

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
NAAS Rating (2026): 4.49
Maths 2026; 11(2): 133-138
© 2026 Stats & Maths
<https://www.mathsjournal.com>
Received: 02-12-2025
Accepted: 05-02-2026

Dr. Mateusz Kowalski
Assistant Professor, Department
of Statistics, Faculty of
Mathematics and Applied
Sciences, Baltic Institute of
Statistical Research, Gdańsk,
Poland

Corresponding Author:
Dr. Mateusz Kowalski
Assistant Professor, Department
of Statistics, Faculty of
Mathematics and Applied
Sciences, Baltic Institute of
Statistical Research, Gdańsk,
Poland

Ratio, exponential, and dual-auxiliary-variable estimators for population variance under stratified random sampling: A literature review

Mateusz Kowalski

DOI: <https://www.doi.org/10.22271/math.2026.v11.i2b.2619>

Abstract

Auxiliary information has long been used to sharpen the precision of survey estimators, and the estimation of population variance is no exception. Building on the ratio and exponential-type estimators first formalised for variance by Isaki and by Bahl and Tuteja, a substantial literature has grown around estimators that draw on one, two, or more auxiliary variables under simple, two-phase, and stratified random sampling designs. This review traces that development, from the foundational stratified ratio framework introduced by Kadilar and Cingi through later refinements incorporating dual auxiliary variables, non-conventional measures such as quartile deviation and extreme values, and randomized-response adaptations. Particular attention is given to how mean squared error is typically derived under a first-order approximation, how competing estimators are benchmarked using percentage relative efficiency, and how validation is usually carried out on a recurring small set of real and simulated populations. The review closes by identifying a structural gap: dual-auxiliary-variable variance estimators are well developed under simple random sampling, and single-auxiliary-variable variance estimators are well developed under stratified sampling, but comparatively few studies combine both features for the variance-estimation problem specifically, as distinct from mean estimation, where such combinations are more common.

Keywords: Population variance, auxiliary variable, ratio estimator, exponential estimator, stratified random sampling, mean squared error, percentage relative efficiency, dual auxiliary information, survey sampling, finite population

Introduction

Estimating the variance of a finite population is a routine but consequential task in survey practice. Analysts rely on variance estimates to judge the precision of a survey, to compare sub-groups, and to plan sample sizes for future rounds of data collection. Unlike the population mean, which has attracted the bulk of methodological attention in sampling theory, variance estimation has developed somewhat unevenly, with bursts of innovation followed by long stretches where existing estimators were simply re-applied to new settings.

The use of auxiliary information—a variable correlated with the study variable but cheaper or easier to observe—goes back to Cochran's classical treatment of ratio and regression estimation^[1]. Cochran's ratio estimator for the mean was later adapted to the variance problem, and Isaki's 1983 paper is generally treated as the first systematic account of a ratio-type variance estimator built on a single auxiliary variable^[2]. Bahl and Tuteja introduced an alternative construction in 1991: rather than a simple ratio of sample to population auxiliary totals, their estimator embeds the ratio inside an exponential function^[3]. This exponential form turned out to have favourable bias properties in several settings and became the template for a long line of exponential-type estimators.

Two further developments pushed the literature ahead. Researchers began asking whether a second auxiliary variable could be folded into the estimator, since surveys often collect more than one auxiliary measure. And Kadilar and Cingi extended ratio-type estimation from simple random sampling to stratified random sampling, a design used far more often in practice

Because it lets a survey control precision within meaningful sub-populations such as regions, age bands, or farm sizes. Their 2003, 2005, and 2006 papers laid out combined and separate ratio estimators for the mean and variance under stratification and remain the most cited starting point for later stratified work [4, 6].

This review brings these two threads together. It surveys how ratio-type and exponential-type estimators for population variance have evolved from single-auxiliary, simple-random-

sampling constructions toward designs using two auxiliary variables, non-conventional measures of the auxiliary distribution, and stratified sampling frames. The aim is not to propose a new estimator but to organise the existing literature by design type and auxiliary-variable structure, summarise how these estimators are typically evaluated, and point out where the combination of stratified sampling with dual auxiliary information remains comparatively thin for variance, as distinct from mean, estimation.

