Nonlinear Analysis: Modelling and Control, Vol. 30, No. 3, 551–572
https://doi.org/10.15388/namc.2025.30.41781

Improved methods for estimation in repeated surveys: Combining time series and calibration

Dalius Pumputis

Department of Mathematical Statistics,
Vilnius Gediminas Technical University (VILNIUS TECH),
Saulėtekis ave. 11, LT-10223 Vilnius
dalius.pumputis@vilniustech.lt

Received: November 25, 2024 / **Revised:** April 23, 2025 / **Published online:** April 30, 2025

Abstract. The paper investigates new estimation techniques for repeated surveys, focusing on improving the precision of finite population parameter estimates at the current time t by incorporating auxiliary time series and calibration methods. Repeated surveys generate temporally correlated estimates, which time series models capture effectively. Calibration further enhances estimation by adjusting estimators with auxiliary data, reducing variance, and improving precision. Several new estimators of a time-dependent finite population characteristic (usually the mean, which is used in various statistical analyses) at time t are developed and evaluated under diverse scenarios, considering factors such as the correlation between the errors of the target and auxiliary time series, sampling variance, number of surveys, and model complexity. Numerical results demonstrate that calibrated estimators, particularly those incorporating time series adjustments, achieve superior accuracy in high-correlation settings. Regression-based estimator also shows robust performance across varying conditions, while traditional estimators relying solely on survey data are less precise.

Keywords: auxiliary information, calibration estimators, repeated surveys, time series models.

1 Introduction

Time series analysis in repeated surveys has emerged as a key methodology for capturing dynamic finite population changes, optimizing survey designs, and achieving more precise parameter estimation. Jessen [8] integrated sampling methods for farm surveys, emphasizing the efficiency and bias reduction achievable through geographical stratification and repeated sampling techniques. This foundational work set the stage for the development of rotational and repeated sampling frameworks. Patterson [11] extended this by formalizing methods for sampling on successive occasions, introducing partial replacement of units, and examining efficiency under varying correlations across time, a concept later generalized and elaborated by Eckler [5] in the context of rotation sampling for minimizing variance in population mean estimates. Yates [18] expanded on these methodologies, offering comprehensive guidelines for implementing time series in large-scale censuses and surveys. These early contributions provided the basis for designing survey methodologies that minimize variance and bias of estimators across successive survey waves. Our study builds on these principles by focusing on time series estimation techniques that further enhance precision of estimators in repeated surveys.

Further advancements were driven by Blight and Scott [1], who introduced stochastic models for finite population means evolving as linear Markov processes, and Scott and Smith [14], who applied standard time series methods to repeated surveys, exploring overlapping and nonoverlapping survey designs. These contributions evolved into comprehensive signal extraction methodologies under possibly nonstationary conditions with real-data examples illustrating their application (see Scott et al. [15]). Theoretical and empirical efficiency comparisons, such as those by Jones [9], revealed the superiority of time series estimators under specific conditions. Tam [17] extended the dynamic modeling framework to finite populations, demonstrating its adaptability through maximum likelihood estimation techniques.

Steel and McLaren [16] explored the interplay between survey design and estimation methods, particularly focusing on rotation patterns and their impact on trend and seasonal estimates. Ismail et al. [7] employed simulations to validate the superiority of time series estimators under specific survey and model conditions. Additionally, they used Egyptian annual unemployment rate data in a numerical comparison of standard and time series estimators to illustrate their practical performance. The estimation techniques developed in our paper build upon these methodological foundations, particularly in utilizing time series models to improve the precision of repeated survey estimators. The incorporation of auxiliary series in the proposed estimators aligns with a modeling framework that enhances the efficient utilization of survey data over time.

Additionally, regression-based and calibration methods have been explored to refine estimation. Särndal et al. [13] considered regression estimators that improve finite population mean estimation by incorporating auxiliary data. Deville and Särndal [4] formalized calibration estimators, demonstrating their ability to adjust survey design weights to match known auxiliary totals. Our study leverages these concepts to propose new estimators that combine regression and calibration techniques with time series models, thereby achieving greater efficiency and reduced estimation variance.

More recently, Merkouris [10] proposed a novel composite estimation method for repeated surveys with rotating panels, improving precision through the simultaneous calibration of overlapping samples. This approach enhances efficiency without requiring micromatching, making it more practical than existing methods. Pfeffermann [12] examines key theoretical and applied advancements in time series analysis of repeated survey data over the past 40 years. It concludes with insights into future challenges and potential developments in the field. Our paper contributes to this ongoing discussion by introducing and evaluating novel estimators designed to integrate auxiliary information for better precision.

In this article, we address the problem of improving time series estimation in repeated surveys by developing and analyzing new estimators that efficiently incorporate past survey data and auxiliary time series. These estimators, derived from time series forecasting principles, leverage regression and calibration techniques to enhance estimation precision and reduce variance of the estimators. Section 2 introduces the mathematical derivations for these estimators with a focus on incorporating auxiliary series. Section 3 presents a numerical comparison of the estimators under various simulation settings, exploring the impact of correlation, the number of surveys, sampling variance, and complexities of time series models. The results of numerical modeling form the basis for the conclusions presented in Section 4. Additional details regarding the specifications of the simulation models are provided in the Appendix.

It is important to note that this study does not focus on individual sampling units but rather on aggregated survey estimates over time. Instead of analyzing microlevel data, the methodology relies on the sampling variance to capture the uncertainty in survey estimates. The estimators are derived using time series models that treat survey estimates as stochastic processes rather than collections of individual responses. This approach aligns with the study's objective of improving estimation precision while optimizing the use of available recent and past aggregated data.

2 Estimation

At each time point $i \in T$, where $T = \{1, 2, \dots, t\}$, we consider a finite population denoted by $U_i = \{1, 2, \dots, N_i\}$, where N_i is the population size at time i. The population may change over time due to the addition or removal of units. Let z_i be the study variable defined on U_i, representing the characteristic of interest, and let x_i be an auxiliary variable available for all units in U_i.

The finite population parameter of interest at time i, denoted by θ_i, is a function of the study variable over U_i such as a mean. A probability sampling design is applied at each time point $i \in T$ to select a sample from U_i, and a design-based estimator $\widehat{\theta}_i$ is used to estimate θ_i. The samples at each time point are obtained from nonoverlapping surveys, meaning that sampled units differ across time periods.

In this context, the sequence θ_i, $i \in T$, can be viewed as a realization of a stationary time series defined on $\mathbb{Z}$, where each θ_i represents a finite population parameter (such as a mean) at time i. Similarly, we consider a corresponding auxiliary time series γ_i for $i \in T$, where each γ_i is derived from the auxiliary variable x_i over U_i, for example, the population mean of x_i at time i. Given that the auxiliary variables x_i are known for all $i \in T$, the associated values γ_i are available as well.

We consider that the time series follow the ARMA (autoregressive moving average) model, which is a fundamental approach in time series analysis, combining autoregressive (AR) and moving average (MA) components. The $\mathrm{ARMA}(p, q)$ model is expressed as [2]

$$\theta_i = c + \sum_{j=1}^{p} \lambda_{\theta,\,j}\theta_{i-j} + \sum_{j=1}^{q} \psi_{\theta,\,j}\varepsilon_{\theta,\,i-j} + \varepsilon_{\theta,\,i}. \tag{1}$$

In the equation for the model, θ_i is the value at time i, c is a constant, $\lambda_{\theta,\,j}$, for $j = 1, \dots, p$, are the autoregressive coefficients, $\psi_{\theta,\,j}$, for $j = 1, \dots, q$, are the moving average coefficients, and $\varepsilon_{\theta,\,i}$ is white noise, a random error term with zero mean, constant variance, and no autocorrelation. The autoregressive term of the model captures the relationship between the current value of the time series and its past values, while the moving average term models the relationship between the current value and past errors.

By setting $p = 0$ or $q = 0$ in model (1), we obtain the moving average $\mathrm{MA}(q)$ or autoregressive $\mathrm{AR}(p)$ model, respectively.

We consider here only stationary $\mathrm{ARMA}(p, q)$ time series for which the roots of the characteristic equation

$$1 - \sum_{j=1}^{p} \lambda_{\theta,j} g^j = 0$$

must lie outside the unit circle. Stationarity of a stochastic process means that its unconditional joint probability distribution does not change when shifted in time. Consequently, parameters such as the mean and variance do not change over time.

Residuals, defined as $\hat{\varepsilon}^*_{\theta,i} = \theta_i - \widehat{\widehat{\theta}}_i$, for $i \in T$, are the differences between the observed values θ_i and the predicted values $\widehat{\widehat{\theta}}_i$ obtained from the fitted $\mathrm{ARMA}(p, q)$ model. These predicted values are generated by applying the estimated $\mathrm{ARMA}(p, q)$ model to the sequence θ_i, $i \in T$, providing one-step-ahead forecasts based on past observations and errors. A good model produces residuals resembling white noise, indicating that all patterns in the data have been captured. The AR component captures dependencies from past values, while the MA component models dependencies from past errors.

The applied time series models reflect key properties of repeated surveys by capturing temporal dependence, accounting for sampling errors, integrating auxiliary information, and enabling forecasting. Since repeated surveys produce correlated estimates over time, $\mathrm{ARMA}(p, q)$ models accommodate this structure through autoregressive and moving average components. Measurement noise and sampling variability are explicitly modeled to improve precision. The stationarity assumption ensures stability in estimation, while the models also adapt to survey designs, making them well-suited for repeated survey analysis.

