



Krzysztof Szymoniak-Książek

 <https://orcid.org/0000-0002-5003-2014>

Department of Statistics,
Econometrics and Mathematics
Faculty of Management
University of Economics in Katowice
Katowice, Poland
szymoniak-ksiazek@ue.katowice.pl

Wojciech Gamrot

 <https://orcid.org/0000-0001-5617-2600>

Department of Statistics,
Econometrics and Mathematics
Faculty of Management
University of Economics in Katowice
Katowice, Poland
wojciech.gamrot@ue.katowice.pl

Controlling for survey costs when estimating covariances between Pathak estimators in fixed-budget sequential sampling

Accepted by Editor Ewa W. Ziemba | Received: October 7, 2024 | Revised: March 18, 2025; April 4, 2025 | Accepted: April 14, 2025 | Published: May 23, 2025.

© 2025 Author(s). This article is licensed under the Creative Commons Attribution-NonCommercial 4.0 license (<https://creativecommons.org/licenses/by-nc/4.0/>)

Abstract

Aim/purpose – This work aims to assess stochastic properties of finite population mean estimators under Pathak's fixed-cost sequential sampling scheme by deriving their covariance and proposing its estimator. This scheme enables strict control over survey-related monetary expenditures when individual data acquisition costs differ and survey budget excesses are unacceptable. It facilitates economically efficient use of available resources. Our work is aimed at extending its applicability.

Design/methodology/approach – This paper's main theoretical result concerning the covariance of Pathak mean value estimators and the unbiasedness of its proposed estimator was derived analytically. A short simulation study supports this.

Findings – The proposed estimator for the covariance between Pathak's estimators of population averages is unbiased under the sequential fixed-cost sampling scheme, which prevents survey budget excesses when the per-unit costs of acquiring information are known in advance but not homogeneous.

Research implications/limitations – Our findings facilitate accuracy assessment for complex estimators of population parameters that utilize auxiliary information under restricted-budget sequential sampling and, in particular, for ratio, product, and regression estimators. They are potentially applicable to any quantities of interest that could be expressed in terms of averages, including, among others, unemployment rates, poverty

Cite as: Szymoniak-Książek, K., & Gamrot, W. (2025). Controlling for survey costs when estimating covariances between Pathak estimators in fixed-budget sequential sampling. *Journal of Economics and Management*, 47, 211-228. <https://doi.org/10.22367/jem.2025.47.09>

measures, inflation indices, and GDP indicators. Their application is justified if the per-unit costs of acquiring information vary.

Originality/value/contribution – Our results allowed us to assess various economic variables' distribution characteristics while controlling for randomly varying survey costs. This facilitates planning and managing sample surveys in an economically efficient way.

Keywords: Pathak's estimator, sequential sampling, fixed-cost sampling, sum-quota sampling.

JEL Classification C15, C83.

1. Introduction

Private and public entities perform various decision processes in their routine functioning. This requires gathering information about their economic surroundings. Therefore, it is indispensable to assess the values of various quantities that characterize finite populations of households, enterprises, producers, and consumers. At the same time, it is usually too time-consuming or costly to obtain complete information about every population unit. This leads to the necessity of drawing random samples and carrying out statistical inference. For example, linear (Yitzhaki, 1996) and nonlinear (Yatchew, 1998; Pilot et al., 2024) regression models are often used to estimate or predict various quantities of interest. Time series analysis helps examine data collected at regular intervals and assess inflation (Dossche & Everaert, 2005; Stock & Watson, 2016), unemployment (Katrís, 2020), gross domestic product (Elbatal et al., 2025) or other macroeconomic indicators. In these endeavors, the role of statistical methods is usually subordinated to wider economic investigations. They are simply useful tools for an economist.

Sometimes, this relationship between economics and statistics becomes reversed. Economic considerations play an essential role in the planning of statistical undertakings. A statistical survey may be considered a production process, where estimates for various interest quantities may be regarded as products. A specific feature of such endeavors is the emphasis on quality rather than quantity of output. While numerous estimates may be simultaneously produced in a single survey, it is their accuracy rather than quantity that counts. Accurate estimates are sold for a high price. Inaccurate estimates are hardly helpful. At the same time, surveys require scarce economic resources to be expended. Interviewers must be hired, money must be paid, and time must be spent. The principle of economic efficiency postulates the formation of a favorable relationship between expenditures and outcomes. Often, it is understood as maximizing the

difference between efforts and their effects, known as profit (Czarny & Rapacki, 2002; Nasiłowski, 1994). Typically, the goal is to achieve a given effect using the smallest possible amount of available resources or achieve the best possible result using a specific amount. In the literature, one can also find other approaches to efficiency, such as productive efficiency (Sickles & Zelenyuk, 2019), allocative efficiency (Rutherford, 2002), or Pareto efficiency (Bishop, 2009). These considerations usually lead to formulating some optimization problem to be solved to ensure economic efficiency. It expresses numerical trade-offs between outcomes and expenditures. The survey cost always plays an important role in these calculations, either as a part of the criterion function to be optimized, as a restriction to be respected, or both (Groves, 1989). In practice, the cost depends on many data acquisition and processing factors. It may be influenced by geographic location, data acquisition method, the size of the objects under study, and many other factors. It is difficult to enumerate all these factors. As a result, simplified cost models are usually considered. In these models, the costs of observing individual population units are usually assumed constant in the whole population or its subsets. These assumptions dominate the existing literature (Gamrot, 2014). More detailed models where the costs are allowed to vary among individual units are relatively infrequent. Estimation procedures that explicitly account for this possibility are rarely discussed in the literature, except for Pathak's (1976) and Kremers' (1986) works. This may be considered an important research gap.

