A measurement error Rao–Yu model for regional prevalence estimation over time using uncertain data obtained from dependent survey estimates

Jan Pablo Burgard¹ · Joscha Krause¹ · Domingo Morales²

Received: 10 July 2020 / Accepted: 17 April 2021 / Published online: 3 May 2021
© The Author(s) 2021

Abstract

The assessment of prevalence on regional levels is an important element of public health reporting. Since regional prevalence is rarely collected in registers, corresponding figures are often estimated via small area estimation using suitable health data. However, such data are frequently subject to uncertainty as values have been estimated from surveys. In that case, the method for prevalence estimation must explicitly account for data uncertainty to allow for reliable results. This can be achieved via measurement error models that introduce distribution assumptions on the noisy data. However, these methods usually require target and explanatory variable errors to be independent. This does not hold when data for both have been estimated from the same survey, which is sometimes the case in official statistics. If not accounted for, prevalence estimates can be severely biased. We propose a new measurement error model for regional prevalence estimation that is suitable for settings where target and explanatory variable errors are dependent. We derive empirical best predictors and demonstrate mean-squared error estimation. A maximum likelihood approach for model parameter estimation is presented. Simulation experiments are conducted to prove the effectiveness of the method. An application to regional hypertension prevalence estimation in Germany is provided.

✉ Joscha Krause
krause@uni-trier.de

Jan Pablo Burgard
burgardj@uni-trier.de

Domingo Morales
d.morales@umh.es

¹ Department of Economic and Social Statistics, Trier University, Universitätsring 15, 54296 Trier, Germany

² Operations Research Center, University Miguel Hernández de Elche, Avenida de la Universidad s/n, 03202 Elche, Spain

Keywords Dependent errors · Empirical best prediction · Hierarchical model · Parametric bootstrap · Small area estimation · Temporal data

Mathematics Subject Classification 62F10 · 62F35 · 62F40 · 62F86 · 62P10

1 Introduction

In recent years, the demand for detailed population health indicators has increased considerably. Policy-makers and researchers have realized the importance of monitoring disease frequencies on geographic regions and demographic groups over time. These insights mark important elements of public health reporting and are key to the provision of comprehensive health care programs. However, regional prevalence is rarely collected in registers due to confidentiality restrictions. Instead, such figures are often estimated via methods of small area estimation (SAE) using suitable health data. See Rao and Molina (2015) and Jiang and Lahiri (2006), or Pfeiffermann (2002) and Pfeiffermann (2013) for comprehensive overviews on SAE. Health data records are used as auxiliary variables in regression models that generate predictions for domain characteristics of interest. This approach allows for efficient estimates when the model has sufficient explanatory power (Münnich et al. 2019; Burgard et al. 2020b). Some recent papers in small area estimation for regional estimation are Trandafir et al. (2020), Bernal et al. (2020), Mills et al. (2020) and Saeedi et al. (2019), as well as Adin et al. (2019).

Due to the confidentiality of the data and the sensitive nature of health variables, area-level models are good options for predicting regional prevalence. Fay and Herriot (1979) proposed the original area-level model, which regresses regional direct estimates of target variables on covariates using cross-sectional data. This model was later extended and applied to multivariate data by Benavent and Morales (2016), Arima et al. (2017), Ubaidillah et al. (2019), Burgard et al. (2021) and Esteban et al. (2020). Since analyzing the development of domain characteristics over time is often relevant in empirical studies, generalizations of the area-level model to the temporal context soon appeared. Among them, the papers of Pfeiffermann and Burck (1990), Rao and Yu (1994), Ghosh et al. (1996), Datta et al. (1999), You and Rao (2000) and Datta et al. (2002) provided generalizations of this model by considering time-series data across domains, multi-level geographic segmentation, Bayesian modeling and robust approaches. More recently, Esteban et al. (2012), Marhuenda et al. (2013), Morales et al. (2015), Jedrejczak and Kubacki (2017), Boubeta et al. (2017) and Benavent and Morales (2021) applied temporal area-level models to poverty mapping.

In SAE models, health variables are often subject to uncertainty. Their values are typically estimated from periodic large-scale health surveys, such as the *National Health and Nutrition Examination Survey* (NHANES) in the US (National Center for Health Statistics 2020) or *Gesundheit in Deutschland aktuell* (GEDA) in Germany (Robert Koch Institute 2012; Lange et al. 2015). Accordingly, the data records are not accurate, but associated with errors. In that case, the regression models must explicitly account for data uncertainty to allow for reliable results. These issues motivated researchers over the years to develop SAE methods that allow for uncertain data in both

dependent and explanatory variables. Such SAE methods are often called *measurement error* models. The basic idea of these methods is to introduce distribution assumptions on the errors and make statistical inference under this premise. On that note, Ybarra and Lohr (2008), Burgard et al. (2020a) and Burgard et al. (2021) provided extensions of the area-level model that explicitly account for explanatory variable errors. A related approach was developed by Torabi et al. (2009) for unit-level observations. However, despite accounting for errors on both target and explanatory variables, these methods are still limited in their applicability. They require the target variable errors to be independent from the explanatory variable errors.

This condition is not fulfilled when some explanatory variables are estimated from the same survey as the target variable, which may occur in practice. That is to say, if we do not have enough data from administrative records, the alternative of using data from other surveys or from the same survey allows proposing new model-based estimates that may improve the direct estimates.

For prevalence estimation, the above-described scenario is relevant. Explanatory variables for health mapping are often health-related indicators as well. Since surveys like NHANES or GEDA are primary data sources for public health reporting in their respective countries, it may happen that both target variable and some explanatory variables are estimated from them. If the resulting dependencies between target and explanatory variable errors are ignored, prevalence estimates based on these records may be biased or inefficient.

We propose a new measurement error area-level model for regional prevalence estimation that allows for dependent errors on target and explanatory variables. With this feature, it is the first SAE method that is suitable for settings where data records for both model components have been estimated on a single survey. The model is an extension of the Rao–Yu model proposed by Rao and Yu (1994), which implies that it allows for either regional time-series data (domains and periods), or nested geographic segmentation (domains and subdomains). Thus, we refer to it as the *measurement error Rao–Yu* (MERY) model. We use it to estimate the development of regional prevalence over time based on uncertain data with dependent errors. Yet, we also show that the MERY model contains many existing measurement error SAE models as particular or limits cases, which makes it applicable to a wide range of empirical settings.

We first present a detailed description of the MERY model. A special emphasis is placed on the conditional component distributions, as the main innovation of the method is to allow for dependent errors on the variables. Next, the best predictors (BPs) and empirical best predictors (EBPs) of the domain characteristics under the model are derived. We subsequently address mean-squared error (MSE) estimation. The MSE of the BP is derived analytically. We further present a set of parametric bootstrap estimators for the MSE of the EBP. A maximum likelihood (ML) approach based on the Fisher scoring algorithm (Jennrich and Sampson 1976; Longford 1987) for model parameter estimation is stated. The methodology is tested in simulation experiments to demonstrate its effectiveness. An empirical application to temporal health mapping in Germany is provided. We estimate the hypertension prevalence for Nielsen regions crossed by age groups (domains), over three periods, using survey estimates from GEDA for the years 2009, 2010 and 2012 (Robert Koch Institute 2013, 2014a, b). We remark that the domains do not correspond to any geographic

segmentation in administrative health records of German statistical offices. Therefore, it is not possible to apply the standard SAE methodology based on area-level models, where the auxiliary variables are taken from out-of-sample data sources. In addition, a quarter of the domains have a sample size of less than 271 and the minimum sample size is equal to 51. Therefore, the estimation of the hypertension prevalence by domains and time periods can be treated by using SAE statistical methodology.

We show that applying the new MERY model in this situation allows for considerable efficiency gains relative to direct estimation.

The remainder of the paper is organized as follows. Section 2 introduces the MERY model and derives the BPs. Section 3 states the MSE of the BP and gives parametric bootstrap estimators for the MSE of the EBP. Section 4 addresses model parameter estimation. Section 5 contains the simulation experiments. Section 6 presents the empirical application to temporal health mapping. Section 7 closes with some conclusive remarks. A supplemental material file with five appendices is provided. Appendix 1 contains the proof for the main theorem. Appendix 2 presents several MSE calculations. Appendix 3 derives the Fisher information matrix required for the ML approach for model parameter estimation. Finally, Appendices 4 and 5 show additional results for simulation and application.

2 Model

2.1 Formulation

Let U be a finite population that is segmented into D disjoint sets $U_d, d = 1, \dots, D$. Suppose that every set U_d is further segmented into disjoint subsets $U_{dt}, d = 1, \dots, D, t = 1, \dots, m_d$. For ease of exposition, we assume that the number of subsets per set is constant with $m_d = T, d = 1, \dots, D$. This is to say, $U = \bigcup_{d=1}^D \bigcup_{t=1}^T U_{dt}$. In practice, the structure may correspond to either a population with a two-stage hierarchical geographic segmentation (domain–subdomain), or a population with geographic segmentation that is observed over several time periods (domain–period). This section introduces the MERY model by using the domain–subdomain terminology. The domain–period interpretation is straightforward. Let μ_{dt} be a characteristic of interest in set d and subset t , for instance the mean or the total of some target variable. The objective is to estimate μ_{dt} for all $d = 1, \dots, D, t = 1, \dots, T$. Suppose that for each μ_{dt} a corresponding direct estimator y_{dt} is available from a target survey $S \subset U$. By analogy with population structure, we assume that $S = \bigcup_{d=1}^D \bigcup_{t=1}^T S_{dt}$ with $S_{dt} \subset U_{dt}$ and $|S_{dt}| > 0, d = 1, \dots, D, t = 1, \dots, T$.

The MERY model is composed of three stages. In the first stage a model, called *sampling* model, is used to represent the sampling error of direct estimators. The sampling model indicates that direct estimators, y_{dt} 's, are unbiased and can be expressed as

$$y_{dt} = \mu_{dt} + e_{dt}, \quad d = 1, \dots, D, \quad t = 1, \dots, T,$$

where the sampling errors, e_{dt} 's, are independent and normally distributed with known variances. This is to say, $e_{dt} \sim N(0, \sigma_{edt}^2)$, where σ_{edt}^2 is known. In the second stage,

called *linking* model, the true domain characteristics, μ_{dt} 's, are assumed to vary linearly with a number p of domain-level auxiliary variables. The linking model is

$$\mu_{dt} = \tilde{\mathbf{x}}_{0,dt}\boldsymbol{\beta}_0 + \tilde{\mathbf{x}}_{1,dt}\boldsymbol{\beta}_1 + u_{1,d} + u_{2,dt}, \quad d = 1, \dots, D, \quad t = 1, \dots, T,$$

where $\tilde{\mathbf{x}}_{0,dt}$ and $\tilde{\mathbf{x}}_{1,dt}$ are a $1 \times p_0$ and $1 \times p_1$ row vectors, respectively, containing the true aggregated (subdomain-level) values of $p = p_0 + p_1$ auxiliary variables for domain d . The terms $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$ are the corresponding column vectors of regression coefficients. The $u_{1,d}$'s and $u_{2,dt}$'s are model errors (domain and subdomain random effects), assumed to be independent and identically distributed (i.i.d.) from $N(0, \sigma_1^2)$ and $N(0, \sigma_2^2)$, respectively, with variance σ_1^2 and σ_2^2 unknown, and independent of the e_{dt} 's. We assume that $\tilde{\mathbf{x}}_{0,dt}$ is known and equal to $\mathbf{x}_{0,dt}$, $d = 1, \dots, D$, $t = 1, \dots, T$. Further, we assume that the $\tilde{\mathbf{x}}_{1,dt}$'s are unknown random vectors that are predicted from the same survey as μ_{dt} or from different surveys. This is to say, the direct estimators y_{dt} and $\mathbf{x}_{1,dt}$, of μ_{dt} and $\tilde{\mathbf{x}}_{1,dt}$, respectively, can be calculated from the same or from a different survey sample. In the third stage, we consider the *measurement error* model

$$\tilde{\mathbf{x}}'_{1,dt} = \mathbf{x}'_{1,dt} + \mathbf{v}_{dt}, \quad \mathbf{v}_{dt} \sim N_{p_1}(\mathbf{0}, \boldsymbol{\Sigma}_{vdt}), \quad d = 1, \dots, D, \quad t = 1, \dots, T, \quad (1)$$

where $\mathbf{x}_{1,dt}$ is a row vector containing the direct estimators of $\tilde{\mathbf{x}}_{1,dt}$. The term $\mathbf{v}_{dt}$ is a column vector of random measurement errors, $\boldsymbol{\Sigma}_{vdt}$ is the known covariance matrix of $\mathbf{v}_{dt}$ (typically taken as a design-based covariance matrix estimate) and $\mathbf{v}_{11}, \dots, \mathbf{v}_{DT}$, $\mathbf{x}_{1,11}, \dots, \mathbf{x}_{1,DT}$ are independent. Model (1) assumes that the vector of true aggregated values $\tilde{\mathbf{x}}'_{1,dt}$ is random and it is the sum of two independent random vectors. The first one contains the predictions of the realizations of $\tilde{\mathbf{x}}'_{1,dt}$ and the second one is a random error term. Further, the model assumes that the joint distribution of the $(p_1 + 1) \times 1$ column vector $\boldsymbol{\varepsilon}_{dt} = (e_{dt}, \mathbf{v}'_{dt})'$ is multivariate normal with zero mean and covariance matrix

$$\boldsymbol{\Sigma}_{dt} = \begin{pmatrix} \sigma_{edt}^2 & \sigma'_{12dt} \\ \boldsymbol{\sigma}_{12dt} & \boldsymbol{\Sigma}_{vdt} \end{pmatrix}.$$

where $\boldsymbol{\sigma}_{12dt}$ is a $p_1 \times 1$ vector containing the covariances between e_{dt} and the components of $\mathbf{v}_{dt}$. The covariances of $\boldsymbol{\sigma}_{12dt}$, associated to auxiliary variables calculated from a different sample, are zero. The MERY model can be expressed as single model in the form

$$y_{dt} = \mathbf{x}_{0,dt}\boldsymbol{\beta}_0 + \mathbf{x}_{1,dt}\boldsymbol{\beta}_1 + \boldsymbol{\beta}'_1\mathbf{v}_{dt} + u_{1,d} + u_{2,dt} + e_{dt}, \quad d = 1, \dots, D, \quad t = 1, \dots, T. \quad (2)$$

or in the simpler form

$$y_{dt} = \mathbf{x}_{dt}\boldsymbol{\beta} + \boldsymbol{\beta}'_1\mathbf{v}_{dt} + u_{1,d} + u_{2,dt} + e_{dt}, \quad d = 1, \dots, D, \quad t = 1, \dots, T,$$

where $\mathbf{x}_{dt} = (\mathbf{x}_{0,dt}, \mathbf{x}_{1,dt})$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}'_0, \boldsymbol{\beta}'_1)'$. We finally assume that $\mathbf{x}_{1,dt}$, $u_{1,d}$, $u_{2,dt}$, $\boldsymbol{\varepsilon}_{dt} = (e_{dt}, \mathbf{v}'_{dt})'$, $d = 1, \dots, D$, $t = 1, \dots, T$, are mutually independent,

but we only introduce inference procedures conditionally on $(\mathbf{x}_{11}, \dots, \mathbf{x}_{DT})$. Under Model (2), it holds that

$$\text{var}(\beta'_1 \mathbf{v}_{dt}) = \beta'_1 \Sigma_{vdt} \beta_1, \quad \text{cov}(\beta'_1 \mathbf{v}_{dt}, e_{dt}) = \beta'_1 \sigma_{12dt}.$$