In this article, we focus on repeated surveys conducted at regular time intervals from a finite population to estimate the current value θ_t, assuming that γ_t is known. Scott and Smith [14] and Scott et al. [15] analyzed the last survey and time series estimators

$$\widehat{\theta}_t = \theta_t + e_t, \quad e_t \sim N\bigl(0, S_\theta^2\bigr), \tag{2}$$

$$\widehat{\theta}_{TS,t} = \left(1 - \frac{S_\theta^2}{\nu_\theta^2}\right)\widehat{\theta}_t + \frac{S_\theta^2}{\nu_\theta^2}\widehat{\widehat{\theta}}_t, \tag{3}$$

where $\widehat{\widehat{\theta}}_t$ is the best linear forecast of $\widehat{\theta}_t$ based on a fitted $\mathrm{ARMA}(p, q)$ model to the sequence of past estimates $\{\widehat{\theta}_i\}_{i \in T}$, utilizing past observed values and residuals. The sampling variance S_θ^2 reflects the variability of the design-based estimator $\widehat{\theta}_t$ due to sampling and is defined as the variance of the sampling errors e_t, i.e., $S_\theta^2 = \mathrm{Var}(e_t)$, where the errors e_i, $i \in T$, are assumed to be independent. ν_θ^2 is the variance of the mentioned linear forecast based on previous repeated surveys.

Estimators (2) and (3) were later studied by Ismail et al. [7].

Estimates from repeated surveys contain autocorrelated errors and sampling variability, making traditional approaches like the last survey estimator (2) inefficient. Time series models, particularly ARMA-based approaches, provide a structured way to incorporate past estimates and improve precision by reducing variance and filtering noise.

The need for alternative estimators arises because classical time series estimators, such as estimator (3), do not utilize auxiliary information that can further enhance accuracy. The proposed estimators extend time series methods by integrating auxiliary series through regression-type and calibration-based techniques. While calibration is used to optimize the estimation process, the issue remains one of time series estimation – specifically, how to best use past and auxiliary data to obtain more accurate and stable estimates over time.

Estimating θ_t with high accuracy is essential in repeated surveys, yet direct estimators often suffer from high variance due to sampling errors. While the last survey estimator (2) provides an unbiased but highly variable estimate, the time series estimator (3) improves efficiency by incorporating past survey data. However, it may still be suboptimal under certain correlation structures or when auxiliary information is available but not explicitly utilized. To address these limitations, our study proposes enhanced estimators that integrate time series modeling with calibration and regression-based techniques, leading to improved precision, as demonstrated in Sections 3.

The following subsections introduce four estimators for θ_t that include the use of auxiliary series.

2.1 Regression-type estimator

In Särndal et al.'s book [13], a regression estimator is given for the finite population mean of a variable, say z, and is expressed as

$$\widehat{\mu}_z = \bar{z} + \frac{s_{zx}}{s_x^2}(\mu_x - \bar{x}), \tag{4}$$

where, under a simple random sample, $\bar{z}$ and $\bar{x}$ denote the sample means of the study variable z and the auxiliary variable x, respectively; s_{zx} represents the sample covariance between z and x; s_x^2 corresponds to the sample variance of x; and μ_x refers to the finite population mean of x.

By replacing $\bar{z}$, s_{zx}, s_x^2, μ_x, and $\bar{x}$ in (4) with $y_{\theta,t}$, $\mathrm{Cov}(\{y_{\theta,i}\}_{i\in T}, \{y_{\gamma,i}\}_{i\in T})$, $\mathrm{Var}(\{y_{\gamma,i}\}_{i\in T})$, γ_t, and $y_{\gamma,t}$, respectively, we obtain the following regression-type estimator for θ_t:

$$\widehat{\theta}_{R,t} = y_{\theta,t} + \frac{\mathrm{Cov}\, g(\{y_{\theta,i}\}_{i\in T}, \{y_{\gamma,i}\}_{i\in T})}{\mathrm{Var}(\{y_{\gamma,i}\}_{i\in T})}(\gamma_t - y_{\gamma,t}), \tag{5}$$

where the values $\{y_{\theta,i}\}_{i\in T}$ and $\{y_{\gamma,i}\}_{i\in T}$ are the last survey estimates of $\{\theta_i\}_{i\in T}$ and $\{\gamma_i\}_{i\in T}$, respectively. Both series $\{y_{\theta,i}\}_{i\in T}$ and $\{y_{\gamma,i}\}_{i\in T}$ are based on the same repeated surveys.

2.2 Calibrated estimator

We employ the calibration approach (see Deville and Särndal [4]) to derive a calibrated estimator of θ_t. Note that θ_t can be estimated simply by averaging the values $\{y_{\theta,i}\}_{i\in T}$

(see Box and Jenkins [2]):

$$\widehat{\theta}_t = \frac{1}{t}\sum_{i=1}^{t} y_{\theta,\,i}. \tag{6}$$

This approach is best suited for situations where $\theta_t, \theta_{t-1}, \ldots \theta_1$ are assumed to be stable and relatively unchanged over time.

Rewrite (6) as

$$\widehat{\theta}_t = \sum_{i=1}^{t} d_i y_{\theta,\,i}, \quad d_i = \frac{1}{t}.$$

The weights $d_i = 1/t$, $i = 1, \ldots, t$, can be modified using auxiliary time series to obtain an estimator with a smaller variance. We define here the calibrated estimator of θ_t of the following shape:

$$\widehat{\theta}_{C,\,t} = \sum_{i=1}^{t} w_i y_{\theta,\,i}, \tag{7}$$

where the new (calibrated) weights w_i

- minimize the distance measure

$$D(\mathbf{w}, \boldsymbol{d}) = \sum_{i=1}^{t} \frac{(w_i - d_i)^2}{d_i}; \tag{8}$$

- satisfy the calibration equation

$$\sum_{i=1}^{t} w_i y_{\gamma,\,i} = \gamma_t. \tag{9}$$

The calibration equation (9) is treated as the requirement to use the new weights in order to obtain the exact estimate of the known value γ_t. That is, using the new calibrated weights to estimate the auxiliary known value γ_t, it would be estimated without error. Thus, in the case of quite a high correlation between the study and auxiliary time series, it is natural to expect that the estimate of θ_t will be more accurate when the calibrated weights w_i, $i = 1, \ldots, t$, are applied in (7).

The weights w_i, $i = 1, \ldots, t$, of estimator (7) are given by the following lemma.

Lemma 1. *The weights w_i, $i = 1, \ldots, t$, of estimator* (7)*, which minimize the distance measure* (8) *and satisfy Eq.* (9)*, are given by*

$$w_i = \frac{1}{t}\left(1 + y_{\gamma,\,i}\left(t\gamma_t - \sum_{j=1}^{t} y_{\gamma,\,j}\right)\left(\sum_{j=1}^{t} y_{\gamma,\,j}^2\right)^{-1}\right).$$

Proof. The derivation of the weights w_i, for $i = 1, \ldots, t$, follows a similar approach to that of the calibrated estimator of the finite population total (see Deville and Särndal [4]), beginning with the definition of the Lagrange function

$$\Lambda = \Lambda(\mathbf{w}, \eta) = \sum_{i=1}^{t} \frac{(w_i - d_i)^2}{d_i} - \eta\left(\sum_{i=1}^{t} w_i y_{\gamma,\, i} - \gamma_t\right).$$

Equating the partial derivatives $\partial\Lambda/\partial w_i$, $i = 1, \dots, t$, to zero leads us to the expressions

$$w_i = d_i\left(1 + \frac{1}{2}\eta y_{\gamma,\, i}\right), \quad i = 1, \dots, t. \tag{10}$$

After inserting them into calibration equation (9), we find the solution

$$\eta = 2\left(\gamma_t - \sum_{i=1}^{t} d_i y_{\gamma,\, i}\right)\left(\sum_{i=1}^{t} d_i y_{\gamma,\, i}^2\right)^{-1}.$$

The insertion of this expression into (10) yields the result stated in the lemma. □

2.3 Calibrated last survey estimator

According to the Wold decomposition (see Brockwell and Davis [3]), every stationary ARMA time series can be represented as an infinite moving average (MA) model. Taking the main t components from this decomposition, the last survey estimator (2) for θ_t can be approximated by

$$\widehat{\theta}_t^{(WD)} = \sum_{i=0}^{t-1} \beta_i \varepsilon_{\theta,\, t-i}^*, \tag{11}$$

where $\beta_0, \beta_1, \dots, \beta_{t-1}$ are the weights in a moving average process, $\beta_0 = 1$, $\varepsilon_{\theta,\, t}^*, \varepsilon_{\theta,\, t-1}^*, \dots, \varepsilon_{\theta,\, 1}^*$, are the residuals derived from a time series model that best fits the set $\{y_{\theta,\, i}\}_{i\in T}$ according to the Box–Jenkins methodology (see Box and Jenkins [2]).

As defined in (11), this form of the estimator suggests the following shape and definition of a new calibrated estimator:

$$\widehat{\theta}_{CLS,\, t} = \sum_{i=0}^{t-1} w_{C,i} \varepsilon_{\theta,\, t-i}^*, \tag{12}$$

where the calibrated weights $w_{C,i}$

- minimize the distance measure

$$D_C(\mathbf{w}, \boldsymbol{\beta}) = \sum_{i=0}^{t-1} \frac{(w_{C,i} - \beta_i)^2}{\beta_i}; \tag{13}$$

- satisfy the calibration equation

$$\sum_{i=0}^{t-1} w_{C,i} \varepsilon_{\gamma,\, t-i}^* = \gamma_t, \tag{14}$$

where $\varepsilon_{\gamma,\, t}^*, \varepsilon_{\gamma,\, t-1}^*, \dots, \varepsilon_{\gamma,\, 1}^*$ are the residuals derived from a time series model that best fits the set $\{y_{\gamma,\, i}\}_{i\in T}$.

The solution to this problem is presented in the following corollary, which follows from Lemma 1 by observing that the weights β_i, for $i = 0, \dots, t-1$, and the residuals $\varepsilon^*_{\gamma,\,i}$, for $i = t, \dots, 1$, correspond to the weights d_i and the estimates $y_{\gamma,\,i}$, respectively, for $i = 1, \dots, t$.