This paper aims to fill this gap by extending Pathak's (1976) theoretical results concerning the estimation of finite population average under the sequential fixed-cost sampling scheme, also known as quota sampling. The scheme is carried out by repeatedly drawing individual elements to the sample until a predetermined survey budget is exhausted. This rules out budget excesses. The sample drawn this way is not representative. Expensive units may be underrepresented. Despite this difficulty, Pathak proposed an estimator of the population mean free from systematic error and provided an explicit formula for its variance. Additionally, he proposed an unbiased estimator of this variance. In this paper, Pathak's ideas are generalized by providing an explicit formula for the covariance between the population mean estimators and proposing an unbiased estimator for this covariance. The results of this undertaking will be interesting to survey practitioners who want to maintain strict control over the cost of data acquisition from random samples and, at the same time, intend to use auxiliary information on additional study variables to accurately estimate economic quantities of interest.

The remainder of this paper is structured as follows. After the introduction, a theoretical background will be presented. Subsequently, the Pathak scheme will be formally defined, followed by a theorem that generalizes the results achieved by Pathak. It will then be proven. A simulation study that demonstrates the applicability of our theoretical result will then be presented, and its economic applications will be briefly discussed.

2. Literature review

The statistical literature discusses various schemes of sample selection (Bracha, 1996; Cochran, 1977; Hedayat & Sinha, 1991; Lehtonen & Pahkinen, 2004). The sampling scheme involves selecting units one by one from a population, with predetermined probabilities of selection for each unit at every draw (Wywiał, 2010). Sampling aims to estimate population parameters, which are certain functions of study variables in the whole population.

A popular group of sampling methods is sequential schemes. According to Tillé (2006), a sequential scheme is a sampling scheme that operates on a previously sorted list of units so that the units are processed one by one, starting with the first one. At each k -th step, the decision about the number of occurrences of the k -th unit in the sample is made independently of the subsequent units on the list.

Most frequently, sequential sampling schemes facilitate the minimization of sampling cost for a desired quality of statistical inference (in particular, accuracy of estimates). This approach is taken, among others, by Champernowne (1953), Fan et al. (1962), Chromy (1979), and Pekasiewicz (2001, 2002). However, sequential sampling may also be used so that the survey cost becomes a restriction imposed on the solution rather than an optimized criterion function. A pioneering work devoted to this subject is the famous paper of Pathak (1976), who considers a fixed-cost sequential sampling scheme in which the cost of drawing individual units into the sample is not constant.¹ In this scheme, elements are sampled until the total cost of examining the characteristics of the drawn elements reaches the budget. A precise definition will be presented later in this paper. It is worth noting that Pathak's scheme is a sequential sampling scheme according to Tillé's definition. However, it does not precisely correspond to specific other definitions. In particular, Steczkowski (1995) stated that a sequential sampling scheme involves randomly selecting individual units or small

¹ It is worth noting that the term "fixed-cost" means the predetermined survey budget. However, this does not mean the samples drawn have the exact total cost. Kremers (1986) used the alternative term "sum-quota sampling" for this scheme.

groups of units from the general population and determining each time whether the information gathered so far is sufficient to make a specific decision. Nevertheless, the only criterion checked at each sampling step in Pathak's scheme is the total cost of examining the individual elements. Other notable definitions of sequential sampling can be found, for example, in the works of Fisz (1954), Ghosh et al. (1997), and Kończak (2007). Furthermore, Pathak proposed a population mean estimator, which, somewhat surprisingly, takes a simple form of the sample average (excluding the last element). He proved that this estimator is unbiased. To derive this estimator and prove its unbiasedness, he employed a procedure known as Rao–Blackwellization (Blackwell, 1947; Kolmogorov, 1950; Rao, 1945). It involves deriving an estimator using the conditional expected value for some sufficient statistics. Basu (1958) and Hajek (1959) previously addressed the issue of the sufficiency of the statistics being utilized. In this context, the works of Basu & Ghosh (1967) and Basu (1969) are also noteworthy. Pathak also utilized the Rao–Blackwellization procedure to construct an unbiased estimator of the variance of the mean estimator. In the subsequent section, the results of Pathak's work will be presented, followed by a new theorem on the covariance of the mean estimators that extends Pathak's findings.

3. Methodology

Stochastic properties of estimators are generally derived using probability calculus methods. Two main approaches are distinguished. The first is the model-based approach, which assumes that the values of the characteristic being studied in the population are realizations of random variables whose distributions are described by a probabilistic model known as the superpopulation model (Krzciuk & Żądło, 2014; Valliant et al., 2000; Żądło, 2015). However, this paper considers the second approach, the randomization approach. It postulates that the values of the characteristic being studied are fixed (that is, non-random) for all units, and the source of randomness is the selection of these units into the sample (Szreder, 2010).² Our main result was obtained on this basis. Conditional expected values were calculated, and analytical results were derived. A numerical experiment supports the findings. It was carried out by repeatedly drawing samples from a given population in controlled conditions and examining the empirical distribution of resulting estimates.