The variance of y_{dt} , given $\mathbf{x}_{dt}$, is

$$\text{var}(y_{dt} | \mathbf{x}_{dt}) = \sigma_1^2 + \sigma_2^2 + \beta'_1 \Sigma_{vdt} \beta_1 + 2\beta'_1 \sigma_{12dt} + \sigma_{edt}^2 \triangleq \sigma_{ydt}^2, \\ d = 1, \dots, D, \quad t = 1, \dots, T.$$

Thus, the distribution of y_{dt} , given $\mathbf{x}_{dt}$, is $y_{dt} | \mathbf{x}_{dt} \sim N(\mathbf{x}_{dt} \boldsymbol{\beta}, \sigma_{ydt}^2)$. Let us define the vectors

$$\mathbf{y}_d = \text{col}_{1 \leq t \leq T} (y_{dt}), \quad \mathbf{u}_{2,d} = \text{col}_{1 \leq t \leq T} (u_{2,dt}), \quad \mathbf{e}_d = \text{col}_{1 \leq t \leq T} (e_{dt}), \quad \mathbf{v}_d = \text{col}_{1 \leq t \leq T} (\mathbf{v}_{dt}), \\ \boldsymbol{\varepsilon}_d = \text{col}_{1 \leq t \leq T} (\boldsymbol{\varepsilon}_{dt}),$$

and matrices $\mathbf{Z}_{1d} = \mathbf{1}_T$, $\mathbf{Z}_{2d} = \mathbf{I}_T$, $\mathbf{V}_{2d} = \sigma_2^2 \mathbf{I}_T$,

$$\mathbf{X}_d = \text{col}_{1 \leq t \leq T} (\mathbf{x}_{dt}), \quad \mathbf{B}_d = \text{diag}_{1 \leq t \leq T} (\beta'_1), \quad \Sigma_{vd} = \text{diag}_{1 \leq t \leq T} (\Sigma_{vdt}), \quad \Sigma_d = \text{diag}_{1 \leq t \leq T} (\Sigma_{dt}).$$

where $\mathbf{1}_T$ is a $T \times 1$ vector with all elements equal to 1 and $\mathbf{I}_T$ is the $T \times T$ identity matrix. At the domain level, Model (2) is

$$\mathbf{y}_d = \mathbf{X}_d \boldsymbol{\beta} + \mathbf{B}_d \mathbf{v}_d + \mathbf{Z}_{1d} u_{1,d} + \mathbf{Z}_{2d} \mathbf{u}_{2,d} + \mathbf{e}_d, \quad d = 1, \dots, D,$$

where $u_{1,d} \sim N(0, \sigma_1^2)$, $\mathbf{u}_{2,d} \sim N_T(\mathbf{0}, \mathbf{V}_{2d})$, $\boldsymbol{\varepsilon}_d \sim N_{T(p_1+1)}(\mathbf{0}, \Sigma_d)$ are mutually independent. Further, we have $\mathbf{v}_d \sim N_{Tp}(\mathbf{0}, \Sigma_{vd})$. The variance matrix of $\mathbf{y}_d$, given $\mathbf{X}_d$, is

$$\mathbf{V}_d = \sigma_1^2 \mathbf{1}_T \mathbf{1}'_T + \text{diag}_{1 \leq t \leq T} (\beta'_1 \Sigma_{vdt} \beta_1 + 2\beta'_1 \sigma_{12dt} + \sigma_2^2 + \sigma_{edt}^2).$$

Define $M = DT$, the vectors $\mathbf{y} = \text{col}_{1 \leq d \leq D} (\mathbf{y}_d)$,

$$\mathbf{u}_1 = \text{col}_{1 \leq d \leq D} (u_{1,d}), \quad \mathbf{u}_2 = \text{col}_{1 \leq d \leq D} (\mathbf{u}_{2,d}), \quad \mathbf{e} = \text{col}_{1 \leq d \leq D} (\mathbf{e}_d), \quad \mathbf{v} \\ = \text{col}_{1 \leq d \leq D} (\mathbf{v}_d), \quad \boldsymbol{\varepsilon} = \text{col}_{1 \leq d \leq D} (\boldsymbol{\varepsilon}_d),$$

and matrices $\mathbf{V}_1 = \sigma_1^2 \text{diag}_{1 \leq d \leq D} (\mathbf{1}_T \mathbf{1}'_T)$, $\mathbf{V}_2 = \sigma_2^2 \mathbf{I}_M$, $\mathbf{Z}_1 = \text{diag}_{1 \leq d \leq D} (\mathbf{1}_T)$, $\mathbf{Z}_2 = \mathbf{I}_M$,

$$\mathbf{X} = \text{col}_{1 \leq d \leq D} (\mathbf{X}_d), \quad \mathbf{B} = \text{diag}_{1 \leq d \leq D} (\mathbf{B}_d), \quad \Sigma_v = \text{diag}_{1 \leq t \leq D} (\Sigma_{vd}), \quad \Sigma = \text{diag}_{1 \leq t \leq D} (\Sigma_d).$$

In matrix form, Model (2) can be alternatively written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{B}\mathbf{v} + \mathbf{Z}_1\mathbf{u}_1 + \mathbf{Z}_2\mathbf{u}_2 + \mathbf{e},$$

where $\mathbf{u}_1 \sim N_D(\mathbf{0}, \mathbf{V}_1)$, $\mathbf{u}_2 \sim N_{DT}(\mathbf{0}, \mathbf{V}_2)$, $\boldsymbol{\varepsilon} \sim N_{DT(p_1+1)}(\mathbf{0}, \boldsymbol{\Sigma})$ are mutually independent. The variance matrix of $\mathbf{y}$, given $\mathbf{X}$, is

$$\begin{aligned} \mathbf{V} = \text{diag}(\mathbf{V}_d) &= \sigma_1^2 \text{diag}(\mathbf{1}_T \mathbf{1}_T') + \text{diag} \left(\text{diag}(\boldsymbol{\beta}'_1 \boldsymbol{\Sigma}_{vdt} \boldsymbol{\beta}_1 \right. \\ &\quad \left. + 2\boldsymbol{\beta}'_1 \boldsymbol{\sigma}_{12dt} + \sigma_2^2 + \sigma_{edt}^2) \right). \end{aligned}$$

Hereafter, we state some further conditional distributions (CDs) resulting under the model. They allow for a clearer picture of the dependencies between specific model components, as the main innovation of our contribution is to allow for SAE from dependent sample estimates. Recall that if $\mathbf{y}$ is normally distributed with partitioned mean vector $\boldsymbol{\mu} = (\boldsymbol{\mu}'_s, \boldsymbol{\mu}'_r)'$, where $\boldsymbol{\mu}_s$ and $\boldsymbol{\mu}_r$ are partitions of $\boldsymbol{\mu}$, as well as covariance matrix

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_s & \mathbf{V}_{sr} \\ \mathbf{V}_{rs} & \mathbf{V}_r \end{pmatrix},$$

then the distribution of $\mathbf{y}_r$, given $\mathbf{y}_s$, is also multivariate normal, i.e., $\mathbf{y}_r | \mathbf{y}_s \sim N(\boldsymbol{\mu}_{r|s}, \mathbf{V}_{r|s})$, with mean vector and covariance matrix

$$\boldsymbol{\mu}_{r|s} = \boldsymbol{\mu}_r + \mathbf{V}_{rs} \mathbf{V}_s^{-1} (\mathbf{y}_s - \boldsymbol{\mu}_s), \quad \mathbf{V}_{r|s} = \mathbf{V}_r - \mathbf{V}_{rs} \mathbf{V}_s^{-1} \mathbf{V}_{sr}.$$

At the subdomain level, we have the following CDs. The CD of e_{dt} , given $\mathbf{v}_{dt}$, is $N(\mu_{e|vdt}, \sigma_{e|vdt}^2)$, where

$$\begin{aligned} \mu_{e|vdt} &= \mu_{edt} + \boldsymbol{\sigma}'_{12dt} \boldsymbol{\Sigma}_{vdt}^{-1} (\mathbf{v}_{dt} - \boldsymbol{\mu}_{vdt}) = \boldsymbol{\sigma}'_{12dt} \boldsymbol{\Sigma}_{vdt}^{-1} \mathbf{v}_{dt}, \\ \sigma_{e|vdt}^2 &= \sigma_{edt}^2 - \boldsymbol{\sigma}'_{12dt} \boldsymbol{\Sigma}_{vdt}^{-1} \boldsymbol{\sigma}_{12dt}. \end{aligned}$$

The CD of $\mathbf{v}_{dt}$, given e_{dt} , is $N_p(\boldsymbol{\mu}_{v|edt}, \boldsymbol{\Sigma}_{v|edt})$, where

$$\begin{aligned} \boldsymbol{\mu}_{v|edt} &= \boldsymbol{\mu}_{vdt} + \boldsymbol{\sigma}_{12dt} \sigma_{edt}^{-2} (e_{dt} - \mu_{edt}) = \sigma_{edt}^{-2} e_{dt} \boldsymbol{\sigma}_{12dt}, \\ \boldsymbol{\Sigma}_{v|edt} &= \boldsymbol{\Sigma}_{vdt} - \sigma_{edt}^{-2} \boldsymbol{\sigma}_{12dt} \boldsymbol{\sigma}'_{12dt}. \end{aligned}$$

The conditional expectation and variance of y_{dt} , given $\mathbf{x}_{dt}$ and $\mathbf{v}_{dt}$, are

$$\begin{aligned} E[y_{dt} | \mathbf{x}_{dt}, \mathbf{v}_{dt}] &= \mathbf{x}_{dt} \boldsymbol{\beta} + \mathbf{v}'_{dt} \boldsymbol{\beta}_1 + E[e_{dt} | \mathbf{v}_{dt}] = \mathbf{x}_{dt} \boldsymbol{\beta} + \mathbf{v}'_{dt} \boldsymbol{\beta}_1 + \boldsymbol{\sigma}'_{12dt} \boldsymbol{\Sigma}_{vdt}^{-1} \mathbf{v}_{dt}, \\ \text{var}(y_{dt} | \mathbf{x}_{dt}, \mathbf{v}_{dt}) &= \text{var}(u_{1,d}) + \text{var}(u_{2,d}) + \text{var}(e_{dt} | \mathbf{v}_{dt}) \\ &= \sigma_1^2 + \sigma_2^2 + \sigma_{edt}^2 - \boldsymbol{\sigma}'_{12dt} \boldsymbol{\Sigma}_{vdt}^{-1} \boldsymbol{\sigma}_{12dt}. \end{aligned}$$

The CD of y_{dt} , given $\mathbf{x}_{dt}$ and $\mathbf{v}_{dt}$, is $y_{dt}|\mathbf{x}_{dt}, \mathbf{v}_{dt} \sim N(\mu_{y|vdt}, \sigma_{y|vdt}^2)$, where

$$\mu_{y|vdt} = E[y_{dt}|\mathbf{x}_{dt}, \mathbf{v}_{dt}], \quad \sigma_{y|vdt}^2 = \text{var}(y_{dt}|\mathbf{x}_{dt}, \mathbf{v}_{dt}).$$

The CD of y_{dt} , given $\mathbf{x}_{dt}$ and $u_{1,d}$, is $y_{dt}|\mathbf{x}_{dt}, u_{1,d} \sim N(\mu_{y|u_1dt}, \sigma_{y|u_1dt}^2)$, where

$$\mu_{y|u_1dt} = \mathbf{x}_{dt}\boldsymbol{\beta} + u_{1,d}, \quad \sigma_{y|u_1dt}^2 = \boldsymbol{\beta}'_1 \boldsymbol{\Sigma}_{vdt} \boldsymbol{\beta}_1 + 2\boldsymbol{\beta}'_1 \boldsymbol{\sigma}_{12dt} + \sigma_2^2 + \sigma_{edt}^2.$$

