,787.54 to 48,922.39 across the four follow-up settings. The corresponding coefficients of variation reported in the supplied chapter are approximately 3.697%, 3.776%, 3.699% and 4.001%.

XI. CONCLUSION

This study developed a joint regression estimator for estimating the finite-population mean under double sampling with nonresponse and two auxiliary variables. The proposed estimator combines the Hansen–Hurwitz nonresponse-adjusted study-variable mean with regression corrections based on the differences between first- and second-phase auxiliary means.

The first-order bias of the proposed estimator is zero under the stated design-based approximation when the regression coefficients are treated as fixed population quantities. The first-order MSE was derived as a quadratic function of the two regression coefficients. Minimization of this quadratic form yielded the joint



regression coefficients obtained from the inverse of the auxiliary covariance matrix.

The theoretical analysis demonstrates that the covariance between the two auxiliary variables plays an important role in determining their joint contribution. When the auxiliary covariance matrix is positive definite and $\theta > \theta'$, the Hessian matrix is positive definite and the resulting coefficients correspond to the minimum first-order MSE.

For the empirical illustration involving 1,023 workers, tax as the study variable, grade level and gross payment as auxiliary variables, the estimated joint regression coefficients were approximately -174.1125 and 0.1179669 . The calculated first-order MSEs of the proposed estimator were $41,787.54$, $43,572.55$, $41,815.78$ and $48,922.39$ for $k = 2, 3, 4, 5$, respectively. In the supplied numerical comparison, these MSE values were below the corresponding source-reported MSE values of the comparison estimators.

Overall, the study provides a theoretical framework for incorporating two correlated auxiliary variables jointly within a double-sampling design with nonresponse. Further empirical validation using unit-level datasets and simulation studies would be useful for examining the estimator under different sample sizes, nonresponse rates, correlation structures and population distributions.

REFERENCES

1. Bahl, S., & Tuteja, R. K. (1991). Ratio and product type exponential estimator. *Information and Optimization Sciences*, 12(1), 159–163.
2. *Double Sampling*. Master's Dissertation, Federal University of Technology, Akure.
3. Faweya et al. (2026). *An Improved Exponential Ratio-Type Estimator with Two Auxiliary Variables in Double Sampling with Nonresponse*.
4. Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2019). *Survey Methodology* (3rd ed.). Wiley.
5. Hansen, M. H., & Hurwitz, W. N. (1946). The problem of non-response in sample surveys. *Journal of the American Statistical Association*, 41(236), 517–529.
6. Hazra, R. (2015). Combination of two ratio type estimator for estimating population mean using auxiliary variable with double sampling in presence of non-response. *International Journal of Scientific and Research Publications*, 5(2).
7. Khan, L., & Khan, I. (2022). Exponential ratio type estimators of population mean using two auxiliary variables under nonresponse. *The Light Journal of Computational and Numerical Sciences*, 01(01), 59–75.
8. Oguagbaka, S. K., Okoli, O. C., & Aronu, C. O. (2024). Ratio estimator for double sampling procedure with nonresponse: An empirical study. *International Journal of Basic and Applied Science*, 12(4), 148–158.
9. Ozel Kadilar, G. (2016). A new exponential type estimator for the population mean in simple random sampling. *Journal of Modern Applied Statistical Methods*, 15(2).
10. Särndal, C. E., Swensson, B., & Wretman, J. (2003). *Model Assisted Survey Sampling*. Springer.
11. Singh, R., Kumar, M., & Chaudhary, M. K. (2009). Estimation of mean in presence of nonresponse using exponential estimator.