² It is worth noting that there is also the model-design approach (Niemiro & Wesołowski, 2012).

4. Pathak's results

Let us consider a finite population $U = \{1, \dots, N\}$. The i -th element of the population corresponds to the value of the studied characteristic y_i and to the individual cost of investigation c_i . Let $\mathbf{y} = (y_1, y_2, \dots, y_N)$ denote the vector of values of the given variable and $\mathbf{c} = (c_1, c_2, \dots, c_N)$ denote the vector of costs. Assume that the costs of investigating each unit are known, and let B be the research budget. Furthermore, let:

$$\max_{i \neq j \in U} (c_i + c_j) < B,$$

$$\sum_{i=1}^N c_i > B,$$

$$\min_{i \in U} c_i > 0.$$

The goal is to estimate the population's mean:

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i.$$

Pathak proposed the following sampling scheme.

Pathak's scheme [Pathak, 1976]

Population units are drawn one by one with equal probabilities without replacement until the accumulated cost reaches or exceeds B . The sample is $\mathbf{s} = (s_1, s_2, \dots, s_{M+1})$ where each s_i is one of the U_i and the stopping variable $M + 1$ is defined as follows: $M + 1 = r$ if and only if $\sum_{i=1}^{r-1} c_{s_i} < B$ and $\sum_{i=1}^r c_{s_i} \geq B$.

To simplify and standardize notation, let $y_{s_i} = y_{(i)}$. Additionally, Pathak proposed an unbiased estimator of the population mean.

Theorem 1 [Pathak, 1976]

The unbiased estimator of the population mean $\bar{y}$ based on a sample drawn using the Pathak scheme is given by the formula:

$$\bar{y}_M = \frac{1}{M} \sum_{i=1}^M y_{(i)}. \quad (1)$$

It is worth noting that the probabilities of individual elements being included in the sample are usually different. This is due to the fact that the costs of

selecting individual units for the sample are not the same. Despite this, which may be surprising, the estimator of the population mean is unbiased. Additionally, Pathak provided a formula for the variance of the population mean estimator and proposed an unbiased estimator for this variance. This may be stated as follows:

Theorem 2 [Pathak, 1976]

a) The variance of the population mean estimator $\bar{y}_M$ is given by the formula:

$$Var(\bar{y}_M) = \frac{1}{2N(N-1)} \sum_{j \neq k=1}^N (y_j - y_k)^2 e_{jk}, \quad (2)$$

where:

$$e_{jk} = E\left(\frac{1}{M} \middle| s_1 = j, s_2 = k\right) - \frac{1}{N}.$$

b) Additionally assuming that:

$$A = \frac{1}{M-1} \left[\frac{1}{M} - \frac{1}{N} \right]$$

we obtain an unbiased estimator of $Var(\bar{y}_M)$, in the form:

$$\hat{V}_{xy} = A \cdot \sum_{i=1}^M (y_{(i)} - \bar{y}_M)^2. \quad (3)$$

Calculating the variance of the population mean estimator may be impossible due to computational complexity, even for very small populations. This highlights the importance of simplicity. The variance estimator plays a key role in this context. It does not depend on inclusion probabilities, and its calculation from the sample is not time-consuming. Proof of the above results may be found in Pathak's (1976) work.

5. Extending Pathak's results

Let us introduce an additional variable $\mathbf{x} = (x_1, x_2, \dots, x_N)$. From this point on, the cost vector $\mathbf{c}$ consists of elements $c_1, c_2, \dots, c_N$ indicating the cost of investigating the value of variable y_i as well as x_i . Similarly, let $x_{s_i} = x_{(i)}$. According to Theorem 2 the statistic:

$$\bar{x}_M = \frac{1}{M} \sum_{i=1}^M x_{(i)}$$

is an unbiased estimator of the population mean:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i.$$

The following theorem extends Pathak's work.

Theorem 3

a) The covariance of Pathak's population mean estimators $\bar{x}_M$ and $\bar{y}_M$ is given by the formula:

$$\text{Cov}(\bar{x}_M, \bar{y}_M) = \frac{1}{2N(N-1)} \sum_{j \neq k=1}^N (x_j - x_k)(y_j - y_k) e_{jk}, \quad (4)$$

where:

$$e_{jk} = E\left(\frac{1}{M} \middle| s_1 = j, s_2 = k\right) - \frac{1}{N}.$$

b) Additionally assuming that:

$$A = \frac{1}{M-1} \left[\frac{1}{M} - \frac{1}{N} \right]$$

we obtain an unbiased estimator of $\text{Cov}(\bar{x}_M, \bar{y}_M)$ in the form:

$$\hat{C}_{xy} = A \cdot \sum_{i=1}^M (x_{(i)} - \bar{x}_M)(y_{(i)} - \bar{y}_M), \quad (5)$$