The CD of y_{dt} , given $\mathbf{x}_{dt}$ and $u_{2,d}$, is $y_{dt}|\mathbf{x}_{dt}, u_{2,d} \sim N(\mu_{y|u_2dt}, \sigma_{y|u_2dt}^2)$, with

$$\mu_{y|u_2dt} = \mathbf{x}_{dt}\boldsymbol{\beta} + u_{2,d}, \quad \sigma_{y|u_2dt}^2 = \boldsymbol{\beta}'_1 \boldsymbol{\Sigma}_{vdt} \boldsymbol{\beta}_1 + 2\boldsymbol{\beta}'_1 \boldsymbol{\sigma}_{12dt} + \sigma_1^2 + \sigma_{edt}^2.$$

At the domain level, we have the following CDs. The CD of y_d , given $\mathbf{X}_d$ and $\mathbf{v}_d$, is $y_d|\mathbf{X}_d, \mathbf{v}_d \sim N_T(\boldsymbol{\mu}_{y|vd}, \boldsymbol{\Sigma}_{y|vd})$, where

$$\begin{aligned} \boldsymbol{\mu}_{y|vd} &= E[\mathbf{y}_d|\mathbf{X}_d, \mathbf{v}_d] = \mathbf{X}_d\boldsymbol{\beta} + \mathbf{B}_d\mathbf{v}_d + E[\mathbf{e}_d|\mathbf{v}_d] \\ &= \mathbf{X}_d\boldsymbol{\beta} + \mathbf{B}_d\mathbf{v}_d + \text{diag}(\boldsymbol{\sigma}'_{12dt} \boldsymbol{\Sigma}_{vdt}^{-1}) \mathbf{v}_d, \\ &= \mathbf{X}_d\boldsymbol{\beta} + \mathbf{C}_d\mathbf{v}_d, \quad \mathbf{C}_d = \mathbf{B}_d + \text{diag}(\boldsymbol{\sigma}'_{12dt} \boldsymbol{\Sigma}_{vdt}^{-1}) = \text{diag}(\boldsymbol{\beta}'_1 + \boldsymbol{\sigma}'_{12dt} \boldsymbol{\Sigma}_{vdt}^{-1}), \\ \boldsymbol{\Sigma}_{y|vd} &= \text{var}(\mathbf{y}_d|\mathbf{X}_d, \mathbf{v}_d) = \text{var}(u_{1,d}\mathbf{1}_T) + \text{var}(\mathbf{u}_{2,d}) + \text{var}(\mathbf{e}_d|\mathbf{v}_d) \\ &= \sigma_1^2 \mathbf{1}_T \mathbf{1}_T' + \text{diag}(\sigma_2^2 + \sigma_{edt}^2 - \boldsymbol{\sigma}'_{12dt} \boldsymbol{\Sigma}_{vdt}^{-1} \boldsymbol{\sigma}_{12dt}). \end{aligned}$$

The CD of y_d , given $\mathbf{X}_d$ and $u_{1,d}$, is $y_d|\mathbf{X}_d, u_{1,d} \sim N_T(\boldsymbol{\mu}_{y|u_1d}, \boldsymbol{\Sigma}_{y|u_1d})$, for

$$\boldsymbol{\mu}_{y|u_1d} = \mathbf{X}_d\boldsymbol{\beta} + u_{1,d}\mathbf{1}_T, \quad \boldsymbol{\Sigma}_{y|u_1d} = \text{diag}(\boldsymbol{\beta}'_1 \boldsymbol{\Sigma}_{vdt} \boldsymbol{\beta}_1 + 2\boldsymbol{\beta}'_1 \boldsymbol{\sigma}_{12dt} + \sigma_2^2 + \sigma_{edt}^2).$$

The CD of y_d , given $\mathbf{X}_d$ and $u_{2,d}$, is $y_d|\mathbf{X}_d, u_{2,d} \sim N_T(\boldsymbol{\mu}_{y|u_2d}, \boldsymbol{\Sigma}_{y|u_2d})$, for

$$\boldsymbol{\mu}_{y|u_2d} = \mathbf{X}_d\boldsymbol{\beta} + u_{2,d}, \quad \boldsymbol{\Sigma}_{y|u_2d} = \sigma_1^2 \mathbf{1}_T \mathbf{1}_T' + \text{diag}(\boldsymbol{\beta}'_1 \boldsymbol{\Sigma}_{vdt} \boldsymbol{\beta}_1 + 2\boldsymbol{\beta}'_1 \boldsymbol{\sigma}_{12dt} + \sigma_{edt}^2).$$

2.2 Empirical best prediction

In this section, we obtain the EBP of the μ_{dt} 's under the MERY model. For this, we first derive the respective BP under the assumption of known model parameters. Thereafter, the EBP is obtained from substituting the model parameters with consistent estimators. The BP is calculated according to the subsequent theorem.

Theorem 1 Under model (2), the BP of $\boldsymbol{\mu}_d = \mathbf{X}_d\boldsymbol{\beta} + \mathbf{v}'_d\boldsymbol{\beta} + \mathbf{1}_T u_{1,d} + u_{2,d}$ is

$$\hat{\boldsymbol{\mu}}_d^{bp} = E[\boldsymbol{\mu}_d|\mathbf{X}_d, \mathbf{y}_d] = \mathbf{X}_d\boldsymbol{\beta} + E[\mathbf{B}_d\mathbf{v}_d + u_{1,d}\mathbf{1}_T + u_{2,d}|\mathbf{X}_d, \mathbf{y}_d]$$

$$= \mathbf{X}_d \boldsymbol{\beta} + \mathbf{B}_d \hat{\mathbf{v}}_d^{bp} + \mathbf{1}_T \hat{u}_{1,d}^{bp} + \hat{u}_{2,d}^{bp},$$

where

$$\begin{aligned}\hat{\mathbf{v}}_d^{bp} &= E[\mathbf{v}_d | \mathbf{X}_d, \mathbf{y}_d] = (\boldsymbol{\Sigma}_{vd}^{-1} + \mathbf{C}'_d \boldsymbol{\Sigma}_{y|vd}^{-1} \mathbf{C}_d)^{-1} \mathbf{C}'_d \boldsymbol{\Sigma}_{y|vd}^{-1} (\mathbf{y}_d - \mathbf{X}_d \boldsymbol{\beta}), \\ \hat{u}_{1,d}^{bp} &= E[u_{1,d} | \mathbf{X}_d, \mathbf{y}_d] = a_{1d}^{-1} \sum_{t=1}^T a_{1dt} (y_{dt} - \mathbf{x}_{dt} \boldsymbol{\beta}) = a_{1d}^{-1} \mathbf{a}'_{1d} (\mathbf{y}_d - \mathbf{X}_d \boldsymbol{\beta}), \\ \hat{u}_{2,d}^{bp} &= E[u_{2,d} | \mathbf{X}_d, \mathbf{y}_d] = (\boldsymbol{\Sigma}_{y|u_{2d}}^{-1} + \mathbf{I}_T)^{-1} \boldsymbol{\Sigma}_{y|u_{2d}}^{-1} (\mathbf{y}_d - \mathbf{X}_d \boldsymbol{\beta}),\end{aligned}$$

and

$$\mathbf{a}_{1d} = \text{col}_{1 \leq t \leq T} (a_{1dt}), \quad a_{1d} = \frac{1}{\sigma_1^2} + \sum_{t=1}^T a_{1dt}, \quad a_{1dt} = \frac{1}{\boldsymbol{\beta}'_1 \boldsymbol{\Sigma}_{vdt} \boldsymbol{\beta}_1 + 2\boldsymbol{\beta}'_1 \boldsymbol{\sigma}_{12dt} + \sigma_2^2 + \sigma_{ed}^2}.$$

The proof can be found in Appendix 1 of the supplemental material. Define the full parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}', \sigma_1^2, \sigma_2^2)'$. Observe that $\hat{\boldsymbol{\mu}}_d^{bp} = \hat{\boldsymbol{\mu}}_d^{bp}(\boldsymbol{\theta})$. The EBP of $\boldsymbol{\mu}_d$ is obtained by replacing $\boldsymbol{\theta}$ with a consistent estimator $\hat{\boldsymbol{\theta}}$, that is, $\hat{\boldsymbol{\mu}}_d^{ebp} = \hat{\boldsymbol{\mu}}_d^{bp}(\hat{\boldsymbol{\theta}})$. Please note that we describe how to obtain $\hat{\boldsymbol{\theta}}$ in Sect. 4. The MERY model is quite general. It allows for explanatory variables with no errors, errors that are independent of the sampling errors (out-of-sample errors) and errors that depend on the sampling errors (in-sample errors). If x_{dtk} is measured without error, we have $\sigma_{dtk}^2 = \text{var}(x_{dtk}) = 0$ and zero covariances. For the case that all explanatory variables are measured without error, that is, $\boldsymbol{\Sigma}_{vdt} = \mathbf{0}_p$, which implies that $\boldsymbol{\varepsilon}_{dt} = (e_{dt}, \mathbf{0})'$, $d = 1, \dots, D$, $t = 1, \dots, T$, then $\hat{\boldsymbol{\mu}}_d^{ebp}$ corresponds to the Rao–Yu predictor proposed by Rao and Yu (1994). When x_{dtk} is an out-of-sample direct estimator, then all its covariances with in-sample direct estimators are zero. For $\sigma_1^2 = 0$ and $\boldsymbol{\sigma}_{12dt} = \mathbf{0}$, $d = 1, \dots, D$, $t = 1, \dots, T$, then $\hat{\boldsymbol{\mu}}_d^{ebp}$ is equivalent to the EBP under the measurement error Fay–Herriot model studied by Ybarra and Lohr (2008) as well as Burgard et al. (2020a).

3 Mean-squared error estimation

In this section, we elaborate on MSE estimation for the EBP under the MERY model. Let us first decompose the MSE into components. By summing and subtracting $\hat{\boldsymbol{\mu}}_d^{bp}$, we obtain

$$\begin{aligned}(\hat{\boldsymbol{\mu}}_d^{ebp} - \boldsymbol{\mu}_d)(\hat{\boldsymbol{\mu}}_d^{ebp} - \boldsymbol{\mu}_d)' &= ((\hat{\boldsymbol{\mu}}_d^{ebp} - \hat{\boldsymbol{\mu}}_d^{bp}) + (\hat{\boldsymbol{\mu}}_d^{bp} - \boldsymbol{\mu}_d))((\hat{\boldsymbol{\mu}}_d^{ebp} - \hat{\boldsymbol{\mu}}_d^{bp}) + (\hat{\boldsymbol{\mu}}_d^{bp} - \boldsymbol{\mu}_d))' \\ &= (\hat{\boldsymbol{\mu}}_d^{bp} - \boldsymbol{\mu}_d)(\hat{\boldsymbol{\mu}}_d^{bp} - \boldsymbol{\mu}_d)' + (\hat{\boldsymbol{\mu}}_d^{ebp} - \hat{\boldsymbol{\mu}}_d^{bp})(\hat{\boldsymbol{\mu}}_d^{ebp} - \hat{\boldsymbol{\mu}}_d^{bp})' \\ &\quad + (\hat{\boldsymbol{\mu}}_d^{ebp} - \hat{\boldsymbol{\mu}}_d^{bp})(\hat{\boldsymbol{\mu}}_d^{bp} - \boldsymbol{\mu}_d)' + (\hat{\boldsymbol{\mu}}_d^{bp} - \boldsymbol{\mu}_d)(\hat{\boldsymbol{\mu}}_d^{ebp} - \hat{\boldsymbol{\mu}}_d^{bp})'.\end{aligned}$$

By taking expectations, we get

$$\begin{aligned} MSE(\hat{\mu}_d^{ebp}) &= MSE(\hat{\mu}_d^{bp}) + E[(\hat{\mu}_d^{ebp} - \hat{\mu}_d^{bp})(\hat{\mu}_d^{ebp} - \hat{\mu}_d^{bp})'] \\ &\quad + E[(\hat{\mu}_d^{ebp} - \hat{\mu}_d^{bp})(\hat{\mu}_d^{bp} - \mu_d)'] + E[(\hat{\mu}_d^{bp} - \mu_d)(\hat{\mu}_d^{ebp} - \hat{\mu}_d^{bp})'] \\ &= \mathbf{G}_{1d}(\theta) + \mathbf{G}_{2d}(\theta) + \mathbf{G}_{3d}(\theta) + \mathbf{G}_{4d}(\theta). \end{aligned}$$

In what follows, we present two parametric bootstrap estimators of $MSE(\hat{\mu}_d^{ebp})$. The first is a standard estimator that approximates all four MSE components via the Monte Carlo method. The second estimator is a bias-corrected estimator that uses an analytical approximation for $\mathbf{G}_{1d}(\theta)$. We will show via simulation in Sect. 5.4 that including the analytical formula improves MSE estimation over the standard estimator considerably.

3.1 Standard parametric bootstrap estimator

This section presents the standard estimator, denoted by mse_{1d}^* , which approximates all four MSE components numerically. It is calculated with the subsequent algorithm.