Corollary 1. *The weights* $w_{C,i}$, $i = 0, \dots, t-1$, *of estimator* (12), *which minimize the distance measure* (13) *and satisfy Eq.* (14), *are given by*

$$w_{C,i} = \beta_i \left(1 + \varepsilon^*_{\gamma,\,t-i} \left(\gamma_t - \sum_{j=0}^{t-1} \beta_j \varepsilon^*_{\gamma,\,t-j} \right) \left(\sum_{j=0}^{t-1} \beta_j \left(\varepsilon^*_{\gamma,\,t-j}\right)^2 \right)^{-1} \right).$$

2.4 Calibrated time series estimator

The time series estimator (3) may be approximated by expressing $\widehat{\theta}_t$ and its forecast $\widehat{\widehat{\theta}}_t$ in terms of the residuals $\varepsilon^*_{\theta,\,t}, \varepsilon^*_{\theta,\,t-1}, \dots, \varepsilon^*_{\theta,\,1}$. Thus, in the general case, $\hat{y}_{\theta,\,t}$ can be viewed as a linear function f of ε^*_θ's. The approximation to (3) then becomes

$$\widehat{\theta}^{(WD)}_{TS,\,t} = \left(1 - \frac{S^2_\theta}{\nu^2_\theta} \right) \sum_{i=0}^{t-1} \beta_i \varepsilon^*_{\theta,\,t-i} + \frac{S^2_\theta}{\nu^2_\theta} f\left(\varepsilon^*_{\theta,\,t}, \varepsilon^*_{\theta,\,t-1}, \dots, \varepsilon^*_{\theta,\,1}\right).$$

After some simplifications, $\widehat{\theta}^{(WD)}_{TS,\,t}$ can be written in the linear form:

$$\widehat{\theta}^{(WD)}_{TS,\,t} = \sum_{i=0}^{t-1} \beta_{TS,\,i} \varepsilon^*_{\theta,\,t-i}, \tag{15}$$

where the weights $\beta_{TS,\,i}$, $i = 1, \dots, t-1$, depend on the function f as well as the weights β_i, for $i = 0, 1, \dots, t-1$, and the variances S^2_θ and ν^2_θ.

Below are three examples of time series models for $\{\theta_i\}_{i \in \mathbb{Z}}$, along with the corresponding weights $\beta_{TS,\,0}, \beta_{TS,1}, \dots, \beta_{TS,\,t-1}$, derived from each model.

Example 1. Consider the $\mathrm{AR}(1)$ model for the process $\{\theta_i\}$:

$$\theta_i = \lambda_\theta \theta_{i-1} + \varepsilon_{\theta,\,i}, \quad |\lambda_\theta| < 1, \quad \varepsilon_{\theta,\,i} \sim N\left(0, \sigma^2_\theta\right), \quad i \in \mathbb{Z}.$$

Next, we analyze the model that governs the time series $\{y_{\theta,\,i}\}$, defined as

$$y_{\theta,\,i} = \theta_i + e_i, \quad e_i \sim N\left(0, S^2_\theta\right), \quad i \in \mathbb{Z}.$$

Using the backshift operator B, where $B\theta_i = \theta_{i-1}$, the $\mathrm{AR}(1)$ model for $\{\theta_i\}_{i \in \mathbb{Z}}$ can be expressed as

$$\phi(B)\theta_i = \varepsilon_{\theta,\,i}, \quad \text{where } \phi(B) = 1 - \lambda_\theta B.$$

Applying $\phi(B)$ to both sides of the equation for $y_{\theta,\,i}$, we obtain

$$\phi(B) y_{\theta,\,i} = \phi(B)\theta_i + \phi(B) e_i.$$

Since $\phi(B)\theta_i = \varepsilon_{\theta,\,i}$, we have

$$\phi(B)y_{\theta,\,i} = \varepsilon_{\theta,\,i} + \phi(B)e_i. \tag{16}$$

The right-hand side of Eq. (16), $\varepsilon_{\theta,\,i} + \phi(B)e_i$, is a linear combination of white noise processes, which is therefore a stationary process. According to the Wold decomposition theorem (see Brockwell and Davis [3]), this process can be represented as an infinite moving average (MA) process. However, since $\phi(B) = 1 - \lambda_\theta B$ is a first-order polynomial, the autocovariance function cuts of after lag 1. This implies that the process in the right-hand side of Eq. (16) is an $\mathrm{MA}(1)$ process, say $(1 + \psi_{y_\theta} B)u_i$, where ψ_{y_θ} is the moving average coefficient, and u_i is white noise with variance σ_u^2.

Now we can write model (16) as

$$(1 - \lambda_\theta B)y_{\theta,\,i} = (1 + \psi_{y_\theta} B)u_i. \tag{17}$$

This proves that $y_{\theta,\,i}$, for $i \in \mathbb{Z}$, follows an $\mathrm{ARMA}(1,1)$ model.

From the left-hand side of Eq. (17) it is clear that the autoregressive parameter for this $\mathrm{ARMA}(1,1)$ model matches that of $\{\theta_i\}_{i\in T}$, i.e., $\lambda_{y_\theta} = \lambda_\theta$. The moving average parameter ψ_{y_θ} and the white noise variance σ_u^2 of this process can be determined by equating the autocovariance function on the right-hand side of Eq. (16) with that of $(1 + \psi_{y_\theta} B)u_i$ for lags $k = 0, 1$ and solving the resulting system of two nonlinear equations

$$\begin{aligned}(1 + \psi_{y_\theta}^2)\sigma_u^2 &= \sigma_\theta^2 + \bigl(1 + \lambda_\theta^2\bigr)S_\theta^2,\\ -\psi_{y_\theta}\sigma_u^2 &= \lambda_\theta S_\theta^2.\end{aligned}$$

Since $\{y_{\theta,\,i}\}_{i\in\mathbb{Z}}$ follows an $\mathrm{ARMA}(1,1)$ model, the time series estimator (3) can be written as

$$\begin{aligned}\widehat{\theta}_{TS,\,t} &= \left(1 - \frac{S_\theta^2}{\nu_\theta^2}\right)y_{\theta,\,t} + \frac{S_\theta^2}{\nu_\theta^2}(\lambda_{y_\theta} y_{\theta,\,t-1} + \psi_{y_\theta}\varepsilon^*_{\theta,\,t-1})\\ &\approx \left(1 - \frac{S_\theta^2}{\nu_\theta^2}\right)\sum_{i=0}^{t-1}\beta_i\varepsilon^*_{\theta,\,t-i} + \frac{S_\theta^2}{\nu_\theta^2}(\lambda_{y_\theta} y_{\theta,\,t-1} + \psi_{y_\theta}\varepsilon^*_{\theta,\,t-1}),\end{aligned} \tag{18}$$

where the symbol $\approx$ signifies that the second equation provides an approximation of the first. The exact form includes $y_{\theta,\,t}$, while the approximation replaces it with a weighted sum of past residuals, $\sum_{i=0}^{t-1}\beta_i\varepsilon^*_{\theta,\,t-i}$. This substitution is useful to define the new calibrated estimator of θ_t.

After algebraic manipulations, formula (18) simplifies to

$$\widehat{\theta}_{TS,\,t} \approx \widehat{\theta}_{TS,\,t}^{(WD)} = \sum_{i=0}^{t-1}\beta_{TS,\,i}\varepsilon^*_{\theta,\,t-i},$$

where

$$\beta_{TS,\,0} = 1 - \frac{S_\theta^2}{\nu_\theta^2},$$

$$\beta_{TS,\,i} = \left(1 - \frac{S_\theta^2}{\nu_\theta^2}\right)\beta_{t-i} + \frac{S_\theta^2}{\nu_\theta^2}\lambda_{y_\theta}\beta_{t-i-1}, \quad i = 1, \dots, t-2,$$
$$\beta_{TS,\,t-1} = \left(1 - \frac{S_\theta^2}{\nu_\theta^2}\right)\beta_1 + \frac{S_\theta^2}{\nu_\theta^2}(\lambda_{y_\theta} + \psi_{y_\theta}).$$

Example 2. Similarly as in Example 1, $y_{\theta,\,i} = \theta_i + e_i$, $e_i \sim N(0, S_\theta^2)$, $i \in \mathbb{Z}$, follow $\mathrm{ARMA}(2,2)$ model if $\{\theta_i\}_{i\in\mathbb{Z}}$ are from stationary $\mathrm{AR}(2)$ model. Denote the autoregressive and moving average coefficients of this process by $\lambda_{y_\theta,1}$, $\lambda_{y_\theta,2}$, $\psi_{y_\theta,1}$, $\psi_{y_\theta,2}$, respectively. Then estimator (3) becomes

$$\begin{aligned}\widehat{\theta}_{TS,\,t} &= \left(1 - \frac{S_\theta^2}{\nu_\theta^2}\right)y_{\theta,\,t} \\ &\quad + \frac{S_\theta^2}{\nu_\theta^2}(\lambda_{y_\theta,1}y_{\theta,\,t-1} + \lambda_{y_\theta,2}y_{\theta,\,t-2} + \psi_{y_\theta,1}\varepsilon^*_{\theta,\,t-1} + \psi_{y_\theta,2}\varepsilon^*_{\theta,\,t-2}) \\ &\approx \left(1 - \frac{S_\theta^2}{\nu_\theta^2}\right)\sum_{i=0}^{t-1}\beta_i\varepsilon^*_{\theta,\,t-i} \\ &\quad + \frac{S_\theta^2}{\nu_\theta^2}(\lambda_{y_\theta,1}y_{\theta,\,t-1} + \lambda_{y_\theta,2}y_{\theta,\,t-2} + \psi_{y_\theta,1}\varepsilon^*_{\theta,\,t-1} + \psi_{y_\theta,2}\varepsilon^*_{\theta,\,t-2}).\end{aligned}$$