Proof

b) Let $t = \frac{1}{N} x_{(1)} y_{(1)} + \frac{N-1}{N} x_{(1)} y_{(2)}$. Calculating the expected value of t , we obtain:

$$E(t) = E\left(\frac{1}{N} x_{(1)} y_{(1)} + \frac{N-1}{N} x_{(1)} y_{(2)}\right) = \frac{1}{N^2} \sum_{i=1}^N x_i y_i + \frac{1}{N^2} \sum_{\substack{j,k=1 \\ j \neq k}}^N x_j y_k = \bar{x} \bar{y}.$$

Moreover, let $T_M = (M, \{s_1, s_2, \dots, s_M\}, s_{M+1})$. Calculating the conditional expected value of t with respect to T_M (denoted by W) we obtain:

$$W = E(t|T_M) = \bar{x}_M \bar{y}_M - A \sum_{i=1}^M (x_{(i)} - \bar{x}_M)(y_{(i)} - \bar{y}_M).$$

Next, let us use the property of the conditional expected value:

$$E(W) = E(E(t|T_M)) = E(t) = \bar{x} \bar{y}.$$

Therefore, W is an unbiased estimator of $\bar{x}\bar{y}$. Note that $\hat{C}_{xy} = \bar{x}_M\bar{y}_M - W$. Then $E(\hat{C}_{xy}) = E(\bar{x}_M\bar{y}_M) - \bar{x}\bar{y} = E(\bar{x}_M\bar{y}_M) - E(\bar{x}_M)E(\bar{y}_M) = Cov(\bar{x}_M, \bar{y}_M)$ which means that $\hat{C}_{xy}$ is an unbiased estimator of $Cov(\bar{x}_M, \bar{y}_M)$.

a) Using the property of conditional expected value: $E(\hat{C}_{xy}) = E(E(\hat{C}_{xy}|T_M))$ and the deformation formula (Zhang et al., 2012), we then obtain:

$$E(\hat{C}_{xy}) = E \left(E \left(\Delta \cdot \frac{1}{2M} \sum_{\substack{j,k=1 \\ j \neq k}}^M (x_{(j)} - x_{(k)})(y_{(j)} - y_{(k)}) \middle| T_M \right) \right).$$

Due to invariance with respect to permutations, we obtain:

$$Cov(\bar{x}_M, \bar{y}_M) = E \left(\left[\frac{1}{M} - \frac{1}{N} \right] \frac{1}{2} (x_{(1)} - x_{(2)})(y_{(1)} - y_{(2)}) \right).$$

Let us express the above formula using an indicator function:

$$Cov(\bar{x}_M, \bar{y}_M) = E \left(\frac{1}{2} \sum_{\substack{j,k=1 \\ j \neq k}}^N (x_j - x_k)(y_j - y_k) \left[\frac{1}{M} - \frac{1}{N} \right] \alpha_{jk} \right),$$

where: $\alpha_{jk} = 1$, if $s_1 = j$ i $s_2 = k$ and $\alpha_{jk} = 0$ otherwise.

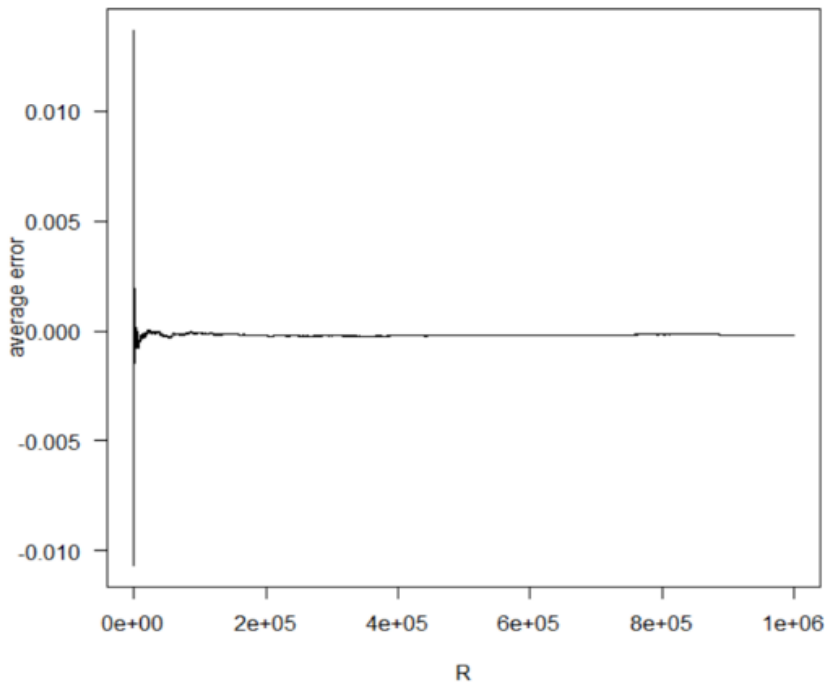
Since $E\left(\frac{1}{M} \alpha_{jk}\right) = E\left(E\left(\frac{1}{M} \alpha_{jk} \middle| s_1, s_2\right)\right) = \frac{1}{N(N-1)} E\left(\frac{1}{M} \middle| s_1 = j, s_2 = k\right)$ and $E(\alpha_{jk}) = P(s_1 = j, s_2 = k) = \frac{1}{N(N-1)}$ we obtain formula (4). This completes the proof.