1. By using the data $(\mathbf{y}_d, \mathbf{X}_d)$, $d = 1, \dots, D$, calculate $\hat{\theta}$ and $\hat{\mathbf{V}}_{2d} = \hat{\sigma}_2^2 \mathbf{T}_T$.
2. Repeat B times
 - 2.1 For $d = 1, \dots, D$, generate $u_{1,d}^{*(b)} \sim N(0, \hat{\sigma}_1^2)$, $\mathbf{u}_{2,d}^{*(b)} \sim N_T(\mathbf{0}, \hat{\mathbf{V}}_{2d})$, $\mathbf{e}_d^{*(b)} = (\mathbf{e}_d^{*(b)'}', \mathbf{v}_d^{*(b)'}')' \sim N_{T(p_1+1)}(\mathbf{0}, \boldsymbol{\Sigma}_d)$, $\boldsymbol{\mu}_d^{*(b)} = \mathbf{X}_d \hat{\boldsymbol{\beta}} + \hat{\mathbf{B}}_d \mathbf{v}_d^{*(b)} + \mathbf{1}_T u_{1,d}^{*(b)} + \mathbf{u}_{2,d}^{*(b)}$ and $\mathbf{y}_d^{*(b)} = \boldsymbol{\mu}_d^{*(b)} + \mathbf{e}_d^{*(b)}$.
 - 2.2 By using the data $(\mathbf{y}_d^{*(b)}, \mathbf{X}_d)$, $d = 1, \dots, D$, calculate the estimator $\hat{\theta}^{*(b)}$.
 - 2.3 Calculate the EBPs $\hat{\mu}_d^{ebp*(b)}$, $d = 1, \dots, D$, under the model of Step 2.1.
3. For $d = 1, \dots, D$, obtain the $T \times T$ matrix of estimators of $MSE(\hat{\mu}_d^{ebp})$; i.e.,

$$mse_{1d}^* = \frac{1}{B} \sum_{b=1}^B (\hat{\mu}_d^{ebp*(b)} - \boldsymbol{\mu}_d^{*(b)})(\hat{\mu}_d^{ebp*(b)} - \boldsymbol{\mu}_d^{*(b)})'.$$

3.2 Bias-corrected parametric bootstrap estimator

This section presents the bias-corrected estimator, denoted by mse_{2d}^* , that uses an analytical approximation for $\mathbf{G}_{1d}(\theta)$. Observe that $\mathbf{G}_{1d}(\theta)$ denotes the MSE of the BP in the course of the resampling process. Hence, we first derive the analytical expression for $MSE(\hat{\mu}_d^{bp} | \mathbf{X}_d)$ under the assumption of known model parameters θ . Afterward, we use this expression in another parametric bootstrap algorithm that yields us the bias-corrected estimator. Please note that we only present the final results of the corresponding mathematical developments. For further details regarding the derivation, see Appendix 2 of the supplemental material. Under model (2), the mean-squared error,

$MSE(\hat{\mu}_d^{bp} | X_d)$, of $\hat{\mu}_d^{bp}$ is

$$\begin{aligned} MSE(\hat{\mu}_d^{bp} | X_d) &= E[(\hat{\mu}_d^{bp} - \mu_d)(\hat{\mu}_d^{bp} - \mu_d)' | X_d] \\ &= E[\hat{\mu}_d^{bp} \hat{\mu}_d^{b p'} | X_d] + E[\mu_d \mu_d' | X_d] - E[\mu_d \hat{\mu}_d^{b p'} | X_d] - E[\hat{\mu}_d^{bp} \mu_d' | X_d] \\ &\triangleq S_1 + S_2 - S_3 - S_4, \end{aligned}$$

where the first summand is

$$\begin{aligned} S_1 &= X_d \beta \beta' X_d' + B_d (\Sigma_{vd}^{-1} + C_d' \Sigma_{y|vd}^{-1} C_d)^{-1} C_d' \Sigma_{y|vd}^{-1} V_d \Sigma_{y|vd}^{-1} C_d (\Sigma_{vd}^{-1} \\ &\quad + C_d' \Sigma_{y|vd}^{-1} C_d)^{-1} B_d' \\ &\quad + B_d (\Sigma_{vd}^{-1} + C_d' \Sigma_{y|vd}^{-1} C_d)^{-1} C_d' \Sigma_{y|vd}^{-1} V_d a_d a_{1d}^{-1} \mathbf{1}_T' \\ &\quad + B_d (\Sigma_{vd}^{-1} + C_d' \Sigma_{y|vd}^{-1} C_d)^{-1} C_d' \Sigma_{y|vd}^{-1} V_d \Sigma_{y|u_2d}^{-1} (I_T + \Sigma_{y|u_2d}^{-1})^{-1} \\ &\quad + \mathbf{1}_T a_{1d}^{-1} a_d' V_d \Sigma_{y|vd}^{-1} C_d (\Sigma_{vd}^{-1} + C_d' \Sigma_{y|vd}^{-1} C_d)^{-1} B_d' \\ &\quad + \mathbf{1}_T a_{1d}^{-2} a_d' V_d a_{1d} \mathbf{1}_T' + \mathbf{1}_T a_{1d}^{-1} a_d' V_d \Sigma_{y|u_2d}^{-1} (I_T + \Sigma_{y|u_2d}^{-1})^{-1} \\ &\quad + (I_T + \Sigma_{y|u_2d}^{-1})^{-1} \Sigma_{y|u_2d}^{-1} V_d \Sigma_{y|vd}^{-1} C_d (\Sigma_{vd}^{-1} + C_d' \Sigma_{y|vd}^{-1} C_d)^{-1} B_d' \\ &\quad + (I_T + \Sigma_{y|u_2d}^{-1})^{-1} \Sigma_{y|u_2d}^{-1} V_d a_d a_{1d}^{-1} \mathbf{1}_T' + (I_T + \Sigma_{y|u_2d}^{-1})^{-1} \Sigma_{y|u_2d}^{-1} V_d \Sigma_{y|u_2d}^{-1} (I_T \\ &\quad + \Sigma_{y|u_2d}^{-1})^{-1}, \end{aligned}$$

the second summand is

$$S_2 = X_d \beta \beta' X_d' + B_d \Sigma_{vd} B_d' + \sigma_1^2 \mathbf{1}_T \mathbf{1}_T' + \sigma_2^2 I_T,$$

and the third summand is

$$\begin{aligned} S_3 &= X_d \beta \beta' X_d' + B_d \left(\Sigma_{vd} B_d' + \text{diag}(\sigma_{12dt})_{1 \leq t \leq T} \right) \\ &\quad \Sigma_{y|vd}^{-1} C_d (\Sigma_{vd}^{-1} + C_d' \Sigma_{y|vd}^{-1} C_d)^{-1} B_d' \\ &\quad + B_d \left(\Sigma_{vd} B_d' + \text{diag}(\sigma_{12dt})_{1 \leq t \leq T} \right) a_{1d} a_{1d}^{-1} \mathbf{1}_T' \\ &\quad + B_d \left(\Sigma_{vd} B_d' + \text{diag}(\sigma_{12dt})_{1 \leq t \leq T} \right) \Sigma_{y|u_2d}^{-1} (\Sigma_{y|u_2d}^{-1} + I_T)^{-1} \\ &\quad + \sigma_1^2 \mathbf{1}_T \mathbf{1}_T' \Sigma_{y|vd}^{-1} C_d (\Sigma_{vd}^{-1} + C_d' \Sigma_{y|vd}^{-1} C_d)^{-1} B_d' \\ &\quad + \sigma_1^2 \mathbf{1}_T \mathbf{1}_T' a_{1d} a_{1d}^{-1} \mathbf{1}_T' + \sigma_1^2 \mathbf{1}_T \mathbf{1}_T' \Sigma_{y|u_2d}^{-1} (\Sigma_{y|u_2d}^{-1} + I_T)^{-1} \\ &\quad + \sigma_2^2 \Sigma_{y|vd}^{-1} C_d (\Sigma_{vd}^{-1} + C_d' \Sigma_{y|vd}^{-1} C_d)^{-1} B_d' + \sigma_2^2 a_{1d} a_{1d}^{-1} \mathbf{1}_T' \\ &\quad + \sigma_2^2 \Sigma_{y|u_2d}^{-1} (\Sigma_{y|u_2d}^{-1} + I_T)^{-1}. \end{aligned}$$

For the last component, it holds that $S_4 = S'_3$. After calculating the MSE of the BP, we use this result to estimate $MSE(\hat{\mu}_d^{ebp})$ with a bias-corrected parametric bootstrap estimator. Define $G_{1d}(\hat{\theta}) \triangleq S_1(\hat{\theta}) + S_2(\hat{\theta}) - S_3(\hat{\theta}) - S_4(\hat{\theta})$, where the known model parameters θ are substituted by the estimate $\hat{\theta}$. The bias-corrected estimator is obtained as follows.

1. By using the data (y_d, X_d) , $d = 1, \dots, D$, calculate $\hat{\theta}$ and $\hat{V}_{2d} = \hat{\sigma}_2^2 T_T$.
2. Repeat B times
 - 2.1 For $d = 1, \dots, D$, generate $u_{1,d}^{*(b)} \sim N(0, \hat{\sigma}_1^2)$, $u_{2,d}^{*(b)} \sim N_T(\mathbf{0}, \hat{V}_{2d})$, $\epsilon_d^{*(b)} = (e_d^{*(b)'} v_d^{*(b)'})' \sim N_{T(p_1+1)}(\mathbf{0}, \Sigma_d)$, $\mu_d^{*(b)} = X_d \hat{\beta} + \hat{B}_d v_d^{*(b)} + \mathbf{1}_T u_{1,d}^{*(b)} + u_{2,d}^{*(b)}$ and $y_d^{*(b)} = \mu_d^{*(b)} + e_d^{*(b)}$.
 - 2.2 By using the data $(y_d^{*(b)}, X_d)$, $d = 1, \dots, D$, calculate the estimator $\hat{\theta}^{*(b)}$.
 - 2.3 Calculate the BPs $\hat{\mu}_d^{*bp(b)} = \hat{\mu}_d^{*bp(b)}(y_d^{*(b)}, \hat{\theta})$ and EBPs $\hat{\mu}_d^{*ebp(b)} = \hat{\mu}_d^{*ebp(b)}(y_d^{*(b)}, \hat{\theta}^{*(b)})$, $d = 1, \dots, D$, under the model of Step 2.1.
3. For $d = 1, \dots, D$, obtain the $T \times T$ matrix of estimators of $MSE(\hat{\mu}_d^{ebp})$; i.e.,

$$\begin{aligned}
 E_*[G_{1d}(\hat{\theta}^*)] &\approx \frac{1}{B} \sum_{b=1}^B G_{1d}(\hat{\theta}^{*(b)}) = H_{1d*}. \\
 E_*[(\hat{\mu}_d^{*ebp} - \mu_d^{*bp})(\hat{\mu}_d^{*ebp} - \mu_d^{*bp})'] &\approx \\
 \frac{1}{B} \sum_{b=1}^B (\hat{\mu}_d^{*ebp(b)} - \mu_d^{*bp(b)})(\hat{\mu}_d^{*ebp(b)} - \mu_d^{*bp(b)})' &= G_{2d*}, \\
 E_*[(\hat{\mu}_d^{*ebp} - \mu_d^{*bp})(\hat{\mu}_d^{*bp} - \mu_d^*)'] &\approx \\
 \frac{1}{B} \sum_{b=1}^B (\hat{\mu}_d^{*ebp(b)} - \mu_d^{*bp(b)})(\hat{\mu}_d^{*bp(b)} - \mu_d^{*(b)})' &= G_{3d*}, \\
 E_*[(\hat{\mu}_d^{*bp} - \mu_d^*)(\hat{\mu}_d^{*ebp} - \mu_d^{*bp})'] &\approx \\
 \frac{1}{B} \sum_{b=1}^B (\hat{\mu}_d^{*bp(b)} - \mu_d^{*(b)})(\hat{\mu}_d^{*ebp(b)} - \mu_d^{*bp(b)})' &= G_{4d*}.
 \end{aligned}$$

4. For $d = 1, \dots, D$, obtain the $T \times T$ matrices of bootstrap estimators of $MSE(\hat{\mu}_d^{ebp})$.

$$\begin{aligned}
 MSE_{2d}^* &= 2G_{1d}(\hat{\theta}) - E_*[G_{1d}(\hat{\theta}^*)] + E_*[(\hat{\mu}_d^{*ebp} - \mu_d^{*bp})(\hat{\mu}_d^{*ebp} - \mu_d^{*bp})'] \\
 &\quad + E_*[(\hat{\mu}_d^{*ebp} - \mu_d^{*bp})(\hat{\mu}_d^{*bp} - \mu_d^*)'] + E_*[(\hat{\mu}_d^{*bp} - \mu_d^*)(\hat{\mu}_d^{*ebp} - \mu_d^{*bp})'],
 \end{aligned}$$

which can be approximated by the Monte Carlo method, i.e.,

$$mse_{2d}^* = 2G_{1d}(\hat{\theta}) - H_{1d*} + G_{2d*} + G_{3d*} + G_{4d*}.$$

Observe that the difference between mse_{1d}^* and mse_{2d}^* lies within the approximation of $\mathbf{G}_{1d}(\boldsymbol{\theta})$. While mse_{1d}^* uses a purely numerical approximation over the Monte Carlo iterations, mse_{2d}^* uses the bias-corrected term $2\mathbf{G}_{1d}(\hat{\boldsymbol{\theta}}) - \mathbf{H}_{1d*}$. Here, both components $\mathbf{G}_{1d}(\hat{\boldsymbol{\theta}})$ and $\mathbf{H}_{1d*}$ are calculated based on the analytical formula for $MSE(\hat{\boldsymbol{\mu}}_d^{bp} | \mathbf{X}_d)$. This allows for an efficiency advantage over the standard estimator.