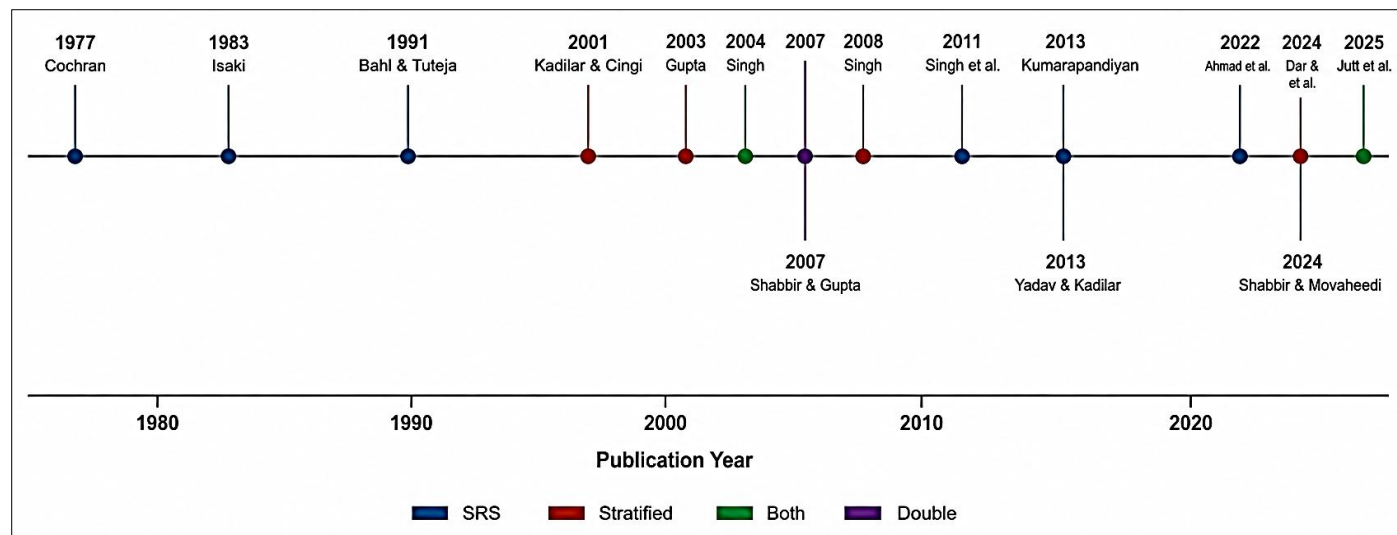


Fig 1: Chronology of the sixteen reviewed studies on ratio- and exponential-type variance estimators, coloured by sampling design (1977-2025)

The timeline shows two visible clusters of activity. A first wave runs from 2003 to 2013, establishing the stratified extension (Kadilar and Cingi; Shabbir and Gupta; Espejo and colleagues) alongside the earliest dual-auxiliary-variable estimators (Singh and colleagues; Subramani and Kumarapandiyan; Yadav and Kadilar). A second, denser wave clusters within 2023-2025-Ahmad and colleagues, Daraz and colleagues, Shabbir and Movaheedi, and Javed and colleagues-indicating renewed and fairly concentrated interest in this estimator family after a quieter stretch through the later 2010s.

Foundations: Ratio and exponential-type variance estimators

Two ideas anchor almost all of the literature reviewed here. The first is the ratio idea itself: if the study variable y and an auxiliary variable x move together, the sample ratio of their totals can be scaled by a known population quantity of x to correct the naive sample variance. Isaki formalised this for the variance problem, deriving the bias and mean squared error (MSE) of a ratio-type variance estimator up to a first-order Taylor approximation and showing conditions under which it outperforms the ordinary sample variance [2].

The second idea is Bahl and Tuteja's exponential transformation [3]. Instead of a direct ratio of auxiliary

quantities, their estimator places the ratio-type correction term in an exponent, which tends to dampen the influence of large sampling fluctuations in the auxiliary variable. Because the exponential form can be expanded in a Taylor series much like the ratio form, its bias and MSE are derived using the same first-order approximation machinery, which is why so much of the later literature reads similarly in mathematical structure even as the specific estimator changes.