6. Numerical study

To demonstrate the suitability of presented results for estimation, a simulation experiment was conducted. A population of 100 elements was considered. The variables under investigation were denoted by $\mathbf{x} = (x_1, x_2, \dots, x_{100})$ and $\mathbf{y} = (y_1, y_2, \dots, y_{100})$, while the cost of examining individual units was represented by the vector $\mathbf{c} = (c_1, c_2, \dots, c_{100})$. For the purpose of simulation the values (x_i, y_i) were generated using a pseudo-random number generator imitating the $N_2\left([0,0], \begin{bmatrix} 1 & \lambda \\ \lambda & 1 \end{bmatrix}\right)$ distribution. The variables $\mathbf{x}$ and $\mathbf{y}$ were sampled until the correlation coefficient was approximately equal to λ (with a precision of 0.001). As a result the correlation coefficient between $\mathbf{x}$ and $\mathbf{y}$ was approximately equal

to a prespecified value of λ . In the simulation, various values of the parameter λ were considered spaced evenly between -1 and 1 with the increment 0.01 . Additionally, the costs of examining individual units, $c_1, c_2, \dots, c_{100}$ were obtained using a pseudo-random number generator imitating the uniform $U(0,100)$ distribution. The research budget B was set to 500 . For the population formed in this way, a sample was repeatedly drawn using Pathak's sampling scheme. Next, unbiased estimators of the population mean were calculated. For a given population this was repeated $R = 10^6$ times. For each sample, $\bar{x}_M$, $\bar{y}_M$ and $\hat{C}_{xy}$ were calculated. The goal was to assess the expected error. Calculating the true covariances was computationally too complex. Therefore, the empirical covariance between $\bar{x}_M$ and $\bar{y}_M$ registered in simulation was used instead. Consequently, the expected error was assessed as the difference between the average value of $\hat{C}_{xy}$ and the covariance between $\bar{x}_M$, $\bar{y}_M$ registered in experiment. In the first simulation, the population was drawn for $\lambda = 0.8$. In Figure 1, the horizontal axis represents the number of sample replications, while the vertical axis shows the average error between the covariance estimator values and the sample covariance.

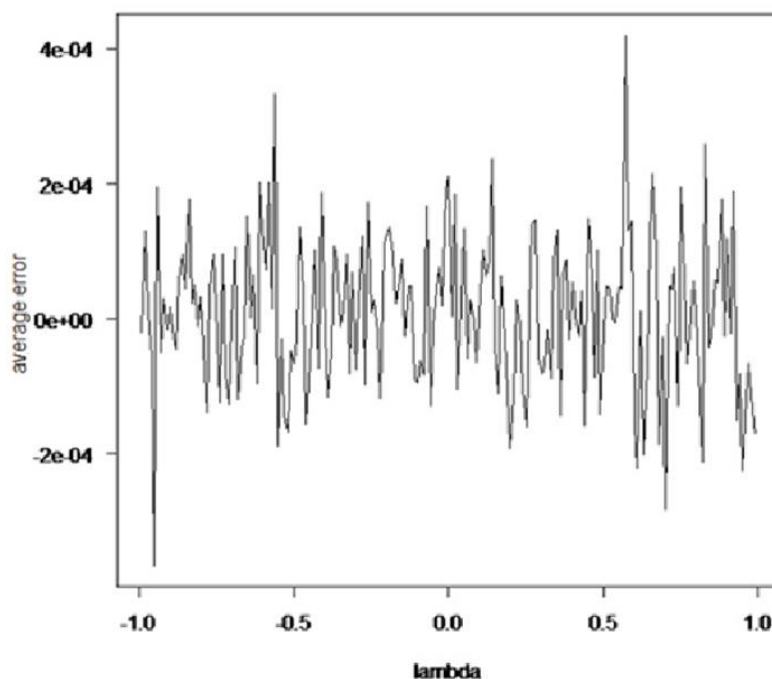
Figure 1. The average error of estimates calculated in successive iterations of the simulation experiment ($R = 2, 3, \dots, 10^6$)



Source: Authors' own calculations.

It is easy to notice that the average error is already close to zero, even with a few sample replications. In the next simulation, the average error was examined depending on the value of λ for $R = 10^6$. The results are presented in Figure 2.

Figure 2. The average error of estimates calculated for varying values of λ



Source: Authors' own calculations.

Regardless of the correlation coefficient λ , the empirical average error of $\hat{C}_{xy}$ is close to zero. This strongly indicates that the estimator is free from systematic error, which aligns with Theorem 3's claim that the estimator is exactly unbiased.

7. Potential economic applications

The fixed-cost sampling technique may find applications in various economic contexts. Numerous practical situations exist where the individual costs of acquiring information from individual population units may vary. Let us mention just a few of them:

1. In financial audits, diverse accounting documents are examined. Usually, the number of documents gathered by large institutions is too high to be exhaustively investigated (Owczarek, 1996, 1998). This forces the auditors to rely on

sampling and only to investigate a subset of these records (Pfaff, 2007). Statistical inference techniques are then employed in order to arrive at conclusions. Meanwhile, individual documents may require more or less effort to scrutinize. This depends on the documents' kind, content, and status and is reflected by the varying per-unit cost of acquiring information. Such a phenomenon is amplified when some needed documents are literally missing, and the required information must be obtained from alternative, more expensive sources (Kadane, 2005).