Under the Fay–Herriot model, this fact was proved in Theorem 3 of González-Manteiga et al. (2008). These authors showed that the biases of the parametric bootstrap estimators mse_{1d}^* and mse_{2d}^* are of orders $O(D^{-1/2})$ and $o(D^{-1})$, respectively, under some model assumptions. The extension of their results to the MERY model is not straightforward, but similar asymptotic results may be expected under the MERY model after adding regularity conditions on the matrices $\boldsymbol{\Sigma}_{vdt}$, $d = 1, \dots, D$, $t = 1, \dots, T$. Simulations of Sect. 5.4 cover the lack of theoretical results and gives some information about the behavior of mse_{1d}^* and mse_{2d}^* under the MERY model.

Finally, note that the properties of the bootstrap estimators mse_{1d}^* and mse_{2d}^* rely on parametric distributional assumptions. Deviations from the normal distribution may produce biases in the MSE estimates. Fortunately, in the application to real data the sample sizes are not very small. Therefore, the asymptotic normality of the direct estimators, y_{dt} and $\mathbf{x}_{1,dt}$, does not allow important deviations from normality.

4 Model parameter estimation

This section presents a ML approach for model parameter estimation. It is based on the Fisher scoring algorithm (Jennrich and Sampson 1976; Longford 1987). Recall that

$$\mathbf{V}_d = \sigma_1^2 \mathbf{1}_T \mathbf{1}_T' + \underset{1 \leq t \leq T}{\text{diag}} (\boldsymbol{\beta}_1' \boldsymbol{\Sigma}_{vdt} \boldsymbol{\beta}_1 + 2\boldsymbol{\beta}_1' \boldsymbol{\sigma}_{12dt} + \sigma_2^2 + \sigma_{edt}^2) \triangleq \sigma_1^2 \mathbf{1}_T \mathbf{1}_T' + \mathbf{H}_d.$$

For calculating $\mathbf{V}_d^{-1}$, we apply the inversion formula

$$(\mathbf{A} + \mathbf{u}\mathbf{v}')^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}'\mathbf{A}^{-1}}{1 + \mathbf{v}'\mathbf{A}^{-1}\mathbf{u}},$$

with $\mathbf{u} = \sigma_1^2 \mathbf{1}_T$, $\mathbf{v}' = \mathbf{1}_T'$, $\mathbf{A} = \mathbf{H}_d$. We obtain

$$\mathbf{V}_d^{-1} = \mathbf{H}_d^{-1} + (1 + T\sigma_1^2)^{-1} \mathbf{H}_d^{-1} \mathbf{1}_T \mathbf{1}_T' \mathbf{H}_d^{-1}.$$

For calculating the derivatives of $\mathbf{V}_d$, we define the $p \times 1$ vector

$$\boldsymbol{\delta}_k = \underset{1 \leq j \leq p}{\text{col}} (\delta_{jk}) = (0, \dots, 0, 1^{(k)}, 0, \dots, 0)'.$$

For $k = 1, \dots, p$, the derivative of $\boldsymbol{\beta}$, with respect to β_k , is $\partial \boldsymbol{\beta} / \partial \beta_k = \boldsymbol{\delta}_k$. For $k = 1, \dots, p_1$, the derivative of $\mathbf{V}_d$, with respect to β_k , is $\partial \mathbf{V}_d / \partial \beta_k = \mathbf{0}$. For $k =$

$p_1 + 1, \dots, p$, we have the following derivatives

$$\begin{aligned}\frac{\partial \boldsymbol{\beta}'_1 \boldsymbol{\Sigma}_{vdt} \boldsymbol{\beta}_1}{\partial \beta_k} &= 2\delta'_k \boldsymbol{\Sigma}_{vdt} \boldsymbol{\beta}_1, \quad \frac{\partial \boldsymbol{\beta}'_1 \boldsymbol{\sigma}_{12dt}}{\partial \beta_k} = \delta'_k \boldsymbol{\sigma}_{12dt} = \text{cov}(v_{dtk}, e_{dt}), \quad \frac{\partial \mathbf{V}_d}{\partial \sigma_2^2} = \mathbf{I}_T, \\ \frac{\partial \mathbf{V}_d}{\partial \sigma_1^2} &= \mathbf{1}_T \mathbf{1}'_T, \\ \frac{\partial \mathbf{V}_d}{\partial \beta_k} &= 2 \text{diag}_{1 \leq t \leq T} (\delta'_k \boldsymbol{\Sigma}_{vdt} \boldsymbol{\beta}_1 + \delta'_k \boldsymbol{\sigma}_{12dt}), \\ \frac{\partial^2 \mathbf{V}_d}{\partial \beta_{k_1} \partial \beta_{k_2}} &= 2 \text{diag}_{1 \leq t \leq T} (\delta'_{k_1} \boldsymbol{\Sigma}_{vdt} \delta_{k_2}).\end{aligned}$$

It holds that $\mathbf{y}|\mathbf{X} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$, with covariance matrix $\mathbf{V} = \text{diag}_{1 \leq d \leq D} (\mathbf{V}_d)$. The vector of model parameters is $\boldsymbol{\theta} = (\boldsymbol{\beta}', \sigma_1^2, \sigma_2^2)'$. The log-likelihood is

$$\begin{aligned}\ell = \ell(\boldsymbol{\theta}; \mathbf{y}, \mathbf{X}) &= -\frac{DT}{2} \log 2\pi - \frac{1}{2} \log |\mathbf{V}| - \frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &= -\frac{DT}{2} \log 2\pi - \frac{1}{2} \sum_{d=1}^D \log |\mathbf{V}_d| \\ &\quad - \frac{1}{2} \sum_{d=1}^D (\mathbf{y}_d - \mathbf{X}_d \boldsymbol{\beta})' \mathbf{V}_d^{-1} (\mathbf{y}_d - \mathbf{X}_d \boldsymbol{\beta}).\end{aligned}$$

In what follows, we apply the formulas

$$\frac{\partial \log |\mathbf{V}|}{\partial \theta} = \text{tr} \left\{ \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta} \right\}, \quad \frac{\partial \mathbf{V}^{-1}}{\partial \theta} = -\mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta} \mathbf{V}^{-1}.$$

The first partial derivatives of ℓ are

$$\begin{aligned}S_k &= \sum_{d=1}^D \delta'_k \mathbf{X}'_d \mathbf{V}_d^{-1} (\mathbf{y}_d - \mathbf{X}_d \boldsymbol{\beta}), \quad k = 1, \dots, p_0, \\ S_k &= \frac{\partial \ell}{\partial \beta_k} = -\frac{1}{2} \sum_{d=1}^D \text{tr} \left(\mathbf{V}_d^{-1} \frac{\partial \mathbf{V}_d}{\partial \beta_k} \right) + \frac{1}{2} \sum_{d=1}^D (\mathbf{y}_d - \mathbf{X}_d \boldsymbol{\beta})' \mathbf{V}_d^{-1} \frac{\partial \mathbf{V}_d}{\partial \beta_k} \mathbf{V}_d^{-1} (\mathbf{y}_d - \mathbf{X}_d \boldsymbol{\beta}) \\ &\quad + \sum_{d=1}^D \delta'_k \mathbf{X}'_d \mathbf{V}_d^{-1} (\mathbf{y}_d - \mathbf{X}_d \boldsymbol{\beta}), \quad k = p_0 + 1, \dots, p, \\ S_{p_1+1} &= \frac{\partial \ell}{\partial \sigma_1^2} = -\frac{1}{2} \sum_{d=1}^D \text{tr} \left(\mathbf{V}_d^{-1} \frac{\partial \mathbf{V}_d}{\partial \sigma_1^2} \right) + \frac{1}{2} \sum_{d=1}^D (\mathbf{y}_d - \mathbf{X}_d \boldsymbol{\beta})' \mathbf{V}_d^{-1} \frac{\partial \mathbf{V}_d}{\partial \sigma_1^2} \mathbf{V}_d^{-1} (\mathbf{y}_d - \mathbf{X}_d \boldsymbol{\beta}), \\ S_{p+2} &= \frac{\partial \ell}{\partial \sigma_2^2} = -\frac{1}{2} \sum_{d=1}^D \text{tr} \left(\mathbf{V}_d^{-1} \frac{\partial \mathbf{V}_d}{\partial \sigma_2^2} \right) + \frac{1}{2} \sum_{d=1}^D (\mathbf{y}_d - \mathbf{X}_d \boldsymbol{\beta})' \mathbf{V}_d^{-1} \frac{\partial \mathbf{V}_d}{\partial \sigma_2^2} \mathbf{V}_d^{-1} (\mathbf{y}_d - \mathbf{X}_d \boldsymbol{\beta}).\end{aligned}$$

The second partial derivatives and the elements of the Fisher information matrix can be found in Appendix 3 of the supplemental material. Let $\theta = (\beta', \sigma_1^2, \sigma_2^2)$ denote the full parameter. In matrix form, the vector of scores and the Fisher information matrix are

$$S(\theta) = \text{col}_{1 \leq k \leq p+a} (S_k(\theta)), \quad F(\theta) = (F_{ij}(\theta))_{i,j=1,\dots,p+2}.$$

The updating formula of the Fisher scoring algorithm, for calculating the ML estimators of the MERY model parameters, is

$$\theta^{(k+1)} = \theta^{(k)} + F^{-1}(\theta^{(k)})S(\theta^{(k)}). \quad (3)$$

Please note that, depending on the data setting, Fisher scoring algorithms can be numerically instable and sensitive regarding starting values. Thus, we recommend using $\theta^{(0)} = \hat{\theta}^{ry}$ as seed, where $\hat{\theta}^{ry}$ is the model parameter estimate obtained under the Rao–Yu model.

5 Simulation experiments

This section presents simulation experiments that are conducted in order to demonstrate the effectiveness of the method in a controlled environment. We investigate three aspects: (i) model parameter estimation, (ii) characteristic prediction, as well as (iii) mean-squared error estimation. Note that additional simulation results can be found in Appendix 4.

5.1 Setup

A simulation study with $R = 500$ iterations, $r = 1, \dots, R$, is conducted. We define $p = p_0 + p_1$ with $p_0 = 1$, $p_1 = 2$. Let $\beta_0 = 10$, $\beta_1 = (2, 2)'$ and $\beta = (\beta'_0, \beta'_1)'$. For $d = 1, \dots, D$, $t = 1, \dots, T$, with $D \in \{50, 100, 200\}$ and $T \in \{5, 10, 20\}$, the auxiliary variables are

$$X_d = \text{col}_{1 \leq t \leq T} (x_{dt}), \quad x_{dt} = (1, x_{dt1}, x_{dt2})', \quad x_{dt1} \sim U(75, 100), \quad x_{dt2} \sim U(75, 100),$$

and the measurement errors are $\varepsilon_{dt} = (e_{dt}, v'_{dt})' \sim N_p(\mathbf{0}, \Sigma_{dt})$, where

$$\Sigma_{dt} = \begin{pmatrix} \sigma_{edt}^2 & \sigma'_{12dt} \\ \sigma_{12dt} & \Sigma_{vdt} \end{pmatrix}, \quad \Sigma_{vdt} = \begin{pmatrix} \sigma_{v1dt}^2 & \rho_{v12}\sigma_{v1dt}\sigma_{v2dt} \\ \rho_{v12}\sigma_{v1dt}\sigma_{v2dt} & \sigma_{v2dt}^2 \end{pmatrix},$$

and $\sigma_{12dt} = (\rho_{e1}\sigma_{edt}\sigma_{v1dt}, \rho_{e2}\sigma_{edt}\sigma_{v2dt})'$. We choose $\rho_{v12} \in \{0, 0.4, 0.6\}$ to resemble the correlation between the covariate measurement errors. Further, we define $\rho_{e1} = \rho_{e2} \in \{0, 0.4, 0.6\}$ for the correlation between e_{dt} and v_{dt} . The realizations for the parameters σ_{v1dt} , σ_{v2dt} , σ_{edt} are drawn for each subdomain from uniform distributions. Here, we distinguish between three error settings: small error (A), moderate

Table 1 Overview of simulation scenarios

Scenario	A.1	A.2	A.3	A.4	A.5
T	5	5	5	10	20
D	50	100	200	50	50
DT	250	500	1000	500	1000
σ_{vjdt}	$U(0, 1)$	$U(0, 1)$	$U(0, 1)$	$U(0, 1)$	$U(0, 1)$
σ_{edt}	$U(8, 10)$	$U(8, 10)$	$U(8, 10)$	$U(8, 10)$	$U(8, 10)$
ρ_{v12}	0	0	0	0	0
ρ_{ej}	0	0	0	0	0
Scenario	B.1	B.2	B.3	B.4	B.5
T	5	5	5	10	20
D	50	100	200	50	50
DT	250	500	1000	500	1000
σ_{vjdt}	$U(6, 8)$	$U(6, 8)$	$U(6, 8)$	$U(6, 8)$	$U(6, 8)$
σ_{edt}	$U(12, 14)$	$U(12, 14)$	$U(12, 14)$	$U(12, 14)$	$U(12, 14)$
ρ_{v12}	0.4	0.4	0.4	0.4	0.4
ρ_{ej}	0.4	0.4	0.4	0.4	0.4
Scenario	C.1	C.2	C.3	C.4	C.5
T	5	5	5	10	20
D	50	100	200	50	50
DT	250	500	1000	500	1000
σ_{vjdt}	$U(10, 12)$	$U(10, 12)$	$U(10, 12)$	$U(10, 12)$	$U(10, 12)$
σ_{edt}	$U(16, 18)$	$U(16, 18)$	$U(16, 18)$	$U(16, 18)$	$U(16, 18)$
ρ_{v12}	0.6	0.6	0.6	0.6	0.6
ρ_{ej}	0.6	0.6	0.6	0.6	0.6

error (B) and large error (C). Let $j \in \{1, 2\}$. We consider a total of 15 simulation scenarios as described in Table 1.