A recurring theme across the studies reviewed here is that the exponential-type construction is rarely used on its own. It is typically combined with a difference-type or regression-type term-creating difference-cum-exponential or ratio-cum-exponential hybrids-with the extra term intended to correct for the residual bias the pure ratio or exponential form leaves behind. Espejo, Singh, Pineda, and Nadarajah, and later Shabbir and Gupta, both worked in this hybrid space, proposing classes of estimators that nest the simpler ratio and exponential forms as special cases and deriving optimal values for the constants that define the class [8, 9]. This generalised-class approach-define a broad family with tunable constants, then show that minimising MSE over those constants recovers or beats existing estimators-recurs throughout the literature and is one of its more durable methodological patterns.

Table 1: Chronological summary of reviewed estimator classes

Study (Year)	Ref.	Sampling Design	Auxiliary Structure	Estimator Family
Cochran (1977)	[1]	SRS (general)	Single	Classical ratio/regression (foundational)
Isaki (1983)	[2]	SRS	Single	Ratio-type variance estimator
Bahl & Tuteja (1991)	[3]	SRS	Single	Exponential-type estimator
Kadilar & Cingi (2003)	[4]	Stratified	Single	Combined/separate ratio estimator
Kadilar & Cingi (2005)	[5]	Stratified	Single	Refined ratio-type constant
Kadilar & Cingi (2006)	[6]	SRS & Stratified	Single	Ratio-type variance estimator
Singh & Vishwakarma (2007)	[7]	Double (two-phase)	Single, observed twice	Modified exponential ratio/product
Shabbir & Gupta (2007)	[8]	Stratified	Single	Generalised ratio-type class
Espejo <i>et al.</i> (2008)	[9]	Stratified (equilibrated)	Single	Optimal allocation + ratio-type
Singh <i>et al.</i> (2011)	[10]	SRS	Dual	Exponential estimator, two auxiliary variables
Subramani & Kumarapandian (2013)	[11]	SRS	Single (known parameters)	CV + median-based estimator
Yadav & Kadilar (2013)	[12]	SRS	Dual	Ratio-cum-dual-to-ratio class
Ahmad <i>et al.</i> (2023)	[13]	SRS	Dual	Enhanced two-auxiliary-variable estimator
Daraz <i>et al.</i> (2024)	[14]	Stratified	Single (min/max transform)	Difference-cum-ratio-type exponential
Shabbir & Movaheedi (2024)	[15]	SRS (randomized response)	Single	Generalised randomised-response variance estimator
Javed <i>et al.</i> (2025)	[16]	SRS & Stratified	Dual (non-conventional measures)	Difference-cum-exponential (mean-focused)

Extension to stratified random sampling

Simple random sampling is analytically convenient, but most large surveys are stratified: the population is divided into relatively homogeneous strata, and independent samples are drawn from each. Kadilar and Cingi's 2003 paper in *Biometrical Journal* was the first to carry the ratio-estimator machinery into this setting in a systematic way, defining both combined estimators, which pool information across strata before forming the ratio, and separate estimators, which form a ratio within each stratum and then combine^[4]. Their follow-up papers extended the same logic to variance estimation and to a wider family of ratio-type constants^[5, 6].

Working in the same stratified setting, Espejo and colleagues proposed what they termed an equilibrated stratified sampling approach, adjusting the allocation of sample sizes across strata to further reduce the mean squared error of the variance estimator, while Shabbir and Gupta focused on refining the ratio-type constant itself rather than the sampling allocation^[8, 9]. Both papers illustrate a point that runs through the stratified literature: because a stratified design already reduces variance relative to simple random sampling, the additional gain from an auxiliary-variable estimator has to be weighed against the sampling design's own contribution, which is not always made explicit in the efficiency comparisons reported.

More recently, Daraz, Wu, Alomair, and Aldoghan combined the difference-cum-ratio-type exponential construction with a stratified design, using the minimum and maximum values of the auxiliary variable within each stratum as the auxiliary summary rather than the stratum mean^[14]. Their 2024 paper reported gains in percentage relative efficiency (PRE) across three real datasets, extending the stratified variance literature into non-conventional auxiliary-measure territory that had previously been explored mostly under simple random sampling.