2. In other investigations, gathering data from individual population units may require specialized technical, scientific, or legal assessments. In particular, the preparation of wildlife resource inventories required for planning agricultural, environmental, and climate-related state interventions may be more or less costly depending on the size, composition, and location of the plots of land or other immovable environmental assets to be examined. The same applies to investigations of economic losses resulting from natural catastrophes such as floodings or insect infestations to distribute state aid or for other reasons.
3. Similar situations may arise when the value of natural resources (such as forests or mineral deposits) needs to be assessed (Breidt & Fuller, 1993). Their examination may involve drilling or taking measurements. The effort needed to perform these activities may depend on various local conditions and may be expected to vary.
4. The non-homogeneity of costs may also be expected in any in-place investigation conducted in spatially distributed populations of objects. These may be industrial facilities, households, stores, or individual material objects such as buildings or bridges. Being located in various places, these objects and institutions may be more or less challenging to reach, which is reflected by a variable data acquisition cost. In particular, the surveys of demand for public transport services may require placing investigators or measuring equipment in certain places at specific moments, where (and when) these places may be more or less accessible (Benedetti et al., 2015).
5. In some populations of interest, different means of contact may be available for various population units. For example, in customer satisfaction surveys, some individuals may be accessible by email, while others are only reachable by telephone or direct contact. This translates into varying data-gathering costs.
6. In all these situations, the varying cost of observing data values may be interpreted literally, expressed in monetary units, and confronted with a fixed monetary budget for the survey. However, it may also happen that the restriction concerns other economic resources, such as the amount of time available for investigation, the quantity of competent personnel, or contractu-

ally restricted means of contact. Data acquisition from individual units may require varying amounts of engaging, consuming, or expending these assets. When the survey is based on random samples and differences in observation costs are not considered, the total cost will also be random. If a fixed sample size is chosen based on some prediction of average cost, then the survey's total cost may considerably exceed the available budget. This may force survey planners to search for emergency funding, which may be challenging to arrange. If these efforts fail, the two remaining options will be to reduce the sample somehow or to repeat the sample drawing until its cost is satisfactory. Both of these practices are highly questionable from the methodological point of view, and resorting to them may radically undermine the reliability of the survey. Survey planners may foresee these problems. They may try to avoid them by choosing a smaller sample, hoping that the corresponding cost safety margin will suffice to cover eventual excesses over the expected cost. However, such a cautious approach is hampered by the cost distribution, which usually has long tails, and a safety margin guaranteeing a complete elimination of budget excesses would have to be very wide, essentially forcing a significant reduction of the sample (Gamrot, 2014). This might harm the survey's credibility from the opposite side leading to a loss of estimation accuracy. Moreover, it would be likely that significant funds are not spent (when the sample turns out to be relatively less expensive). This breeds the risk of funding reductions for future editions of the survey. It may also lead to accusations that the funds are not used effectively, which are not without merit. Applying fixed-cost sampling techniques, like the Pathak scheme eliminates all these problems. Sequential sampling enables an economically efficient use of available funds and budgetary stability.

8. Conclusions

Sequential fixed-cost sampling offers significant economic and practical advantages over other sampling procedures in situations where individual data-gathering costs are not homogenous. After all, the planning and execution of sample surveys and the analysis of their results are services. They may be commercially traded, and they often are. The ability to rule out budget excesses and, at the same time, efficiently allocate available resources may improve the quality of these services or reduce their cost. This may significantly improve the institution's competitive position offering survey-related services against its less advanced contenders. Given these benefits, one might wonder why this procedure is unpopular. It appears that one of the reasons for this state of affairs is its com-

plexity. In contrast to popular sampling schemes, such as stratified or multistage sampling, the assessment of stochastic properties, even for estimates of some simple population parameters such as means and totals, has generally been considered difficult under sequential sampling. This has hampered efforts to use available information to produce accurate estimates.

The results presented in this paper will facilitate the overcoming of these barriers. The ability to estimate the covariance between Pathak estimators directly from the sample will facilitate the construction of complex estimators of population parameters, such as ratio estimators, product estimators, and regression estimators that utilize available information on auxiliary variables (Bracha, 1996; Särndal et al., 1992, pp. 226-374). More importantly, it will also provide a comfortable way of assessing their accuracy through the use of the Taylor linearization technique (e.g., Demnati & Rao, 2002; Gamrot, 2012; Wolter, 2007). The exploration of these possibilities will naturally extend the results of our study. It should also be noted that our research is restricted to a simple procedure that treats only one kind of economic resource as restricted. It may be generalized by considering a more sophisticated stopping condition involving the exhaustion of more than one scarce resource (for example, available funds and time). This would be another natural extension of our work. However, such an investigation exceeds the scope of the current paper: it will require a separate treatment.

Author contributions

Krzysztof Szymoniak-Książek – 60% (collection and/or assembly of data, data analysis and interpretation, writing the article, critical revision of the article).

Wojciech Gamrot – 40% (research concept and design, writing the article, critical revision of the article, final approval of the article).