The A-scenarios include almost no covariate measurement errors with no dependencies. In these cases, the true model is very close to the original Rao–Yu model. The B- and C-scenarios contain moderate and large covariate measurement errors with dependencies. For $d = 1, \dots, D$, $t = 1, \dots, T$, the random effects are $u_{1,d} \sim N(0, \sigma_1^2)$, $\mathbf{u}_{2,d} \sim N_T(\mathbf{0}, \sigma_2^2 \mathbf{I}_T)$, where $\sigma_1 = 7$, $\sigma_2 = 2$. Define the vectors and matrices

$$\mathbf{u}_{2,d} = \text{col}_{1 \leq t \leq T} (u_{2,dt}), \quad \mathbf{e}_d = \text{col}_{1 \leq t \leq T} (e_{dt}), \quad \mathbf{v}_d = \text{col}_{1 \leq t \leq T} (v_{dt}), \quad \mathbf{B}_d = \text{diag}(\beta'_1).$$

For $d = 1, \dots, D$, the target vectors are given by

$$\mathbf{y}_d = \mathbf{X}_d \boldsymbol{\beta} + \mathbf{B}_d \mathbf{v}_d + \mathbf{1}_d u_{1,d} + \mathbf{u}_{2,d} + \mathbf{e}_d, \quad d = 1, \dots, D.$$

We apply a fixed covariate setting. The covariate realizations are generated once and held fixed over all Monte Carlo iterations. Note that the same applies to the error covariance matrices Σ_{dt} . However, the realizations for v_d , $u_{1,d}$, $u_{2,d}$ and e_d , $d = 1, \dots, D$, are drawn in every iteration from their respective distributions. This implies that both y_d and μ_d , $d = 1, \dots, D$ are also generated in each iteration. We study the performance of two methods:

- EBP based on the Rao–Yu (RY) model with REML estimation (R-package *saery*),
- EBP based on the MERY model with ML estimation.

5.2 Results of model parameter estimation

This section studies the behavior of the ML fitting algorithm presented in Sect. 4. We use the root mean-squared error (RMSE) and the bias as performance measures. For $\tau \in \{\beta_0, \beta_1, \beta_2, \sigma_1^2, \sigma_2^2\}$, they are given by

$$RMSE(\hat{\tau}) = \left(\frac{1}{R} \sum_{r=1}^R (\hat{\tau}^{(r)} - \tau)^2 \right)^{1/2}, \quad BIAS(\hat{\tau}) = \frac{1}{R} \sum_{r=1}^R (\hat{\tau}^{(r)} - \tau).$$

Since $\beta_1 = (2, 2)'$, we can average the performance measures for the regression parameters. We start with the RMSE, which is displayed in Table 2. We observe that the model parameter estimates under the MERY model are always more efficient than under the RY model. The respective RMSE is smaller for all scenarios and for all model parameters. In the A-scenarios, MERY and RY perform very similarly with respect to regression parameter estimation. This was expected, since scenarios A generate the target data under the MERY model and not under the RY model. So, the simulation results showing that the MERY model performs slightly better compared to the RY model in the presence of small covariate measurement error are coherent with the theory.

In the B- and C-scenarios, the MERY model is clearly more efficient as it uses the additional information on the measurement error distributions. The biggest performance differences are in variance parameter estimation. Here, the RY model cannot distinguish between random effect variation and variation caused by the measurement errors. This leads to highly inefficient subdomain variance parameter estimates. The MERY model, on the other hand, shows a very stable estimation performance, as it can distinguish between the individual sources of variation.

We continue with the bias. The results are summarized in Table 3. For regression parameter estimation, the results are mixed. The MERY model is less biased with respect to intercept estimation, but often more biased in slope parameter estimation. However, the relative bias of slope parameter estimates is at most 2.6%, which is a reasonable magnitude under measurement error settings. Regarding variance parameter estimation, the MERY model is considerably less biased than the RY model. Especially on the subdomain level, which is also the level of the measurement errors, the RY model vastly overestimates the random effect variance. This is due to its incapability to distinguish between measurement errors and random effect variation. The MERY

Table 2 RMSEs of MERY (left) and RY (right) model parameter estimation

Scenario	MERY				RY			
	β_0	β_1	σ_1^2	σ_2^2	β_0	β_1	σ_1^2	σ_2^2
A.1	8.92	0.0724	23.77	4.10	8.98	0.0725	25.47	5.31
A.2	6.05	0.0463	21.39	4.10	6.07	0.0464	22.24	5.31
A.3	4.88	0.0378	23.70	2.88	4.88	0.0379	24.10	3.73
A.4	6.21	0.0469	23.53	4.26	6.22	0.0471	24.81	5.55
A.5	4.58	0.0342	22.96	2.75	4.59	0.0343	24.05	3.74
B.1	18.82	0.1851	32.65	41.52	23.39	0.2002	39.18	578.81
B.2	16.05	0.1271	24.46	41.52	18.45	0.1392	28.29	578.81
B.3	9.13	0.0835	23.10	25.13	11.49	0.0911	25.18	579.25
B.4	13.55	0.1327	23.81	30.45	16.70	0.1464	26.75	567.69
B.5	8.90	0.0879	22.60	20.28	11.64	0.0977	24.28	580.54
C.1	25.07	0.2685	55.78	108.84	34.68	0.3153	86.21	1841.41
C.2	18.87	0.1854	49.46	108.84	28.76	0.2264	66.98	1841.41
C.3	14.06	0.1439	38.97	90.64	19.22	0.1571	47.83	1851.93
C.4	21.71	0.1934	39.52	123.69	29.64	0.2285	47.61	1839.74
C.5	14.18	0.1404	26.83	88.09	20.28	0.1589	29.63	1848.66

model, on the other hand, shows a much weaker tendency of overestimation because it uses the additional distribution information. The remaining bias is likely due to the likelihood function being flat under measurement errors.

5.3 Results of subdomain characteristic prediction

This section studies the performance of the EBP $\hat{\mu}_{dt}^{ebp}$ based on the MERY model, derived in Sect. 2.2, as well as the EBLUP $\hat{\mu}_{dt}^{ebp0}$ based on the RY model. We consider RMSE, RRMSE, absolute bias and relative absolute bias as measures. For $\hat{\mu}_{dt}^{(i)} \in \{\hat{\mu}_{dt}^{ebp(i)}, \hat{\mu}_{dt}^{ebp0(i)}\}$, $d = 1, \dots, D$, $t = 1, \dots, T$, we have

$$RMSE_{dt} = \left(\frac{1}{R} \sum_{r=1}^R (\hat{\mu}_{dt}^{(r)} - \mu_{dt}^{(r)})^2 \right)^{1/2}, \quad \bar{\mu}_{dt} = \frac{1}{R} \sum_{r=1}^R \mu_{dt}^{(r)}, \quad RRMSE_{dt} = \frac{RMSE_{dt}}{|\bar{\mu}_{dt}|},$$

$$ABIAS_{dt} = \frac{1}{R} \sum_{r=1}^R |\hat{\mu}_{dt}^{(r)} - \mu_{dt}^{(r)}|, \quad RABIAS_{dt} = \frac{ABIAS_{dt}}{|\bar{\mu}_{dt}|}.$$

For a compact presentation, we average the performance measures over the subdomains. The averaged results are presented in Table 4. We observe that the EBP based on the MERY model is superior to the EBLUP based on the RY model in all measures and scenarios. In the A-scenarios, the performance difference is very small, which is consistent with the theoretical developments from Sect. 2. In the B- and C-scenarios,

Table 3 Biases of MERY (left) and RY (right) model parameter estimation

Scenario	MERY				RY			
	β_0	β_1	σ_1^2	σ_2^2	β_0	β_1	σ_1^2	σ_2^2
A.1	-0.652	-0.0007	0.937	0.089	-0.654	-0.0005	1.009	0.608
A.2	-0.622	-0.0041	0.855	0.089	-0.622	-0.0041	0.890	0.608
A.3	-0.676	0.0010	0.948	-0.018	-0.676	0.0010	0.964	0.641
A.4	-0.673	0.0002	0.940	0.118	-0.674	0.0004	0.992	0.607
A.5	-0.676	0.0009	0.918	0.051	-0.677	0.0009	0.962	0.696
B.1	-0.291	-0.0316	0.771	9.320	-0.614	-0.0037	1.021	144.702
B.2	-0.379	-0.0244	0.820	9.320	-0.602	-0.0053	1.007	144.702
B.3	-0.566	-0.0088	0.780	5.362	-0.698	0.0025	0.896	144.813
B.4	-0.392	-0.0236	0.785	6.515	-0.645	-0.0020	0.941	141.923
B.5	-0.600	-0.0059	0.844	4.145	-0.743	0.0064	0.920	145.134
C.1	-0.057	-0.0514	1.137	26.226	-0.596	-0.0049	1.455	460.353
C.2	-0.223	-0.0384	0.944	26.226	-0.741	0.0058	1.171	460.353
C.3	-0.331	-0.0294	1.096	21.660	-0.626	-0.0040	1.117	462.981
C.4	-0.175	-0.0422	0.946	29.912	-0.572	-0.0082	0.998	459.934
C.5	-0.375	-0.0250	0.757	21.022	-0.692	0.0021	0.884	462.165

the MERY model shows more distinct efficiency advantages. The inclusion of the additional information from the subdomain covariance matrices Σ_{dt} in combination with the more efficient model parameter estimates allows for more accurate results. Further, we see that the predictions under the MERY are also less biased than under the RY model in terms of the absolute bias.

5.4 Results of mean-squared error estimation

This section studies the parametric bootstrap procedures for MSE estimation in Sect. 3. In particular, we look at two estimators mse_1^* and mse_2^* , where the first is the standard parametric bootstrap estimator (Sect. 3.1) and the second is the bias-corrected estimator (Sect. 3.2). We consider bias, relative bias and RMSE as performance measures:

$$BIAS_{dt} = \frac{1}{R} \sum_{r=1}^R (mse_{dt}^{*(r)} - mse_{dt}), \quad mse_{dt} = \frac{1}{R} \sum_{r=1}^R (\hat{\mu}_{dt}^{(r)} - \mu_{dt}^{(r)})^2,$$

$$RBIAS_{dt} = \frac{1}{R} \sum_{r=1}^R \frac{(mse_{dt}^{*(r)} - mse_{dt})}{mse_{dt}}, \quad RMSE_{dt} = \left(\frac{1}{R} \sum_{r=1}^R (mse_{dt}^{*(r)} - mse_{dt})^2 \right)^{1/2}.$$

Note that mse_{dt} , $d = 1, \dots, D$, $t = 1, \dots, T$, is the MSE of the EBP over Monte Carlo iterations, which we take as approximation for the real value $MSE(\hat{\mu}_{dt})$. For a

Table 4 Prediction results for MERY (left) and RY (right) models

Scenario	MERY			RY				
	RMSE	RRMSE	ABIAS	RBIAS	RMSE	RRMSE	ABIAS	RBIAS
A.1	4.228	0.0118	3.393	0.0095	4.278	0.0120	3.436	0.0096
A.2	4.155	0.0115	3.317	0.0092	4.178	0.0116	3.334	0.0093
A.3	4.183	0.0117	3.345	0.0093	4.189	0.0117	3.351	0.0094
A.4	3.594	0.0100	2.873	0.0080	3.638	0.0101	2.907	0.0081
A.5	3.119	0.0087	2.492	0.0069	3.129	0.0087	2.499	0.0070
B.1	10.426	0.0291	8.324	0.0233	10.468	0.0293	8.342	0.0233
B.2	10.257	0.0285	8.197	0.0228	10.315	0.0286	8.255	0.0229
B.3	10.217	0.0285	8.174	0.0228	10.359	0.0289	8.284	0.0231
B.4	10.229	0.0285	8.184	0.0228	10.328	0.0288	8.252	0.0230
B.5	10.146	0.0282	8.115	0.0226	10.346	0.0288	8.277	0.0230
C.1	13.011	0.0364	10.331	0.0289	13.459	0.0376	10.770	0.0301
C.2	12.757	0.0354	10.145	0.0282	13.441	0.0373	10.732	0.0298
C.3	12.476	0.0349	9.972	0.0279	13.430	0.0375	10.739	0.0300
C.4	12.755	0.0356	10.168	0.0283	13.348	0.0372	10.683	0.0298
C.5	12.478	0.0347	9.969	0.0277	13.420	0.0373	10.740	0.0299

more compact presentation, we use averaged measures over the subdomains, as in the previous section. For the number of bootstrap replicates, we use $B = 500$.

The results are summarized in Table 5. Note that the columns MSE and mse^* correspond to the means of mse_{dt} and mse_{dt}^* over all subdomains of the respective scenario. From Table 5, we observe that in the A- and B-scenarios, the performance of both MSE estimators is decent. The relative bias ranges between 0.1 and 7.2%, which is a reasonable magnitude for a mixed model with a multi-level random effect structure. However, we see that the bias of the MSE estimator mse_2^* is always smaller than for the standard estimator mse_1^* except for Scenario A.5. Further, by looking at the RMSE, it can be concluded that the second MSE estimator is more efficient than the first. In the C-scenarios, which resemble large measurement errors, the standard estimator mse_1^* become visibly biased. Its relative bias ranges from 10.0 to 24.6%. The bias-corrected estimator mse_2^* , on the other hand, is not as much affected by the large errors. Its relative bias only ranges from 0.6 to 1.9%. In addition to that, by looking at the RMSE, a significant efficiency advantage by the second estimator becomes evident. Accordingly, including the analytical approximation to the first MSE component improves the estimation accuracy considerably in these settings.