Incorporating dual and multiple auxiliary variables

A separate strand of the literature asks whether a second auxiliary variable improves matters enough to justify the added complexity. Singh, Chauhan, Sawan, and Smarandache proposed an exponential estimator for population variance that draws on two auxiliary variables simultaneously, deriving its bias and MSE and comparing it against single-auxiliary-

variable alternatives under simple random sampling^[10]. Subramani and Kumarapandian took a related but distinct approach, using the known coefficient of variation together with the median of a single auxiliary variable-a known-parameter strategy rather than a second full variable^[11]-while Yadav and Kadilar proposed a ratio-cum-dual-to-ratio class that explicitly combines a ratio-type term with a dual-to-ratio term built from a second auxiliary variable^[12].

Singh and Vishwakarma's earlier work sits adjacent to this strand: rather than two auxiliary variables observed on the same sample, their estimator uses one auxiliary variable observed twice, under a double (two-phase) sampling design, which raises estimation issues similar to the dual-variable case because the estimator has to combine two distinct pieces of auxiliary information^[7].

The most recent addition to this strand, and the one closest to a genuine two-auxiliary-variable, single-phase setting, is Ahmad and colleagues' 2023 paper in *Scientific Reports*^[13]. They constructed an estimator for finite population variance using two auxiliary variables under simple random sampling, derived its bias and MSE to first order, and validated it against several existing estimators using four real datasets together with a simulation study. Their results, like most in this literature, are reported through percentage relative efficiency, and the paper illustrates how quickly this specific sub-area-dual auxiliary variables, variance as the target, simple random sampling as the design-has become populated with closely related estimators, several of which the authors cite as immediate predecessors.

Recent developments and broader extensions

Two further papers show where the literature has moved most recently. Javed, Irfan, Shongwe, Hussain, and Meetei shifted the target from the usual auxiliary summaries to non-conventional measures such as quartile deviation, the inter-quartile range, and the Hodges-Lehmann estimator, arguing that these measures are more robust when the auxiliary variable contains outliers^[16]. Although their focus is the population mean rather than variance, the same non-conventional-measure logic has begun appearing in variance-estimation papers, including some of the stratified work discussed above, suggesting a gradual convergence between

the mean- and variance-estimation literatures in terms of which auxiliary summaries are considered.

Separately, some groups have combined ratio and exponential-type ideas with randomized-response designs, aimed at sensitive-variable surveys rather than ordinary auxiliary-variable settings. Shabbir and Movaheedi's 2024 paper is one example, extending variance estimation under a

generalised randomized-response model ^[15]. While outside the immediate ratio-cum-exponential, dual-auxiliary, stratified-sampling scope of this review, it signals that the first-order MSE-derivation and PRE-comparison toolkit developed for ordinary auxiliary-variable estimation is being exported into other corners of survey methodology.

Table 2: Key estimator forms referenced in this review

Estimator	Ref.	Form (described)
Classical ratio estimator	[1]	Sample mean of y , multiplied by the ratio of the known population auxiliary mean to the sample auxiliary mean
Exponential-type estimator	[3]	Sample mean of y , multiplied by an exponential function of the normalised difference between population and sample auxiliary means
Kadilar-Cingi stratified combined ratio estimator	[4]	Stratified sample mean of y , divided by a weighted combination of stratum auxiliary means and scaled by the corresponding population quantity
Ratio-cum-dual-to-ratio estimator	[12]	A weighted combination of a ratio-type term and a dual-to-ratio term, each built from a different auxiliary variable

Table 2 summarises, in words rather than full notation, the four estimator forms most often referenced by the studies discussed

above; later hybrids in the reviewed literature are generally built by combining or re-weighting these four basic structures.

Table 3: Reference distribution by journal or source

Journal / Source	Number of References
Journal of the American Statistical Association	1
Journal of Information and Optimization Sciences	1
Biometrical Journal	1
Communications in Statistics – Theory and Methods	2
Applied Mathematics and Computation	1
Austrian Journal of Statistics	1
Journal of the Korean Statistical Society	1
Italian Journal of Pure and Applied Mathematics	1
Journal of Modern Applied Statistical Methods	1
Journal of Reliability and Statistical Studies	1
Scientific Reports	1
Heliyon	1
PLoS One	2
Book (Wiley)	1

Two venues, Communications in Statistics – Theory and Methods and PLoS One, account for two references each; the remaining twelve references are spread across twelve distinct journals and one

textbook, which suggests this literature is not concentrated in a single specialist outlet.