By combining similar terms, we obtain the expression

$$\widehat{\theta}_{TS,\,t} \approx \widehat{\theta}^{(WD)}_{TS,\,t} = \sum_{i=0}^{t-1}\beta_{TS,\,i}\varepsilon^*_{\theta,\,t-i},$$

where

$$\begin{aligned}\beta_{TS,\,0} &= 1 - \frac{S_\theta^2}{\nu_\theta^2}, \\ \beta_{TS,\,i} &= \left(1 - \frac{S_\theta^2}{\nu_\theta^2}\right)\beta_{t-i} + \frac{S_\theta^2}{\nu_\theta^2}(\lambda_{y_\theta,1}\beta_{t-i-1} + \lambda_{y_\theta,2}\beta_{t-i-2}), \quad i = 1, \dots, t-3, \\ \beta_{TS,\,t-2} &= \left(1 - \frac{S_\theta^2}{\nu_\theta^2}\right)\beta_2 + \frac{S_\theta^2}{\nu_\theta^2}(\lambda_{y_\theta,1}\beta_1 + \lambda_{y_\theta,2} + \psi_{y_\theta,2}), \\ \beta_{TS,\,t-1} &= \left(1 - \frac{S_\theta^2}{\nu_\theta^2}\right)\beta_1 + \frac{S_\theta^2}{\nu_\theta^2}(\lambda_{y_\theta,1} + \psi_{y_\theta,1}).\end{aligned}$$

Example 3. If $\{\theta_i\}_{i\in\mathbb{Z}}$ follows $\mathrm{MA}(2)$ model, then $y_{\theta,\,i} = \theta_i + e_i$, $e_i \sim N(0, S_\theta^2)$, $i \in \mathbb{Z}$, are defined also by $\mathrm{MA}(2)$ model, but with different coefficients, specifically $\psi_{y_\theta,1}$ and $\psi_{y_\theta,2}$. By substituting the forecast $\widehat{\widehat{\theta}}_t = \psi_{y_\theta,1}\varepsilon^*_{\theta,\,t-1} + \psi_{y_\theta,2}\varepsilon^*_{\theta,\,t-2}$ into (3), we get

$$\begin{aligned}\widehat{\theta}_{TS,\,t} &= \left(1 - \frac{S_\theta^2}{\nu_\theta^2}\right)y_{\theta,\,t} + \frac{S_\theta^2}{\nu_\theta^2}(\psi_{y_\theta,1}\varepsilon^*_{\theta,\,t-1} + \psi_{y_\theta,2}\varepsilon^*_{\theta,\,t-2}) \\ &\approx \left(1 - \frac{S_\theta^2}{\nu_\theta^2}\right)\sum_{i=0}^{t-1}\beta_i\varepsilon^*_{\theta,\,t-i} + \frac{S_\theta^2}{\nu_\theta^2}(\psi_{y_\theta,1}\varepsilon^*_{\theta,\,t-1} + \psi_{y_\theta,2}\varepsilon^*_{\theta,\,t-2}).\end{aligned}$$

A few simplifications yield this formula:

$$\widehat{\theta}_{TS,\,t} \approx \widehat{\theta}^{(WD)}_{TS,\,t} = \sum_{i=0}^{t-1} \beta_{TS,\,i}\varepsilon^*_{\theta,\,t-i},$$

where

$$\beta_{TS,\,0} = 1 - \frac{S^2_\theta}{\nu^2_\theta}, \quad \beta_{TS,\,i} = 0, \quad i = 1, \dots, t-3,$$
$$\beta_{TS,\,t-2} = \psi_{y_\theta,2}, \qquad \beta_{TS,\,t-1} = \psi_{y_\theta,1}.$$

Based on expression (15), the calibrated times series estimator is defined by the formula

$$\widehat{\theta}_{CTS,\,t} = \sum_{i=0}^{t-1} w_{TS,\,i}\varepsilon^*_{\theta,\,t-i}, \tag{19}$$

where the calibrated weights $w_{TS,\,i}$

- minimize the distance measure

$$D_{TS}(\mathbf{w}, \boldsymbol{\beta}) = \sum_{i=0}^{t-1} \frac{(w_{TS,\,i} - \beta_{TS,\,i})^2}{\beta_{TS,\,i}}; \tag{20}$$

- satisfy the calibration equation

$$\sum_{i=0}^{t-1} w_{TS,\,i}\varepsilon^*_{\gamma,\,t-i} = \gamma_t. \tag{21}$$

The solution of this problem is given in the following corollary.

Corollary 2. *The weights $w_{TS,\,i}$, $i = 0, \dots, t-1$, of estimator* (19)*, which minimize the distance measure* (20) *and satisfy Eq.* (21)*, are given by*

$$w_{TS,\,i} = \beta_{TS,\,i}\left(1 + \varepsilon^*_{\gamma,\,t-i}\left(\gamma_t - \sum_{j=0}^{t-1} \beta_{TS,j}\varepsilon^*_{\gamma,\,t-j}\right)\left(\sum_{j=0}^{t-1} \beta_{TS,j}(\varepsilon^*_{\gamma,\,t-j})^2\right)^{-1}\right).$$

The results of Corollary 2 follow from Lemma 1 by noting that the weights $\beta_{TS,\,i}$, for $i = 0, \dots, t-1$, and the residuals $\varepsilon^*_{\gamma,\,i}$, for $i = t, \dots, 1$, correspond to the weights d_i and the estimates $y_{\gamma,\,i}$, respectively, for $i = 1, \dots, t$.

The calibration approach used in this paper is suitable for time series because, similar to survey sampling, minimizing the distance function results in a nearly unbiased estimator. Additionally, the calibration equation adjusts the weights in the linear combination so that the newly assigned weights for the auxiliary series yield exact estimates. Consequently, in settings with high correlation, these calibrated weights produce highly precise calibrated estimators of θ_t.

3 Numerical comparisons

The last survey $\widehat{\theta}_t$ (2), time series $\widehat{\theta}_{TS,\,t}$ (3), regression-type $\widehat{\theta}_{R,\,t}$ (5), calibrated $\widehat{\theta}_{C,\,t}$ (7), calibrated last survey $\widehat{\theta}_{CLS,\,t}$ (12), and calibrated time series $\widehat{\theta}_{CTS,\,t}$ (19) estimators are compared using time series models presented in Table 1, according to which study and auxiliary time series, $\{\theta_i\}$ and $\{\gamma_i\}$, $i = 1, 2, \dots, 700$, are generated, representing the modeled finite population means created under the assumption that the initial values are zero. The implementation and comparison of these estimators are carried out using the R programming language. Detailed model specifications, including model equations, coefficients, and the variance of model residuals, are provided in the Appendix.

The comparison of estimators is also conducted depending on the following factors.

- The sampling variances S^2_θ and S^2_γ of the last survey estimators for θ_t and γ_t, respectively. The values for S^2_θ are 0.49, 1, or 2.25, while those for S^2_γ are 0.7, 1.2, or 2.5.
- Numbers of repeated surveys t representing the number of sample means. The simulation includes survey counts of 15, 30, 60, or 120, where 15 is the minimum, and 120 is the maximum number of repeated surveys. To examine the effect of series size, each selected survey number t is formed by taking the last t values in the series of sample means with a total size of 700, removing the first $700 - t$ values of each series to eliminate the initialization effect. This ensures that the sample means included in smaller series sizes are also part of the larger series sizes. For example, if $t = 30$, the sample means from the series size of 15 are included in the larger series.
- The correlation $\rho(\varepsilon_\theta, \varepsilon_\gamma)$ between the errors $\varepsilon_{\theta,\,700}, \varepsilon_{\theta,\,699}, \dots, \varepsilon_{\theta,\,1}$ and $\varepsilon_{\gamma,\,700}$, $\varepsilon_{\gamma,\,699}, \dots, \varepsilon_{\gamma,\,1}$, which takes values of 0.9 or 0.6, characterizes the relationship between the errors used to generate the series $\theta_{700}, \theta_{699}, \dots, \theta_1$ and $\gamma_{700}, \gamma_{699}, \dots, \gamma_1$ according to the selected models.

For each scenario and each specific combination of values for the factors S^2_θ, S^2_γ, σ^2_θ, σ^2_γ, t, and $\rho(\varepsilon_\theta, \varepsilon_\gamma)$, $M = 10^5$ estimates of $\widehat{\theta}_t$, $\widehat{\theta}_{TS,\,t}$, $\widehat{\theta}_{R,\,t}$, $\widehat{\theta}_{C,\,t}$, $\widehat{\theta}_{CLS,\,t}$, and $\widehat{\theta}_{CTS,\,t}$ are calculated. These estimates are then used to estimate the quartiles of the mean square error (MSE) of the estimators for particular combinations of the mentioned factors, enabling an assessment of estimators' efficiency.

Specifically, for one particular combination of factor values, the MSE of an estimator, say $\widehat{\theta}_t$, is estimated using the formula

$$\mathrm{MSE}(\widehat{\theta}_t) = \frac{1}{M} \sum_{m=1}^{M} (\widehat{\theta}_{t,m} - \theta_t)^2,$$

where $\widehat{\theta}_{t,m}$ denotes the estimate of θ_t obtained from the mth simulated time series $\{y_{\theta,\,i}\}_{i \in T}$.

For example, to obtain the quartiles of the MSE for $t = 15$, $\rho(\varepsilon_\theta, \varepsilon_\gamma) = 0.9$, and the $\mathrm{AR}(1)$ model for γ's in Scenario 1, we consider all MSE estimates across all combinations of the values of S^2_θ, S^2_γ, σ^2_θ, σ^2_γ, and the coefficients of the $\mathrm{AR}(1)$ model used in this scenario. Here the shorthand γ's refers to the set $\{\gamma_i\}_{i \in 1,2,\dots,700}$.

Table 1. Overview of model scenarios.

Scenario	Model for study time series $\{\theta_i\}$	Models for auxiliary time series $\{\gamma_i\}$
1	$\mathrm{AR}(1)$	$\mathrm{AR}(1)$ or $\mathrm{ARMA}(1,1)$
2	$\mathrm{AR}(2)$	$\mathrm{AR}(2)$ or $\mathrm{ARMA}(2,2)$
3	$\mathrm{MA}(2)$	$\mathrm{MA}(2)$ or $\mathrm{AR}(1)$

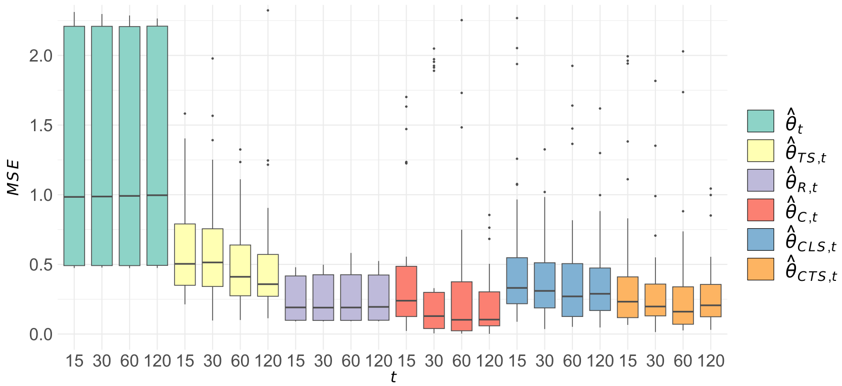

Figure 1. Boxplots of the MSE of estimators under Scenario 1 across values of t, for $\rho(\varepsilon_\theta, \varepsilon_\gamma) = 0.9$, assuming an $\mathrm{AR}(1)$ model for the γ's.

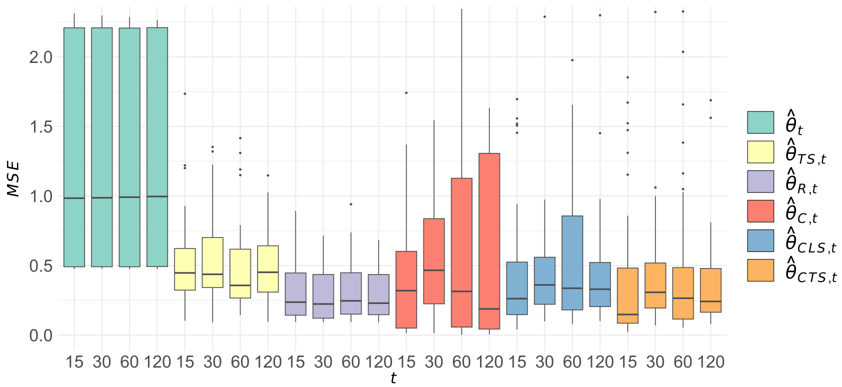

Figure 2. Boxplots of the MSE of estimators under Scenario 1 across values of t, for $\rho(\varepsilon_\theta, \varepsilon_\gamma) = 0.6$, assuming an $\mathrm{AR}(1)$ model for the γ's.

Comparison of estimators across values of t. In Figs. 1–4, we present boxplots of the MSE of estimators under Scenario 1 across values of t and $\rho(\varepsilon_\theta, \varepsilon_\gamma)$ for both $\mathrm{AR}(1)$ and $\mathrm{ARMA}(1,1)$ models for the γ's. As the results across values of t for Scenarios 2 and 3 exhibit similar patterns to those in Scenario 1, we focus our presentation on Scenario 1.

Calibration significantly reduces MSE for all calibrated estimators with its impact increasing as the correlation $\rho(\varepsilon_\theta, \varepsilon_\gamma)$ rises. When $\rho = 0.9$, calibrated estimators like

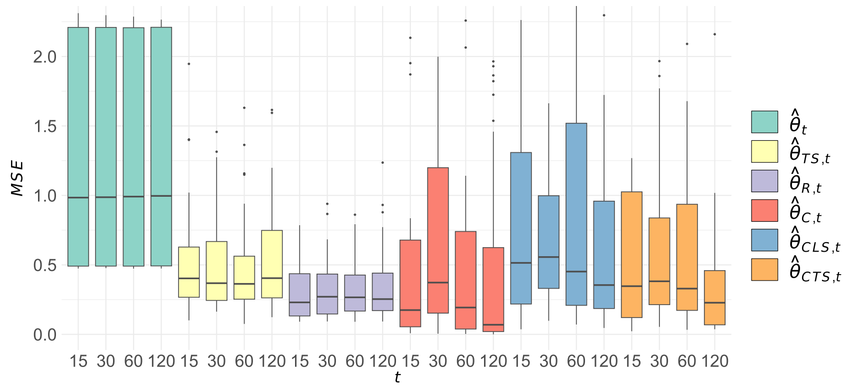

Figure 3. Boxplots of the MSE of estimators under Scenario 1 across values of t, for $\rho(\varepsilon_\theta, \varepsilon_\gamma) = 0.9$, assuming an $\mathrm{ARMA}(1,1)$ model for the γ's.

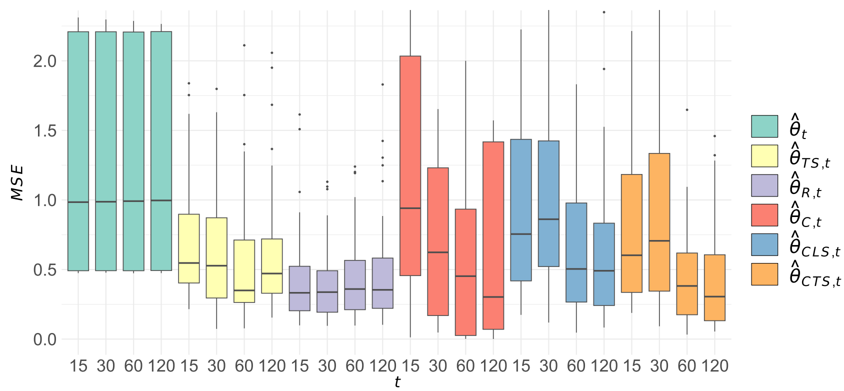

Figure 4. Boxplots of the MSE of estimators under Scenario 1 across values of t, for $\rho(\varepsilon_\theta, \varepsilon_\gamma) = 0.6$, assuming an $\mathrm{ARMA}(1,1)$ model for the γ's.

$\widehat{\theta}_{CTS,\,t}$ and $\widehat{\theta}_{CLS,\,t}$ achieve much lower MSE than $\widehat{\theta}_{TS,\,t}$ and $\widehat{\theta}_t$, highlighting calibration's effectiveness in leveraging auxiliary information.

Among calibrated estimators, $\widehat{\theta}_{CTS,\,t}$ consistently yields the lowest MSE in high-correlation settings by integrating time series and calibration adjustments. The regression estimator $\widehat{\theta}_{R,\,t}$ also performs well, maintaining low MSE across varying correlations by systematically incorporating auxiliary data. As ρ increases, $\widehat{\theta}_{R,\,t}$ achieves MSE values close to the best-performing calibrated estimators.

With more repeated surveys (t), the MSE decreases for most estimators, particularly when $\rho = 0.9$, except for $\widehat{\theta}_t$, which does not depend on t. However, at moderate correlation ($\rho = 0.6$), the reduction is less pronounced as auxiliary data has a weaker influence.

Under both the $\mathrm{AR}(1)$ and $\mathrm{ARMA}(1,1)$ models, the regression estimator $\widehat{\theta}_{R,\,t}$ outperforms the others. The $\mathrm{AR}(1)$ model yields slightly lower MSE for all calibrated and regression estimators, particularly at high values of ρ, suggesting that its simpler structure allows for better utilization of strong correlations. At moderate ρ, the choice of model has minimal effect on MSE, reinforcing that $\mathrm{AR}(1)$ is most beneficial under high correlation.

Comparison of estimators across sampling variances S_θ^2 and S_γ^2. The estimators of θ_t are compared across different combinations of sampling variances S_θ^2, S_γ^2 and correlation $\rho(\varepsilon_\theta, \varepsilon_\gamma)$. Due to the large number of tables, only key results are presented.

Except for $\widehat{\theta}_{C,\,t}$, estimator variability increases with higher sampling variance. Interestingly, $\widehat{\theta}_{C,\,t}$ remains relatively robust, showing similar quartiles across different S_θ^2 and S_γ^2, though they are larger than those of other calibration-based estimators.

$\widehat{\theta}_{CLS,\,t}$ and $\widehat{\theta}_{CTS,\,t}$ perform well with high $\rho(\varepsilon_\theta, \varepsilon_\gamma)$ and simpler auxiliary models but sometimes underperform compared to $\widehat{\theta}_{TS,\,t}$, especially with larger S_θ^2 and S_γ^2. Thus, $\widehat{\theta}_{TS,\,t}$ remains reliable when the auxiliary information used for calibration has a low correlation with the study's time series.

The regression estimator $\widehat{\theta}_{R,\,t}$ shows the smallest MSE quartile differences, indicating consistent, low-variability performance. The last survey estimator's MSE closely matches S_θ^2, aligning with theoretical expectation.

The overall comparison of the estimators. Table 2 displays the overall quartiles of the MSE for estimators under Scenario 1, incorporating both $\mathrm{AR}(1)$ and $\mathrm{ARMA}(1,1)$ models for the γ's across different levels of correlation $\rho(\varepsilon_\theta, \varepsilon_\gamma)$. For example, to obtain the overall quartiles of the MSE for $\rho(\varepsilon_\theta, \varepsilon_\gamma) = 0.9$ under the $\mathrm{AR}(1)$ model for the γ's in Scenario 1, we aggregate all MSE estimates across all combinations of the values of t, S_θ^2, S_γ^2, σ_θ^2, σ_γ^2, and the coefficients of the $\mathrm{AR}(1)$ model used in this scenario.

Calibration, especially in $\widehat{\theta}_{CTS,\,t}$, is most effective when correlation is high (≈ 0.9), minimizing MSE. Under $\mathrm{AR}(1)$, strong dependency enhances calibration, but its advantage diminishes with $\mathrm{ARMA}(1,1)$ or weaker correlation (0.6), increasing MSE variability.

The time series estimator $\widehat{\theta}_{TS,\,t}$ remains stable with low MSE, ensuring reliability across scenarios. While it is consistent, calibrated estimators like $\widehat{\theta}_{CTS,\,t}$ achieve lower MSE in simpler cases.

The regression estimator $\widehat{\theta}_{R,\,t}$ adapts well across correlation levels, maintaining low MSE even with $\mathrm{ARMA}(1,1)$. Meanwhile, the last survey estimator $\widehat{\theta}_t$ has the highest MSE as it relies only on recent data.

Table 3 presents the overall quartiles of the MSE for estimators in Scenario 2, incorporating both $\mathrm{AR}(2)$ and $\mathrm{ARMA}(2,2)$ models for the γ's at two levels of correlation $\rho(\varepsilon_\theta, \varepsilon_\gamma)$. When correlation is high ($\rho(\varepsilon_\theta, \varepsilon_\gamma) = 0.9$), the calibrated time series estimator $\widehat{\theta}_{CTS,\,t}$ achieves the lowest MSE across all quartiles in the $\mathrm{AR}(2)$ model by leveraging correlation for improved accuracy. In the $\mathrm{ARMA}(2,2)$ model, $\widehat{\theta}_{C,\,t}$ performs best in the first quartile, while $\widehat{\theta}_{CTS,\,t}$ remains competitive, particularly in the middle and upper quartiles. The calibrated last survey estimator $\widehat{\theta}_{CLS,\,t}$ performs similarly to $\widehat{\theta}_{CTS,\,t}$ but is slightly less accurate in more complex models. The regression estimator $\widehat{\theta}_{R,\,t}$ shows higher MSE in both models.

When correlation decreases ($\rho(\varepsilon_\theta, \varepsilon_\gamma) = 0.6$), $\widehat{\theta}_{CTS,\,t}$ still performs well in the $\mathrm{AR}(2)$ model, though with slightly higher MSE. In the $\mathrm{ARMA}(2,2)$ model, it maintains the

Table 2. Overall quartiles of the MSE of estimators under Scenario 1 with $\mathrm{AR}(1)$ model (on the left) and $\mathrm{ARMA}(1,1)$ model (on the right) for γ's.

Quartiles	$\widehat{\theta}_t$	$\widehat{\theta}_{TS,t}$	$\widehat{\theta}_{R,t}$	$\widehat{\theta}_{C,t}$	$\widehat{\theta}_{CLS,t}$	$\widehat{\theta}_{CTS,t}$	Quartiles	$\widehat{\theta}_t$	$\widehat{\theta}_{TS,t}$	$\widehat{\theta}_{R,t}$	$\widehat{\theta}_{C,t}$	$\widehat{\theta}_{CLS,t}$	$\widehat{\theta}_{CTS,t}$
	$\rho(\varepsilon_\theta,\varepsilon_\gamma)=0.9$							$\rho(\varepsilon_\theta,\varepsilon_\gamma)=0.9$					
Q_1	0.491	0.293	0.098	0.041	0.169	0.110	Q_1	0.491	0.293	0.152	0.043	0.244	0.154
Q_2	0.989	0.420	0.191	0.164	0.299	0.197	Q_2	0.989	0.420	0.245	0.213	0.486	0.314
Q_3	2.209	0.667	0.421	0.353	0.513	0.369	Q_3	2.209	0.667	0.436	0.750	1.176	0.813
	$\rho(\varepsilon_\theta,\varepsilon_\gamma)=0.6$							$\rho(\varepsilon_\theta,\varepsilon_\gamma)=0.6$					
Q_1	0.491	0.293	0.143	0.057	0.187	0.128	Q_1	0.491	0.293	0.203	0.139	0.341	0.226
Q_2	0.989	0.420	0.233	0.360	0.325	0.240	Q_2	0.989	0.420	0.350	0.563	0.657	0.531
Q_3	2.209	0.667	0.443	0.962	0.608	0.496	Q_3	2.209	0.667	0.568	1.328	1.298	0.924

Table 3. Overall quartiles of the MSE of estimators under Scenario 2 with $\mathrm{AR}(2)$ model (on the left) and $\mathrm{ARMA}(2,2)$ model (on the right) for γ's.

Quartiles	$\widehat{\theta}_t$	$\widehat{\theta}_{TS,t}$	$\widehat{\theta}_{R,t}$	$\widehat{\theta}_{C,t}$	$\widehat{\theta}_{CLS,t}$	$\widehat{\theta}_{CTS,t}$	Quartiles	$\widehat{\theta}_t$	$\widehat{\theta}_{TS,t}$	$\widehat{\theta}_{R,t}$	$\widehat{\theta}_{C,t}$	$\widehat{\theta}_{CLS,t}$	$\widehat{\theta}_{CTS,t}$
	$\rho(\varepsilon_\theta,\varepsilon_\gamma)=0.9$							$\rho(\varepsilon_\theta,\varepsilon_\gamma)=0.9$					
Q_1	0.491	0.323	0.113	0.122	0.150	0.084	Q_1	0.491	0.323	0.122	0.072	0.212	0.133
Q_2	0.995	0.487	0.196	0.343	0.219	0.123	Q_2	0.995	0.487	0.220	0.302	0.347	0.239
Q_3	2.212	0.774	0.426	0.971	0.411	0.222	Q_3	2.212	0.774	0.440	0.805	0.509	0.393
	$\rho(\varepsilon_\theta,\varepsilon_\gamma)=0.6$							$\rho(\varepsilon_\theta,\varepsilon_\gamma)=0.6$					
Q_1	0.491	0.323	0.189	0.192	0.172	0.107	Q_1	0.491	0.323	0.186	0.184	0.237	0.128
Q_2	0.995	0.487	0.274	0.608	0.270	0.196	Q_2	0.995	0.487	0.275	0.802	0.369	0.229
Q_3	2.212	0.774	0.476	2.429	0.461	0.441	Q_3	2.212	0.774	0.480	2.575	0.540	0.467

Table 4. Overall quartiles of the MSE of estimators under Scenario 3 with $\mathrm{MA}(2)$ model (on the left) and $\mathrm{AR}(1)$ model (on the right) for γ's.

Quartiles	$\widehat{\theta}_t$	$\widehat{\theta}_{TS,t}$	$\widehat{\theta}_{R,t}$	$\widehat{\theta}_{C,t}$	$\widehat{\theta}_{CLS,t}$	$\widehat{\theta}_{CTS,t}$	Quartiles	$\widehat{\theta}_t$	$\widehat{\theta}_{TS,t}$	$\widehat{\theta}_{R,t}$	$\widehat{\theta}_{C,t}$	$\widehat{\theta}_{CLS,t}$	$\widehat{\theta}_{CTS,t}$
	$\rho(\varepsilon_\theta,\varepsilon_\gamma)=0.9$							$\rho(\varepsilon_\theta,\varepsilon_\gamma)=0.9$					
Q_1	0.489	0.347	0.098	0.043	0.176	0.125	Q_1	0.489	0.347	0.117	0.042	0.288	0.194
Q_2	0.992	0.474	0.191	0.129	0.275	0.219	Q_2	0.992	0.474	0.201	0.179	0.532	0.413
Q_3	2.210	0.759	0.421	0.440	0.463	0.368	Q_3	2.210	0.759	0.429	0.735	1.111	0.965
	$\rho(\varepsilon_\theta,\varepsilon_\gamma)=0.6$							$\rho(\varepsilon_\theta,\varepsilon_\gamma)=0.6$					
Q_1	0.489	0.347	0.127	0.144	0.243	0.198	Q_1	0.489	0.347	0.118	0.096	0.344	0.260
Q_2	0.992	0.474	0.229	0.457	0.450	0.382	Q_2	0.992	0.474	0.214	0.364	0.552	0.454
Q_3	2.210	0.759	0.454	1.394	0.809	0.826	Q_3	2.210	0.759	0.442	1.256	0.931	0.783

lowest MSE in upper quartiles, while $\widehat{\theta}_{C,t}$ sees a greater increase in MSE, indicating reduced adaptability. The regression estimator $\widehat{\theta}_{R,t}$ remains a reasonable option.

The time series estimator $\widehat{\theta}_{TS,t}$ is stable across conditions, often outperforming $\widehat{\theta}_{C,t}$ in the upper quartile.

Table 4 summarizes the quartiles of MSE for the estimators in Scenario 3, considering both $\mathrm{MA}(2)$ and $\mathrm{AR}(1)$ models for the γ's at two specified correlation levels, $\rho(\varepsilon_\theta,\varepsilon_\gamma)$. In most scenarios, across different correlations and models, the calibrated estimator $\widehat{\theta}_{C,t}$ achieves the lowest MSE values in the first quartile, making it the best choice among calibrated estimators, especially under high correlation. The estimators $\widehat{\theta}_{CLS,\,t}$ and $\widehat{\theta}_{CTS,\,t}$ are similar in quality, though $\widehat{\theta}_{CTS,\,t}$ performs slightly better overall. Both typically outperform the time series estimator $\widehat{\theta}_{TS,\,t}$ but show slightly higher MSE in the upper quartile when using the $\mathrm{AR}(1)$ model for γ's or under moderate correlation. The regression estimator $\widehat{\theta}_{R,\,t}$ remains effective across most conditions, making it versatile. As in other simulation settings, the last survey estimator $\widehat{\theta}_t$ has the highest MSE as it is the simplest estimator, relying only on the last survey sample mean. It serves as the basis for constructing the calibrated last survey estimator $\widehat{\theta}_{CLS,\,t}$.

Real data example. For additional analysis, real data from the Lithuanian Labour Force Survey, conducted quarterly by the State Data Agency between 2011 and 2020 (excluding the third quarter of 2016), are used. The dataset, titled Employment Survey, was obtained from the Lithuanian Open Data Portal [20]. The Labour Force Survey in Lithuania employs a sampling design where a simple random sample of individuals is initially selected from the Population Register. Subsequently, all members of the selected individuals' households are included in the survey, resulting in a cluster sample where each cluster comprises all persons living at the selected address. Thus, each quarterly sample consists of approximately 8 000 households, representing around 1% of the population aged 15 and over [19].

In this example, $\{y_{\theta,\,i}\}_{i\in T}$ and $\{y_{\gamma,\,i}\}_{i\in T}$ represent the last survey estimates, calculated as the quarterly sample unemployment rate and the proportion of sampled individuals aged 15 to 25, respectively. The values $\{\gamma_i\}_{i\in T}$, where $T=\{1,2,\dots,39\}$, represent the true quarterly proportions of individuals aged 15 to 25, obtained from the Population Register [21]. The aim is to estimate θ_{39}, which represents the unemployment rate for the fourth quarter of 2020.

Following the Box and Jenkins [2] procedures for time series model estimation, the sample unemployment rate and the proportion of sampled individuals aged 15 to 25 are found to follow $\mathrm{ARMA}(2,1,2)$ models:

$$\begin{aligned}
\Delta y_{\theta,\,i} &= -0.09\Delta y_{\theta,\,i-1} - 0.98\Delta y_{\theta,\,i-2} - 0.11\varepsilon_{y_\theta,\,i-1} + 0.77\varepsilon_{y_\theta,\,i-2} + \varepsilon_{y_\theta,\,i},\\
\Delta y_{\gamma,\,i} &= -0.45\Delta y_{\gamma,\,i-1} - 0.69\Delta y_{\gamma,\,i-2} + 0.45\varepsilon_{y_\gamma,\,i-1} + 0.27\varepsilon_{y_\gamma,\,i-2} + \varepsilon_{y_\gamma,\,i}
\end{aligned}$$

with $\nu_\theta^2 = 1.127\cdot 10^{-4}$, where $\Delta y_{\theta,\,i} = y_{\theta,\,i} - y_{\theta,\,i-1}$, $\Delta y_{\gamma,\,i} = y_{\gamma,\,i} - y_{\gamma,\,i-1}$, and $\varepsilon_{y_\theta,\,i}$ and $\varepsilon_{y_\gamma,\,i}$ denote the white noise error terms. Although the correlation between $\{y_{\theta,\,i}\}_{i\in T}$ and $\{y_{\gamma,\,i}\}_{i\in T}$ is strong ($\rho=0.83$), the correlation between the corresponding residuals, $\rho(\varepsilon_{y_\theta},\varepsilon_{y_\gamma})=0.47$, is only moderate. The strong correlation between the series $\{y_{\theta,\,i}\}_{i\in T}$ and $\{y_{\gamma,\,i}\}_{i\in T}$ suggests that both the regression-based $\widehat{\theta}_{R,\,t}$ and the calibrated $\widehat{\theta}_{C,\,t}$ estimators are likely to perform well.

Since the dataset under analysis provides access to unit-level sample data, we estimate the variance S_θ^2 by calculating the average of the sampling variances estimated for each i, where $i\in T$. This method yields a variance value of $S_\theta^2 = 1.039\cdot 10^{-5}$. Scott et al. [15] provide some suggestions for estimating the variance S_θ^2 when unit-level sample data is not available. Since unit-level sample data are available, the variances of the estimators are estimated using bootstrap without-replacement [6].

Table 5. Estimates of θ_t and variances of estimators for the real-data case.

Estimators	$\widehat{\theta}_t$	$\widehat{\theta}_{TS,\,t}$	$\widehat{\theta}_{R,\,t}$	$\widehat{\theta}_{C,\,t}$
Estimates of θ_t	0.085	0.086	0.076	0.067
Variance of estimators $\times 10^5$	1.039	0.943	0.209	0.124

Table 5 compares different estimators of $\theta_t = \theta_{39}$ based on their estimates and variances. However, the estimators $\widehat{\theta}_{CLS,\,t}$ and $\widehat{\theta}_{CTS,\,t}$ are excluded from the calculations as the approximation (11) fails to capture the most significant part of the Wold decomposition due to the relatively large absolute value of the autoregressive coefficient, $|\lambda_{\theta,\,2}| = 0.98$, and the short time series length of $t = 39$. Specifically, the decomposition coefficients omitted beyond $t = 39$ remain significant in terms of their absolute values; for example, the next few are -0.148, -0.082, and 0.152. The exclusion of these coefficients leads to a substantial loss of information in the approximation, which in turn negatively affects the effectiveness and stability of the estimators $\widehat{\theta}_{CLS,\,t}$ and $\widehat{\theta}_{CTS,\,t}$ under these conditions.

As seen from Table 5, the highest estimate is given by $\widehat{\theta}_{TS,\,t} = 0.086$, while the lowest is obtained from $\widehat{\theta}_{C,\,t} = 0.067$. In terms of variance, $\widehat{\theta}_t$ has the highest value ($1.039 \cdot 10^{-5}$), indicating lower reliability. The variance of the time series estimator $\widehat{\theta}_{TS,\,t}$ is slightly lower than that of $\widehat{\theta}_t$ as it incorporates the time-dependent structure of the data. The regression estimator $\widehat{\theta}_{R,\,t}$ outperforms the time series estimator but performs worse than the calibrated estimator $\widehat{\theta}_{C,\,t}$, which exhibits the lowest variance ($0.124 \cdot 10^{-5}$), indicating higher stability. These findings are consistent with the results from artificial data, where, in some simulation settings, the calibrated estimator $\widehat{\theta}_{C,\,t}$ shows the lowest first quartile of MSE among all estimators. While the median is commonly used for such comparisons, the lowest first quartile indicates that in at least 25% of simulated cases, $\widehat{\theta}_{C,\,t}$ produces particularly small errors. This implies a relatively high probability that, in real-data applications with only one realization of the time series $\{y_{\theta,\,i}\}_{i\in T}$ and $\{y_{\gamma,\,i}\}_{i\in T}$, the MSE of $\widehat{\theta}_{C,\,t}$ will be the smallest among the considered estimators.

4 Conclusions

Our study introduces a set of newly improved estimators that integrate calibration and regression-based adjustments in repeated surveys. The last survey estimator (2) is commonly used but does not leverage past information, leading to higher variance [14]. Similarly, while the time series estimator (3) provides some improvements, it lacks auxiliary information [7,15]. In contrast, our approach achieves lower MSE and improved accuracy as confirmed through numerical simulations. These findings suggest that regression- and calibration-based methods should be preferred in settings where auxiliary series are available.

Overall, the performance of the estimators is significantly influenced by factors such as the correlation between the errors of the target and auxiliary series, the complexity of the models for θ's and γ's, the sampling variances, and the number of repeated surveys, though not all estimators are affected by all these factors to the same extent.

High correlation ($\rho(\varepsilon_\theta, \varepsilon_\gamma)$) consistently enhances the performance of calibration-based estimators, particularly $\widehat{\theta}_{CTS,\,t}$, which combines calibration with time-series adjustments to achieve the lowest MSE across most scenarios. This estimator is highly adaptable and robust, excelling in both simple (e.g., $\mathrm{AR}(1)$) and relatively complex (e.g., $\mathrm{ARMA}(2,2)$) model settings. In all scenarios, it is evident that the calibrated estimators perform better when the model structure for θ's and γ's is the same. This alignment allows the estimators to more effectively employ the correlation and structural similarity between the study and auxiliary time series, resulting in lower MSE values. When the models differ, these estimators lose some efficiency due to the mismatch in model dynamics, which reduces their ability to fully exploit the correlation between the errors of θ's and γ's.

The calibrated estimator $\widehat{\theta}_{C,\,t}$ achieves low MSE in high-correlation settings, particularly with simpler auxiliary models. However, its sensitivity to correlation strength and model complexity limits its reliability in varied conditions.

The regression estimator $\widehat{\theta}_{R,\,t}$ demonstrates consistent and reliable performance across various scenarios, achieving low MSE with minimal variability. It excels in high-correlation settings and remains robust under moderate correlations and more complex auxiliary models. While occasionally outperformed by calibration-based estimators like $\widehat{\theta}_{CTS,\,t}$ in very high-correlation scenarios, $\widehat{\theta}_{R,\,t}$ is a versatile and stable choice, particularly when the relationship between the target and auxiliary series is weaker.

Analysis of real Lithuanian Labour Force Survey data supports these findings for both the calibrated estimator $\widehat{\theta}_{C,\,t}$ and the regression-based estimator $\widehat{\theta}_{R,\,t}$, showing that they achieve lower variance and greater reliability, particularly in high-correlation settings.

The time series estimator $\widehat{\theta}_{TS,\,t}$ relies solely on the target series and performs consistently across various scenarios, independent of auxiliary series. Its effectiveness stems from employing the time series structure of θ's, making it robust to model specifications. While it may be outperformed by calibration-based estimators in scenarios where auxiliary information is valuable, $\widehat{\theta}_{TS,\,t}$ remains a straightforward and reliable option for estimating θ_t.

In contrast, the last survey estimator $\widehat{\theta}_t$ consistently exhibits the highest MSE, primarily serving as a baseline for comparison and as a basis for constructing the calibrated last survey estimator $\widehat{\theta}_{CLS,\,t}$. While simple, its reliance on sampling variance limits its precision, particularly in scenarios with high sampling variance.

Thus, when choosing an estimator, practitioners should consider the availability and quality of auxiliary information along with the correlation structure and complexity of the underlying models. The calibrated time series estimator $\widehat{\theta}_{CTS,\,t}$ is recommended when high correlation exists between the target and auxiliary series, particularly when both follow similar structural models, as it consistently achieves the lowest MSE. The estimator $\widehat{\theta}_{CLS,\,t}$ is a slightly worse alternative to $\widehat{\theta}_{CTS,\,t}$ but accommodates a simpler structure, resulting in easier calculations. Note that in cases with a strong autoregressive component and limited data length (t), $\widehat{\theta}_{CLS,\,t}$ and $\widehat{\theta}_{CTS,\,t}$ yield less reliable results as the Wold decomposition approximation (11) becomes inaccurate and undermines estimator accuracy. The regression estimator $\widehat{\theta}_{R,\,t}$ is a strong alternative, especially in settings where correlation is moderate or when auxiliary models differ from the target series, as it remains stable and reliable across various conditions. The calibrated estimator $\widehat{\theta}_{C,\,t}$ is well-suited for cases with high correlation and simpler auxiliary models but should be used cautiously in complex settings due to its sensitivity to correlation strength. The time series estimator $\widehat{\theta}_{TS,\,t}$ is ideal when no reliable auxiliary information is available as it provides a robust approach based solely on the target series structure. Finally, the last survey estimator $\widehat{\theta}_t$ should generally be avoided for inference due to its high MSE.

The estimation methods derived in this paper can be applied to time series that arise from real surveys as well as from human activities. They can also be applied to naturally occurring series such as the average daily temperature.

Future research could extend the estimators derived in this paper to nonstationary time series, incorporate machine learning for improved precision, and explore adaptive methods like Bayesian approaches. Applying them to real-world surveys and complex designs would validate their practicality, while multivariate extensions and uncertainty quantification could enhance robustness and reliability. The development of variance estimation methods for calibrated estimators with a nonlinear weight structure is also a major challenge.

Author contributions. Dalius Pumputis carried out all research, analysis, and writing. The author has read and approved the published version of the manuscript.

Conflicts of interest. The author declares no conflicts of interest.

Acknowledgment. The author would like to thank the journal editors and anonymous reviewers for their valuable comments and suggestions, which greatly improved the quality of the paper.

Appendix. Detailed model specifications

Scenario 1

Models for study time series $\{\theta_i\}_{i\in\{1,\dots,700\}}$:

$$\mathrm{AR}(1){:}\quad \theta_i = \lambda_\theta\theta_{i-1} + \varepsilon_{\theta,\,i}, \quad \varepsilon_{\theta,\,i} \sim N(0, \sigma_\theta^2),$$

where $\lambda_\theta \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$.

Models for auxiliary time series $\{\gamma_i\}_{i\in\{1,\dots,700\}}$:

$$\mathrm{AR}(1){:}\quad \gamma_i = \lambda_\gamma\gamma_{i-1} + \varepsilon_{\gamma,\,i}, \quad \varepsilon_{\gamma,\,i} \sim N(0, \sigma_\gamma^2),$$

with $\lambda_\gamma \in \{0.2, 0.4, 0.6, 0.8, 0.95\}$, and

$$\mathrm{ARMA}(1,1){:}\quad \gamma_i = \lambda_\gamma\gamma_{i-1} + \psi_\gamma\varepsilon_{\gamma,\,i-1} + \varepsilon_{\gamma,\,i}, \quad \varepsilon_{\gamma,\,i} \sim N(0, \sigma_\gamma^2),$$

with $(\lambda_\gamma,\, \psi_\gamma)$ selected from $\{(0.2, 0.1), (0.4, 0.3), (0.6, 0.5), (0.8, 0.7), (0.95, 0.9)\}$.

Scenario 2

Models for study time series $\{\theta_i\}_{i\in\{1,\dots,700\}}$:

$$\text{AR}(2):\quad \theta_i = \lambda_{\theta,1}\theta_{i-1} + \lambda_{\theta,2}\theta_{i-2} + \varepsilon_{\theta,i},\quad \varepsilon_{\theta,i} \sim N(0,\sigma_\theta^2),$$

where $(\lambda_{\theta,1}, \lambda_{\theta,2})$ are from $\{(0.1, 0.8), (0.3, 0.6), (0.5, 0.4), (0.7, 0.2), (0.9, 0.05)\}$.

Models for auxiliary time series $\{\gamma_i\}_{i\in\{1,\dots,700\}}$:

$$\text{AR}(2):\quad \gamma_i = \lambda_{\gamma,1}\gamma_{i-1} + \lambda_{\gamma,2}\gamma_{i-2} + \varepsilon_{\gamma,i},\quad \varepsilon_{\gamma,i} \sim N(0,\sigma_\gamma^2),$$

where $(\lambda_{\gamma,1}, \lambda_{\gamma,2}) \in \{(0.2, 0.7), (0.4, 0.5), (0.6, 0.3), (0.8, 0.1), (0.95, 0.01)\}$, and

$$\begin{aligned}\text{ARMA}(2,2):\quad & \gamma_i = \lambda_{\gamma,1}\gamma_{i-1} + \lambda_{\gamma,2}\gamma_{i-2} + \psi_{\gamma,1}\varepsilon_{\gamma,i-1} + \psi_{\gamma,2}\varepsilon_{\gamma,i-2} + \varepsilon_{\gamma,i},\\ & \varepsilon_{\gamma,i} \sim N(0,\sigma_\gamma^2),\end{aligned}$$

with tuples from $\{(0.2, 0.7, 0.1, -0.15), (0.4, 0.5, 0.3, -0.25), (0.6, 0.3, 0.5, -0.45), (0.8, 0.1, 0.7, -0.65), (0.95, 0.01, 0.9, -0.85)\}$.

Scenario 3

Models for study time series $\{\theta_i\}_{i\in\{1,\dots,700\}}$:

$$\text{MA}(2):\quad \theta_i = \psi_{\theta,1}\varepsilon_{\theta,i-1} + \psi_{\theta,2}\varepsilon_{\theta,i-2} + \varepsilon_{\theta,i},\quad \varepsilon_{\theta,i} \sim N(0,\sigma_\theta^2),$$

with $(\psi_{\theta,1}, \psi_{\theta,2}) \in \{(0.1, 0.8), (0.3, 0.6), (0.5, 0.4), (0.7, 0.2), (0.9, 0.05)\}$.

Models for auxiliary time series $\{\gamma_i\}_{i\in\{1,\dots,700\}}$:

$$\text{MA}(2):\quad \gamma_i = \psi_{\gamma,1}\varepsilon_{\gamma,i-1} + \psi_{\gamma,2}\varepsilon_{\gamma,i-2} + \varepsilon_{\gamma,i},\quad \varepsilon_{\gamma,i} \sim N(0,\sigma_\gamma^2),$$

where $(\psi_{\gamma,1}, \psi_{\gamma,2}) \in \{(0.2, 0.7), (0.4, 0.5), (0.6, 0.3), (0.8, 0.1), (0.95, 0.01)\}$, and

$$\text{AR}(1):\quad \gamma_i = \lambda_\gamma\gamma_{i-1} + \varepsilon_{\gamma,i},\quad \varepsilon_{\gamma,i} \sim N(0,\sigma_\gamma^2),$$

where $\lambda_\gamma \in \{0.1, 0.2, 0.3, 0.4, 0.95\}$.

In all scenarios, the values of σ_θ^2 and σ_γ^2 are set to 0.49, 1, 2.25 and 0.7, 1.2, 2.5, respectively.

References

1. B.J.N. Blight, A.J. Scott, A stochastic model for repeated surveys, *J. R. Stat. Soc., Ser. B*, **35**(1): 61–66, 1973, `https://doi.org/10.1111/j.2517-6161.1973.tb00936.x`.
2. G.E.P. Box, G.M. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden-Day, Oakland, CA, 1976, `https://books.google.lt/books?id=1WVHAAAAMAAJ`.
3. P.J. Brockwell, R.A. Davis, *Time Series: Theory and Methods*, Springer, New York, 1987, `https://doi.org/10.1007/978-1-4899-0004-3`.

4. J.C. Deville, C.-E. Särndal, Calibration estimators in survey sampling, *J. Am. Stat. Assoc.*, **87**(418):376–382, 1992, https://doi.org/10.2307/2290268.
5. A.R. Eckler, Rotation sampling, *Ann. Math. Stat.*, **26**(4):664–685, 1955, https://doi.org/10.1214/aoms/1177728427.
6. S. Gross, Median estimation in sample surveys, in *Proceedings of the Section on Survey Research Methods*, American Statistical Association, Alexandria, VA, 1980, pp. 181–184, http://www.asasrms.org/Proceedings/papers/1980_037.pdf.
7. M.A. Ismail, H.A. Auda, Y.A. Elzafrany, On time series analysis for repeated surveys, *J. Stat. Theory Appl.*, **17**(4):587–596, 2018, https://doi.org/10.2991/jsta.2018.17.4.1.
8. R.J. Jessen, Statistical investigation of a sample survey for obtaining farm facts, Research Bulletin 304, Iowa Agricultural Experiment Station, 1942.
9. R.G. Jones, The efficiency of time series estimators for repeated surveys, *Aust. J. Stat.*, **21**(1): 45–56, 1979, https://doi.org/10.1111/j.1467-842X.1979.tb01118.x.
10. T. Merkouris, A new approach to composite estimation for repeated surveys with rotating panels, *J. Off. Stat.*, **40**(3):409–424, 2024, https://doi.org/10.1177/0282423X241254193.
11. H.D. Patterson, Sampling on successive occasions with partial replacement of units, *J. R. Stat. Soc., Ser. B*, **12**(2):241–255, 1950, https://doi.org/10.1111/j.2517-6161.1950.tb00058.x.
12. D. Pfeffermann, Time series modelling of repeated survey data for estimation of finite population parameters, *J. R. Stat. Soc., Ser. A, Stat. Soc.*, **185**(4):1757–1777, 2022, https://doi.org/10.1111/rssa.12950.
13. C.-E. Särndal, B. Swensson, J. Wretman, *Model Assisted Survey Sampling*, Springer, New York, 1992, https://link.springer.com/book/9780387406206.
14. A.J. Scott, T.M.F. Smith, Analysis of repeated surveys using time series methods, *J. Am. Stat. Assoc.*, **69**(347):674–678, 1974, https://doi.org/10.2307/2286000.
15. A.J. Scott, T.M.F. Smith, R.G. Jones, The application of time series methods to the analysis of repeated surveys, *Int. Stat. Rev.*, **45**(1):13–28, 1977, https://doi.org/10.2307/1403000.
16. D. Steel, C. McLaren, Design and analysis of repeated surveys, Working Paper 11-08, Centre for Statistical and Survey Methodology, University of Wollongong, 2008, http://ro.uow.edu.au/cssmwp/10.
17. S.M. Tam, Analysis of repeated surveys using a dynamic linear model, *Int. Stat. Rev.*, **55**(1):63–73, 1987, https://doi.org/10.2307/1403271.
18. F. Yates, *Sampling Methods for Censuses and Surveys,* 3rd ed., Charles Griffin, London, 1960, https://archive.org/details/samplingmethodsf0000fran/page/n5/mode/2up.
19. Employment and unemployment (Labour force survey). Statistics Lithuania, https://ec.europa.eu/eurostat/cache/metadata/EN/employ_esqrs_lt.htm.
20. Lietuvos atvirų duomenų portalas, 2025, https://data.gov.lt/.
21. Population Register, https://www.registrucentras.lt/en/.