Figure 1 shows the convergence behavior of mse_2^* . In particular, it shows the distribution of $mse_{2dt}^*(B) - mse_{2dt}^*(500)$, $d = 1, \dots, D$, $t = 1, \dots, T$, when the parametric bootstrap estimate $mse_{2dt}^*(B)$ is calculated based on $B = 1, \dots, 500$ replicates. The light gray band displays the inner 90% of differences over all subdomains, while the gray band marks the inner 70%. We see that the mean stabilizes at around $B = 150$, implying that 150 bootstrap replicates are sufficient to obtain decent MSE estimates in terms of overall performance. However, if we require that the inner 90% differences vary less than 1% of $MSE(\hat{\mu})$, which would roughly be ± 0.48 in Scenario 1, $B = 300$ is an appropriate choice.

6 Application

This section presents an empirical application to health mapping on the example of regional age-referenced hypertension prevalence in Germany. Note that further results are located in Appendix 5 of the supplemental material.

6.1 Data and specification

We use the health survey GEDA (Robert Koch Institute 2013, 2014a, b) from 2009 ($n_1 = 21,262$), 2010 ($n_2 = 22,050$) and 2012 ($n_3 = 19,294$) as data basis. It contains citizens of age 18+ from the German population that are interviewed via CATI in a representative telephone sample. The data sets contain health-related information of participants on the unit level. For further information on the survey as well as its sampling design and response rates, see Robert Koch Institute (2013) and Robert Koch Institute (2014a, b).

We define the domains as cross combinations of the seven Nielsen regions of Germany and the age groups I: 18–29, II: 30–39, III: 40–49, IV: 50–59, V: 60–69, VI:

Table 5 Results of mean-squared error estimation

Scenario	mse_1^*			mse_2^*			BIAS	RBIAS	RMSE	mse^*	BIAS	RBIAS	RMSE
	MSE	mse^*	BIAS	MSE	mse^*	BIAS							
A.1	17.97	19.23	1.26	17.97	18.24	0.26	0.01	0.01	6.15				
A.2	17.38	17.97	0.59	17.38	17.08	-0.30	-0.02	-0.02	4.93				
A.3	17.60	17.06	-0.54	17.60	17.55	-0.05	-0.00	-0.00	4.06				
A.4	13.01	13.60	0.59	13.01	12.61	-0.40	-0.03	-0.03	4.94				
A.5	9.82	10.01	0.19	9.82	9.56	-0.26	-0.03	-0.03	3.86				
B.1	110.22	116.89	6.67	110.22	109.86	-0.36	-0.00	-0.00	27.86				
B.2	105.87	113.50	7.63	105.87	109.12	3.25	0.03	0.03	24.00				
B.3	105.09	108.20	3.11	105.09	105.45	0.36	0.00	0.00	19.77				
B.4	105.20	110.04	4.84	105.20	105.31	0.11	0.00	0.00	20.17				
B.5	103.57	106.17	2.60	103.57	103.35	-0.22	-0.00	-0.00	18.52				
C.1	170.30	212.21	41.91	170.30	167.04	-3.26	-0.02	-0.02	71.04				
C.2	163.74	188.45	24.72	163.74	162.38	-1.35	-0.00	-0.00	56.63				
C.3	156.52	173.24	16.72	156.52	157.11	0.59	0.00	0.00	38.73				
C.4	163.65	192.56	28.91	163.65	163.65	0.00	0.00	0.00	53.77				
C.5	156.56	172.14	15.58	156.56	155.54	-1.01	-0.01	-0.01	37.98

Fig. 1 Convergence behavior of MSE estimators per subdomain

70–79 and VII: 80+. Accordingly, we have $D = 49$ domains that are observed in $T = 3$, periods, which implies $M = 147$ domain–time sets in total. The Nielsen population structure does not correspond to the one used by the German statistical offices, and therefore, it is difficult to collect aggregated auxiliary variables to apply the standard SAE approach based on area-level models. However, we can obtain auxiliary data from the health survey GEDA and apply the statistical methodology based on the MERY model. For this sake, we calculate a set of Horvitz–Thompson estimators for the regional hypertension prevalence as target variable and other health indicators that we use as explanatory variables. Further, we exploit the fact that the hierarchical structure of the MERY model in Sect. 2.1 can be interpreted as a collection of domains that is observed over multiple time periods.

The objective is to estimate the hypertension prevalence per each domain d and time period t . That is to say, we define the number of individuals with high blood pressure as the model component μ_{dt} , $d = 1, \dots, D$, $t = 1, \dots, T$, in the sense of Sect. 2.1, and divide by the known domain–time size N_{dt} afterward. In order to predict the total of *hypertensive* people, we use the estimated totals of people with *diabetes* and *increased blood lipids* per domain and time period as covariates, $\tilde{\mathbf{x}}_{1,dt} = (\tilde{x}_{1,1dt}, \tilde{x}_{1,2dt})$ in the notation of Sect. 2.1, since these diagnoses are well known to be associated with hypertension (see, e.g., Lastra et al. 2014). To determine whether a person suffers from any of these medical conditions in the sample, we rely on the 12-month prevalence profiles used by Robert Koch Institute (2012).

The main advantage of the new MERY model is that it allows for SAE based on uncertain data obtained from dependent survey estimates. In order to provide a suitable data basis for our method, we calculate Horvitz–Thompson estimators for the number of citizens per domain–time in the population that are associated with any of the three medical conditions. These results yield us the model components y_{dt} and $\mathbf{x}_{1,dt}$ that are

Table 6 Data quartiles across domains and time periods

	min	q0.25	q0.50	q0.75	max
n_{dt}	51	271	421	554	1165
$cv(y_{dt})$	5.92	9.06	12.54	19.8	48.41
$cv(x_{1,1dt})$	10.84	19.32	27.65	41.39	72.45
$cv(x_{1,2dt})$	7.65	10.95	15.09	22.76	76.64
$\rho(y_{dt}, x_{1,1dt})$	−0.0621	0.0878	0.1458	0.2082	0.3128
$\rho(y_{dt}, x_{1,2dt})$	−0.3006	0.1061	0.1598	0.2031	0.3500
$\rho(x_{1,1dt}, x_{1,2dt})$	−0.1378	0.0519	0.1108	0.1565	0.3228

direct estimators of μ_{dt} and $\tilde{x}_{1,dt}$, respectively. Note that the estimators are obtained from the same set of sample observations per domain and time period. In standard measurement error models for SAE, this would not be possible as y_{dt} and $x_{1,dt}$ are not allowed to have a covariances different from zero. The MERY model, on the other hand, allows for covariances between the direct estimators. In fact, one could argue that it even encourages dependency between the errors, as the covariances between y_{dt} and $x_{1,dt}$ mark additional information that can be used to improve the prediction. Recall that in Sect. 2.1, normality was assumed for these components. For the application, the applied Horvitz–Thompson estimators satisfy this assumption asymptotically, as demonstrated by Hájek (1960), Berger (1998) and Chen and Rao (2007).

We further calculate direct estimators of the variances and covariances of y_{dt} and $x_{1,dt}$. In this way, we construct the matrices Σ_{dt} and Σ_{vdt} for every domain and time period, in accordance with Sect. 2.1. This can be done using the unit-level observations in GEDA. We calculate direct estimators of the variances $\text{var}(y_{dt})$ and $\text{var}(x_{1,jdt})$, $j = 1, 2$, based on the unit-level survey data. For $j = 1, 2$, the estimators of the covariances between the direct target and covariate estimators are obtained from Wood (2008)

$$\text{cov}(y_{dt}, x_{1,jdt}) = \frac{1}{2} \left[(\text{var}(y_{dt}) + \text{var}(x_{1,jdt}) - \text{var}(y_{dt} - x_{1,jdt})) \right],$$

$$d = 1, \dots, D, \quad t = 1, \dots, T.$$

The estimator of the covariance between the direct covariate estimators $x_{1,1dt}$ and $x_{1,2dt}$ is obtained analogously.

Table 6 presents the quartiles, across domains and time periods, of the sample sizes n_{dt} , the coefficients of variation, $cv(y_{dt})$, $cv(x_{1,1dt})$ and $cv(x_{1,2dt})$, of the direct estimators of the totals of the objective variable (hypertension) and of the two auxiliary variables (diabetes and lipids), and the corresponding correlation coefficients $\rho(y_{dt}, x_{1,1dt})$, $\rho(y_{dt}, x_{1,2dt})$, $\rho(x_{1,1dt}, x_{1,2dt})$. This table shows that some domain–time sample sizes are small and that the direct estimators of the totals of the considered medical variables are correlated. The introduced MERY model is the first one that can deal with this situation.

The full specification of the MERY model is

$$y_{dt} = \beta_0 + x_{1,1dt}\beta_1 + x_{1,2dt}\beta_2 + v_{1,1dt}\beta_1 + v_{1,2dt}\beta_2 + u_{1,d} + u_{2,d} + e_{dt}, \\ d = 1, \dots, D, \quad t = 1, \dots, T.$$

where y_{dt} , $x_{1,1dt}$ and $x_{1,2dt}$ are the Horvitz–Thompson estimators of the number of individuals with high blood pressure, diabetes and increased blood lipids per domain d and time period t , respectively. We apply the ML algorithm from Sect. 4 for model parameter estimation. Further, we use the subsequent parametric bootstrap procedure in order to approximate the standard deviation of the model parameter estimates in order to evaluate their significations.

1. By using the data (y_d, X_d) , $d = 1, \dots, D$, calculate $\hat{\theta}$ and $\hat{V}_{2d} = \hat{\sigma}_2^2 T_T$.
2. Repeat B times
 - 2.1 For $d = 1, \dots, D$, generate $u_{1,d}^{*(b)} \sim N(0, \hat{\sigma}_1^2)$, $u_{2,d}^{*(b)} \sim N_T(\mathbf{0}, \hat{V}_{2d})$, $\epsilon_d^{*(b)} = (\mathbf{e}_d^{*(b)'} , \mathbf{v}_d^{*(b)'})' \sim N_{T(P_1+1)}(\mathbf{0}, \Sigma_d)$, $\mu_d^{*(b)} = X_d \hat{\beta} + \hat{B}_d \mathbf{v}_d^{*(b)} + \mathbf{1}_T u_{1,d}^{*(b)} + \mathbf{u}_{2,d}^{*(b)}$ and $\mathbf{y}_d^{*(b)} = \mu_d^{*(b)} + \mathbf{e}_d^{*(b)}$.
 - 2.2 By using the data $(\mathbf{y}_d^{*(b)}, X_d)$, $d = 1, \dots, D$, calculate the estimator $\hat{\theta}^{*(b)}$.
3. For $\tau \in \{\beta_0, \beta_1, \beta_2, \sigma_1^2, \sigma_2^2\}$, calculate the bootstrap estimator of $sd(\hat{\tau})$; i.e.,

$$sd^*(\hat{\tau}) = \sqrt{\frac{1}{B} \sum_{b=1}^B \left(\hat{\tau}^{(b)} - \frac{1}{B} \sum_{k=1}^B \hat{\tau}^{(k)} \right)^2}.$$

In accordance with Sect. 5.4, we rely on $B = 300$ bootstrap replicates. For significance testing, we use a t -test with test statistic $t(\hat{\tau}) = \hat{\tau}/sd^*(\hat{\tau})$ that is asymptotically distributed as standard normal. The resulting values are located in the probability density function of the standard normal to find the respective p values. Confidence intervals for the model parameter estimates are calculated by $\hat{\tau} \pm t_{(m, 1-\alpha/2)} sd^*(\hat{\tau})$, where $t_{(\cdot)}$ is the corresponding quantile of the t -distribution with $m = DT - 3$ degrees of freedom and significance level α . Please note that the mathematically correct distribution would be the normal distribution, as maximum likelihood estimates are asymptotically normal distributed. However, the t -distribution provides slightly more conservative test decisions and confidence intervals. This seems reasonable to us in a small area setting, where asymptotic arguments, due to small sample sizes, are critical. In this particular application, the confidence intervals are almost the same, as the number of degrees of freedom is considerably large, such that the corresponding t -distribution already approximates the normal distribution quite well.

Table 7 displays the results of model parameter estimation obtained from the ML algorithm as well as the parametric bootstrap procedure. For the sake of presentability, we show the random effect standard deviations, $\hat{\sigma}_1$ and $\hat{\sigma}_2$, rather than the respective variances. It can be seen that all model parameters except for the intercept are significant on a 1% level. The intercept is not significantly different from zero. The

Table 7 Results of model parameter estimation

Parameter	Estimate	SD	<i>p</i> value	95%-Confidence Int.
Intercept	5954.6601	4366.3330	0.15742	[− 2674.233; 14,583.550]
Diabetes	1.2516	0.1277	0.00000	[0.999; 1.504]
Lipids	0.8884	0.0502	0.00000	[0.789; 0.988]
σ_1	1300.8130	421.8853	0.00344	[467.069; 2134.557]
σ_2	1111.9911	325.9147	0.00118	[467.908; 1756.075]

Fig. 2 Estimated hypertension prevalence over age groups

signs of the regression parameters β_1 and β_2 are plausible, as diabetes and increased blood lipids are positively correlated with hypertension. Furthermore, the domain and domain–time-level random effects are clearly visible in the distribution of y_{dt} .