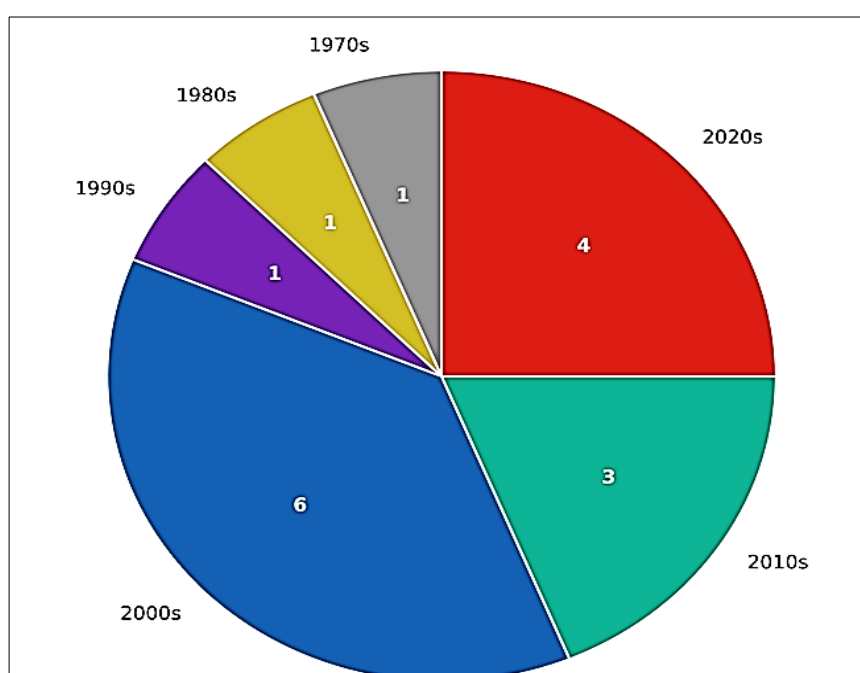


Fig 3: Reviewed references by decade of publication ($n = 16$)

Six of the sixteen references, more than a third, come from the 2000s, when Kadilar and Cingi's stratified framework and its immediate refinements were published. Activity picks up

again in the 2020s, with four references, after a comparatively quiet 2010s that produced three.

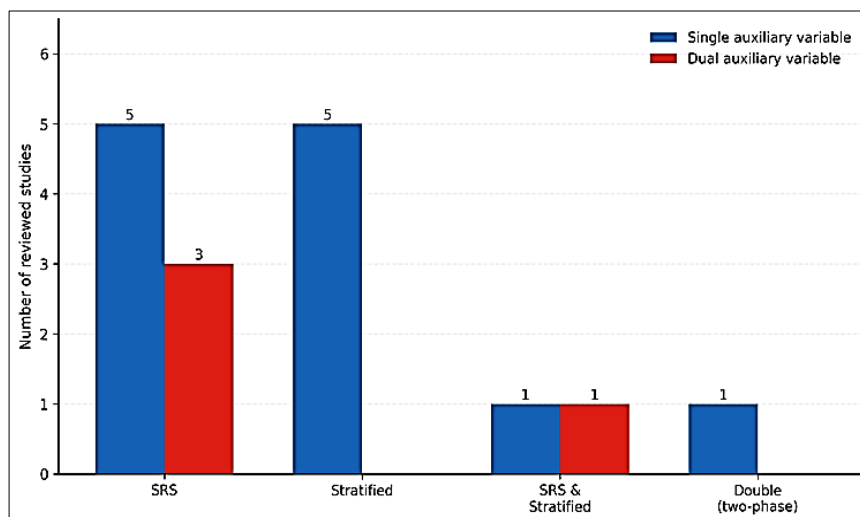


Fig 2: Number of reviewed studies by sampling design and auxiliary-variable structure ($n = 16$)

Figure 2 cross-tabulates the same sixteen studies by sampling design and by whether the estimator draws on one or two auxiliary variables. Under simple random sampling, single- and dual-auxiliary-variable estimators are both reasonably represented (five and three studies). Under a purely stratified design, however, every reviewed study uses a single auxiliary

variable; the one dual-auxiliary-variable study that touches stratified sampling (Javed and colleagues) does so alongside simple random sampling rather than within a stand-alone stratified design, and targets the population mean rather than the variance. This asymmetry is the empirical basis for the gap raised in the Synthesis below.

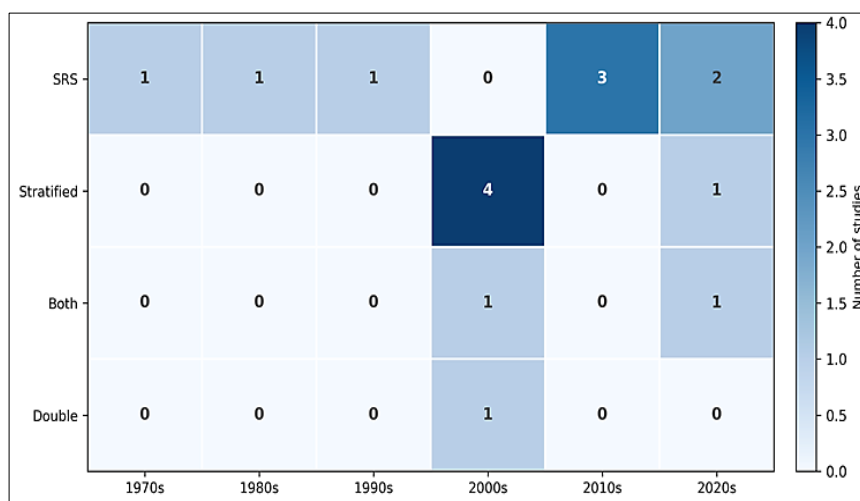


Fig 4: Sampling design by decade of publication ($n = 16$)

Figure 4 recasts the same sixteen studies as a design-by-decade matrix. The 2000s cell for stratified designs (four studies) is the single densest cell in the matrix, reflecting the concentrated burst of Kadilar-and-Cingi-lineage work discussed above, while simple random sampling is the only design with a presence in every decade from the 1970s onward. No decade before the 2020s shows more than one study outside the simple-random and stratified rows, underlining how thin the double-sampling and combined-design literature remains throughout the period covered here.

Synthesis and research gaps

Read together, these studies show a fairly clear line of methodological descent. The exponential transformation introduced by Bahl and Tuteja proved flexible enough to be combined with almost every subsequent innovation-stratification, a second auxiliary variable, non-conventional

auxiliary measures, even randomized-response designs-without requiring a fundamentally different derivation each time^[3]. That flexibility is also a limitation of sorts: because so many papers follow the same first-order-approximation, PRE-comparison template, it is hard to judge from the literature alone how much of the reported efficiency gain is a general property of the exponential-cum-difference construction and how much depends on the particular datasets used for illustration, since most papers validate against a small, overlapping set of textbook and simulated populations rather than a common benchmark suite.

Figure 2 also makes a structural gap visible. Dual-auxiliary-variable estimators for variance are well developed under simple random sampling-Singh, Chauhan, Sawan, and Smarandache; Yadav and Kadilar; and Ahmad and colleagues all work in this setting^[10, 12, 13]. Single-auxiliary-variable estimators for variance are well developed under stratified

sampling, following the Kadilar and Cingi line through Shabbir and Gupta, Espejo and colleagues, and Daraz and colleagues [4, 6, 8, 9, 14]. What is comparatively rare is the direct combination of the two: a variance estimator that uses two genuinely distinct auxiliary variables within an explicitly stratified design, evaluated with the same first-order MSE and PRE tools used elsewhere in this literature. Daraz and colleagues come closest, but their auxiliary information is a transformation of a single variable within each stratum rather than two independent auxiliary variables [14]. On the mean-estimation side, dual-auxiliary-variable stratified estimators are more common, which suggests the gap on the variance side is a matter of the literature not having caught up rather than any technical barrier-the same difference-cum-exponential machinery used for the mean should, in principle, extend to the stratified, dual-auxiliary variance problem in a fairly direct way.

A second, more practical observation is that almost all of the reviewed papers report efficiency gains in relative terms-percentage relative efficiency against a baseline estimator-rather than in terms of confidence-interval width or design effect, which makes the practical size of the improvement harder to judge for a reader deciding whether a more complex estimator is worth the extra computation for a specific survey.

Conclusion

This review has traced how ratio- and exponential-type estimators for population variance have grown from Isaki's and Bahl and Tuteja's original single-auxiliary-variable, simple-random-sampling constructions into a much broader family spanning stratified designs, dual and non-conventional auxiliary measures, and randomized-response settings. The Kadilar and Cingi line of work remains the anchor point for anything done under stratification, while the dual-auxiliary-variable strand has so far developed mostly within simple random sampling.

The clearest gap identified here is the shortage of estimators that combine an explicitly stratified design with two genuinely separate auxiliary variables for the variance-estimation problem specifically. Given how directly the underlying first-order MSE derivation carries over from the mean to the variance case, and from simple random to stratified sampling, this looks like a tractable next step for work in this area, along with more consistent benchmarking across studies so that reported efficiency gains can be compared on a common footing rather than population by population.

Declarations

Conflict of interest: The author(s) declare no conflict of interest.

Funding: This review received no specific funding.

References

1. Cochran WG. Sampling Techniques. 3rd ed. New York: John Wiley & Sons; 1977.
2. Isaki CT. Variance estimation using auxiliary information. *J Am Stat Assoc.* 1983;78(381):117-123.
3. Bahl S, Tuteja RK. Ratio and product type exponential estimators. *J Inf Optim Sci.* 1991;12(1):159-164.
4. Kadilar C, Cingi H. Ratio estimators in stratified random sampling. *Biometrical J.* 2003;45(2):218-225.

5. Kadilar C, Cingi H. A new ratio estimator in stratified random sampling. *Commun Stat Theory Methods.* 2005;34(3):597-602.
6. Kadilar C, Cingi H. Ratio estimators for the population variance in simple and stratified random sampling. *Appl Math Comput.* 2006;173(2):1047-1058.
7. Singh HP, Vishwakarma GK. Modified exponential ratio and product estimators for finite population mean in double sampling. *Austrian J Stat.* 2007;36(3):217-225.
8. Shabbir J, Gupta S. On improvement in variance estimation using auxiliary information. *Commun Stat Theory Methods.* 2007;36(12):2177-2185.
9. Espejo MR, Singh HP, Pineda MD, Nadarajah S. Optimal estimation of population variance using equilibrated stratified sampling from infinite populations. *J Korean Stat Soc.* 2008;37(4):375-383.
10. Singh R, Chauhan P, Sawan N, Smarandache F. Improved exponential estimator for population variance using two auxiliary variables. *Ital J Pure Appl Math.* 2011;28:101-108.
11. Subramani J, Kumarapandiyan G. Estimation of variance using known coefficient of variation and median of an auxiliary variable. *J Mod Appl Stat Methods.* 2013;12(1):58-64.
12. Yadav SK, Kadilar C. A class of ratio-cum-dual to ratio estimator of population variance. *J Reliab Stat Stud.* 2013;6(2):29-34.
13. Ahmad S, Adichwal NK, Aamir M, Shabbir J, Alsadat N, Elgarhy M, Ahmad H. An enhanced estimator of finite population variance using two auxiliary variables under simple random sampling. *Sci Rep.* 2023;13(1):21444.
14. Daraz U, Wu J, Alomair MA, Aldoghan LA. New classes of difference-cum-ratio-type exponential estimators for a finite population variance in stratified random sampling. *Heliyon.* 2024;10(12):e33402.
15. Shabbir J, Movahedi Z. Use of generalized randomized response model for enhancement of finite population variance: A simulation approach. *PLoS One.* 2024;19(12):e0315658.
16. Javed M, Irfan M, Shongwe SC, Hussain MA, Meetei MZ. Difference-cum-exponential-type estimators for estimation of finite population mean in survey sampling. *PLoS One.* 2025;20(1):e0313712.