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

- Basu, D. (1958). On sampling with and without replacement. *Sankhyā: The Indian Journal of Statistics (1933-1960)*, 20, 287-294. <https://www.jstor.org/stable/25048396?seq=1>
- Basu, D. (1969). Role of sufficiency and likelihood principle in sample survey theory. *Sankhyā: The Indian Journal of Statistics (1933-1960)*, 31, 441-454. <https://www.jstor.org/stable/25049610?seq=1>

- Basu, D., & Ghosh, J. K. (1967). Sufficient statistics in sampling from a finite universe. *Bulletin of the International Statistical Institute*, 42, 850-859.
- Benedetti, R., Piersimoni, F., & Postiglione, P. (2015). Spatial sampling designs. In *Sampling spatial units for agricultural surveys. Advances in spatial science* (pp. 149-196). Springer. https://doi.org/10.1007/978-3-662-46008-5_7
- Bishop, M. (2009). *Essential economics: An A-Z guide*. Bloomberg Press.
- Blackwell, D. (1947). Conditional expectation and unbiased sequential estimation. *The Annals of Mathematical Statistics*, 18(1), 105-110. <https://doi.org/10.1214/aoms/1177730497>
- Bracha, C. (1996). *Teoretyczne podstawy metody reprezentacyjnej* [Foundations of survey sampling]. PWN.
- Breidt, F. J., & Fuller, W. A. (1993). Regression weighting for multiphase samples. *Sankhyā: The Indian Journal of Statistics, Series B (1960-2002)*, 55(Series B), 297-309. <https://www.jstor.org/stable/25052799?seq=1>
- Champernowne, D. G. (1953). Cost calculations for sequential sampling schemes. *The Incorporated Statistician*, 4(4), 197-205. <https://doi.org/10.2307/2986526>
- Chromy, J. R. (1979). Sequential sample selection methods. In *Proceedings of the American Statistical Association, Section on Survey Research Methods* (pp. 401-406). ASA. http://www.asasrms.org/Proceedings/papers/1979_081.pdf
- Cochran, W. G. (1977). *Sampling techniques*. Wiley.
- Czarny, B., & Rapacki, R. (2002). *Podstawy ekonomii* [The basics of economics]. PWE.
- Demnati, A., & Rao, J. N. K. (2002). Linearization variance estimators for survey data. In *SSC Annual Meeting, May 2002, Proceedings of the Survey Methods Section* (pp. 87-92). Statistical Society of Canada. https://ssc.ca/sites/default/files/survey/documents/SSC2002_A_Demnati.pdf
- Dossche, M., & Everaert, G. (2005). *Measuring inflation persistence: a structural time series approach* (Working Paper, No. 70). European Central Bank. <https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp495.pdf>
- Elbatal, I., Sarwar, M., Jamal, F., Daniyal, M., Hussain, Z., & Ghorbal, A. B. (2025). Modelling on gross domestic product annual growth rate data by using time series, machine learning, and probability models. *Journal of Radiation Research and Applied Sciences*, 18(2), 101481. <https://doi.org/10.1016/j.jrras.2025.101481>
- Fan, C. T., Muller, M. E., & Rezucha, I. (1962). Development of sampling plans by using sequential (item by item) selection techniques and digital computers. *Journal of the American Statistical Association*, 57(298), 387-402. <https://doi.org/10.1080/01621459.1962.10480667>
- Fisz, M. (1954). *Rachunek prawdopodobieństwa i statystyka matematyczna* [Probability theory and mathematical statistics]. PWN.
- Gamrot, W. (2012). Estimation of finite population kurtosis under two-phase sampling for nonresponse. *Statistical Papers*, 53, 887-894. <https://doi.org/10.1007/s00362-011-0392-3>

- Gamrot, W. (2014). *Estymacja wartości przeciętnej uwzględniająca koszt pozyskania danych* [Estimation of the average value considering the cost of data acquisition]. Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach.
- Ghosh, M., Mukhopadhyay, N., & Sen, P. K. (1997). *Sequential estimation*. John Wiley & Sons. <https://doi.org/10.1002/9781118165928>
- Groves, R. M. (1989). *Survey errors and survey costs*. John Wiley & Sons.
- Hajek, J. (1959). Optimum strategy and other problems in probability sampling. *Časopis pro pěstování matematiky*, 84, 387-423. <http://dml.cz/dmlcz/117317>
- Hedayat, A. S., & Sinha, B. K. (1991). *Design and inference in finite population sampling*. Wiley.
- Kadane, J. B. (2005). Optimal dynamic sample allocation among strata. *Journal of Official Statistics*, 21, 531-541. <https://doi.org/10.1184/R1/6586808.v1>
- Katris, C. (2020). Prediction of unemployment rates with time series and machine learning techniques. *Computational Economics*, 55(2), 673-706. <https://doi.org/10.1007/s10614-019-09908-9>
- Kolmogorov, A. N. (1950). Nesmeshchennyye otsenki [Unbiased estimates]. *Izvestiya Akademii Nauk SSSR. Seriya Matematicheskaya*, 14(4), 303-326. <https://www.mathnet.ru/links/2115df6e678ab1539a76c87d39e55abe/im3274.pdf>
- Kończak, G. (2007). *Metody statystyczne w sterowaniu jakością produkcji* [Statistical methods in quality control of production]. Wydawnictwo Akademii Ekonomicznej w Katowicach.
- Kremers, W. K. (1986). Completeness and unbiased estimation for sum-quota sampling. *Journal of the American Statistical Association*, 81(396), 1070-1073. <https://doi.org/10.1080/01621459.1986.10478375>
- Krzciuk, M., & Żądło, T. (2014). On some tests of fixed effects for linear mixed models. *Studia Ekonomiczne*, 189, 49-57. https://www.ue.katowice.pl/fileadmin/_migrated/content_uploads/5_M.K.Krzciuk_T.Zadlo_On_some_tests_of_fixed..._01.pdf
- Lehtonen, R., & Pahkinen, E. (2004). *Practical methods for design and analysis of complex surveys*. Wiley.
- Nasiłowski, M. (1994). *System rynkowy. Podstawy mikro- i makroekonomii* [Market system. Basics of microeconomics and macroeconomics]. Key Text.
- Niemiro, W., & Wesołowski, J. (2012). Linear estimation and prediction under model-design approach with small area effects. *Statistics*, 46(4), 523-547. <https://doi.org/10.1080/02331888.2010.542656>
- Owczarek, K. (1996). Badanie próbki jako narzędzie wspomagające rewizję [Sample testing as a tool to support auditing]. *Rachunkowość*, 9, 453-460.
- Owczarek, K. (1998). Próbkowanie według zmiennych – metoda szacowania różnicy i proporcji [Variable sampling – difference and ratio estimation method]. *Rachunkowość*, 11, 586-592.
- Pathak, K. (1976). Unbiased estimation in fixed cost sequential sampling scheme. *Annals of Statistics*, 4(5), 1012-1017. <https://doi.org/10.1214/aos/1176343601>