6.2 Results

The estimated prevalence for 2009 over all domains is 26.1%. The estimated prevalence for 2010 is 26.4%, and for 2012, we have 27.1%. The estimated evolution of the hypertension prevalence over age groups for all three considered periods is displayed in Fig. 2. We see that the hypertension prevalence is low in the youngest cohorts (18–39). From age 40–79, a steep monotonous increase is evident for all three periods. For individuals of age 80+, the curve flattens on a high level in 2010 and 2012. An interesting observation is that in 2009, the prevalence for the eldest cohort is slightly lower than for the second eldest. A possible explanation is that for age 80+, individuals without hypertension have higher life expectancy than individuals with hypertension

Fig. 3 Empirical best predictions $\hat{\mu}_{dt}/N_{dt}$ and direct estimates y_{dt}/N_{dt}

due to a lesser risk of suffering a stroke or other cardiovascular events (Kassai et al. 2005).

Figure 3 shows the empirical best predictions dependent on the respective direct estimates. The left plot displays the point estimates. The right plot marks the comparison of the EBP RMSEs and the direct estimator standard deviations. We have used the bias-corrected parametric bootstrap estimator presented in Sect. 3.2, to estimate the EBP RMSEs of the domain–time characteristic predictions.

From the plot, we see that the direct estimates vary randomly around the prevalence predictions. This is in line with theory, as the EBP based on the MERY model exploits the design unbiasedness of the direct estimators with the error components being zero on expectation. Furthermore, it indicates that we do not have systematic prediction bias, which is further supported by the simulation results in Sect. 5.3. From the right plot, we see that the MERY predictions are always more efficient than the direct estimates. The EBP RMSEs are always smaller than the corresponding standard deviations of the direct estimators. Accordingly, we can conclude that the application of the MERY model allows for a considerable gain in accuracy. On the other hand, we also remark that most coefficients of variation of the direct estimators are below 20%, since the sample sizes in several domains are quite high. For some Statistical Offices, 20% marks a quality threshold that admits a publication of the results. However, the MERY still manages to improve these results uniformly, which is why its application is recommendable in this setting.

Figure 4 displays the estimated regional hypertension prevalence over the seven Nielsen regions in percent for 2010. The figures are with respect to the total regional populations (without age referencing). We see that the highest prevalences are located in the east of Germany, which is the former territory of the German Democratic

Fig. 4 Estimated regional prevalence for all age groups, 2010

Republic. The lowest prevalences are located in Baden-Württemberg and Bavaria. The estimated distribution is plausible, as similar patterns have been found for hypertension for instance by Burgard et al. (2020b). Further, similar distributions are known for strongly correlated conditions, such as diabetes (Schipf et al. 2014).

7 Conclusions

This paper introduces a new small area model for the estimation of regional population characteristics that allows for explanatory variables with measurement errors that are possibly correlated with the errors of the target variable. The approach enriches the modeling process by admitting auxiliary information that comes from administrative records, from survey samples other than the target survey, or even from that same target survey. The MERY model is a generalization of the Rao–Yu model that can be applied to both temporal data (domains and periods) or geographically nested data (domains and subdomains). With these features, it is broadly applicable in practice and marks an important contribution to both the SAE and the measurement error model literature.

Concerning model parameter estimation, the paper proposes the maximum likelihood method by implementing a Fisher scoring algorithm. For predicting the measurement errors, the random effects, as well as the target population quantities, the paper derives best predictors and calculates the corresponding EBPs. After calculating the variances and covariances of the random effects and errors, the MSEs of the BPs

are approximated and applied to construct parametric bootstrap procedures with bias correction term. Several simulation experiments are performed to study the behavior of the fitting algorithm, the EBPs and the parametric bootstrap estimators of the MSEs. In particular, the efficiency gain derived from the modeling of measurement errors and their correlations is shown and a recommendation about the number of bootstrap replicates is given. An interesting finding is that if the measurements errors are ignored, then the MSEs of predictors are underestimated, which leads to the predictions seeming more accurate than they actually are (Appendix 4 of supplemental material). The new statistical methodology is applied to the health survey GEDA from 2009, 2010 and 2012. The estimation target is the regional hypertension prevalence in Germany. In order to predict hypertension, the employed covariates, diabetes and increased blood lipids prevalence, are also taken from the health survey GEDA. The new MERY model is the only model in the statistical literature that admits this possibility. Finally, the obtained EBPs are shown to be more precise than the corresponding direct estimates.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11749-021-00776-w>.

Acknowledgements This research is supported by the Spanish Grant PGC2018-096840-B-I00, by the Grant “Algorithmic Optimization (ALOP)—Graduate School 2126” funded by the German Research Foundation, as well as the Grant “RIFOSS—Research Innovation for Official and Survey Statistics” funded by the German Federal Statistical Office.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability The data used for this study are retrieved from the German health survey *Gesundheit in Deutschland aktuell (GEDA)*. The related data sets are available on request at the Robert Koch Institute, Germany. The authors have no permission to share them with third parties.

Code availability The R-code for this paper is accessible as online supplement.

Declarations

Conflict of interest The authors declare that they have no conflict interest to report.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adin A, Lee D, Goicoa T, Ugarte MD (2019) A two-stage approach to estimate spatial and spatio-temporal disease risks in the presence of local discontinuities and clusters. *Stat Methods Med Res* 28(9):2595–2613
- Arima S, Bell WR, Datta GS, Franco C, Liseo B (2017) Multivariate Fay–Herriot Bayesian estimation of small area means under functional measurement error. *J R Stat Soc Ser A* 180(4):1191–1209

- Benavent R, Morales D (2016) Multivariate Fay–Herriot models for small area estimation. *Comput Stat Data Anal* 94:372–390
- Benavent R, Morales D (2021) Small area estimation under a temporal bivariate area-level linear mixed model with independent time effects. *Stat Methods Appl* 30:195–222. <https://doi.org/10.1007/s10260-020-00521-x>
- Berger YG (1998) Rate of convergence to normal distribution for the Horvitz–Thompson estimator. *J Stat Plan Inference* 67(2):209–226
- Bernal RTI, de Carvalho QH, Pell JP, Leyland AH, Dundas R, Barreto ML, Malta DC (2020) A methodology for small area prevalence estimation based on survey data. *Int J Equity Health* 19:124
- Boubeta M, Lombardía MJ, Morales D (2017) Poisson mixed models for studying the poverty in small areas. *Comput Stat Data Anal* 107(2):32–47
- Burgard JP, Esteban MD, Morales D, Pérez A (2020a) A Fay–Herriot model when auxiliary variables are measured with error. *TEST* 29:166–195. <https://doi.org/10.1007/S11749-019-00649-3>
- Burgard JP, Krause J, Münnich R (2020b) An elastic net penalized small area model combining unit- and area-level data for regional hypertension prevalence estimation. *J Appl Stat*. <https://doi.org/10.1080/02664763.2020.1765323>
- Burgard JP, Esteban MD, Morales D, Pérez A (2021) Small area estimation under a measurement error bivariate Fay–Herriot model. *Stat Methods Appl* 30:79–108. <https://doi.org/10.1007/s10260-020-00515-9>
- Chen J, Rao JNK (2007) Asymptotic normality under two-phase sampling designs. *Stat Sin* 17(3):1047–1064
- Datta GS, Lahiri P, Maiti T, Lu KL (1999) Hierarchical Bayes estimation of unemployment rates for the US States. *J Am Stat Assoc* 94(448):1074–1082
- Datta GS, Lahiri P, Maiti T (2002) Empirical Bayes estimation of median income of four-person families by state using time series and cross-sectional data. *J Stat Plan Inference* 102(1):83–97
- Esteban MD, Morales D, Pérez A, Santamaría L (2012) Small area estimation of poverty proportions under area-level time models. *Comput Stat Data Anal* 56:2840–2855
- Esteban MD, Lombardía MJ, López-Vizcaíno E, Morales D, Pérez A (2020) Small area estimation of proportions under area-level compositional mixed models. *TEST* 29(3):793–818
- Fay RE, Herriot RA (1979) Estimates of income for small places: an application of James–Stein procedures to census data. *J Am Stat Assoc* 74(366):269–277
- Ghosh M, Nangia N, Kim D (1996) Estimation of median income of four-person families: a Bayesian time series approach. *J Am Stat Assoc* 91(43):1423–1431
- González-Manteiga W, Lombardía MJ, Molina I, Morales D, Santamaría L (2008) Analytic and bootstrap approximations of prediction errors under a multivariate Fay–Herriot model. *Comput Stat Data Anal* 52:5242–5252
- Hájek J (1960) Limiting distributions in simple random sampling from a finite population. *Publ Math Inst Hung Acad Sci* 5:361–374
- Jedrejczak A, Kubacki J (2017) Estimation of small area characteristics using multivariate Rao–Yu model. *Stat Trans* 18(4):725–742
- Jennrich RI, Sampson PF (1976) Newton–Raphson and related algorithms for maximum likelihood variance component estimation. *Technometrics* 18(1):11–17
- Jiang J, Lahiri P (2006) Mixed model prediction and small area estimation. *TEST* 15(1):1–96
- Kassai B, Boissel JP, Cucherat M, Boutitie F, Gueyffier F (2005) Treatment of high blood pressure and gain in event-free life expectancy. *Vasc Health Risk Manag* 1(2):163–169
- Lange C, Jentsch F, Allen J, Hoebel J, Kratz AL, von der Lippe E, Müters S, Schmich P, Thelen J, Wetzstein M, Fuchs J, Ziese T (2015) Data resource profile: German health update (GEDA)—the health interview survey for adults in Germany. *Int J Epidemiol* 44(2):442–450
- Lastra G, Syed S, Kurukulasuriya L, Manrique C, Sowers J (2014) Type 2 diabetes mellitus and hypertension: an update. *Endocrinol Metab Clin North Am* 43(1):103–122
- Longford NT (1987) A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika* 74(4):817–827
- Marhuenda Y, Molina I, Morales D (2013) Small area estimation with spatio-temporal Fay–Herriot models. *Comput Stat Data Anal* 58:308–325
- Mills CW, Johnson G, Huang TTK, Balk D, Wyka K (2020) Use of small-area estimates to describe county-level geographic variation in prevalence of extreme obesity among US adults. *JAMA Netw Open* 3(5):e204289

- Morales D, Pagliarella MC, Salvatore R (2015) Small area estimation of poverty indicators under partitioned area-level time models. *SORT Stat Oper Res* 39(1):19–34
- Münnich R, Burgard JP, Krause J (2019) Adjusting selection bias in German health insurance records for regional prevalence estimation. *Popul Health Met* 17:13
- National Center for Health Statistics (2020) National health and nutrition examination survey: 1999–2020 survey content brochure. Online. https://www.cdc.gov/nchs/data/nhanes/survey_contents.pdf
- Pfeffermann D (2002) Small area estimation—new developments and directions. *Int Stat Rev* 70(1):55–76
- Pfeffermann D (2013) New important developments in small area estimation. *Stat Sci* 28(1):40–68
- Pfeffermann D, Burck L (1990) Robust small area estimation combining time series and cross-sectional data. *Surv Methodol* 16(2):217–237
- Rao JNK, Molina I (2015) Small area estimation. Wiley series in survey methodology. Wiley, Hoboken
- Rao JNK, Yu M (1994) Small-area estimation by combining time-series and cross-sectional data. *Can J Stat* 22(4):511–528
- Robert Koch Institute (2012) Daten und Fakten: Ergebnisse der Studie “Gesundheit in Deutschland aktuell 2010”. Online. https://www.gbe-bund.de/pdf/GEDA_2010_Gesamtausgabe.pdf
- Robert Koch Institute (2013) German Health Update 2010 (GEDA 2010). Public use file third version. Online. <https://doi.org/10.7797/27-200910-1-1-3>
- Robert Koch Institute (2014a) German Health Update 2009 (GEDA 2009). Public use file second version. Online. <https://doi.org/10.7797/26-200809-1-1-2>
- Robert Koch Institute (2014b) German Health Update 2012 (GEDA 2012). Public use file first version. Online. <https://doi.org/10.7797/29-201213-1-1-1>
- Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, Colagiuri S, Guariguata L, Motala AA, Ogurtsova K, Shaw JE, Bright D, Williams R (2019) Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: results from the international diabetes federation diabetes atlas, 9th edition. *Diab Res Clin Pract* 157:107843
- Schipf S, Ittermann T, Tamayo T, Holle R, Schunk M, Maier W, Meisinger C, Thorand B, Kluttig A, Greiser KH, Berger K, Müller G, Moebus S, Slomiany U, Rathmann W, Völzke H (2014) Regional differences in the incidence of self-reported type 2 diabetes in Germany: Results from five population-based studies in Germany (Diab-core consortium). *J Epidemiol Community Health* 68:1088–1095
- Torabi M, Datta GS, Rao JNK (2009) Empirical Bayes estimation of small area means under a nested error linear regression model with measurement errors in the covariates. *Scand J Stat* 36(2):355–368
- Trandafir PC, Adin A, Ugarte MD (2020) Space-time analysis of ovarian cancer mortality rates by age groups in Spanish provinces (1989–2015). *BMC Public Health* 20:1244. <https://doi.org/10.1186/s12889-020-09267-3>
- Ubaidillah A, Notodiputro KA, Kurnia A, Wayan I (2019) Multivariate Fay–Herriot models for small area estimation with application to household consumption per capita expenditure in Indonesia. *J Appl Stat* 46(15):2845–2861
- Wood J (2008) On the covariance between related Horvitz–Thompson estimators. *J Offic Stat* 24(1):53–78
- Ybarra LMR, Lohr SL (2008) Small area estimation when auxiliary information is measured with error. *Biometrika* 95:919–931
- You Y, Rao JNK (2000) Hierarchical Bayes estimation of small area means using multi-level models. *Surv Methodol* 26(2):173–181

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.