- Pekasiewicz, D. (2001). Ilorazowe testy sekwencyjne dla frakcji w badaniach małych obszarów [Ratio sequential tests for proportions in small area studies]. *Wiadomości Statystyczne*, 46(10), 1-10.
- Pekasiewicz, D. (2002). Ilorazowe testy sekwencyjne dla frakcji dla prób nieprostych [Ratio sequential tests for proportions for complex samples]. *Acta Universitatis Lodziensis. Folia Oeconomica*, 162(16), 86-95. http://cejsh.icm.edu.pl/cejsh/element/bwmeta1.element.hdl_11089_6679/c/86-95.pdf
- Pilot, K., Ganczarek-Gamrot, A., & Kania, K. (2024). Dealing with anomalies in day-ahead market prediction using machine learning hybrid model. *Energies*, 17(17), 4436. <https://doi.org/10.3390/en17174436>
- Pfaff, J. (2007). *Wpływ rewizji finansowej na wiarygodność sprawozdania finansowego* [Impact of financial auditing on the credibility of financial statements]. Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach.
- Rao, C. R. (1945). Information and accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, 37(3), 81-91. <https://gwern.net/doc/statistics/1992-rao.pdf>
- Rutherford, D. (2002). *Routledge dictionary of economics* (2nd ed.). Routledge. <https://doi.org/10.4324/9780203000540>
- Särndal, C. E., Swensson, B., & Wretman, J. H. (1992). *Model assisted survey sampling*. Springer Verlag.
- Sickles, R., & Zelenyuk, V. (2019). *Measurement of productivity and efficiency: Theory and practice*. Cambridge University Press. <https://doi.org/10.1017/9781139565981>
- Steczowski, J. (1995). *Metoda reprezentacyjna w badaniach zjawisk ekonomiczno-społecznych* [Representative method in the study of economic and social phenomena]. Wydawnictwo Naukowe PWN.
- Stock, J. H., & Watson, M. W. (2016). Core inflation and trend inflation. *Review of Economics and Statistics*, 98(4), 770-784. https://doi.org/10.1162/REST_a_00608
- Szreder, M. (2010). Losowe i nielosowe próby w badaniach statystycznych [Random and non-random samples in statistical studies]. *Przegląd Statystyczny*, LVII(4), 168-174.
- Tillé, Y. (2006). *Sampling algorithms*. Springer. <https://doi.org/10.1007/0-387-34240-0>
- Valliant, R., Dorfman, A. H., & Royall, R. M. (2000). *Finite population sampling and interference: A prediction approach* (Wiley Series in Survey Methodology). Wiley.
- Wolter, K. M. (2007). *Introduction to variance estimation* (Springer Series of Statistics). Springer. <https://doi.org/10.1007/978-0-387-35099-8>
- Wywiał, J. L. (2010). *Wprowadzenie do metody reprezentacyjnej* [Introduction to the representative method]. Wydawnictwo Akademii Ekonomicznej w Katowicach.
- Yatchew, A. (1998). Nonparametric regression techniques in economics. *Journal of Economic Literature*, 36(2), 669-721. <https://www.jstor.org/stable/2565120>
- Yitzhaki, S. (1996). On using linear regressions in welfare economics. *Journal of Business & Economic Statistics*, 14(4), 478-486. <https://doi.org/10.2307/1392256>

Zhang, Y., Wu, H., & Cheng, L. (2012). Some new deformation formulas about variance and covariance. In *2012 Proceedings of International Conference on Modelling, Identification and Control, Wuhan, China* (pp. 984-989).

Żądło, T. (2015). *Statystyka małych obszarów w badaniach ekonomicznych: podejście modelowe i mieszane* [Small area statistics in economic research: Model-based and mixed approaches]. Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach.