UNIVERSITÄT
TRIER

Model-Based Prediction and Estimation Using Incomplete Survey Data

Doctoral thesis submitted in partial fulfillment of the requirements for the degree

Doctor rerum politicarum (Dr. rer. pol.)

to the Faculty IV at Trier University by

Anna-Lena Wölwer

Born 7 March 1992 in Mayen

Supervisors	
Prof. Dr. Ralf Münnich	Trier University, Trier, Germany
Prof. Dr. Domingo Morales	Universidad Miguel Hernández, Elche, Spain

Date of submission	09.06.2022
Date of defense	09.12.2022
Month of publication	January 2023

Code and examples for selected parts of this thesis are available at https://github.com/alwoelwer/Doc_Thesis_AnnaLena_Woelwer.

Acknowledgements

I would like to thank my first supervisor Prof. Dr. Ralf Münnich for the wonderful years of collaboration since my undergraduate studies and for his support and exchange of ideas on the research presented in this thesis. Furthermore, my thanks go to my second supervisor Prof. Dr. Domingo Morales for the great collaboration and insightful discussions. I would also like to thank my former colleagues at the Department of Economic and Social Statistics at Trier University, especially Dr. Jan Pablo Burgard, for the informal academic exchange and the unforgettable time spent together. My special thanks go to Daniel and my family for their loving support.

The thesis was supported by a doctoral scholarship from the *Studienstiftung des deutschen Volkes*, for which I am very grateful. For the research contributions presented in Chapters 3 and 4, I would like to thank the Federal Statistical Office of Germany for its support within the research project *Research Innovation for Official and Survey Statistics (RIFOSS)*.

Anna-Lena Wölwer

Education

10/2018 – 12/2022 **PhD student**
Trier University, Economic and Social Statistics Department

08/2018 **M.Sc. Economics – European Political Economy**
Trier University
Thesis: *Imputation of Missing German Rent Prices on NUTS3*

10/2017 **M.Sc. Survey Statistics**
Trier University
Thesis: *Multivariate Fay-Herriot Models under Missings*
European Master in Official Statistics (EMOS) Certificate

09/2015 – 01/2016 **Study Abroad (Erasmus)**
University of Southampton, England
Subjects: Social Statistics and Demography

09/2014 **B.Sc. Economics**
Trier University
Thesis: *Comparison of Different Estimation Methods for Gravity Models*

03/2011 **Allgemeine Hochschulreife**
Megina-Gymnasium Mayen

Academic career

10/2019 – 09/2022 **PhD scholarship**
Awarded by *Studienstiftung des deutschen Volkes*

10/2017 – 12/2021 **Research assistant**
Trier University, Economic and Social Statistics Department

Contents

German Summary

Survey Daten können unter verschiedensten Blickwinkeln als unvollständig bzw. als partiell fehlend angesehen werden und es gibt verschiedene Möglichkeiten, mit diesen Daten in der Prädiktion und Schätzung interessierender volkswirtschaftlicher Größen umzugehen. In dieser Arbeit werden zwei ausgewählte Forschungskontexte vorgestellt, in denen die Prädiktion bzw. Schätzung unter unvollständigen Survey Daten untersucht wird. Diese Kontexte sind zum einen die Untersuchung zusammengesetzter Schätzer im deutschen Mikrozensus (Kapitel 3 und 4) und zum anderen Erweiterungen multivariater Fay-Herriot (MFH) Modelle (Kapitel 5 und 6), die bei Small Area Problemen Anwendung finden.

Zusammengesetzte Schätzer sind Schätzmethoden, die die Stichprobenüberlappung in rotierenden Panel Surveys wie dem deutschen Mikrozensus zur Stabilisierung der Schätzung interessierender Größen (z.B. Erwerbsstatistiken) nutzen. Durch die partiellen Stichprobenüberlappungen in rotierenden Panel Surveys liegen immer nur für einen Teil der Befragten Informationen aus vorangegangenen Erhebungen vor. Die resultierenden Daten sind damit partiell fehlend.

MFH Modelle sind modellbasierte Schätzmethoden, die mit aggregierten Survey Daten arbeiten, um im Vergleich zu klassischen Schätzmethoden präzisere Schätzergebnisse für Small Area Probleme zu erhalten. In den Modellen werden mehrere interessierende Größen gleichzeitig modelliert. Die Survey-Schätzwerte dieser Größen, die in MFH Modelle als Input eingehen, sind häufig partiell fehlend. Wenn die interessierenden Domains nicht explizit im Stichprobendesign berücksichtigt wurden, kann es sein, dass die Größe der Stichprobe, die auf sie entfällt, so klein ist, dass entweder gar keine Schätzwerte berechnet werden können oder aber die Schätzwerte von statistischen Ämtern nicht veröffentlicht werden, da ihre Varianzen zu groß ist.

Nach einem Überblick zu theoretischen und methodischen Grundlagen der Survey Statistik in Kapitel 2 stellt Kapitel 3 die Generierung eines Längsschnittdatensatz vor, auf Basis dessen in Kapitel 4 designbasierte Simulationsstudien zum Einsatz von zusammengesetzten Schätzern im Mikrozensus durchgeführt werden. Für diese Studien wird ein Längsschnittdatensatz mit monatlichen Erwerbsinformationen benötigt, der die deutsche Wohnbevölkerung abbildet. Auf Grundlage des SIAB Datensatzes[1] (Antoni et al., 2019) werden Prädiktionsmodelle für monatliche Erwerbsübergänge geschätzt, mit denen monatliche Erwerbsinformationen im RIFOSS Datensatz[2], einem halb-synthetischen Querschnittdatensatz der deutschen Wohnbevölkerung, generiert werden. Für die Prädiktionsmodelle werden mehrere generalisierte additive Modelle, die jeweils auf Substichproben der aufbereiteten SIAB Daten geschätzt werden (Subagging), zu optimal gewichteten Ensemble Modellen (Stacking) verbunden. In der Optimierung der Ensemble Gewichte wird eine in diesem Kapitel vorgestellte Erweiterung des Brier Scores verwendet.

[1] Factually anonymous Version of the Sample of Integrated Labour Market Biographies (SIAB-Regionalfile) Version 7517 v1.

[2] Version RIFOSS_GG_v0.1.1_vanilla_ice_cream.

In Kapitel 4 wird der Einsatz von zusammengesetzten Schätzern für Erwerbsstatistiken im deutschen Mikrozensus untersucht. Das Design des deutschen Mikrozensus wurde 2020 wesentlichen Änderungen unterzogen, welche neue Möglichkeiten für den Einsatz dieser Schätzmethoden schaffen. In dem Kapitel wird analysiert, welche Einsatzmöglichkeiten sich für zusammengesetzte Schätzer aus dem neuen Mikrozensus Design ergeben. Beispielsweise bieten sich verschiedene Stichprobenüberlappungen zu vorangegangenen Zeitpunkten für die Nutzung in den zusammengesetzten Schätzern an. Zusätzlich werden Anpassungen der Formeln der zusammengesetzten Schätzer für die sich aus dem Mikrozensus Design ergebenden regional heterogenen Stichprobenüberlappungen vorgestellt. In einer designbasierten Simulationsstudie, deren Basis der in Kapitel 3 erstellte Datensatz ist, wird die Performanz der angepassten Methoden unter verschiedenen Sets an Stichprobenüberlappungen für verschiedene Erwerbsstatistiken verglichen.

Im Fokus von Kapitel 5 und 6 stehen unvollständige aggregierte Survey-Schätzwerte, die zur Small Area Schätzung in MFH Modellen verwendet werden. Mit den Beiträgen der beiden Kapitel ist es möglich, die unter den jeweiligen Modellen sogenannten besten Prädiktoren multivariater Domain-Indikatoren zu berechnen, auch wenn die in die Modelle eingehenden Survey-Schätzwerte partiell fehlen.

Kapitel 5 beschäftigt sich mit den besten Prädiktoren von (potenziell nicht-linearen) Indikatoren unter MFH Modellen. Ein nicht-linearer Indikator kann beispielsweise die Erwerbslosenrate sein. Für diese und andere Indikatoren werden häufig Plug-in Schätzer genutzt. Als Alternative werden in Kapitel 5 die MSE-optimalen Prädiktoren von Domain-Indikatoren unter MFH Modellen untersucht. Diese sind als mehrdimensionale Integrale gegeben, die sich im generellen Fall nicht analytisch berechnen lassen. Es werden deswegen verschiedene Methoden zur Approximation dieser Integrale verglichen. Zur MSE Schätzung werden parametrische Bootstrap Prozeduren vorgestellt. In modellbasierten Simulationsstudien werden die verschiedenen Approximationen evaluiert und ihre Performanz mit der Performanz der entsprechenden Plug-in Prädiktoren verglichen. Des Weiteren werden die MSE Schätzer evaluiert. Die vorgestellte Methode wird anhand der Schätzung der Erwerbslosenrate in Kreuzkombinationen spanischer Provinzen mit Alters- und Geschlechtsklassen illustriert. Dazu werden öffentlich zugängliche Mikrodaten der spanischen Arbeitskräfteerhebung verwendet.

MFH Modelle können nur auf Grundlage der Domain-Informationen geschätzt werden, für die Survey-Schätzwerte für alle abhängigen Variablen vorliegen, was ihre Anwendbarkeit in der Praxis wesentlich beschränkt. In Kapitel 6 wird eine Generalisierung von MFH Modellen für partiell fehlende Werte, genannt MMFH Modelle, vorgestellt. Für die MMFH Modelle werden Algorithmen zur ML und REML Parameterschätzung gegeben und die Formeln für die besten Prädiktoren unter dem Modell sowie deren MSE Schätzer hergeleitet. In einer modellbasierten Simulationsstudie werden die vorgestellten MMFH Algorithmen und Formeln validiert. Des Weiteren wird ihre Performanz mit der Performanz der entsprechenden univariaten und multivariaten Fay-Herriot Modelle verglichen. Eine illustrative Anwendung basierend auf öffentlich zugänglichen Daten des U.S. Zensus Büros zeigt die praktische Notwendigkeit der vorgestellten Methode sowie ihre Anwendbarkeit.

Kapitel 7 fasst die Beiträge und Ergebnisse der Arbeit abschließend zusammen.

List of Figures

List of Tables

List of Algorithms

List of Abbreviations

AK	Estimator with parameters A and K, variants: AK1, AK2, AK3
ARBias	Absolute Relative Bias
(X)ASU	Jobseeker Histories in SIAB
AVar	Approximated Variance
B	Brier Score
Bagging	Bootstrap Aggregating
BeH	Employment Histories in SIAB
BFH	Bivariate Fay-Herriot
BHF	Battese-Harter-Fuller
BLP	Best Linear Predictor
BLUE	Best Linear Unbiased Estimator
BLUP	Best Linear Unbiased Predictor
BP	Best Predictor
BSS	Brier Skill Score
Bw	Weighted Brier Score
Change_t1	Change to previous time point
Change_ty	Change to previous year
Corr	Correlation
Cov	Covariance
CPS	Current Population Survey
EBLUE	Empirical Best Linear Unbiased Estimator
EBLUP	Empirical Best Linear Unbiased Predictor
EBP	Empirical Best Predictor
EM	Expectation-Maximization
EU	European Union
EUROSTAT	European Statistical Office
FH	Fay-Herriot
FS	Fisher-Scoring
GAM	Generalized Additive Model
GH	Gauss-Hermite
GLM	Generalized Linear Model
GLS	Generalized Least Squares
GREG	Generalized Regression
HT	Horvitz-Thompson
IAB	Institute for Employment Research
ICT	Survey on Information and Communication Technologies
IEB	Integrated Employment Biographies
ILO	International Labour Organization
INE	Spanish Statistical Office
LeH	Benefit Recipient Histories in SIAB
LF	Labour Force

LFS	Labour Force Survey
LHG	Unemployment Benefit II Recipient Histories in SIAB
IRLS	Iteratively Re-Weighted Least Squares
LMM	Linear Mixed Model
MC	Monte Carlo
MCA	Monte Carlo with Antithetic Variates
MFH	Multivariate Fay-Herriot
MINQUE	Minimum Norm Quadratic Unbiased Estimation
ML	Maximum Likelihood
MMFH	Multivariate Fay-Herriot with Partially Missing Direct Estimates
MR	Modified Regression, variants: MR1, MR2, MR3, MRR
MSE	Mean Squared Error
(X)MTH	Participants-in-Measures History Files in SIAB
NA	Not available
NR	Newton-Raphson
NUTS	Classification of Territorial Units for Statistics
Q	Quarter
Q-1	Overlap of a quarter to the previous quarter
Q-4	Overlap of a quarter to the same quarter a year before
QMC	Quasi Monte Carlo
QMCH	Quasi Monte Carlo with Halton sequence
QMCS	Quasi Monte Carlo with Sobol sequence
QP	Quadratic Programming
`R`	Programming software `R` (R Core Team, 2020)
RBias	Relative Bias
RC	Regression Composite
RDiff	Relative Difference
REML	Restricted Maximum Likelihood
RGB	Rotation Group Bias
RIFOSS	Research Innovation for Official and Survey Statistics
RMSE	Relative Mean Squared Error
RRMSE	Relative Root Mean Squared Error
SAE	Small Area Estimation
SAIPE	Small Area Income and Poverty Estimates
SB	Stratified Brier Score
SGB	Social Code Book
SIAB	Sample of Integrated Employment Biographies
SILC	Statistics on Income and Living Conditions
SLFS	Spanish Labour Force Survey
Stacking	Stacked Generalization
Subagging	Subsample Aggregating
Syn	Synthetic
U.S.	United States
Var	Variance

Notation

Unless stated differently: A scalar is denoted as x, a vector as $\boldsymbol{x}$, a matrix as $\boldsymbol{X}$. Vectors are defined as column vectors.

As is common in the survey and small area literature, e.g. Searle et al. (2006, p. 139) and Särndal et al. (1992, p. 226), we often times do not formally distinguish between random variables and their realisations. For instance, we often do not formally distinguish an estimator from its estimates or a model from its predictions.

General symbols

$\boldsymbol{M}^{-1}$	Inverse of matrix $\boldsymbol{M}$
$\boldsymbol{M}^{\top}$	Transpose of matrix $\boldsymbol{M}$
$\det(\boldsymbol{M})$	Determinant of matrix $\boldsymbol{M}$
$\mathrm{tr}(\boldsymbol{M})$	Trace of matrix $\boldsymbol{M}$
$\boldsymbol{I}_x$	$x \times x$ identity matrix
col	Matrix operator stacking by columns
diag	Matrix operator stacking the elements row-wise to a block-diagonal matrix
row	Matrix operator stacking by rows
$\sim N(x, y)$	Normal distribution with mean x and variance y
$\sim N_m(\boldsymbol{x}, \boldsymbol{Y})$	m-variate normal distribution with mean vector $\boldsymbol{x} \in \mathbb{R}^m$ and covariance matrix $\boldsymbol{Y} \in \mathbb{R}^{m \times m}$
$\sim Unif(x, y)$	Uniform distribution in interval $[x, y]$
$\mathrm{E}[X]$	Expectation of X
$\mathrm{AVar}(X)$	Approximated variance of X
$\mathrm{Bias}(X)$	Bias of X
$\mathrm{Cov}(X, Y)$	Covariance of X and Y
$\mathrm{Corr}(X, Y)$	Correlation of X and Y
$\mathrm{MSE}(X)$	Mean squared error of X
$\mathrm{Var}(X)$	Variance of X
I	Indicator function taking values in $\{0, 1\}$
ℓ	Log-likelihood
$\mathcal{O}$, $\mathcal{O}_P$, o, o_P	Landau-Bachmann notation, similar to Jiang (2007, Appendix C.3), Morales et al. (2021, p. 523)
$\Pr(x)$	Probability of an event x
$\lvert x \rvert$	Absolute value of x
$\hat{x}$, $\widehat{x}$	Estimator/estimate/predictor/prediction of x
$\#\big(\mathcal{X}\big)$	Cardinality of a set $\mathcal{X}$

Chapter-specific notation The chapters of the thesis deal with different theories and concepts, each of which has its own typical notation. To maintain the recognition value of the basic literature and the readability of the individual chapters, certain quantities are redefined and replaced in individual chapters. The following lists contain selected symbols that are frequently used in the individual chapters. Symbols that are defined and used only in individual sections are not included.

Survey Sampling (Section 2.2, 2.3, Chapter 4)

A	Parameter of AK estimator
D	Number of domains
d_k (d_{kt})	Design weight of unit $k \in U$ (at time t)
g_k	Correction weight of unit $k \in s$
K	Parameter of (A)K estimator
N	Population size
n	Sample size
p	Number of auxiliary variables
$\mathfrak{p}$	Sampling design
q	Number of additional auxiliary variables in composite estimators
S	Set-valued random variable
s (s_t)	Sample (at time t)
$s_t \cap s_{t'}$	Overlapping sample in t, t'
$s_t \setminus s_{t'}$	Non-overlapping sample in t, t'
$\mathscr{S}$	Set of all possible realisations of S
t, t', t''	Time points, $t'' < t' < t$
U	Finite population of size N
w_k (w_{kt})	Calibration weight of unit $k \in s$ $(k \in s_t)$
$\boldsymbol{x}_k$ $(\boldsymbol{x}_{kt})$	Vector of auxiliary values of unit $k \in U$ (at time t) of length p
x_{kj}	Value of the j-th auxiliary variable of unit $k \in U$
Y	Variable of interest
y_k (y_{kt})	Value of Y of unit $k \in U$ (at time t)
$\boldsymbol{z}_{kt}$	Vector of additional auxiliary values of unit $k \in s_t$ of length q
z_{kt}	Value of additional auxiliary variable of unit $k \in s_t$
α	Parameter of RC estimator
$\boldsymbol{\beta}$	Vector of fixed effects
θ	Quantity of interest
$\theta_{tt'}$	Sample overlap between s_t and $s_{t'}$
π_k (π_{kt})	First-order inclusion probability of unit $k \in U$ (at time t)
π_{kl} (π_{klt})	Second-order inclusion probability of units $k, l \in U$ (at time t)
$\boldsymbol{\tau}_{\boldsymbol{x}}$ $(\boldsymbol{\tau}_{\boldsymbol{x}_t})$	Population totals of p auxiliary variables (at time t)
τ_{xj}	Population total of the j-th auxiliary variable
τ_y (τ_{y_t})	Population total of Y (at time t)
$\boldsymbol{\tau}_{\boldsymbol{z}_t}$	Population totals of q additional auxiliary variables at time t
τ_{z_t}	Population total of an additional auxiliary variable at time t

Small area estimation (Section 2.4, Chapters 5, 6)

$\mathcal{A}_d$	Additional set for domain d in MMFH model
$b_{\hat{\boldsymbol{\delta}}}$	Bias in estimation of variance parameters $\boldsymbol{\delta}$

D	Number of domains
$\mathcal{D}$	Domain set
$\boldsymbol{e}$ ($\boldsymbol{e}_d$, e_d)	Residual vector (in domain d), in FH models: Sampling errors
$\boldsymbol{F}$	Fisher information matrix
f	Univariate or m-variate normal density function
$\boldsymbol{G}$ ($\boldsymbol{G}_d$)	Covariance matrix of $\boldsymbol{u}$ ($\boldsymbol{u}_d$) in LMMs
$\boldsymbol{G}_1, \boldsymbol{G}_2, \boldsymbol{G}_3$ $(\boldsymbol{G}_{1d}, \boldsymbol{G}_{2d}, \boldsymbol{G}_{3d})$	Multivariate MSE components (in domain d)
G_d	Parameter of interest in domain d
g	Arbitrary function
g_1, g_2, g_3 (g_{1d}, g_{2d}, g_{3d})	MSE components (in domain d)
$\boldsymbol{H}$	Hessian matrix
h	Length of $\boldsymbol{u}$ in LMMs
$\boldsymbol{l}, \boldsymbol{m}$ ($\boldsymbol{l}_d, \boldsymbol{m}_d$)	Vectors for defining μ (μ_d) in LMMs
$\breve{M} = \sum_{d=1}^{D} \breve{m}_d$	Number of observed direct estimates
m ($\breve{m}_d$)	Number of (dependent) variables (observed in domain d)
n	Observation length
p (p_k)	Number of fixed effects $\boldsymbol{\beta}$ ($\boldsymbol{\beta}_k$)
$\mathcal{Q}_d$	Additional set for domain d in MMFH model
q	Number of variance parameters $\boldsymbol{\delta}$ in LMMs or $\boldsymbol{\theta}$ in FH models
$\boldsymbol{R}$ ($\boldsymbol{R}_d$)	Covariance matrix of $\boldsymbol{e}$ ($\boldsymbol{e}_d$) in LMMs
R_d	Unemployment rate in domain d
S_{1d}, S_{2d}	MSE components in domain d
T	Number of function evaluations
t (t_d)	Functional form of predictor of μ (μ_d)
U (U_d)	Finite population (in domain d)
$\boldsymbol{u}$ ($\boldsymbol{u}_d$, u_d)	Random effects (in domain d)
$\boldsymbol{V}$ ($\boldsymbol{V}_d$)	Covariance matrix of $\boldsymbol{y}$ ($\boldsymbol{y}_d$)
$\boldsymbol{V}_e$ ($\boldsymbol{V}_{ed}$)	Covariance matrix of $\boldsymbol{e}$ ($\boldsymbol{e}_d$)
$\boldsymbol{V}_u$ ($\boldsymbol{V}_{ud}$)	Covariance matrix of $\boldsymbol{u}$ ($\boldsymbol{u}_d$)
w_t	Weights for function evaluation t
$\boldsymbol{X}$ ($\boldsymbol{X}_d$)	Auxiliary matrix (in domain d)
$\boldsymbol{x}_d$ ($\boldsymbol{x}_{dk}$)	Auxiliary vector in domain d (for variable k)
$\boldsymbol{x}_t$	Vector of nodes for function evaluation t, of length m
$\boldsymbol{y}$ ($\boldsymbol{y}_d$, y_d)	Observations of variable(s) of interest (in domain d), in FH models: Direct estimates
$\boldsymbol{Z}$ ($\boldsymbol{Z}_d$)	Matrix for random effects structure (in domain d) in LMMs
$\boldsymbol{\beta}$ ($\boldsymbol{\beta}_k$)	Vector of fixed effects (for variable k)
γ_d	Shrinkage factor in domain d in FH model
$\boldsymbol{\delta}$	Vector of variance parameters in LMMs
$\boldsymbol{\theta}$	Vector of variance parameters
$\boldsymbol{\Lambda}_d$	Diagonal matrix of $\boldsymbol{\lambda}_d$ where all rows with row sum equal to zero are deleted
$\boldsymbol{\lambda}_d$ (λ_{dk})	Additional vector for domain d in MMFH model (for variable k)

$\boldsymbol{\mu}$ ($\boldsymbol{\mu}_d$)	Vector of parameters of interest (in domain d)
μ_{dk}	k-th parameter of interest in domain d
μ (μ_d)	Parameter of interest (in domain d)
$\boldsymbol{\psi}$	Vector of unknown model parameters
ρ_{edkl}	Correlation of sampling errors of variables k and l in domain d
ρ_{kl}	Correlation of random effects of variables k and l
σ^2_{ed} (σ^2_{edk})	Sampling error variance (of variable k) in domain d
σ^2_u (σ^2_{uk})	Random effect variance (of variable k)

In the MMFH model, we define two additional versions for certain quantities to account for missing values. For a quantity $\boldsymbol{r}_d$, e.g. a vector, $\check{\boldsymbol{r}}_d$ corresponds to $\boldsymbol{r}_d$ reduced to the observed variables of interest in domain d and $\acute{\boldsymbol{r}}_d$ corresponds to $\boldsymbol{r}_d$, where all elements referring to missing direct estimates are set to zero.

Modelling employment histories (Chapter 3)

B	Brier score
BSS_j	Brier skill score for category j
Bw	Weighted Brier score
$\mathcal{D}$	Arbitrary dataset consisting of n observation tuples
$\boldsymbol{d} = (d_1, \ldots, d_J)^\top$	Category weight vector for Bw
f	Arbitrary function, a model
g	Monotonic link function for the exponential family of distributions
J	Number of categories
L	Loss function
M	Number of individual models
n (n_j)	Number of observation (of category j)
p	Number of auxiliary variables
$\hat{p}(Y_i = j \vert \boldsymbol{x}_i)$	Predicted probability that Y_i takes category j given $\boldsymbol{x}_i$
SB	Stratified Brier score
w_m	Ensemble weight for model $\hat{f}^*_m$
$\boldsymbol{x}_i$	Vector of auxiliary variables for i of length p
$\boldsymbol{X}$	Matrix of auxiliary information
Y_i	Random variable with expectation μ_i
$\boldsymbol{\beta}$	Vector of fixed effects
η (η_j)	Linear or additive predictor (for category j)
μ_i	Expectation of Y_i
π_j (π_{ij})	Probability of category j (for unit i)

Chapter 1

Introduction

Survey data can be viewed as incomplete or partially missing from a variety of perspectives and there are different ways of dealing with this kind of data in the prediction and the estimation of economic quantities. In this thesis, we present two selected research contexts in which the prediction or estimation of economic quantities is examined under incomplete survey data.

These contexts are first the investigation of composite estimators in the German Microcensus (Chapters 3 and 4) and second extensions of multivariate Fay-Herriot (MFH) models (Chapters 5 and 6), which are applied to small area problems.

Composite estimators are estimation methods that take into account the sample overlap in rotating panel surveys such as the German Microcensus in order to stabilise the estimation of the statistics of interest (e.g. employment statistics). Due to the partial sample overlaps, information from previous samples is only available for some of the respondents, so the data are partially missing.

MFH models are model-based estimation methods that work with aggregated survey data in order to obtain more precise estimation results for small area problems compared to classical estimation methods. In these models, several variables of interest are modelled simultaneously. The survey estimates of these variables, which are used as input in the MFH models, are often partially missing. If the domains of interest are not explicitly accounted for in a sampling design, the sizes of the samples allocated to them can, by chance, be small. As a result, it can happen that either no estimates can be calculated at all or that the estimated values are not published by statistical offices because their variances are too large. In the following, we give a more detailed description of the chapters of this thesis.

Chapter 2: Fundamentals of Survey Estimation
Chapter 2 gives a brief overview of the theoretical and methodological concepts needed for the developments in the later chapters. These include foundations of design-based and model-based survey estimation, Monte Carlo simulation methods, and Fay-Herriot models, which belong to the class of model-based area-level small area estimation methods.

Chapter 3: Generation of a Longitudinal Employment Dataset for Simulations
Chapter 3 deals with the extension of a cross-sectional dataset with longitudinal employment information, which then serves as the simulation population for the studies in Chapter 4. For the research questions in Chapter 4, a longitudinal dataset is needed that represents the German resident population including monthly employment information such that the design of the German Microcensus and the production of employment statistics can be replicated in its full regional depth and temporal dimension. For that purpose, we use

the RIFOSS and SIAB dataset: The RIFOSS dataset[1] is a semi-synthetic cross-sectional dataset of the German resident population. The SIAB dataset[2] is a random sample drawn from the Integrated Employment Biographies of the Institute for Employment Research (Antoni et al., 2019).

In Chapter 3, we describe the editing, aggregation, and validation of the SIAB dataset. Based on this data, we calculate prediction models for monthly employment categories. With the prediction models, we generate longitudinal employment information in the RIFOSS dataset. We also discuss the evaluation of probability predictions for imbalanced categorical data such as the employment status. In the course of the discussion, we propose an extension of the Brier score, the so-called weighted Brier score, to account for the imbalanced categories. For the prediction models, we use several generalised additive models, each estimated on subsamples of the processed SIAB data (subagging). The models are combined into optimally weighted ensemble models (stacking). In the optimisation of the ensemble weights, the proposed weighted Brier score is used as a loss function, which constitutes a quadratic optimisation problem. With the final ensemble models, we generate longitudinal employment information in the RIFOSS dataset. We validate the generated data with information from the SIAB dataset and aggregate statistics published by the German statistical institute (Destatis).

Chapter 4: Composite Estimation in the German Microcensus
The Microcensus is the largest annual household survey of official statistics in Germany. Every year around 1% of all German households participate in the survey and answer questions related to their working and living conditions. The sampling design of the survey underwent major changes beginning in 2020, including a modified rotation scheme and the integration of different household surveys, which were previously conducted separately.

The rotation design of the Microcensus results in sample overlaps of different time points. The sample overlaps constitute incomplete survey data: For each time point, information from previous samples is available for only a subset of the respondents in a sample. The rotation pattern of the sampling design determines the magnitude of the overlaps. Composite estimators use the partial overlaps with previous samples in the estimation process. Particularly for employment statistics, the inclusion of this additional information can lead to more efficient estimators since the employment status is typically rather stable over time.

The new rotation design of the Microcensus creates new opportunities for the application of composite estimators. After an overview of different composite estimators and their applications, we describe the Microcensus design and its changes in 2020, and analyse how composite estimators can be applied in this survey. Thereafter, we present adjustments of the composite estimators to account for the regionally heterogeneous sample overlaps resulting from the Microcensus design. In a design-based simulation study based on the dataset created in Chapter 3 we compare the performance of the adjusted composite

[1] Version RIFOSS_GG_v0.1.1_vanilla_ice_cream.

[2] Factually anonymous Version of the Sample of Integrated Labour Market Biographies (SIAB-Regionalfile) Version 7517 v1.

estimators. We evaluate the estimators for various employment statistics like monthly and quarterly totals and changes of the number of employed and unemployed at different regional levels. Furthermore, we evaluate them for different sets of sample overlap information.

Chapter 5: Empirical Best Prediction in Multivariate Fay-Herriot Models
Keeping everything else fixed, the variance of design-based methods continues to increase as domain sizes decrease due to small sample sizes. That is, for small domains, design-based estimation methods typically have high variances. The challenge of producing estimates based on small sample sizes is referred to as a small area estimation problem. Different model-based small area methods have been developed and studied for this problem. MFH models are area-level small area methods which take aggregate survey information as input and model several dependent variables at the same time. They have been an active field of research in recent years. In Chapters 5 and 6, we present two additional methodological developments for these models. Taken together, the combination of the contributions presented in these two chapters allows for the approximation of the empirical best predictors (BPs) of multi-variable, potentially non-linear domain indicators under the MFH models, even in domains for which some survey estimates are missing.

Chapter 5 focuses on multi-variable domain indicators. An example of such an indicator is the domain unemployment rate, defined as the number of unemployed divided by the sum of employed and unemployed. For such indicators, so-called plug-in estimators are frequently used. They are calculated by substituting the indicator input values by their survey estimates. However, plug-in estimators do not consider the joint distribution of their inputs. Furthermore, plug-in estimators are not unbiased for non-linear indicators and asymptotic unbiasedness cannot be assumed when dealing with small sample sizes.

As an alternative to plug-in predictors in small area problems, we study BPs of multi-variable domain indicators in MFH models in Chapter 5. By definition, the BPs are the estimators with minimum mean squared error (MSE) in the class of all model-unbiased predictors and therefore advantageous to the model-based plug-in predictors. As the BPs do not have a closed form, we analyse different techniques for approximating their integrals. Furthermore, we present parametric bootstrap procedures for estimating the MSE of the approximations. In several model-based simulation studies, each replicating the estimation of domain unemployment rates, we evaluate the approximations and compare their performance with that of the corresponding plug-in predictors. In addition, we evaluate the proposed MSE estimators, including recommendations for the number of repetitions in the parametric bootstrap procedures. The chapter concludes with an illustrative application of the approach to publicly available Spanish labour force data aimed at estimating the unemployment rates of Spanish provinces crossed by age classes and sex.

Chapter 6: Multivariate Fay-Herriot Models under Missing Direct Estimates
MFH models take domain survey estimates of the dependent variables as input. However, they can only use information from domains with fully available survey estimates, both for parameter estimation and for the calculation of BPs under the model. This limits

the applicability of MFH models for practical purposes, where the multi-variable survey estimates are often partially missing.

In Chapter 6, we present the MFH model under missings (MMFH) as a generalisation of the MFH model. The MMFH model is capable of considering all those domains for which at least one of the survey estimates is available, both for estimating the model parameters and calculating BPs under the model. We present Fisher-Scoring algorithms for model parameter estimation, both using the maximum likelihood and restricted maximum likelihood approach. Furthermore, we present formulas for the MSE estimation, both for the BPs under the model and for synthetic MFH predictors. The proposed MMFH algorithms and formulas are evaluated in model-based simulation studies. We compare their performance to that of different competing Fay-Herriot models. The chapter closes with an illustrative application of the MMFH model to publicly available data from the American Community Survey. The target is the estimation of the median annual income of the population with Hispanic or Latino origin in 2010 and 2011 for U.S. counties.

Chapter 7: Summary and Conclusions
We summarise the contributions and findings of the thesis in Chapter 7, including an outlook for potential future research.

Chapter 2

Fundamentals of Survey Estimation

2.1 Introduction

In this chapter we briefly introduce some basic concepts which are needed for the developments and investigations in the subsequent chapters. It is structured as follows. Section 2.2 presents basic concepts of survey estimation including the distinction of design-based and model-based approaches and descriptions of design-based estimators and their properties. The Horvitz-Thompson estimator and the generalised regression estimator are described in Section 2.3. The design-based concepts and estimators are in the focus of Chapters 3 and 4. The following Section 2.4 gives an introduction to the theory of linear mixed models, specifically tailored to model-based small area estimation techniques with the Fay-Herriot model in particular. Multivariate versions of the Fay-Herriot model and their extensions are the research subjects of Chapters 5 and 6.

2.2 Basic concepts of survey estimation

Surveys are used to obtain information on a target population by means of population *samples*. A *sampling design* determines the procedure by which a sample of population units is chosen from that target population. The elements of a sample are the *sampling units*. For example, in the German Microcensus the target population is the resident population of Germany, the sampling units are clusters of persons in households, for example street sections, and the samples are drawn via a one-stage cluster sampling design (Destatis, 2021). In order to make inferences about the target population based on a concrete sample, *estimation methods* are applied. For a comprehensive overview of the general theory of survey sampling and estimation, we refer to Cochran (1977), Lohr (2010), and Särndal et al. (1992).

2.2.1 Design-based, model-assisted, and model-based approach

There are essentially two approaches according to which survey information can be analysed: The design-based (and model-assisted) and the model-based approach. In Chapters 3 and 4, the focus is on the design-based (and model-assisted) approach, while the model-based approach is applied in Chapters 5 and 6. A general overview of both approaches is given in Skinner and Wakefield (2017), on which the following description is based. We refer to Lehtonen and Veijanen (2009), Morales et al. (2021), and Rao and Molina (2015) for a discussion of the approaches with a focus on small area estimation.

The *design-based* approach focuses on fixed and finite populations. The *parameters/statistics of interest*, e.g. the unemployment rate in certain domains, are interpreted as fixed and typically unknown quantities which can be estimated using population samples. The only randomness considered in this approach is the process of drawing samples according to a sampling design. Each sample is seen as a random realisation of the sampling process, the distribution of all possible samples is determined by the sampling design. Design-based estimators are designed and evaluated with respect to the randomisation process of the sampling design. For example, the properties design-unbiasedness and design-consistency, covered in Section 2.2.3, are defined with respect to the distribution of all possible samples under a specific design. Classical examples of design-based estimators are the Horvitz-Thompson estimator, covered in Section 2.3.1, and the Hájek estimator, which is e.g. described in Särndal et al. (1992).

Model-assisted estimators include a statistical model which links the target information in a sample to additional auxiliary information. In addition to the sampling information, there is usually further information available that can be used for the estimation process. For example, population statistics on the number of persons for different demographic and regional levels could be available from the last Census and incorporated into the estimation process. Model-assisted estimators are only assisted by a model, not model-dependent, as certain design-based features hold, at least asymptotically, regardless of the (implicit) choice of the model. The estimators are often more efficient than estimators which do not incorporate additional information. Furthermore, they can be used to ensure consistent estimates, for example of different surveys. A prominent example of a model-assisted estimator is the generalised regression estimator, covered in Section 2.3.2. It is for example used in the German Microcensus with key figures from the current population update as auxiliary information (Destatis, 2020b). Often, the term design-based is used to refer to both kinds of methods, design-based and model-assisted, which is what we will do in the following.

In the *model-based* approach, a population is interpreted as a realisation from a *super-population* model. Thereby, the target quantities are seen as random variables. Similar to model-assisted estimators, model-based estimators also include models linking the target survey information to additional data. Model-based estimators are *model-dependent*, i.e. they are derived under concrete model assumptions and their features are evaluated with respect to these assumptions. In practical applications, the model assumptions of a model-based estimator have to be carefully evaluated. When the model assumptions are not met by the data at hand, model-based estimators can exhibit huge biases. However, there are applications where model-based estimators have significant advantages over design-based estimators.

One particular application area of model-based estimators is *small area estimation*, covered in Section 2.4.1 It targets the analysis of survey data when sample sizes are small and the variances of design-based estimators are high. In these small area problems, model-based estimators can have substantially smaller mean squared errors than design-based methods. General information on small area estimation techniques, the theory of linear mixed models, and Fay-Herriot models as particular model-based small area models are covered in Section 2.4. They build the theoretical basis for the developments based on multivariate Fay-Herriot models, which are presented in Chapters 5 and 6.

2.2.2 Sampling design and inclusion probabilities

In the following, the focus is on the design-based approach. The description is based on Särndal et al. (1992). Let there be a fixed and finite population $U = \{1, \dots, N\}$, $N \in \mathbb{N}$. A *sampling design*, denoted by $\mathfrak{p}$, returns the probability of selecting a specific sample $s \subseteq U$ from that population, denoted by $\mathfrak{p}(s)$. A sample s is a set of population units of size n. It can be regarded as a realisation of a set-valued random variable S. In this thesis, only sampling designs where each population unit appears at most once in a sample are considered, called *sampling without replacement*, which implies $n \leq N$. Sampling design $\mathfrak{p}$ determines the distribution of S, $\Pr(S = s) = \mathfrak{p}(s)$. The set of all possible realisations of S is denoted as $\mathscr{S}$. We have $\mathfrak{p}(s) \geq 0$, $\forall s \in \mathscr{S}$, and $\sum_{s \in \mathscr{S}} \mathfrak{p}(s) = 1$.

Under a specific sampling design, each unit of the population can be assigned the probability of being included in a sample drawn according to that design. The *first-order inclusion probability* of a population unit k is given by (Särndal et al., 1992, Equation 2.4.2)

$$\pi_k = \Pr(k \in S) = \sum_{s \ni k} \mathfrak{p}(s), \quad \forall k \in U, \tag{2.1}$$

where $\sum_{s \ni k}$ denotes the sum of all possible samples in which unit k is included. The *second-order inclusion probability* of population units k and l is given by

$$\pi_{kl} = \Pr(k \& l \in S) = \sum_{s \ni k \& l} \mathfrak{p}(s), \quad \forall k, l \in U, \tag{2.2}$$

where $\sum_{s \ni k \& l}$ denotes the sum of all possible samples in which both units k and l are included. We have $\pi_{kk} = \pi_k$, $\forall k \in U$. In this thesis, only *probability sampling designs* are considered, implying that $\pi_k > 0$ holds, $\forall k \in U$ (Särndal et al., 1992, p. 32). That is, all units of the target population U have a known, fixed, positive probability of being included in a sample. Furthermore, we always consider $\pi_{kl} > 0$, $\forall k, l \in U$, i.e. the design is *measurable* (Särndal et al., 1992, p. 33).

The *design weights* of the population units are defined as the inverse of the first-order inclusion probabilities, given by (Lehtonen & Veijanen, 2009, p. 222)

$$d_k = 1/\pi_k, \quad \forall k \in U. \tag{2.3}$$

In praxis, the design weights of the sampling units are typically additionally adjusted, e.g. to account for *non-response*. As an example, we refer to Afentakis and Bihler (2005, Section 2.2) for the description of the non-response adjustment in the German Microcensus. In Section 2.3.2, calibration estimators are presented, with the generalised regression estimator in particular, which can be used for non-response adjustment.

2.2.3 Basic properties of estimators

Survey samples are used to estimate various unknown quantities of the target population, like totals, means, or ratios of different variables. Consider a variable of interest Y, where $(y_1, \dots, y_N)^\top$ are the values of that variable for the units of population U. We are interested in a real-valued function of the variable of interest, given by $\theta = f((y_1, \dots, y_N)^\top)$, which is called a *statistic/quantity/parameter*. For example, the variable of interest can be an dummy variable with value 1 if a person in population U is unemployed and 0 otherwise. A corresponding statistic of interest could be the total or the proportion of unemployed persons.

A function which maps the data of sample s, and potentially additional data, to a concrete estimate of θ is called *estimator* and denoted as $\hat{\theta}$. In the design-based context, the properties of an estimator $\hat{\theta}$ are evaluated with respect to its distribution under a sampling design $\mathfrak{p}$. As $\hat{\theta} = \hat{\theta}(S)$ is a statistic over the random set S, $\hat{\theta}(S)$ itself is a random variable. The following description of the properties of estimators is based on Särndal et al. (1992, Section 2.7).

The *design-expectation* and *design-variance* of an estimator $\hat{\theta}$ are given by (Särndal et al., 1992, Definition 2.5.1, Section 2.7)

$$\mathrm{E}[\hat{\theta}] = \sum_{s \in \mathscr{S}} \mathfrak{p}(s)\hat{\theta}(s), \tag{2.4}$$

$$\mathrm{Var}(\hat{\theta}) = \mathrm{E}\left[(\hat{\theta} - \mathrm{E}[\hat{\theta}])^2\right] = \sum_{s \in \mathscr{S}} \mathfrak{p}(s)\Big(\hat{\theta}(s) - \mathrm{E}[\hat{\theta}]\Big)^2. \tag{2.5}$$

In accordance to the design-variance of one estimator, the *design-covariance* of two estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ is given by

$$\mathrm{Cov}(\hat{\theta}_1, \hat{\theta}_2) = \mathrm{E}\left[(\hat{\theta}_1 - \mathrm{E}[\hat{\theta}_1])(\hat{\theta}_2 - \mathrm{E}[\hat{\theta}_2])\right] = \sum_{s \in \mathscr{S}} \mathfrak{p}(s)\Big((\hat{\theta}_1 - \mathrm{E}[\hat{\theta}_1])(\hat{\theta}_2 - \mathrm{E}[\hat{\theta}_2])\Big). \tag{2.6}$$

The *design-bias* of an estimator $\hat{\theta}$ is given by

$$\mathrm{Bias}(\hat{\theta}) = \mathrm{E}[\hat{\theta}] - \theta. \tag{2.7}$$

Estimator $\hat{\theta}$ is said to be *design-unbiased* if $\mathrm{Bias}(\hat{\theta}) = 0$. The design-based *mean squared error* (MSE) is used as a measure of precision and defined as

$$\begin{aligned}
\mathrm{MSE}(\hat{\theta}) &= \mathrm{E}\left[(\hat{\theta} - \theta)^2\right] \\
&= \mathrm{E}\left[(\hat{\theta} - \mathrm{E}(\hat{\theta}))^2\right] + 2\Big(\mathrm{E}[\hat{\theta}] - \theta\Big)\,\mathrm{E}\left[\hat{\theta} - \mathrm{E}[\hat{\theta}]\right] + \Big(\mathrm{E}[\hat{\theta}] - \theta\Big)^2 \\
&= \mathrm{Var}(\hat{\theta}) + (\mathrm{Bias}(\hat{\theta}))^2.
\end{aligned} \tag{2.8}$$

As $\mathrm{E}\left[\hat{\theta} - \mathrm{E}[\hat{\theta}]\right] = \mathrm{E}[\hat{\theta}] - \mathrm{E}\left[\mathrm{E}[\hat{\theta}]\right] = \mathrm{E}[\hat{\theta}] - \mathrm{E}[\hat{\theta}] = 0$, by the law of iterated expectation, the cross-product term vanishes. From (2.8), we see that for unbiased estimators the MSE equals the variance. The MSE takes into account both variance and bias of an estimator and is especially useful when comparing estimators which are not unbiased. It is useful to consider the bias in addition to the MSE when evaluating estimators as the bias is for example essential for the construction and validity of confidence intervals (Särndal et al., 1992, p. 164).

Two additional features of estimators are asymptotic unbiasedness and consistency, the description of which is based on Särndal et al. (1992, Section 5.3). Let there be n independently and identically distributed random variables $\boldsymbol{\iota}_{(n)} = (\iota_1, \ldots, \iota_n)^\top$. An estimator $\hat{\theta}_{(n)} = \hat{\theta}(\boldsymbol{\iota}_{(n)})$ of quantity θ is *asymptotically unbiased* if

$$\lim_{n\to\infty} \mathrm{E}[\hat{\theta}_{(n)}] = \theta \tag{2.9}$$

and *consistent* if

$$\lim_{n\to\infty} \Pr\Big(|\hat{\theta}_{(n)} - \theta| > \varepsilon\Big) = 0, \tag{2.10}$$

for any fixed $\varepsilon > 0$. For large n, an asymptotically unbiased estimator can be considered approximately unbiased and a consistent estimator can be considered to be concentrated around the true parameter θ. To apply both concepts to finite populations, one has to consider the limes for both n and N increasing to infinity. We refer to Särndal et al. (1992, Section 5.3) for further details.

Just as the design-based properties of estimators are defined over the randomisation process of the sampling design, the model-based properties of estimators are defined over the randomisation process of the underlying super-population model. In this thesis, it should be clear from the context whether the focus is on the design- or model-based properties of estimators, which is why the prefixes *design-* and *model-* are often omitted.

2.2.4 Monte Carlo simulation studies

In concrete applications with complex survey designs, it can be difficult to analytically derive the theoretical properties of estimators, like their MSE for specific target quantities. Both, design- and model-based properties of estimators, under different data, sampling designs, and model scenarios, can be approximated using *Monte Carlo simulation studies*.

In a Monte Carlo simulation, the randomisation process of a sampling design (design-based simulation study) or the randomisation process of the super-population model (model-based simulation study) is simulated repeatability using random numbers. The properties of the estimators can be inferred over the simulated randomisation process by analysing their Monte Carlo distribution. For a valid analysis, it should be ensured that the considered randomisation process has been simulated often enough to approximate the corresponding quantities sufficiently well, e.g. by comparing the results for different numbers of repetitions. We refer to Gentle (2003) for a general overview of random number generation and Monte Carlo methods and Burgard et al. (2020a) for an overview of Monte Carlo methods tailored to survey statistics.

In the simulation studies presented in this thesis, we mainly focus on the approximation of different variants of the bias or MSE of an estimator. In a Monte Carlo simulation study, the Bias and MSE of an estimator $\hat{\theta}$ can be approximated by

$$\text{Bias}(\hat{\theta}) = \frac{1}{MC} \sum_{mc=1}^{MC} \hat{\theta}^{(mc)} - \theta^{(mc)}, \quad \text{MSE}(\hat{\theta}) = \frac{1}{MC} \sum_{mc=1}^{MC} (\hat{\theta}^{(mc)} - \theta^{(mc)})^2, \tag{2.11}$$

where MC is the number of simulation repetitions. $\theta^{(mc)}$ and $\hat{\theta}^{(mc)}$ refer to the values of the quantity of interest and its estimate in iteration mc.

In a design-based study, such as the study presented in Chapter 4, the true values are considered fix, i.e. $\theta^{(mc)} = \theta$, $\forall mc \in \{1, \ldots, MC\}$. In each Monte Carlo iteration, a random sample is drawn from the population according to a specific sampling design and different estimators are applied to the sample to estimate θ. In a model-based simulation study, such as the studies presented in Chapters 5 and 6, the true values are considered random. In each Monte Carlo iteration, a population is drawn as a realisation of an underlying super-population model and $\theta^{(mc)}$ is a random variable.

Throughout this thesis, we use `R` package `baseR` with functions `runif` and `rnorm` to simulate random draws from the uniform and normal distribution. To simulate multivariate normal draws from independent univariate draws, we use Algorithm 5.1, which we will look at in more detail in Section 5.2.

2.3 Horvitz-Thompson and generalised regression estimator

2.3.1 Horvitz-Thompson estimator

Two of the most well-known survey estimators are the Horvitz-Thompson estimator and the generalised regression estimator, which are described in the following. We take the population total of a variable of interest Y as the parameter of interest, denoted by $\tau_y = \sum_{k \in U} y_k$.

The *Horvitz-Thompson* (HT) estimator, named after Horvitz and Thompson (1952), is a design-unbiased and design-consistent (Isaki & Fuller, 1982) estimator of τ_y, which, based on a sample s, is given by

$$\hat{\tau}_y^{\text{HT}} = \sum_{k \in s} \frac{y_k}{\pi_k} = \sum_{k \in s} y_k d_k. \tag{2.12}$$

The following description of the HT estimator is based on Särndal et al. (1992, Section 2.8). The HT estimator is often called *π-estimator* as it produces estimates as a π-weighted sums of the sample values, i.e. weighted by the inclusion probabilities. The variance of the estimator is given by

$$\text{Var}(\hat{\tau}_y^{\text{HT}}) = \sum_{k\in U}\sum_{l\in U}(\pi_{kl} - \pi_k\pi_l)\frac{y_k}{\pi_k}\frac{y_l}{\pi_l}. \tag{2.13}$$

Given that for the joint inclusion probability $\pi_{kl} > 0$, $\forall k, l \in U$, holds, an unbiased estimator of the variance is given by

$$\widehat{\text{Var}}(\hat{\tau}_y^{\text{HT}}) = \sum_{k\in s}\sum_{l\in s}(1 - \frac{\pi_k\pi_l}{\pi_{kl}})\frac{y_k}{\pi_k}\frac{y_l}{\pi_l}. \tag{2.14}$$

The proof is given in Särndal et al. (1992, Result 2.8.1). For the formulas of the HT point and variance estimation for other quantities of interest, like ratios, we refer to Särndal et al. (1992).

Using the description of the HT estimator, we would like to draw attention to another property of the typical survey statistics notation. As seen in (2.12), the notation typically does not distinguish between an estimator and the resulting estimate. However, it should be clear that properties such as unbiasedness, variance, or MSE always refer to the estimator, not the estimated value.

For certain sampling designs, the formulas of the HT estimator simplify considerably, especially the form of the second-order inclusion probabilities. Compare e.g. the respective formulas under stratified random sampling in Särndal et al. (1992). In practice, computing the second-order inclusion probabilities π_{kl}, which are needed for the variance estimation, can be difficult, especially under complex survey designs. We do not cover variance estimation methods in this thesis and refer to Särndal et al. (1992, Chapter 11), Wolter (1985), and Münnich (2008) for overviews of variance approximation and estimation techniques.

Another design-unbiased estimator is the Hájek-type estimator, which can be especially useful for sampling designs where $\hat{N} = \sum_{k\in s} d_k$ is random. We do not consider Hájek-type estimators in the following chapters and refer to Morales et al. (2021, Section 2.5) and Särndal et al. (1992, Result 5.7.1) for its formulas and further information.

Using the HT estimator, we would like to introduce some additional terms related to estimators: Domain estimation, multivariate estimation, and the distinction between direct and indirect estimators.

Domain estimation

When conducting a survey, the interest is usually not only in statistics for the entire target population U, but also in statistics for a variety of sub-populations. The sub-populations are referred to as *areas* or *domains*, terms that are used synonymously in the following. Domains can be defined by regional aspects, referring to specific time periods, demographic information such as age and sex classes, or combinations of the three. An overview of domain-specific estimation is given in Särndal et al. (1992, Chapter 10). We consider that population U can be partitioned into D sub-populations, called domains, denoted by U_d of size N_d, $d = 1\ldots,D$, with (Särndal et al., 1992, Equation 10.2.1)

$$U = \bigcup_{d=1}^{D} U_d, \quad N = \sum_{d=1}^{D} N_d. \tag{2.15}$$

Assume we are interested in estimating $\tau_{y_d} = \sum_{k\in U_d} y_k$, the total of a variable Y in domain d. The domain membership of each sampling unit in the sample s is known. Therefore, and as the domains are non-overlapping, s can be partitioned into domain-specific sub-samples s_d, with $s_d = s \cap U_d$, $d = 1,\ldots,D$. The domain-specific HT estimates of τ_{y_d} are given by

$$\hat{\tau}_{y_d}^{\mathrm{HT}} = \sum_{k\in s_d} y_k d_k, \quad d = 1,\ldots,D. \tag{2.16}$$

Direct and indirect estimators

Especially in the context of small area estimation, a distinction is made between *direct estimators* and *indirect estimators*. In this context, we refer to Lehtonen and Veijanen (2009, Section 2.2.3) for a detailed description of the two. The HT estimator (2.16) is a direct estimator as it uses only the sample information of Y available in domain d, given by s_d, to calculate (2.16). For a direct estimator like the HT estimator, domain-specific estimation corresponds to treating each domain as a population of its own. The formulas of the variance (2.13) and the estimated variance (2.14) of the HT estimator can therefore be applied accordingly for domain estimation. *Indirect estimators* additionally use sample information of the variable of interest from other domains by the use of implicit or explicit statistical models. The generalised regression estimator, introduced in Section 2.3.2, can be calculated as a direct or indirect estimator.

Multiple quantities of interest

Surveys are used to compute a whole range of different statistics for a wide variety of variables and domain sets. When the sample data based on which different estimators are calculated is (partially) overlapping, their sampling errors are correlated. Särndal et al. (1992, Section 5.4) give an overview of the HT estimator for multiple quantities of interest. We consider m variables of interest with population values $(y_{k1},\ldots,y_{km})^\top$, $\forall k \in U$. The quantities of interest are the total values of these variables and given by $\tau_{y_j} = \sum_{k\in U} y_{kj}$, $j = 1,\ldots,m$. The HT estimator (2.12) can be calculated for each of the m variables resulting in $\hat{\tau}_{y_j}^{\mathrm{HT}}$, $j = 1,\ldots,m$. The covariance of any two of the HT estimators is given by (2.6). The covariance matrix of the m HT estimators can be set up with the variances of the m estimators on the diagonal and off-diagonal elements

$$\mathrm{Cov}(\hat{\tau}_{y_j}^{\mathrm{HT}}, \hat{\tau}_{y_p}^{\mathrm{HT}}) = \sum_{k\in U}\sum_{l\in U} (\pi_{kl} - \pi_k\pi_l) y_{kj} d_k y_{lp} d_l, \quad j,p = 1,\ldots,m. \tag{2.17}$$

An unbiased estimator for the variance and covariance is derived similarly to (2.14) by inserting the observations of a sample s instead of the population U.

In concrete applications, estimating the covariance between two estimators can be difficult, especially when the estimators only have some, but not all, sampling elements in common as in the rotating panel surveys covered in Chapter 4. We refer to Berger and Priam (2016) and the references therein for covariance estimation under rotating panel designs.

The covariances of multiple estimators play an important role for composite estimators in rotating panel surveys, which are covered in Chapter 4. Furthermore, the covariances between different estimators are taken as input in multivariate Fay-Herriot models, which are covered in in Chapters 5 and 6.

2.3.2 Generalized regression estimator

In a survey sample s usually different variables are observed for all units. For example, in the German Microcensus (Destatis, 2021) not only the employment status of a persons, but also their age and sex are observed. In addition, information from previous survey samples can be available. Furthermore, data from outside the survey, such as register information or population totals from the current population update, may also be available at the estimation stage. All this information can be used to form estimators which often exhibit higher precision, i.e. lower variance/MSE, than the HT estimator. Furthermore, the additional information can be used to ensure that different statistics add up to known population values or that the estimates of the same statistic in different surveys coincide. A prominent example of such an estimator is the generalised regression estimator, which is a special case of a calibration estimator. In the following, calibration estimators are introduced with the generalised regression estimator in particular.

Consider that in a sample s, in addition to the variable of interest Y, p additional variables are observed, which are henceforth called *auxiliary variables/auxiliaries/covariates*. That is, sample s contains pairs $(y_k, \boldsymbol{x}_k)$, where $\boldsymbol{x}_k = (x_{k1}, \ldots, x_{kp})^\top$, $\forall k \in s$. The population totals of the p variables, $\boldsymbol{\tau}_{\boldsymbol{x}} = (\tau_{x1}, \ldots, \tau_{xp})^\top$, are assumed to be known. Following the description in Särndal (2007), a *calibration estimator* of population total τ_y is given by

$$\hat{\tau}_y^{\text{calib}} = \sum_{k \in s} y_k w_k, \tag{2.18}$$

where *calibration weights* w_k, $k \in s$, are the solution to the optimisation problem

$$\begin{aligned} &\min_{w_k, k \in s} \sum_{k \in s} D(w_k, d_k) \\ &\text{subject to } \sum_{k \in s} \boldsymbol{x}_k w_k = \boldsymbol{\tau}_{\boldsymbol{x}}. \end{aligned} \tag{2.19}$$

The constraints in (2.19) are called *calibration constraints*. D is a strictly convex and continuously differentiable *distance function* for which $D(w_k, d_k) \geq 0$, $D(d_k, d_k) = 0$, and $\partial D(d_k, d_k)/\partial w_k = 0$, $\forall k \in s$, hold. Calibration weights w_k can be written as the product of design weights d_k and *correction weights* g_k, $w_k = d_k g_k$, $\forall k \in s$.

Särndal (2007, p. 99) summarises important features of calibration estimators: The optimisation of the calibration weights is independent of the choice of the variables of interest. To put it differently, a calibration estimator returns a set of calibration weights w_k, $\forall k \in s$, regardless of the variable(s) of interest. These weights satisfy the calibration constraints and can be used to compute all linearly weighted estimates. The calibration estimator is not design-unbiased, but design-consistent, i.e. the contribution of the bias to MSE of the estimator is asymptotically insignificant (Kott, 2006). Therefore, the calibration estimator is also called nearly design-unbiased by Särndal (2007, p. 99).

The calibration constraints in (2.19) can be used for different purposes. They can be used to ensure that different surveys return the same estimates of specific statistics or that regional estimates add up to certain national estimates. Calibration estimators can therefore be used to ensure the credibility of different estimates, which is considered as an important feature for the quality of official survey estimates (Särndal, 2007, p. 100). This is also the reasons why in practice the vector of totals $\boldsymbol{\tau_x}$ used in the calibration constraints often not only includes (assumed) known, but also estimated totals. The inclusion of estimated calibration totals, however, adds additional uncertainty to the calibration estimator, which has to be considered in its variance estimation. We refer to Berger et al. (2009) and Dever and Valliant (2010) for further information on the variance estimation under estimated totals. The calibration estimator can be severely biased when the totals of the p covariates are only approximated (Särndal et al., 1992, Remark 6.4.3). In addition to ensuring consistency with the calibration totals, calibration estimators often have lower variance than design-based estimators which do not make use of additional auxiliary information (Särndal, 2007, p. 101).

There are different distance functions D available, each of which leads to a different type of calibration estimator. In this thesis, we only use calibration estimators with the chi-square distance function, i.e. $D(w_k, d_k) = (w_k - d_k)^2/(d_k q_k)$, where $q_k > 0$ is an additional scaling factor. As in most applications, we take the scaling factor $q_k = 1$, $\forall k \in s$, and do not consider it further. We refer to Deville and Särndal (1992) for examples of calibration estimation with $q_k \neq 1$. For information on other calibration estimators, we refer to Deville and Särndal (1992), Kott (2009), and Särndal (2007). The estimates resulting from a calibration estimator with chi-square distance function correspond to the estimates of the standard linear *generalised regression* (GREG) estimator, which is described further on. Although the calibration estimator with chi-square distance function and the standard GREG estimator give the same survey estimates, they are grounded in very different philosophies. While the former focuses on the calibration to chosen totals, the GREG estimator is motivated by a regression model. We refer to Särndal (2007) for a detailed comparison of the two philosophies. In the following, the GREG estimator is presented in its different forms.

The standard linear GREG estimator (Cassel et al., 1976; Isaki & Fuller, 1982; Särndal, 1980; Wright, 1983) is given by

$$\begin{aligned}\hat{\tau}_y^{\text{GREG}} &= \sum_{k\in s} y_k w_k = \sum_{k\in s} y_k g_k d_k \\ &= \hat{\tau}_y^{\text{HT}} + (\boldsymbol{\tau_x} - \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}^{\text{HT}})^\top \hat{\boldsymbol{\beta}} \\ &= \sum_{k\in U} \boldsymbol{x}_k^\top \hat{\boldsymbol{\beta}} + \sum_{k\in s} w_k (y_k - \boldsymbol{x}_k^\top \hat{\boldsymbol{\beta}})\end{aligned} \tag{2.20}$$

with

$$g_k = 1 + (\boldsymbol{\tau_x} - \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}^{\text{HT}})^\top \left(\sum_{k\in s} \frac{\boldsymbol{x}_k \boldsymbol{x}_k^\top}{\pi_k} \right)^{-1} \boldsymbol{x}_k, \tag{2.21}$$

and the vector of estimated *coefficients/fixed effects* of length p

$$\hat{\boldsymbol{\beta}} = \left(\sum_{k\in s} \frac{\boldsymbol{x}_k \boldsymbol{x}_k^\top}{\pi_k} \right)^{-1} \sum_{k\in s} \frac{\boldsymbol{x}_k y_k}{\pi_k}. \tag{2.22}$$

The matrix of the auxiliary information is assumed to be of full rank such that its cross-product can be inverted. When it is not of full rank, the *generalised inverse* can be used instead, the computation of which is for example shown in Gentle (2007, Section 3.6).

The calibration form of the GREG is given by the first row of (2.20) with correction weights (2.21). It emphasises the incorporation of auxiliary information to receive a single set of weights w_k, $\forall k \in s$, satisfying the calibration constraints. For this formula, there is no distributional assumption and no statistical model involved. In contrast, the regression form of the GREG, given in the last two rows of (2.20) with weighted least squares coefficients (2.22), emphasises fitting a statistical model which reflects well the correlations of the variable of interest and the set of covariates. For a description of the statistical model which can be used to motivate the GREG estimator, we refer to Särndal et al. (1992, pp. 225–228).

Due to the complex nature of the GREG estimator, only an approximate variance formula based on a Taylor approximation can be given by (Särndal et al., 1992, Result 6.6.1)

$$\text{AVar}(\hat{\tau}_y^{\text{GREG}}) = \sum_{k\in U} \sum_{l\in U} (\pi_{kl} - \pi_k \pi_l) \frac{(y_k - \boldsymbol{x}_k^\top \boldsymbol{\beta})}{\pi_k} \frac{(y_l - \boldsymbol{x}_l^\top \boldsymbol{\beta})}{\pi_l}, \tag{2.23}$$

which can be estimated by

$$\begin{aligned}\widehat{\text{Var}}(\hat{\tau}_y^{\text{GREG}}) &= \sum_{k\in s} \sum_{l\in s} \left(1 - \frac{\pi_k \pi_l}{\pi_{kl}}\right) \frac{g_k (y_k - \boldsymbol{x}_k^\top \hat{\boldsymbol{\beta}})}{\pi_k} \frac{g_l (y_l - \boldsymbol{x}_l^\top \hat{\boldsymbol{\beta}})}{\pi_l} \\ &= \sum_{k\in s} \sum_{l\in s} \left(1 - \frac{\pi_k \pi_l}{\pi_{kl}}\right) w_k (y_k - \boldsymbol{x}_k^\top \hat{\boldsymbol{\beta}}) w_l (y_l - \boldsymbol{x}_l^\top \hat{\boldsymbol{\beta}}),\end{aligned} \tag{2.24}$$

where again $w_k = g_k d_k$, $\forall k \in s$. The variance formula of the HT (2.13) and GREG estimator (2.23) are quite similar. They differ only in that for the GREG estimator the *residuals* $(y_k - \boldsymbol{x}_k^\top \hat{\boldsymbol{\beta}})$ are inserted instead of y_k in the formula of the HT estimator (2.13), $\forall k \in s$. In the same way, also the formulas of the covariances of GREG estimates can be given similar to the HT formula (2.17).

Even though the GREG estimator can be motivated by the regression approach, it is a model-assisted estimator. The chosen set of auxiliary information constitutes an *assisting model* in the sense of the regression form of the GREG. The design-based properties of the estimator hold regardless of how well the chosen assisting model fits the data. Breidt and Opsomer (2017) and Robinson and Särndal (1983) show that the GREG estimator is asymptotically design-unbiased and design-consistent under mild conditions. Although the choice of the assisting model does not influence the design-based properties of the GREG estimator or the validity of its variance formulas, it heavily influences the magnitude of its variance. From (2.23) it can be seen that the smaller the residuals, i.e. the better the assisting model fits the data, the lower the variance of the GREG estimator.

The GREG estimator can be used for domain-specific estimation, either as a direct or an indirect estimator. If the GREG estimator is applied as a direct estimator, each domain is treated as a population of its own, similar to the domain estimation of the HT estimator (2.16). If, however, the GREG estimator is calculated based on the joint information of several domains and hence the estimated vector of coefficients $\hat{\boldsymbol{\beta}}$ is the same for all these domains, it is called an indirect estimator. We refer to Lehtonen and Veijanen (2009) for an overview of different direct and indirect GREG estimators.

There are two additional features of the presented GREG estimator which are relevant for practical applications. Särndal (2007, Section 5) emphasises that practitioners prefer the calibration weights w_k, $\forall k \in s$, to be positive and extreme values to be avoided. In theory, the design weights reflect the inverse of the probability of a unit to be included in a random sample drawn according to a sampling design. Negative weights are not compatible with this. Extreme calibration weights, i.e. a large ratio of the largest to the smallest w_k, is also seen as problematic, especially for domain estimation where sample sizes are small and thus extreme weights have a high influence. The calculation of the calibration weights $w_k = g_k d_k$ with g_k according to (2.21), $\forall k \in s$, can have negative weights as outcomes. In particular, an increasing number of calibration constraints and small sample sizes may lead to negative weights. Statistical offices therefore frequently use additional techniques to avoid negative or extreme weights when working with the GREG estimator. Särndal (2007, Section 5) and Park and Fuller (2005) list several techniques which can be used to avoid negative or extreme weights. To avoid extreme calibration weights, Münnich et al. (2012b) add *box-constraints* to the calibration estimator and present a semi-smooth Newton method for solving the resulting calibration problem. In the German Microcensus, an iterative algorithm, presented in Nieuwenbroek and Boonstra (2002), is used which adds dampening factors to the GREG equations (Afentakis & Bihler, 2005).

2.4 Area-level small area estimation

2.4.1 Introduction

Surveys are designed to provide estimates of various statistics of interest for various domain levels. The design-based direct estimators presented in Section 2.3 work well for domain estimation when the sample sizes of the domains are large. With large sample sizes, the variances of the (approximately) design-unbiased estimators are small. However, when considering smaller domains, e.g. statistics for cross-combinations of regionally fine-grained domains by age, sex, and nationality, the variance of the corresponding design-based estimators is typically larger due to small sample sizes.

Domains for which design-based direct estimators do not suffice to provide reliable estimates due to small sample sizes are called *small areas* or *small domains*. The research area of *small area estimation* (SAE) deals with methods and techniques that are designed for such situations. For small areas, indirect estimators that jointly use information from several domains and potential additional information in a common model can give domain estimators with a lower MSE than the corresponding direct estimators. This concept is referred to as *borrowing strength*. The joint modelling approach with data from several domains is said to increase the *effective sample size* of the domain estimates. A comprehensive overview of SAE techniques is given in Rao and Molina (2015). Morales et al. (2021) provide an overview with special focus on mixed model theory and `R` (R Core Team, 2020) implementations. Münnich et al. (2013) provide a compressed German summary. SAE is an area of ongoing research. Ghosh (2020), Jiang and Lahiri (2006), Pfeffermann (2002, 2013), and Rao and Molina (2015) provide overview articles of developments related to SAE. We name a few areas of recent research, without being exhaustive. There have been contributions on the recognition of measurement errors, e.g. in Burgard et al. (2020b, 2021a), Burgard et al. (2019a); the influence of sampling designs, e.g. in Burgard et al. (2014), Münnich and Burgard (2012), and Zimmermann (2018); the recognition of non-linear relationships in model-based estimators, e.g. in Wagner et al. (2017), and the analysis on the design-based MSEs of model-based estimators in Lahiri and Pramanik (2019) and Pfeffermann and Ben-Hur (2019).

There are different approaches of SAE, summarised e.g. in Morales et al. (2021) and Rao and Molina (2015). In the further course of this chapter as well as in Chapters 5 and 6, we focus on model-based SAE methods. Model-based SAE methods are applied in various contexts including official statistics. For example, in the U.S. Census Bureau's *Small Area Income and Poverty Estimates* (SAIPE) programme, described e.g. in Bell and Robinson (2020), federal funds are allocated based on model-based small area estimates of income and poverty for counties and school districts. Because of the importance of SAE to official statistics, it has been designated as one of seven research areas of the U.S. Census Bureau, which provides an overview of current projects in U.S. Bureau of the Census (2020). Zimmermann (2019) evaluated the applicability of SAE for the German Census 2021. The World Bank applies SAE for poverty mapping in many countries. For example, National Statistics Bureau of Bhutan and the World Bank (2010) examined poverty mapping in Buthan. A summary of different projects related to the application of SAE techniques in official statistics is given in Kordos (2014).

In Chapters 5 and 6, we present contributions to multivariate Fay-Herriot models, which belong to the class of area-level model-based small area methods. The model-based small area methods, which we cover in this thesis, are based on the theory of linear mixed models. In the following, an overview of the theoretical basics of linear mixed models is given. Based on this theory, the Fay-Herriot model is introduced in Section 2.4.3.

2.4.2 Linear mixed model

Model

Mixed models constitute a flexible modelling framework that allows both *fixed effects* and *random effects* in a model. Searle et al. (2006, Section 1.3) and Demidenko (2013, Chapter 1) give a detailed overview of the different philosophies of fixed and random effects and examples of their applications. While fixed effects represent unobserved constant parameters, random effects constitute realisations of random variables, which themselves are functions of fixed parameters. By combining the two concepts of effects, mixed models can be seen as a compromise between Bayesian and frequentist approaches (Demidenko, 2013, Section 1.4). The use of random effects allows for the inclusion of potentially complex correlation structures into a model. For example, in a longitudinal dataset each data row can refer to a specific person and a specific time point. The observations corresponding to a particular person are often assumed to be dependent. A random effect at the person-level can account for this dependence structure in the data by assuming that all observations of a specific person are subject to the same realisation of a random variable and thus correlated. In the SAE context, random effects allow to incorporate domain-specific differences in a model.

We focus on *linear mixed models* (LMMs). For the notation and description, we follow Rao and Molina (2015), Morales et al. (2021), and Henderson (1975), unless stated otherwise. The general linear mixed model is given by

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{e}, \tag{2.25}$$

where $\boldsymbol{y}$ is a vector of observations of length n, $\boldsymbol{X}$ is a known $n \times p$ matrix of auxiliary information, $\boldsymbol{\beta}$ is the vector of fixed effects associated with the p auxiliary variables, $\boldsymbol{Z}$ is a known $n \times h$ matrix for the covariance structure of the random effects, $\boldsymbol{u}$ is a vector of random effects of length h, and $\boldsymbol{e}$ is a vector of residuals of length n. Matrices $\boldsymbol{X}$ and $\boldsymbol{Z}$ are assumed to be of full (column) rank. $\boldsymbol{X}$ usually incorporates a column with ones as the first column to account for an intercept in the model. We assume an intercept column in all models, unless stated otherwise.

In LMMs, both fixed and random effects are linearly related to observations $\boldsymbol{y}$. In model (2.25), there are two sources of variation, random effects $\boldsymbol{u}$ and residuals, also called random errors, $\boldsymbol{e}$. We consider that

$$\mathrm{E}\begin{bmatrix}\boldsymbol{u}\\ \boldsymbol{e}\end{bmatrix} = \begin{pmatrix}\mathbf{0}\\ \mathbf{0}\end{pmatrix} \quad \text{and} \quad \mathrm{Cov}\begin{pmatrix}\boldsymbol{u}\\ \boldsymbol{e}\end{pmatrix} = \begin{pmatrix}\boldsymbol{G} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{R}\end{pmatrix}. \tag{2.26}$$

That is, the two sources of variation are uncorrelated, have expectation of zero, and their covariance matrices are given by $\boldsymbol{G}$ and $\boldsymbol{R}$ respectively. The variance parameters that make up $\boldsymbol{G}$ and $\boldsymbol{R}$ are denoted by a vector $\boldsymbol{\delta}$ of length q. Under model (2.25), the covariance matrix of $\boldsymbol{y}$ is given by $\boldsymbol{V} = \mathrm{Var}(\boldsymbol{y}) = \boldsymbol{R} + \boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}^\top$. The expected values of $\boldsymbol{y}$ under LMM (2.25) can be written in two ways. The *conditional expectation* is the expectation conditioned on the realisations of the random effects and given by $\mathrm{E}[\boldsymbol{y}|\boldsymbol{u}] = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u}$. By iterated expectation, the *unconditional expectation* is given by $\mathrm{E}[\boldsymbol{y}] = \mathrm{E}[\mathrm{E}[\boldsymbol{y}|\boldsymbol{u}]] = \boldsymbol{X}\boldsymbol{\beta}$. In applications of mixed models in SAE, the focus is often on the conditional expectation $\mathrm{E}[\boldsymbol{y}|\boldsymbol{u}]$.

Prediction

In SAE, we are interested in making predictions in terms of the conditional expectation. In the general case, we are interested in linear combinations of type

$$\mu = \boldsymbol{l}^\top\boldsymbol{\beta} + \boldsymbol{m}^\top\boldsymbol{u}, \tag{2.27}$$

for some fixed vector $\boldsymbol{l}$ of length p and vector $\boldsymbol{m}$ of length h (Rao & Molina, 2015, Section 5.2). In order to predict μ, we need to predict the realisations of the random effects $\boldsymbol{u}$. In this thesis, we focus on frequentist approaches to predict (2.27). We note that also Bayesian approaches can be applied and refer to Morales et al. (2021, Section 16.7) and Rao and Molina (2015, Chapter 10) for a description of the hierarchical Bayes approach in SAE and Rao and Molina (2015, Chapter 9) for a description of the empirical Bayes approach to SAE.

Depending on the availability of information, we differentiate between the following predictors. For detailed derivations of the predictors, we refer to Morales et al. (2021, Chapter 6), Searle et al. (2006, Chapter 7), and Henderson (1975), on which the following description is based. We start with assuming that both $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ are known. A predictor is called *best* when its corresponding MSE to the true value is minimal in its class of predictors (Searle et al., 2006, Section 7.2). The *best predictor* (BP) of μ and $\boldsymbol{u}$ is given by their conditional expectation given the data,

$$\hat{\mu}^{\mathrm{BP}} = \mathrm{E}[\mu|\boldsymbol{y}], \quad \hat{\boldsymbol{u}}^{\mathrm{BP}} = \mathrm{E}[\boldsymbol{u}|\boldsymbol{y}]. \tag{2.28}$$

The *best linear predictor* (BLP) of μ and $\boldsymbol{u}$ is given by

$$\hat{\mu}^{\mathrm{BLP}} = t(\boldsymbol{\delta}, \boldsymbol{y}) = \boldsymbol{l}^\top\boldsymbol{\beta} + \boldsymbol{m}^\top\hat{\boldsymbol{u}}^{\mathrm{BLP}}, \tag{2.29}$$

and

$$\hat{\boldsymbol{u}}^{\mathrm{BLP}} = \hat{\boldsymbol{u}}^{\mathrm{BLP}}(\boldsymbol{\delta}) = \boldsymbol{G}\boldsymbol{Z}^\top\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}). \tag{2.30}$$

The term *linear* refers to the linearity of $\hat{\mu}^{\mathrm{BLP}}$ and $\hat{\boldsymbol{u}}^{\mathrm{BLP}}$ in $\boldsymbol{y}$. Under normally distributed $\boldsymbol{e}$ and $\boldsymbol{u}$, predictors $\hat{\boldsymbol{u}}^{\mathrm{BLP}}$ and $\hat{\mu}^{\mathrm{BLP}}$ are also $\hat{\boldsymbol{u}}^{\mathrm{BP}}$ and $\hat{\mu}^{\mathrm{BP}}$ (Henderson, 1975).

Note that in LMM and SAE theory, especially for the MSE formulas presented later, it is common to stress certain dependencies between different terms and random variables. This is why the predictors of μ are often written as a function t of variance components $\boldsymbol{\delta}$ and observations $\boldsymbol{y}$ like $\hat{\mu}^{\text{BLP}} = t\left(\boldsymbol{\delta}, \boldsymbol{y}\right)$ in (2.29).

Generally, the vector of fixed effects $\boldsymbol{\beta}$ is unknown. It can be estimated by generalised least squares (GLS) giving the *best linear unbiased estimator* (BLUE)

$$\hat{\boldsymbol{\beta}}^{\text{BLUE}} = \hat{\boldsymbol{\beta}}^{\text{BLUE}}\left(\boldsymbol{\delta}\right) = \left(\boldsymbol{X}^\top \boldsymbol{V}^{-1} \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{V}^{-1} \boldsymbol{y}. \tag{2.31}$$

With normally distributed random effects and residuals, $\hat{\boldsymbol{\beta}}^{\text{BLUE}}$ also corresponds to the maximum likelihood solution of $\boldsymbol{\beta}$ (Morales et al., 2021, p. 113). The linearity of $\hat{\boldsymbol{\beta}}^{\text{BLUE}}$ is again given with respect to $\boldsymbol{y}$. Inserting $\hat{\boldsymbol{\beta}}^{\text{BLUE}}$ in (2.29) and (2.30) yields the *best linear unbiased predictor* (BLUP) (Henderson (1975), proof in Henderson (1963))

$$\hat{\mu}^{\text{BLUP}} = t\left(\boldsymbol{\delta}, \boldsymbol{y}\right) = \boldsymbol{l}^\top \hat{\boldsymbol{\beta}}^{\text{BLUE}} + \boldsymbol{m}^\top \hat{\boldsymbol{u}}^{\text{BLUP}} \tag{2.32}$$

with

$$\hat{\boldsymbol{u}}^{\text{BLUP}} = \hat{\boldsymbol{u}}^{\text{BLUP}}\left(\boldsymbol{\delta}\right) = \boldsymbol{G} \boldsymbol{Z}^\top \boldsymbol{V}^{-1} \left(\boldsymbol{y} - \boldsymbol{X} \hat{\boldsymbol{\beta}}^{\text{BLUE}}\right). \tag{2.33}$$

Predictor $\hat{\mu}^{\text{BLUP}}$ is unbiased as $\mathrm{E}\left[\hat{\mu}^{\text{BLUP}}\right] = \mathrm{E}\left[\mu\right]$.

Alternatively, $\hat{\boldsymbol{u}}^{\text{BLUP}}$ and $\hat{\boldsymbol{\beta}}^{\text{BLUE}}$ can be derived by the *mixed model equations* (Henderson (1975), proof in Henderson (1963)), denoted by additional superscript $*$, taking

$$\hat{\mu}^{\text{BLUP}\,*} = \boldsymbol{l}^\top \hat{\boldsymbol{\beta}}^{\text{BLUE}\,*} + \boldsymbol{m}^\top \hat{\boldsymbol{u}}^{\text{BLUP}\,*} \tag{2.34}$$

with

$$\begin{pmatrix} \boldsymbol{X}^\top \boldsymbol{R}^{-1} \boldsymbol{X} & \boldsymbol{X}^\top \boldsymbol{R}^{-1} \boldsymbol{Z} \\ \boldsymbol{Z}^\top \boldsymbol{R}^{-1} \boldsymbol{X} & \boldsymbol{Z}^\top \boldsymbol{R}^{-1} \boldsymbol{Z} + \boldsymbol{G}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}}^{\text{BLUE}\,*} \\ \hat{\boldsymbol{u}}^{\text{BLUP}\,*} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}^\top \boldsymbol{R}^{-1} \boldsymbol{y} \\ \boldsymbol{Z}^\top \boldsymbol{R}^{-1} \boldsymbol{y} \end{pmatrix}. \tag{2.35}$$

The mixed model equations avoid the computation of $\boldsymbol{V}^{-1}$ in (2.33) and can therefore be computationally simpler when $\boldsymbol{G}$ and $\boldsymbol{R}$ have a (block-)diagonal structure.

The preceding predictors are defined for known variance components $\boldsymbol{\delta}$, i.e. the covariance matrix $\boldsymbol{V}$ is fully specified. Vector $\boldsymbol{\delta}$ is generally unknown and can be replaced by estimates $\hat{\boldsymbol{\delta}} = \hat{\boldsymbol{\delta}}\left(\boldsymbol{y}\right)$. Up to now, except for the noted exceptions, the assumption of normally distributed $\boldsymbol{u}$ and $\boldsymbol{e}$ was not necessary. If both $\boldsymbol{u}$ and $\boldsymbol{e}$ are assumed to be normally distributed, the model is referred to as *gaussian linear mixed model.* We refer to Jiang (2007) for an overview of gaussian and non-gaussian mixed models. In this thesis, only gaussian mixed models are considered, i.e. $\boldsymbol{u} \sim N_n\left(\boldsymbol{0}, \boldsymbol{G}\right)$ and $\boldsymbol{e} \sim N_n\left(\boldsymbol{0}, \boldsymbol{R}\right)$. Therefore, the conditional and marginal distribution of the observations are given by

$$\boldsymbol{y} | \boldsymbol{u} \sim N_n\left(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u}, \boldsymbol{R}\right) \tag{2.36}$$

$$\boldsymbol{y} \sim N_n\left(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{R} + \boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}^\top\right). \tag{2.37}$$

By making distributional assumptions about the random variables of the LMM, likelihood-based methods can be applied to estimate variance parameters $\boldsymbol{\delta}$, namely *Maximum Likelihood* (ML) and *Restricted Maximum Likelihood* (REML), which give consistent parameter estimates. They are described in Section 2.4.2. Inserting ML or REML estimates $\hat{\boldsymbol{\delta}}$ for $\boldsymbol{\delta}$ in the BLUP formula (2.32) yields the *empirical BLUP* (EBLUP). Also $\hat{\boldsymbol{\beta}}^{\text{BLUE}}$ is a function of $\boldsymbol{\delta}$. Plugging in $\hat{\boldsymbol{\delta}}$ for $\boldsymbol{\delta}$ in (2.31) yields the *empirical BLUE* (EBLUE) $\hat{\boldsymbol{\beta}}^{\text{EBLUE}}$, which is henceforth denoted as $\hat{\boldsymbol{\beta}}$.

Mean squared error

We are not only interested in finding a good prediction of the linear combination (2.27), but also want to determine the uncertainty associated with the prediction, given by its MSE. As shown in Rao and Molina (2015, Section 5.2.2), we can define $\boldsymbol{d}^\top = \boldsymbol{l}^\top - \boldsymbol{m}^\top \boldsymbol{G}\boldsymbol{Z}^\top \boldsymbol{V}^{-1}\boldsymbol{X}$ and express the MSE of $\hat{\mu}^{\text{BLUP}}$ as

$$\begin{aligned}\text{MSE}\left(\hat{\mu}^{\text{BLUP}}\right) &= \text{MSE}\left(t\left(\boldsymbol{\delta}, \boldsymbol{y}\right)\right)\\ &= \text{MSE}\left(\hat{\mu}^{\text{BLP}}\right) + \text{Var}\left(\boldsymbol{d}^\top\left(\hat{\boldsymbol{\beta}}^{\text{BLUE}} - \boldsymbol{\beta}\right)\right)\\ &= g_1\left(\boldsymbol{\delta}\right) + g_2\left(\boldsymbol{\delta}\right)\end{aligned} \tag{2.38}$$

with

$$g_1\left(\boldsymbol{\delta}\right) = \boldsymbol{m}^\top\left(\boldsymbol{G} - \boldsymbol{G}\boldsymbol{Z}^\top\boldsymbol{V}^{-1}\boldsymbol{Z}\boldsymbol{G}\right)\boldsymbol{m} \tag{2.39}$$

$$g_2\left(\boldsymbol{\delta}\right) = \boldsymbol{d}^\top\left(\boldsymbol{X}^\top\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{d} \tag{2.40}$$

such that terms $g_1\left(\boldsymbol{\delta}\right)$ and $g_2\left(\boldsymbol{\delta}\right)$ account for the variability of the BLP for known $\boldsymbol{\beta}$ and the variability of estimating $\boldsymbol{\beta}$ respectively.

Compared to the MSE of $\hat{\mu}^{\text{BLUP}}$, the MSE of $\hat{\mu}^{\text{EBLUP}}$ contains an additional component to account for the variability attributed to estimating $\boldsymbol{\delta}$. For estimators $\hat{\boldsymbol{\delta}}$, which are translation-invariant and even like ML/REML, and normally distributed random effects and residuals, which we always assume here, it is given by (Kackar & Harville, 1981)

$$\begin{aligned}\text{MSE}\left(\hat{\mu}^{\text{EBLUP}}\right) &= \text{MSE}\left(t\left(\hat{\boldsymbol{\delta}}, \boldsymbol{y}\right)\right)\\ &= \text{MSE}\left(\hat{\mu}^{\text{BLUP}}\right) + \text{E}\left[\hat{\mu}^{\text{EBLUP}} - \hat{\mu}^{\text{BLUP}}\right]^2\\ &= \text{MSE}\left(\hat{\mu}^{\text{BLUP}}\right) + \text{E}\left[t\left(\hat{\boldsymbol{\delta}}, \boldsymbol{y}\right) - t\left(\boldsymbol{\delta}, \boldsymbol{y}\right)\right]^2.\end{aligned} \tag{2.41}$$

Term $\text{E}\left[\hat{\mu}^{\text{EBLUP}} - \hat{\mu}^{\text{BLUP}}\right]^2$ does not have a closed form expression. Das et al. (2004) developed approximations of (2.41) for general LMMs. By specifying the covariance structure of the LMM, the approximation and estimation of MSE $\left(\hat{\mu}^{\text{EBLUP}}\right)$ can be simplified. After a description of likelihood-based estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ in the following Section 2.4.2, we give an approximation of (2.41) for LMMs with block-diagonal covariance structures in Section 2.4 as we only consider LMMs with block-diagonal covariance structures in this thesis.

Parameter estimation

The standard approaches to estimating variance parameters $\boldsymbol{\delta}$ in LMMs are based on likelihood maximization. One can either apply *maximum likelihood* (ML), which uses the unrestricted likelihood, or *restricted maximum likelihood* (REML), where the restricted likelihood, also called *residual likelihood*, is used. With the term *maximum likelihood* we will refer to both methods, ML and REML. Alternative approaches to estimate the variance components in LMMs include the *Minimum Norm Quadratic Unbiased Estimation* (MINQUE) and method of moments. We refer to Demidenko (2013, Chapter 3) for more information on these. The following description of ML and REML is based on Jiang (2007, Section 1.3).

With (2.37), the distribution of $\boldsymbol{y}$ is n-variate normal with density function

$$f\left(\boldsymbol{y}\right) = \frac{1}{\left(2\pi\right)^{n/2}\det\left(\boldsymbol{V}\right)^{1/2}}\exp\left(-\frac{1}{2}\left(\boldsymbol{y}-\boldsymbol{X\beta}\right)^{\top}\boldsymbol{V}^{-1}\left(\boldsymbol{y}-\boldsymbol{X\beta}\right)\right) \tag{2.42}$$

and log-likelihood

$$\ell\left(\boldsymbol{\beta},\boldsymbol{\delta}\right) = c - \frac{1}{2}\log\left(\det\left(\boldsymbol{V}\right)\right) - \frac{1}{2}\left(\boldsymbol{y}-\boldsymbol{X\beta}\right)^{\top}\boldsymbol{V}^{-1}\left(\boldsymbol{y}-\boldsymbol{X\beta}\right), \tag{2.43}$$

where c is a constant term.

In ML estimation, (2.43) is maximised by choice of $\boldsymbol{\delta}$ and $\boldsymbol{\beta}$. The first derivatives of (2.43) with respect to these parameters are given by

$$\frac{\partial\ell}{\partial\boldsymbol{\beta}} = \boldsymbol{X}^{\top}\boldsymbol{V}^{-1}\boldsymbol{y} - \boldsymbol{X}^{\top}\boldsymbol{V}^{-1}\boldsymbol{X\beta}, \tag{2.44}$$

$$\frac{\partial\ell}{\partial\delta_r} = -\frac{1}{2}\left(\left(\boldsymbol{y}-\boldsymbol{X\beta}\right)^{\top}\boldsymbol{V}^{-1}\frac{\partial\boldsymbol{V}}{\partial\delta_r}\boldsymbol{V}^{-1}\left(\boldsymbol{y}-\boldsymbol{X\beta}\right) - \operatorname{tr}\left(\boldsymbol{V}^{-1}\frac{\partial\boldsymbol{V}}{\partial\delta_r}\right)\right), \tag{2.45}$$

$$r = 1,\ldots,q.$$

The vectors of the first partial derivatives are called *score* vectors.

ML estimation has one particular drawback: It generally does not lead to unbiased estimators of the variance components. To show this with a simple example, Neyman and Scott (1948) presented the problem of estimating the variance parameter σ^2 of a normal distribution based on sample data. The sample observations of the variable of interest are denoted by y_i, $i = 1,\ldots,n$. An unbiased estimator of the variance is given by $s^2 = \left(\sum_{i=1}^{n}\left(y_i - \sum_{i=1}^{n} y_i/n\right)^2\right)/\left(n-1\right)$. The ML estimator, however, yields $\hat{\sigma}^2 = \left(\left(n-1\right)/n\right)s^2$ and hence is biased. Similarly, in general LMMs the loss in degrees of freedom caused by estimating $\boldsymbol{\beta}$ is not accounted for in ML, leading to a biased estimators of $\boldsymbol{\delta}$. When the number of fixed effects is constant, ML estimators are consistent and asymptotically normally distributed (Jiang, 2007, p. 11). In the above mentioned example we see that for a constant number of fixed effects, the difference between s^2 and $\hat{\sigma}^2$ vanishes for $n \to \infty$. We refer to Demidenko (2013, Section 3.6) for more information on the statistical properties of ML estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ for small and large samples.

To avoid the potential bias in ML, especially for small sample sizes, often REML is applied instead. The formulation of REML estimation was presented in Patterson and Thompson (1971), extending the ideas of Nelder (1968) to incomplete block designs. The procedure is e.g. described in Jiang (2007, Section 1.3.2), which is the main source of the following description. The idea of REML is based on transforming the likelihood of the data such that it no longer depends on $\boldsymbol{\beta}$. For that, recall that $\boldsymbol{X}$ is assumed to be of full column-rank p. The transformation is achieved by constructing an $n \times (n-p)$ matrix $\boldsymbol{A}$ of rank $n-p$ such that $\boldsymbol{A}^\top \boldsymbol{X} = \boldsymbol{0}$. Then, set $\boldsymbol{z} = \boldsymbol{A}^\top$ such that $\boldsymbol{z} \sim N_{n-p}\left(\boldsymbol{0}, \boldsymbol{A}^\top \boldsymbol{V} \boldsymbol{A}\right)$. Following (2.37), the distribution of $\boldsymbol{y}$ is n-variate normal and the density function of $\boldsymbol{z}$ is thus given by

$$f\left(\boldsymbol{z}\right) = \frac{1}{\left(2\pi\right)^{(n-p)/2} \det\left(\boldsymbol{A}^\top \boldsymbol{V} \boldsymbol{A}\right)^{1/2}} \exp\left(-\frac{1}{2}\boldsymbol{z}^\top \left(\boldsymbol{A}^\top \boldsymbol{V} \boldsymbol{A}\right)^{-1} \boldsymbol{z}\right) \tag{2.46}$$

such that the log of the restricted likelihood is given by

$$\ell^{\text{REML}}\left(\boldsymbol{\delta}\right) = c - \frac{1}{2}\log\left(\det\left(\boldsymbol{A}^\top \boldsymbol{V} \boldsymbol{A}\right)\right) - \frac{1}{2}\boldsymbol{z}^\top \left(\boldsymbol{A}^\top \boldsymbol{V} \boldsymbol{A}\right)^{-1} \boldsymbol{z}. \tag{2.47}$$

In REML estimation, (2.47) is maximised by choice of $\boldsymbol{\delta}$. Plugging in the REML estimates $\hat{\boldsymbol{\delta}}$ in the GLS equation (2.31) yields REML solutions $\hat{\boldsymbol{\beta}}$. The REML log-likelihood can also be motivated by a Bayesian approach, see Laird and Ware (1982) and Demidenko (2013, Section 2.2.6) respectively. The first partial derivatives of (2.47) are given by

$$\frac{\partial \ell^{\text{REML}}}{\partial \delta_r} = \frac{1}{2}\left(\boldsymbol{y}^\top \boldsymbol{P} \frac{\partial \boldsymbol{V}}{\partial \delta_r} \boldsymbol{P} \boldsymbol{y} - \operatorname{tr}\left(\boldsymbol{P}\frac{\partial \boldsymbol{V}}{\partial \delta_r}\right)\right), \quad r = 1, \ldots, q, \tag{2.48}$$

where we set

$$\begin{aligned} \boldsymbol{P} &= \boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}\left(\boldsymbol{X}^\top \boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^\top \boldsymbol{V}^{-1} \\ &= \boldsymbol{A}\left(\boldsymbol{A}^\top \boldsymbol{V} \boldsymbol{A}\right)^{-1}\boldsymbol{A}^\top. \end{aligned} \tag{2.49}$$

Jiang (2007, Section 1.8) summarised some findings and literature for the asymptotic behaviour of ML and REML estimation. Similar to ML, also REML estimators are consistent and asymptotically follow a normal distribution. Under normality and fixed rank p of design matrix $\boldsymbol{X}$, the ML and REML estimators are asymptotically equivalent. However, Neyman and Scott (1948) show that ML estimators can be inconsistent and REML estimation superior when the number of fixed effects grows with the sample size n. See the discussion in Jiang (2007, Section 1.8) for further information.

We already showed the first partial derivatives of ML and REML with respect to the unknown parameters $\boldsymbol{\psi}$, i.e. $\boldsymbol{\psi} = \left(\boldsymbol{\beta}^\top, \boldsymbol{\delta}^\top\right)^\top$ for ML and $\boldsymbol{\psi} = \boldsymbol{\delta}$ for REML. Recall that the quadratic matrix of the second partial derivatives with respect to $\boldsymbol{\psi}$ is called *Hessian matrix* $\boldsymbol{H}$. The *Fisher information matrix* $\boldsymbol{F}$ is given by the negative expectation of the Hessian matrix, i.e.

$$\boldsymbol{F}\left(\boldsymbol{\psi}\right) = -\operatorname{E}\left[\boldsymbol{H}\right] = -\left(\operatorname{E}\left[\frac{\partial^2 \ell}{\partial \boldsymbol{\psi}\boldsymbol{\psi}^\top}\right]\right). \tag{2.50}$$

For the second partial derivatives we use notation $\partial^2\ell/\partial\boldsymbol{\psi}\boldsymbol{\psi}^\top = \ell_{\boldsymbol{\psi}\boldsymbol{\psi}}$. For ML, we have (Searle et al., 2006, Section 6.3)

$$\boldsymbol{F}^{\mathrm{ML}}(\boldsymbol{\psi}) = \boldsymbol{F}^{\mathrm{ML}}\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{\delta}\end{pmatrix} = -\operatorname{E}\begin{bmatrix}\ell_{\boldsymbol{\beta}\boldsymbol{\beta}} & \ell_{\boldsymbol{\beta}\boldsymbol{\delta}}\\ \ell_{\boldsymbol{\delta}\boldsymbol{\beta}} & \ell_{\boldsymbol{\delta}\boldsymbol{\delta}}\end{bmatrix}, \tag{2.51}$$

where

$$-\operatorname{E}[\ell_{\boldsymbol{\beta}\boldsymbol{\beta}}] = \boldsymbol{X}^\top\boldsymbol{V}^{-1}\boldsymbol{X}, \tag{2.52}$$

$$-\operatorname{E}[\ell_{\delta_r\delta_l}] = \frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}^{-1}\frac{\partial\boldsymbol{V}}{\partial\delta_r}\boldsymbol{V}^{-1}\frac{\partial\boldsymbol{V}}{\partial\delta_l}\right), \quad r,l = 1,\dots,q \tag{2.53}$$

$$-\operatorname{E}[\ell_{\beta_r\delta_l}] = -\operatorname{E}[\ell_{\delta_l\beta_r}] = 0, \quad r = 1,\dots,p,\ l = 1,\dots,q. \tag{2.54}$$

Likewise, for REML we have (Searle et al., 2006, Section 6.6)

$$\boldsymbol{F}^{\mathrm{REML}}(\boldsymbol{\delta}) = -\operatorname{E}\left[\ell^{\mathrm{REML}}_{\boldsymbol{\delta}\boldsymbol{\delta}}\right], \tag{2.55}$$

where

$$-\operatorname{E}\left[\ell^{\mathrm{REML}}_{\delta_r\delta_l}\right] = \frac{1}{2}\operatorname{tr}\left(\boldsymbol{P}\frac{\partial\boldsymbol{V}}{\partial\delta_r}\boldsymbol{P}\frac{\partial\boldsymbol{V}}{\partial\delta_l}\right), \quad r,l = 1,\dots,q. \tag{2.56}$$

We do not only want to estimate the unknown parameters $\boldsymbol{\psi}$, but also determine the variance of the estimates. The large-sample, or asymptotic as $n \to \infty$, dispersion matrix of ML/REML estimates $\hat{\boldsymbol{\psi}}$ is given by

$$\operatorname{AVar}\left(\hat{\boldsymbol{\psi}}\right) = \left(\boldsymbol{F}(\boldsymbol{\psi})\right)^{-1}, \tag{2.57}$$

where AVar denotes the asymptotic covariance matrix, similar to the notation in Rao and Molina (2015, Section 5). The technical details of (2.57) are given in Searle et al. (2006, Appendix S. 7). We have

$$\operatorname{AVar}\left(\hat{\boldsymbol{\beta}}^{\mathrm{ML}}\right) = \operatorname{AVar}\left(\hat{\boldsymbol{\beta}}^{\mathrm{REML}}\right) = \left(\boldsymbol{X}^\top\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1} \tag{2.58}$$

$$\operatorname{AVar}\left(\hat{\boldsymbol{\delta}}^{\mathrm{ML}}\right) = \left(\boldsymbol{F}^{\mathrm{ML}}(\boldsymbol{\psi})\right)^{-1} \tag{2.59}$$

$$\operatorname{AVar}\left(\hat{\boldsymbol{\delta}}^{\mathrm{REML}}\right) = \left(\boldsymbol{F}^{\mathrm{REML}}(\boldsymbol{\psi})\right)^{-1}. \tag{2.60}$$

From the first partial derivatives of ML and REML with respect to $\boldsymbol{\psi}$, it follows that there are interdependencies wherefore the likelihood can generally not be maximised analytically in LMMs. Instead, iterative procedures such as the algorithms *Expectation-Maximization* (EM), *Fisher-Scoring* (FS), and *Newton-Raphson* (NR) can be used. Demidenko (2013, Chapter 2) gives detailed descriptions of the three algorithms in the context of LMMs. NR and FS are closely related. NR uses the negative Hessian matrix, whereas FS uses the information matrix. Using FS instead of NR has the advantage that the information matrix is always positive definite and therefore invertible. This does not hold for the negative Hessian in the NR algorithm. In addition, the findings in Demidenko and Spiegelman

(1997) support the use of using FS instead of NR. Demidenko (2013, p. 85) argues that FS is more robust to outliers and performs better for far-off starting values than NR. Compared to EM, the asymptotic covariance matrix of the estimated parameters is calculated as a by-product in NR and FS. Furthermore, in the context of the later discussed Fay-Herriot models, which build upon the presented LMM theory, it is common to use the FS algorithm. Therefore, in Chapters 5 and 6 we use the FS algorithm, presented in Algorithms 6.1 and 6.2, for the parameter estimation in the presented multivariate Fay-Herriot models.

LMMs with block-diagonal structure

The previous description of LMMs and best predictors was rather general. In this thesis, we focus on data with block-diagonal covariance structures, where the data of each domain represents a block. That is, the observations of a particular domain are assumed to all be subject to the same realisation of a domain-specific random effect and are therefore correlated, whereas observations from different domains are uncorrelated.

Rao and Molina (2015, Section 5.3) displayed how to transfer the general LMM theory to LMMs with block-diagonal covariance structure and are taken as the main source of the following description. Consider D domains of interest. A block-diagonal covariance structure implies that LMM (2.25) can be rewritten as

$$\boldsymbol{y}_d = \boldsymbol{X}_d\boldsymbol{\beta} + \boldsymbol{Z}_d\boldsymbol{u}_d + \boldsymbol{e}_d, \quad d = 1, \dots, D, \tag{2.61}$$

where

$$\begin{aligned} \boldsymbol{y} &= \operatorname*{col}_{1\le d\le D}(\boldsymbol{y}_d) = \left(\boldsymbol{y}_1^\top, \dots, \boldsymbol{y}_D^\top\right)^\top, \quad \boldsymbol{X} = \operatorname*{col}_{1\le d\le D}(\boldsymbol{X}_d), \\ \boldsymbol{Z} &= \operatorname*{diag}_{1\le d\le D}(\boldsymbol{Z}_d), \quad \boldsymbol{u} = \operatorname*{col}_{1\le d\le D}(\boldsymbol{u}_d), \quad \boldsymbol{e} = \operatorname*{col}_{1\le d\le D}(\boldsymbol{e}_d). \end{aligned} \tag{2.62}$$

As the observations of different domains are uncorrelated, we have that $\boldsymbol{V}_d = \operatorname{Var}(\boldsymbol{y}_d) = \boldsymbol{R}_d + \boldsymbol{Z}_d\boldsymbol{G}_d\boldsymbol{Z}_d^\top$ with $\boldsymbol{G} = \operatorname*{diag}_{1\le d\le D} \boldsymbol{G}_d$, $d = 1, \dots, D$. Note that the number of data rows assigned to any domain d is allowed to vary, i.e. the blocks may be of different size.

With block-diagonal LMMs, the linear combinations of interest are domain specific, given by

$$\mu_d = \boldsymbol{l}_d^\top\boldsymbol{\beta} + \boldsymbol{m}_d^\top\boldsymbol{u}_d, \quad d = 1, \dots, D, \tag{2.63}$$

for some fixed vectors $\boldsymbol{l}_d$ of length p and $\boldsymbol{m}_d$ of length h_d. The predictors and estimators of Section 2.4.2 can be re-expressed as

$$\hat{\boldsymbol{\beta}}^{\text{BLUE}} = \hat{\boldsymbol{\beta}}^{\text{BLUE}}(\boldsymbol{\delta}) = \left(\sum_{d=1}^{D} \boldsymbol{X}_d^\top\boldsymbol{V}_d^{-1}\boldsymbol{X}_d\right)^{-1}\left(\sum_{d=1}^{D} \boldsymbol{X}_d^\top\boldsymbol{V}_d^{-1}\boldsymbol{y}_d\right), \tag{2.64}$$

$$\hat{\boldsymbol{u}}_d^{\text{BLUP}} = \hat{\boldsymbol{u}}_d^{\text{BLUP}}(\boldsymbol{\delta}) = \boldsymbol{G}_d\boldsymbol{Z}_d^\top\boldsymbol{V}_d^{-1}\left(\boldsymbol{y}_d - \boldsymbol{X}_d\hat{\boldsymbol{\beta}}^{\text{BLUE}}\right), \tag{2.65}$$

$$\hat{\mu}_d^{\text{BLUP}} = t_d(\boldsymbol{\delta}, \boldsymbol{y}) = \boldsymbol{l}_d^\top\hat{\boldsymbol{\beta}}^{\text{BLUE}} + \boldsymbol{m}_d^\top\hat{\boldsymbol{u}}_d^{\text{BLUP}}, \quad d = 1, \dots, D. \tag{2.66}$$

The empirical predictors of μ_d and $\boldsymbol{u}_d$ and the empirical BLUE of $\boldsymbol{\beta}$ are obtained as described in Section 2.4.2, by plugging the ML/REML parameter estimates $\hat{\boldsymbol{\delta}}$ for $\boldsymbol{\delta}$ into the above formulas.

Also the formulas of the MSE in Section 2.4.2 can be re-expressed under a block-diagonal covariance structure as (Rao & Molina, 2015, Section 5.3.1)

$$\mathrm{MSE}\left(\hat{\mu}_d^{\mathrm{BLUP}}\right) = \mathrm{MSE}\left(t_d\left(\boldsymbol{\delta}, \boldsymbol{y}\right)\right) = g_{1d}\left(\boldsymbol{\delta}\right) + g_{2d}\left(\boldsymbol{\delta}\right), \quad d = 1, \dots, D \tag{2.67}$$

with

$$\begin{aligned} g_{1d}\left(\boldsymbol{\delta}\right) &= \boldsymbol{m}_d^\top \left(\boldsymbol{G}_d - \boldsymbol{G}_d \boldsymbol{Z}_d^\top \boldsymbol{V}_d^{-1} \boldsymbol{Z}_d \boldsymbol{G}_d\right) \boldsymbol{m}_d, && (2.68)\\ g_{2d}\left(\boldsymbol{\delta}\right) &= \boldsymbol{d}_d^\top \left(\boldsymbol{X}_d^\top \boldsymbol{V}_d^{-1} \boldsymbol{X}_d\right)^{-1} \boldsymbol{d}_d, && (2.69)\\ \boldsymbol{d}_d^\top &= \boldsymbol{l}_d^\top - \boldsymbol{m}_d^\top \boldsymbol{G}_d \boldsymbol{Z}_d^\top \boldsymbol{V}_d^{-1} \boldsymbol{X}_d. && (2.70) \end{aligned}$$

In Section 2.4.2, we only gave a rather general expression of $\mathrm{MSE}\left(\hat{\mu}^{\mathrm{EBLUP}}\right)$ as

$$\begin{aligned} \mathrm{MSE}\left(\hat{\mu}^{\mathrm{EBLUP}}\right) &= \mathrm{MSE}\left(t\left(\hat{\boldsymbol{\delta}}, \boldsymbol{y}\right)\right) \\ &= \mathrm{MSE}\left(\hat{\mu}^{\mathrm{BLUP}}\right) + \mathrm{E}\left[\hat{\mu}^{\mathrm{EBLUP}} - \hat{\mu}^{\mathrm{BLUP}}\right]^2 \\ &= \mathrm{MSE}\left(\hat{\mu}^{\mathrm{BLUP}}\right) + \mathrm{E}\left[t\left(\hat{\boldsymbol{\delta}}, \boldsymbol{y}\right) - t\left(\boldsymbol{\delta}, \boldsymbol{y}\right)\right]^2 \end{aligned} \tag{2.71}$$

due to the difficulty of expressing term $\mathrm{E}\left[\hat{\mu}^{\mathrm{EBLUP}} - \hat{\mu}^{\mathrm{BLUP}}\right]^2$ for general LMMs. With a block-diagonal covariance structure, the expression simplifies. Building on the work of Kackar and Harville (1984), Prasad and Rao (1990) presented the following MSE approximation for LMMs with block-diagonal covariance structures

$$\mathrm{MSE}\left(\hat{\mu}_d^{\mathrm{EBLUP}}\right) = \mathrm{MSE}\left(t_d\left(\hat{\boldsymbol{\delta}}, \boldsymbol{y}\right)\right) \approx g_{1d}\left(\boldsymbol{\delta}\right) + g_{2d}\left(\boldsymbol{\delta}\right) + g_{3d}\left(\boldsymbol{\delta}\right) \tag{2.72}$$

with

$$g_{3d}\left(\boldsymbol{\delta}\right) = \mathrm{tr}\left(\left(\frac{\partial \boldsymbol{b}_d^\top}{\partial \boldsymbol{\delta}}\right) \boldsymbol{V}_d \left(\frac{\partial \boldsymbol{b}_d^\top}{\partial \boldsymbol{\delta}}\right)^\top \mathrm{E}\left[(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})^\top\right]\right), \tag{2.73}$$

where $\boldsymbol{b}_d^\top = \boldsymbol{m}_d^\top \boldsymbol{G}_d \boldsymbol{Z}_d^\top \boldsymbol{V}_d^{-1}$, $g_{1d}\left(\boldsymbol{\delta}\right)$ and $g_{2d}\left(\boldsymbol{\delta}\right)$ are given in (2.68) and (2.69), and $g_{1d}\left(\boldsymbol{\delta}\right)$ is the leading term. For the later described Fay-Herriot model, the neglected term in the approximation is of order $o\left(D^{-1}\right)$ for large D.

For an estimator of the MSE, we have to take a look at the expectations of the three terms in (2.72) when plugging in ML/REML estimates $\hat{\boldsymbol{\delta}}$ for $\boldsymbol{\delta}$. For this we lean on Datta and Lahiri (2000). The expectations are given by

$$\begin{aligned} \mathrm{E}\left[g_{1d}\left(\hat{\boldsymbol{\delta}}\right)\right] &\approx g_{1d}\left(\boldsymbol{\delta}\right) + g_{3d}\left(\boldsymbol{\delta}\right) && (2.74)\\ \mathrm{E}\left[g_{2d}\left(\hat{\boldsymbol{\delta}}\right)\right] &\approx g_{2d}\left(\boldsymbol{\delta}\right) && (2.75)\\ \mathrm{E}\left[g_{3d}\left(\hat{\boldsymbol{\delta}}\right)\right] &\approx g_{3d}\left(\boldsymbol{\delta}\right), \quad d = 1, \dots, D. && (2.76) \end{aligned}$$

Datta and Lahiri (2000) build on the work of Prasad and Rao (1990), but additionally recognize term

$$b_{\hat{\boldsymbol{\delta}}}\left(\boldsymbol{\delta}\right) = \mathrm{E}\left[\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}\right] \tag{2.77}$$

as the bias of estimating the variance parameter up to order $o\left(D^{-1}\right)$. We denote by $\nabla g_{1d}\left(\boldsymbol{\delta}\right)$ the gradient of term $g_{1d}\left(\boldsymbol{\delta}\right)$ with respect to the components of $\boldsymbol{\delta}$. Then, we have

$$\begin{aligned}\widehat{\mathrm{MSE}}\left(\hat{\mu}_d^{\mathrm{EBLUP}}\right) &= \widehat{\mathrm{MSE}}\left(t_d\left(\hat{\boldsymbol{\delta}}, \boldsymbol{y}\right)\right)\\ &= g_{1d}\left(\hat{\boldsymbol{\delta}}\right) + g_{2d}\left(\hat{\boldsymbol{\delta}}\right) + 2g_{3d}\left(\hat{\boldsymbol{\delta}}\right) - b_{\hat{\boldsymbol{\delta}}}^{\top}\left(\hat{\boldsymbol{\delta}}\right)\nabla g_{1d}\left(\hat{\boldsymbol{\delta}}\right), \quad d = 1, \ldots, D,\end{aligned} \tag{2.78}$$

as an estimator of $\mathrm{MSE}\left(t_d\left(\boldsymbol{\delta}, \boldsymbol{y}\right)\right)$ with $\mathrm{E}\left[\mathrm{MSE}\left(t_d\left(\hat{\boldsymbol{\delta}}, \boldsymbol{y}\right)\right)\right] \approx \mathrm{MSE}\left(t_d\left(\boldsymbol{\delta}, \boldsymbol{y}\right)\right)$.

As shown in Datta and Lahiri (2000, Appendix A.3), the bias term (2.77) is zero when estimating the parameters by REML and should not be neglected when estimating by ML. When $\boldsymbol{\delta}$ is estimated by ML, the bias term is given by (Rao & Molina, 2015, p. 109)

$$b_{\hat{\boldsymbol{\delta}}}^{\mathrm{ML}}\left(\boldsymbol{\delta}\right) = \frac{1}{2D}\left(\left(\boldsymbol{F}^{\mathrm{ML}}(\boldsymbol{\delta})\right)^{-1} \underset{1\leq r\leq q}{\mathrm{col}}\left(\mathrm{tr}\left(\sum_{d=1}^{D}(\boldsymbol{X}_d^{\top}\boldsymbol{V}_d^{-1}\boldsymbol{X}_d)^{-1}\sum_{d=1}^{D}\boldsymbol{X}_d^{\top}\frac{\partial \boldsymbol{V}_d^{-1}}{\partial \delta_r}\boldsymbol{X}_d\right)\right)\right) \tag{2.79}$$

with

$$(\boldsymbol{F}^{\mathrm{ML}}(\boldsymbol{\delta}))_{r,s} = \frac{1}{2}\sum_{d=1}^{D}\mathrm{tr}\left(\boldsymbol{V}_d^{-1}\frac{\partial \boldsymbol{V}_d}{\partial \delta_r}\right)\left(\boldsymbol{V}_d^{-1}\frac{\partial \boldsymbol{V}_d}{\partial \delta_s}\right), \quad r, s = 1, \ldots, q, \tag{2.80}$$

where $(\boldsymbol{F}^{\mathrm{ML}}(\boldsymbol{\delta}))_{r,s}$ refers to the element in the r-th row and s-th column of $(\boldsymbol{F}^{\mathrm{ML}}(\boldsymbol{\delta}))$.

2.4.3 Fay-Herriot model

Model-based approaches to small area estimation

There are two main types of model-based SAE techniques, *unit-level models* and *area-level models*. Following the works of Battese et al. (1988) and Fay and Herriot (1979), their most prominent variants are referred to as the *Battese-Harter-Fuller* (BHF) and the *Fay-Herriot* (FH) model respectively. Both are special cases of LMMs with block-diagonal covariance structure (2.61).

Unit-level models can be calculated when the sample information is available at the unit-level, i.e. at the level of the sampling units. The sample observations are modelled as domain clusters, where each cluster is subject to a realisation of a domain-specific random effect. The model is therefore also called *nested-error regression* model. For a general description of the model, we refer to Rao and Molina (2015, Chapter 7) and Morales et al. (2021, Chapter 7).

In practice, there are many situations where it is not possible to compute unit-level, but only area-level models. Official statistics make micro-data from surveys available to interested users. Methods summarised under the term *disclosure control* are applied to the micro-data before the publication to minimize the risk of involuntary data release. We refer to Templ (2017) and Willenborg and De Waal (2012) for more information on disclosure control. Such disclosure control measures are designed to prevent a micro-data user from being able to identify individuals in surveys and thus obtain sensitive information about them. Therefore, compared to the full sample, publicly available survey micro-data or data provided to researchers on request usually contains only significantly coarsened information. For example, often only a highly coarsened identifier for regional affiliation is provided in the micro-data. As a result, outside the statistical offices it is often not possible to calculate estimates for the small areas of interest with the micro-data provided. The risk of discloser is much smaller for aggregated information than for micro-data. Therefore, statistical agencies typically provide aggregated sample information for much finer domains than identifiers are available in micro-data. These survey aggregates can be used in area-level models.

Even in cases where unit-level sampling information is available for the domains of interest, unit-level model cannot always be calculated. As argued by Morales et al. (2021, Section 16.1), for linear statistics such as domain means and totals, it is sufficient to have domain averages of auxiliary information as input in unit-level models. However, for non-linear statistics, such as most poverty statistics, a Census file with the auxiliary information is needed for unit-level models. That is, the auxiliary information is needed for each unit of the population, not only for the sampling units. This requirement is rarely met in real-world applications. Even though unit-level models cannot be computed in the situations presented, area-level models like the Fay-Herriot model can often be computed and are presented in the following.

Model

The most prominent area-level small area model, the Fay-Herriot model, was introduced by Fay and Herriot (1979). In the following, we present the Fay-Herriot model in accordance with the notation and description in Morales et al. (2021, Section 16.3). Again, consider that population U can be partitioned into D domains of interest, $U_1, \ldots, U_D$. Let μ_d be the characteristic of interest in domain d. For example, we might be interested in a domain mean or total of a variable, or some non-linear poverty measure. Let y_d be the direct estimate of μ_d calculated by using the data of a survey sample.

The Fay-Herriot model is defined in two stages. The first stage specifies the *sampling model*

$$y_d = \mu_d + e_d, \quad d = 1, \ldots, D, \tag{2.81}$$

where $e_d \overset{\text{ind}}{\sim} N(0, \sigma_{ed}^2)$ and σ_{ed}^2 is assumed to be known such that $y_d | \mu_d \overset{\text{ind}}{\sim} N(\mu_d, \sigma_{ed}^2)$. σ_{ed}^2 is the design-based variance of direct estimator y_d. The sampling model implies that the direct survey estimator y_d is unbiased for characteristic μ_d and that its variance σ_{ed}^2 is estimated without error.

The second stage specifies the relationship of the characteristic of interest μ_d to the domain-specific auxiliary information $\boldsymbol{x}_d = (x_{d1}, \ldots, x_{dp})^\top$ via fixed effects $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ and domain-specific random effects u_d. The *linking model* is given by

$$\mu_d = \boldsymbol{x}_d^\top \boldsymbol{\beta} + u_d, \quad d = 1, \ldots, D, \tag{2.82}$$

where $u_d \overset{\text{iid}}{\sim} N(0, \sigma_u^2)$ such that $\mu_d \overset{\text{ind}}{\sim} N\left(\boldsymbol{x}_d^\top \boldsymbol{\beta}, \sigma_u^2\right)$. The random components u_d and e_d are assumed to be independent. The linking model implies that the same linear relationship holds between the μ_d and the p auxiliary variables for all domains and that domain-specific systematic differences are captured in random effects u_d.

Putting the linking and sampling model together, the Fay-Herriot model is given as a specific form of a block-diagonal LMM in the form

$$y_d = \mu_d + e_d = \boldsymbol{x}_d^\top \boldsymbol{\beta} + u_d + e_d, \quad d = 1, \ldots, D, \tag{2.83}$$

with independent u_d and e_d, where $u_d \overset{\text{iid}}{\sim} N(0, \sigma_u^2)$, $e_d \overset{\text{ind}}{\sim} N(0, \sigma_{ed}^2)$. Rao and Molina (2015, Chapter 6) gave a detailed description of how the LMM quantities in (2.61) are specified to obtain (2.83).

Prediction

Rewriting formulas (2.64), (2.65), and (2.66) for the FH model yields

$$\begin{aligned}
\hat{\boldsymbol{\beta}}^{\text{BLUE}} &= \hat{\boldsymbol{\beta}}^{\text{BLUE}}\left(\sigma_u^2\right) = \left(\sum_{d=1}^{D} \frac{\boldsymbol{x}_d \boldsymbol{x}_d^\top}{\sigma_u^2 + \sigma_{ed}^2}\right)^{-1} \left(\sum_{d=1}^{D} \frac{\boldsymbol{x}_d y_d}{\sigma_u^2 + \sigma_{ed}^2}\right), && (2.84)\\
\hat{u}_d^{\text{BLUP}} &= \frac{\sigma_u^2}{\sigma_u^2 + \sigma_{ed}^2}(y_d - \boldsymbol{x}_d^\top \hat{\boldsymbol{\beta}}^{\text{BLUE}}), && (2.85)\\
\hat{\mu}_d^{\text{BLUP}} &= \boldsymbol{x}_d^\top \hat{\boldsymbol{\beta}}^{\text{BLUE}} + \hat{u}_d^{\text{BLUP}} \\
&= \boldsymbol{x}_d^\top \hat{\boldsymbol{\beta}}^{\text{BLUE}} + \frac{\sigma_u^2}{\sigma_u^2 + \sigma_{ed}^2}(y_d - \boldsymbol{x}_d^\top \hat{\boldsymbol{\beta}}^{\text{BLUE}}) \\
&= \gamma_d y_d + (1 - \gamma_d) \underbrace{\boldsymbol{x}_d^\top \hat{\boldsymbol{\beta}}^{\text{BLUE}}}_{= \text{ synthetic predictor}} && (2.86)
\end{aligned}$$

with *shrinkage factor*

$$\gamma_d = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_{ed}^2}, \quad 0 \leq \gamma_d \leq 1. \tag{2.87}$$

The calculation of the FH model only requires knowledge of domain-specific direct estimates y_d with their design-based variance estimates σ_{ed}^2 and a set of domain-specific auxiliary values $\boldsymbol{x}_d$, $d = 1, \ldots, D$. The BLUP formula (2.86) reveals important features of the FH model. The FH BLUP is a convex combination of two estimators, weighted by shrinkage factor γ_d. Thus, it can be referred to as *composite estimator* of the direct estimator y_d and the synthetic predictor $\boldsymbol{x}_d^\top \hat{\boldsymbol{\beta}}^{BLUE}$. The lower the sampling variance σ_{ed}^2 in relation to the model variance σ_u^2 in a domain, the more weight is put on the direct estimator. To put it differently, in domains with precise direct estimators the FH BLUPs are close to the direct estimates, while in domains with imprecise direct estimators the FH BLUPs are close to the domain-specific synthetic predictions. While direct design-based estimators are (asymptotically) unbiased, they can have a large variance, especially under small underlying sample sizes. By contrast, synthetic predictors can be highly biased, but generally have small variances. The FH predictors gives domain-specific MSE-optimal weighted combination of the two estimator. Thereby, they establish a model-based weighting of the *bias-variance trade-off*, which is e.g. discussed in Münnich et al. (2013) for SAE models. Predictor (2.86) is also design-consistent as $\gamma_d \to 1$ for $\sigma_{ed}^2 \to 0$ (Rao & Molina, 2015, p. 125). However, e.g. Jiang and Lahiri (2006, Section 5) argued that design-consistency is a negligible feature in the context of SAE as SAE problems handle survey estimates with small underlying sample sizes.

We note that the FH model can also be motivated by a Bayesian approach and refer to Morales et al. (2021, Section 16.7), Rao and Molina (2015, Sections 9.2, 10.3), Datta et al. (1996), and Jiang and Lahiri (2006, pp. 4–5) for further details on the connection of Bayesian statistics and the FH model.

The empirical versions of the predictors and estimators in the above formulas are obtained by plugging in ML/REML estimator $\hat{\sigma}_u^2$ for σ_u^2. The EBLUE of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}^{\text{BLUE}}\left(\hat{\sigma}_u^2\right) = \left(\sum_{d=1}^{D} \frac{\boldsymbol{x}_d \boldsymbol{x}_d^\top}{\hat{\sigma}_u^2 + \sigma_{ed}^2}\right)^{-1} \left(\sum_{d=1}^{D} \frac{\boldsymbol{x}_d y_d}{\hat{\sigma}_u^2 + \sigma_{ed}^2}\right). \tag{2.88}$$

Equivalently, the EBLUPs of u_d and μ_d are given by

$$\hat{u}^{\text{EBLUP}} = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \sigma_{ed}^2}(y_d - \boldsymbol{x}_d^\top \hat{\boldsymbol{\beta}}) \tag{2.89}$$

$$\hat{\mu}_d^{\text{EBLUP}} = \underbrace{\boldsymbol{x}_d^\top \hat{\boldsymbol{\beta}}^{\text{BLUE}}}_{=\text{ synthetic predictor}} + \hat{u}_d^{\text{EBLUP}} \tag{2.90}$$

Mean squared error

In Section 2.4.2, we shortly discussed the Datta and Lahiri (2000) and Prasad and Rao (1990) approximation to the MSE of the BLUP for block-diagonal LMMs. For the FH model and parameter estimation via ML/REML, the formulas reduce to (Rao & Molina, 2015, Sections 6.1, 6.2)

$$\text{MSE}\left(\hat{\mu}_d^{\text{EBLUP}}\right) \approx g_{1d}\left(\sigma_u^2\right) + g_{2d}\left(\sigma_u^2\right) + g_{3d}\left(\sigma_u^2\right) \tag{2.91}$$

with

$$g_{1d}\left(\sigma_u^2\right) = \frac{\sigma_u^2 \sigma_{ed}^2}{\sigma_u^2 + \sigma_{ed}^2} = \gamma_d \sigma_{ed}^2, \tag{2.92}$$

$$g_{2d}\left(\sigma_u^2\right) = (1-\gamma_d)^2 \boldsymbol{x}_d^\top \left(\sum_{d=1}^{D} \frac{\boldsymbol{x}_d \boldsymbol{x}_d^\top}{\sigma_{ed}^2 + \sigma_u^2}\right)^{-1} \boldsymbol{x}_d, \tag{2.93}$$

$$g_{3d}\left(\sigma_u^2\right) = \frac{\sigma_{ed}^4}{\left(\sigma_u^2 + \sigma_{ed}^2\right)^3} \text{AVar}\left(\hat{\sigma}_u^2\right). \tag{2.94}$$

and

$$\mathrm{AVar}\left(\hat{\sigma}_u^{2,ML}\right) = \mathrm{AVar}\left(\hat{\sigma}_u^{2,REML}\right) = 2\left(\sum_{d=1}^{D}\left(\sigma_u^2+\sigma_{ed}^2\right)^{-2}\right)^{-1} \tag{2.95}$$

as the asymptotic variances of the ML and REML estimators for σ_u^2. Then, formula (2.78) becomes

$$\begin{aligned}\widehat{\mathrm{MSE}}\left(\hat{\mu}_d^{\mathrm{EBLUP}}\right) &= \widehat{\mathrm{MSE}}\left(t_d\left(\hat{\sigma}_u^2, \boldsymbol{y}\right)\right)\\ &= g_{1d}\left(\hat{\sigma}_u^2\right) + g_{1d}\left(\hat{\sigma}_u^2\right) + 2g_{3d}\left(\hat{\sigma}_u^2\right) - b_{\hat{\sigma}_u^2}^{\top}\left(\hat{\sigma}_u^2\right)\nabla g_{1d}\left(\hat{\sigma}_u^2\right).\end{aligned} \tag{2.96}$$

As we saw in Section 2.4.2, the additional bias correction term is zero for REML and non-zero when estimating the variance components by ML. In the FH model, for ML we have (Datta & Lahiri, 2000, Remark 1)

$$\nabla g_1\left(\sigma_u^2\right) = \left(\frac{\sigma_{ed}^2}{\sigma_u^2+\sigma_{ed}^2}\right)^2 > 0 \tag{2.97}$$

and

$$b_{\hat{\sigma}_u^{2,ML}}\left(\sigma_u^2\right) = -\operatorname{tr}\left(\left(\sum_{d=1}^{D}\frac{\boldsymbol{x}_d\boldsymbol{x}_d^{\top}}{\sigma_u^2+\sigma_{ed}^2}\right)^{-1}\left(\sum_{d=1}^{D}\frac{\boldsymbol{x}_d\boldsymbol{x}_d^{\top}}{\left(\sigma_u^2+\sigma_{ed}^2\right)^2}\right)\right)/\sum_{d=1}^{D}\left(\sigma_u^2+\sigma_{ed}^2\right)^{-2}. \tag{2.98}$$

In Chapter 5, we present the multivariate version of the FH model and investigate the prediction of non-linear domain indicators. In Chapter 6, we extend the multivariate FH model for missing direct estimates.

Chapter 3

Generation of a Longitudinal Employment Dataset for Simulations

3.1 Introduction

The design-based properties of estimators for a specific survey application can be approximated by design-based simulation studies. In such studies, a chosen simulation dataset is treated as the simulation population. In each iteration of the simulation, a sample is drawn from that population according to the sampling design of the specific application. To each sample, different estimators are applied. The design-based properties of these estimators are then inferred over all simulated samples. For more details, we refer to the descriptions of Monte Carlo studies in Section 2.2.4. With such a simulation study, researchers can for example examine the performance of the GREG estimator, presented in Section 2.3.2, under different sets of auxiliary variables and for different target statistics.

Several aspect have to be considered for choosing a simulation dataset for such design-based simulation study. The simulation dataset should contain all the information necessary to adequately simulate the possibly complex sampling design of interest, e.g. with all regional levels. Furthermore, it is crucial that the joint distributions of the variables in the simulation dataset adequately represent their distributions in the target population, i.e. the population for which the study is conducted. Using the GREG example from earlier, we see from the approximated GREG variance (2.23) that the correlation of the target variable with the auxiliary variables plays a crucial role for evaluating the efficiency of the GREG estimator under different set of auxiliary variables.

In order to meet the requirements of a simulation dataset tailored to a specific research question, it is often necessary to modify or extend existing datasets or even create a completely new dataset. In this chapter, we create a simulation dataset tailored to the design-based simulation studies in Chapter 4. In Chapter 4, we evaluate the application composite estimators for the production of employment statistics in the German Microcensus. In order to define the requirements for a simulation dataset for this study, the following information is relevant.

The target population of the German Microcensus is the German resident population, which is structured in strata. In each strata, the sampling units are constituted by clusters of persons in households. For a detailed description of the Microcensus design, we refer to Destatis (2020b); a description of the design with focus on its changes in 2020 and the rotation pattern is given in Section 4.3. The focus of the studies in Chapter 4 is on the estimation of employment statistics by composite estimators, which make use of the rotational design of the Microcensus. The performance of the estimators is significantly affected by the rotation pattern, which is determined by the sampling design, and by the dependencies of the person-specific employment categories over time.

For the simulation dataset for the studies in Chapter 4 we therefore set the following requirements. The dataset should contain person-level information and adequately represent the German resident population with its person and regional structure. In particular, the regional information in the dataset should be fine enough so that the sampling design of the German Microcensus can be simulated in its full regional depth and employment estimators can be calculated down to the NUTS2-level. Since the research focus in Chapter 4 is on estimation procedures which exploit the rotational design of the Microcensus for the production of employment statistics, the dataset must include longitudinal employment categories. More precisely, a longitudinal dataset is needed that contains a monthly employment status for each person in the data for a predefined sequence of months. The employment status should be defined in accordance with the definition of the International Labour Organization (ILO) as this is the definition used in the Microcensus. It refers to individuals being assigned to one of three categories: Employed, unemployed, or not in labour force. At the person level, the monthly employment categories should have realistic patterns since these patterns play an important role in the performance of the estimation methods evaluated in Chapter 4. For example, an employed person may be more likely to be employed in the following month than an unemployed person. At the aggregate level, the longitudinal employment information should reflect the seasonal patterns and trends evident in statistics from the German national statistics institute.

There are two datasets available for generating the simulation dataset for Chapter 4, the RIFOSS dataset, presented in Section 3.3.1, and the SIAB dataset, presented in Section 3.3.2. The RIFOSS dataset is only cross-sectional, but otherwise meets the requirements for the simulation dataset. The SIAB dataset contains register information on longitudinal employment categories, but otherwise does not meet the criteria for the simulation dataset. For example, it do not cover the complete German resident population, does not contain any regional information, and the longitudinal employment categories from the registers are not in accordance to the ILO definition. In this chapter, we aim at overcoming the disadvantages of both datasets by combining them with a model-based approach. For data security reasons, it is not allowed to match the SIAB dataset with other data. Therefore, only prediction models can be used to transfer the longitudinal employment information in the SIAB dataset to the RIFOSS dataset.

The chapter is organised as follows. Section 3.2 provides a brief description of employment categories in terms of the ILO definition. The RIFOSS and SIAB dataset are described in Section 3.3. In Section 3.4, we present how we edit the SIAB dataset such that a consistent longitudinal dataset with monthly employment categories is obtained. This includes the derivation of the employment status according to the ILO definition from the SIAB register information and the validation of it. Based on the processed SIAB dataset, we can fit prediction models for monthly employment transitions. In Section 3.5, we give a short introduction to generalised additive models and ensemble methods. Furthermore, we discuss evaluation criteria for probability predictions and present an extension of the Brier score, called weighted Brier score, tailored to imbalanced categorical data like the employment status. We show that a combination of the ensemble methods subagging and stacking with the proposed Brier score as a loss function in stacking represents quadratic programming. In Section 3.6, we apply the subagging-stacking combination with generalised additive models and the proposed weighted Brier score to model the monthly employment transitions in the SIAB dataset. With the models, we extend the RIFOSS dataset with monthly employment information. We validate the generated data with the original SIAB dataset on the person- and aggregated level and with aggregate statistics from Destatis, the German national statistical institute. Section 3.7 provides a summary and an outlook.

3.2 ILO employment status

Throughout this chapter and Chapter 4, the employment status is defined according to the concept of the ILO. The ILO sets guidelines to provide a general definition of employment to achieve international comparability of employment statistics. In order to obtain comparable labour market statistics within the countries of the European Union (EU), the European Commission sets standards for LFS statistics based on the ILO definition, summarised e.g. in European Commission (2016). In specific EU countries, there can be divergences from these recommendations. In this chapter and Chapter 4, the focus is on the German LFS, which is integrated in the German Microcensus. The employment definition used in the German Microcensus can be found in Destatis (2020b, Section 2.1.3). Rengers (2004) provided a detailed overview of the application of the ILO concept in Germany and its historical development.

Table 3.1 presents an overview of the employment categories. It is based on Table 1 in Rengers (2004). There are three categories of ILO employment: Employed, unemployed, and not in labour force (LF). The definitions of the categories are rather broad and formulated in such a way that they allow for comparable statistics despite country-specific definitions of employed and unemployed persons, e.g. for national accounts. The employed and unemployed persons build the LF, which is also referred to as the active population. If a person neither qualifies as employed nor unemployed it is considered not in LF. Unemployed persons and those not in LF build the non-employed population.

In Germany, there are different definitions and statistics on employment. Next to statistics based on the Microcensus, with included LFS, based on the ILO definition of employment, there are also the employment statistics of the Federal Employment Agency. Fritsch and Lüken (2004), Hartmann and Riede (2005), and Körner and Marder-Puch (2015) summarised the differences between the two statistics. The different definitions are the reason why the employment status according to ILO can only be approximated from the employment information available in the SIAB dataset, which is described in Section 3.4.

Table 3.1: ILO concept of employment, based on Rengers (2004, Table 1)

Labour force (LF) (Currently active population)		Not in LF (Not currently active population)
Employed Persons aged 15 and over which are working in a formal employment relationship with at least one hour per week, even if this was temporarily not exercised during the reporting period (e.g. due to vacation or illness) or self-employed persons or freelancers or soldiers/civilian service members of unpaid family workers or apprentices.	**Unemployed** Persons aged $15 - 74$ which are not employed during the reference week and actively looking for a job in the four weeks prior to the interview (the temporal scope of the activity sought is irrelevant) and can start a new job within two weeks. The involvement of an employment agency for employment or a municipal agency in the search efforts is not required.	**Not in LF** Persons aged 15 and over which are neither considered employed nor considered unemployed.
Employed	Non-employed	

3.3 Data description

3.3.1 RIFOSS dataset

The RIFOSS dataset[1] is a cross-sectional semi-synthetic dataset which was generated at the Economic and Social Statistics department at Trier University. It reflects the German person- and household-level population and contains detailed regional information. It allows to mimic large-scale household surveys such as the German Microcensus in design-based simulation studies.

The generation of the RIFOSS dataset took place in two phases. Münnich et al. (2012a) investigated methodologies for the German Census 2011 and generated a simulation dataset for that purpose. The simulated population was built using anonymized population register data for all of Germany, 2006 Microcensus information, anonymized material from Census tests, and postal code information. Kolb (2013, Section 5.1) and Münnich et al. (2012a, Section 3.1) provided detailed descriptions of the data generation. The aim of the generation was to provide a dataset that reflects the target population of the German Microcensus while preserving the heterogeneity of the data, e.g. between administrative units. In the course of the RIFOSS[2] project, funded by the federal statistical office of Germany,

[1] Version RIFOSS_GG_v0.1.1_vanilla_ice_cream.

[2] RIFOSS: Research Innovation for Official and Survey Statistics.

the dataset was further enriched by 2008 Microcensus Scientific Use File[3] information. Categorical variables were generated using multinomial logistic regression models, further information on these models is given in Section 3.5. The data were aligned to match results of the German Census 2011[4]. A description of the German Census 2011 is given in Statistische Ämter des Bundes und der Länder (2015). For the alignments, calibration methods (Deville & Särndal, 1992) and heuristic optimisation methods such as simulated annealing (van Laarhoven & Aarts, 1987) were used. The RIFOSS dataset was used for the analyses in Rupp (2018) and, in an adjusted form, in Burgard et al. (2020c).

The RIFOSS dataset consists of about 85 million persons, thereof 82.5 million persons at their main residence, which are grouped in about 38 million households. Table 3.2 lists those RIFOSS variables which we use in the further applications to the data in this chapter as well as in Chapter 4. We note that for the extension of the RIFOSS dataset with monthly employment categories, we can only use those variables that are present in both the RIFOSS and the SIAB dataset. The RIFOSS variables are defined in accordance to the Microcensus 2008 Scientific Use File (Destatis & GESIS, 2012); an English description of the labels is given in Research Data Centres of the Statistical Offices of the Federation and the Federal States (2018). To clarify the definitions, the German expressions of some of the variables are added in brackets in Table 3.2. Compared to the original RIFOSS dataset, we reduced variable `EF29`, which corresponds to the employment status according to ILO, to the three ILO categories given in Table 3.1. Table 3.3 displays the absolute and relative frequencies of the employment categories in the RIFOSS dataset. Variable `AGS` is a regional identifier, indicating the federal state (NUTS1-level), government district (NUTS2-level), NUTS3-level, and the municipality identifier. For details on the regional levels of Germany, we refer to Destatis (2020a). Variables `AGS`, `EF3`, `HID`, `EF30`, and `EF570` are not of further concern in this chapter, but used for the implementation of the Microcensus sampling design in Chapter 4.

[3] Available from https://www.forschungsdatenzentrum.de/en/household/microcensus.

[4] Available from https://www.zensus2011.de/DE/Home/Aktuelles/DemografischeGrunddaten.html. Table *Bevölkerung im 100 Meter-Gitter*.

Table 3.2: RIFOSS: Chosen variables

Variable	Information
AGS	Official municipality key (Amtlicher Gemeindeschlüssel)
EF3	Number of district (Auswahlbezirksnummer)
HID	Household identifier
EF29	Employment type
EF30	Population at primary residence
EF44	Age (in years)
EF46	Sex
EF310	Highest school-leaving qualification
EF312	Highest vocational qualification
EF540	Highest grade of educational or vocational training (ISCED97)
EF570	Classified building sizes (building layers) (Anschriftengrößenklasse)

Table 3.3: RIFOSS: Absolute and relative employment frequencies

Persons	Employed	Unemployed	Not in LF
All ages	$43,296,693$ (50.95%)	$2,335,098$ (2.75%)	$39,352,060$ (46.31%)
Aged 15 and over	$43,296,693$ (58.47%)	$2,335,098$ (3.15%)	$28,415,818$ (38.38%)

3.3.2 SIAB dataset

The SIAB dataset is the regional file of the Sample of Integrated Labour Market Biographies 1975-2017 (SIAB-R7517). It is a factually anonymous scientific use file of a 2% sample of the Integrated Employment Biographies (IEB) of the Institute for Employment Research (IAB). The dataset is longitudinal and consists of about 62 million data rows from about 1.8 million persons. The IEB is an administrative dataset comprising all individuals in Germany officially recognized by different administrative sources during a respective time period. A detailed description of the data and its variables is given in Antoni et al. (2019), which is the main source for the following description.

Table 3.4 displays those variables of the SIAB dataset, which are used in the following. Their outcomes and relative frequencies are given in Antoni et al. (2019). The SIAB dataset bundles information from different administrative data sources. Variable source_gr indicates from which data source the information of a specific data row originates. The different data sources are listed in Table 3.5, including the number of data rows associated with each. The person-specific identifier variable persnr enables the matching of a person's data rows in the different data sources. The different data sources and their information are specific to German administrative procedures. We refer to Antoni et al. (2019) for further information.

Variables begepi and endepi are day-specific and together mark a day-specific time interval for each data row. One data row in the SIAB dataset corresponds to an observation of a

Table 3.4: SIAB: Chosen variables

Variable	Information
`persnr`	Individual ID
`quelle_gr`	Source of spell, grouped
`begepi`	Episode start date
`endepi`	Episode end date
`frau`	Gender
`gebjahr`	Year of birth
`ausbildung_gr`	Vocational training, grouped
`ausbildung_imp`	Vocational training, imputed
`schule`	School leaving qualification
`tentgelt_gr`	Daily wage/daily benefit
`teilzeit`	Part-time
`erwstat_gr`	Employment status, grouped

Table 3.5: SIAB: Data sources

Data source (`quelle_gr`)		Information	Nr. of data rows
BeH	Employee History	Employment subject to social security and marginal part-time employment	$38,710,742$
LeH	Benefit Recipient History	Receipt of benefits in accordance with Social Code Book (SGB) III	$6,483,432$
LHG	Unemployment Benefit II Recipient History	Receipt of benefits in accordance with SGB II	$4,225,937$
(X)MTH	Participants-in-Measures History Files	Participation in employment and training measures	$1,785,856$
(X)ASU	Jobseeker Histories	Periods of job search recorded by the Federal Employment Agency or by municipal institutions responsible for implementing SGB II	$11,134,554$

person (`persnr`) in a day-specific period of time (`begepi`, `endepi`) from a specific data source (`quelle_gr`), with data-source specific information, for example on the employment status `erwstat_gr`. It appears that for one person and period of time there can be multiple data rows from different data sources. For example, there could be two data rows from BeH, one row for a full-time and another for a part-time employment, and one data row from (X)ASU. The age of a person can be approximated using their year of birth and the time interval of the respective data row. In the data there are persons aged 17 to 62 years. Next to the variables in Table 3.4, there are many more variables available in the SIAB dataset. However, only those are of interest, which coincide with the RIFOSS variables of Table 3.2 and can be used to model monthly employment transitions.

The SIAB dataset has some features that are particularly relevant to the goal of extending the RIFOSS dataset by monthly employment information with prediction models calculated on the SIAB dataset. (1) The information in the SIAB dataset comes from administrative sources. This means that, in contrast to the RIFOSS dataset, the German resident population is not fully covered in the SIAB dataset. (2) In the SIAB dataset, some information is only available for observations corresponding to specific data sources. For example, variable `schule` is partially available for BeH entries, almost completely available for (X)ASU and (X)MTH entries and not available for LeH and LHG entries, see Antoni et al. (2019, Table 4). (3) There are also inconsistencies in the data due to the way data is collected by different sources. For example, in the dataset it can appear that the education status of a person is decreasing over time. (4) Each row of the SIAB dataset contains information about a person, in a specific time period, from a specific data source. Several data rows from the same person can refer to the same period of time. In order for the dataset to be used to model monthly employment transitions, the data must be aggregated in time so that there is only one data row per person per month, including an employment status according to ILO. (5) There are variables in the SIAB dataset that are related to a person's working life. However, there is no employment status in the data that corresponds to the ILO definition in Table 3.1. We refer to Hartmann and Riede (2005) and Körner and Marder-Puch (2015) for an overview of the differences between the employment definition according to ILO used in the Microcensus and the employment information collected by the Federal Employment Agency which applies to the SIAB dataset.

Because of these features of the SIAB dataset, in the next Section 3.4, we impute missing values and edit the data to have consistent person-specific information. In addition, we aggregate the data to have one data row per person and month, derive a monthly employment status according to the ILO definition, and evaluate whether the resulting employment data reflects realistic employment transitions and seasonal patterns.

3.4 Derivation of a monthly employment status in the SIAB dataset

3.4.1 Data editing

For some of the SIAB dataset sources there is no data available for certain time intervals, compare Antoni et al. (2019, Table 2). Furthermore, there are incomplete observations in data source LHG until the beginning of 2007 (Antoni et al., 2019, p. 18). Therefore, we restrict the dataset to data rows with begepi referring to 2007-2017.

There are missing values in the different variables of the SIAB dataset. For the variables with missing values, Table 3.6 shows the percentage of missings, both in the complete dataset (second column) and for each data source. We see that the occurrence of missing values is closely related to the data sources from which a data rows originates. For example, for the school leaving qualification variable schule, there is no information available in data sources LeH and LHG. For variable frau, there 11 missing values which are not visible from the percentages in Table 3.6. Next to missing values, there are also inconsistencies in the dataset. For example, there are individuals in the data for whom the level of education decreases over time.

Table 3.6: SIAB: Missing values (in %) of chosen variables per data source

Variable	SIAB	BeH	LeH	LHG	(X)MTH	(X)ASU
frau	0.00	0.00	0.00	0.00	0.00	0.00
ausbildung_gr	42.95	38.22	100.00	100.00	15.13	12.35
ausbildung_imp	52.04	4.67	100.00	100.00	100.00	100.00
schule	40.83	35.54	100.00	100.00	8.67	10.37

Note: For variable frau there are 11 missing values
(thereof 2, 6, and 3 for LHG, (X)MTH, and (X)ASU respectively)

To reduce the percentages of missing values and potential data inconsistencies, we apply Algorithm 3.1. The algorithm ensures that for each person in the dataset the nominal variables frau and gebjahr take only one unique value and missing values are filled by the person-specific modus of the variable. For the ordinal education variables ausbildung_gr, ausbildung_imp, and schule, the algorithm ensures that missing values are filled by other information of the same person. Variable begepi, the episode start date, is used to determine the chronological order of the observations. From Table 3.6, we see that some variables are data source specific. For example, variable ausbildung_imp is only available for BeH entries. By the algorithm, if a person has at least one BeH entry where ausbildung_imp is available, this value is used to impute missing values of ausbildung_imp in other data rows of that person. Furthermore, the algorithm ensures time consistency of the ordinal variables. That is, their values can only increase over time.

Table 3.7 shows the percentages of missing values in the SIAB dataset before and after applying Algorithm 3.1. From the table, we see that the application of Algorithm 3.1 significantly reduced the proportion of missing values.

Algorithm 3.1 SIAB: Imputation of missing values, correction of data inconsistencies

```
For each person $p$ in the SIAB dataset do
    1. Take all data rows corresponding to person $p$, denote them as $U_p$, and proceed with $U_p$.
    2. For nominal variable $j \in \{\texttt{frau}, \texttt{gebjahr}\}$ do
        • Set all values of variable $j$ to the modus of variable $j$.
    3. For ordinal variable $j \in \{\texttt{ausbildung\_gr}, \texttt{ausbildung\_imp}, \texttt{schule}\}$ do
        • Fill missing values of variable $j$ by the nearest previously observed non-missing value of variable $j$.
        • Starting from oldest to newest entry in $U_p$, according to begepi: If the value of variable $j$ is lower than the nearest previous value of variable $j$, set it to the nearest previously value of variable $j$.
```

Table 3.7: SIAB: Missing values (in %) before and after applying Algorithm 3.1

Variable	Before imputation	After imputation
`frau`	0.00	0.00
`ausbildung_gr`	42.95	2.57
`ausbildung_imp`	52.04	9.59
`schule`	40.83	2.47

Note: For variable `frau` there are 11 missing values before and after imputation.

3.4.2 Aggregation to monthly data

In the original SIAB dataset, each data row refers to a specific person, day-specific time interval in 2007-2017, and data source. Thereby, for a person there may be several data rows referring to at least one day of a specific month. For modelling monthly employment transitions, the SIAB dataset needs to be aggregated such that there is only one data row per person and month from January 2007 to December 2017. The variables to be included in the aggregated SIAB dataset are listed in Table 3.8, where the variable `ym` indicates the year and month a data row refers to.

Algorithm 3.2 is applied to get from the non-aggregated to the aggregated SIAB dataset, which contains the variables listed in Table 3.8. For each person and month, the algorithm chooses the data rows which cover at least one day of that month for the chosen person, by use of variables `persnr`, `begepi`, and `endepi`. Since for a person, there can be several data rows representing different numbers of days in a month, only those data rows that represent the largest number of days in a month are selected. For example, a person could

Table 3.8: SIAB: Variables in aggregated data with one data row per person and month

Variable	Information
`persnr`	Individual ID
`ym`	Year-month
`quelle_gr`	Source of spell, grouped
`frau`	Gender
`age`	Age (year in `ym` − `gebjahr`)
`ausbildung_gr`	Vocational training, grouped
`ausbildung_imp`	Vocational training, imputed
`schule`	School leaving qualification
`erwstat_gr`	Employment status, grouped
`ILO`	ILO employment status

be employed for 25 days of a month with a corresponding BeH entry and unemployed for 5 days of the same month with a corresponding (X)ASU entry. Then, only the BeH information referring to 25 days of a month is kept, all other information is deleted.

It is possible that there are data rows from different data sources referring to the same period of time and same person. For this case, decision rules are defined in the algorithm to determine which of the information should be deleted and which should be retained. From the ILO definition in Table 3.1, it is clear that information indicating employment is more relevant for determining the ILO employment status than e.g. possible additional job searches of a person. A BeH data row indicating employment specifies that the corresponding person should be classified as employed as defined by the ILO for the time interval to which the data row refers. Therefore, information from the BeH data source is given priority. Following the same logic, there are decision rules in the algorithm that determine which information is retained when there are multiple data rows from the same data source. Assume that for a person there are several data rows, each representing all days of a chosen month, including one BeH entry for a full-time employment, an additional BeH entry for a part-time employment, and an (X)ASU entry. With the decision rules set in the algorithm, only the BeH data row for the full-time employment is kept, all other data rows are removed. After the application of the algorithm, only one row of data per person and month remains.

We would like to draw particular attention to the last piece of the algorithm. It deals with those months in 2007-2017 for which no information is available for a person in the SIAB dataset. When there is no record of an individual in BeH, LeH, LHG, (X)MTH, or (X)ASU in a month, we interpret this as information in itself. From the information of the construction of the SIAB dataset, we conclude that in the months without a SIAB record a person is likely to be classified as not in LF (instead of employed or unemployed) according to the ILO definition. Therefore, also for the months in 2007-2017 with no record of a person in the original SIAB dataset, we add data rows to the aggregated SIAB dataset. The person-specific information for these months, including the year of birth, gender, and education related information, is imputed by the person-specific information of other months. After applying Algorithm 3.2, in the aggregated SIAB dataset there are 132 data rows for each person, one for each month in January 2007 to December 2017.

From the aggregated data, variable `age` in Table 3.8 is calculated as `ym-gebjahr`. In the original SIAB dataset there were only persons aged $17-62$. Therefore, the aggregated data is restricted to rows with `age` in $17-62$. Observations with missing sex (`frau`) are removed from the dataset.

Algorithm 3.2 SIAB: Aggregation to one data row per person and month

```
For each person $p$ in the SIAB dataset do
1. Take all data rows corresponding to person $p$, denote them as $U_p$, and proceed with $U_p$.
2. For all months $m \in \{2007\text{-}01, \dots, 2017\text{-}12\}$, for which at least one day is covered in $U_p$, do
   a) Take all data rows covering at least one day of month $m$, denote them as $U_{pm}$, and proceed with $U_{pm}$.
   b) Keep only those data rows covering the highest available number of days in month $m$.
   c) If the remaining data rows correspond to multiple data sources (quelle_gr), then keep only the rows that belong to the highest prioritized available source. The order from highest to lowest is: BeH, (X)MTH, (X)ASU, LeH, LHG.
   d) If there is more than one data row corresponding to BeH, then
      - If there is at least one data row with teilzeit $= 0$ (full-time), then keep only these data rows.
      - If there is at least one data row with erwstat_gr $= 1$ (employees subject to social security), then keep only these data rows.
      - Keep only data rows with highest value of tentgelt_gr.
   e) If there is more than one data row corresponding to (X)MTH, then
      - If there is at least one data row with erwstat_gr $= 42$ (start an employment), then keep only these data rows.
      - If there is at least one data row with erwstat_gr $= 43$ (career choice and vocation), then keep only these data rows.
      - If there is at least one data row with erwstat_gr $= 44$ (employment-generating measure), then keep only these data rows.
   f) Keep only data rows with lowest number of missing values.
   g) Keep only the first data row.
   h) There remains one data row for person $p$ in month $m$. With this data row, fill in the values of the variables in Table 3.8, missing values are set to $NA$.
3. For months $m \in \{2007\text{-}01, \dots, 2017\text{-}12\}$ for which no day is covered in $U_p$ do
   - Create a data row for person $p$ in month $m$ with variables in Table 3.8, missing values set to $NA$.
   - Find the nearest previously observed month $m^* \in \{2007\text{-}01, \dots, 2017\text{-}12\}$ for person $p$.
   - Fill variables frau, gebjahr, ausbildung_gr, ausbildung_imp, schule by the values observed for person $p$ in month $m^*$.
```

3.4.3 Derivation of employment status according to ILO

We apply Algorithm 3.3 to derive a monthly employment status according to the ILO definition from the aggregated SIAB dataset. The decision rules in the algorithm are defined based on a comparison of the ILO employment definition in Table 3.1 and the available SIAB information. In the first step, it is checked whether there are any indications of employment. If not, it is checked whether there are indications that a person is unemployed. If neither is true, a person in a specific month is categorized as not in LF.

Algorithm 3.3 SIAB: Derivation of ILO employment variable `ILO`

```
Define variable ILO with categories
    1 Employed,
    2 Unemployed,
    3 Not in LF,
by applying the following rules
    • Set ILO = 3.
    • If quelle_gr = $BeH$ then set ILO = 1.
    • If quelle_gr = $(X)MTH$ & erwstat_gr $\in \{42, 44\}$ then set ILO = 1
      (42: Start an employment, 44: Employment-generating measures).
    • If ILO $\neq 1$ & quelle_gr = $(X)ASU$ & erwstat_gr = 21 then set ILO = 2
      (21: Unemployed, registered as a job seeker (ALO)).
```

Figure 3.1 shows the monthly aggregates of the created monthly employment categories in the SIAB dataset. We can see both seasonal patterns and a trend in the data, e.g. number of employed persons increases over time. The absolute and relative frequencies of categories of the derived variable `ILO` are displayed in Table 3.9.

Table 3.9: SIAB: Absolute (relative) frequencies of variable `ILO`

Employed	Unemployed	Not in LF
84, 056, 668 (63.64%)	6, 247, 089 (4.73%)	41, 781, 047 (31.63%)

In the description of the SIAB dataset in Section 3.3.2, we already see that it covers only certain parts of the target population of the German Microcensus, which is the German resident population. In addition, variable `ILO` could only be inferred to some extent from the available information in the dataset. In the following, we therefore address the question of how well the SIAB dataset covers the target population and how well the employment status according to ILO could be derived from the available information. We conduct this analysis first in terms of content and then with a comparison to aggregated data from official statistics.

Using the information on the SIAB dataset in Antoni et al. (2019) and the employment definitions in Section 3.2, we can make some statements about how well the SIAB dataset represents the German resident population and information according to the ILO employment status. The SIAB dataset covers persons employed subject to social

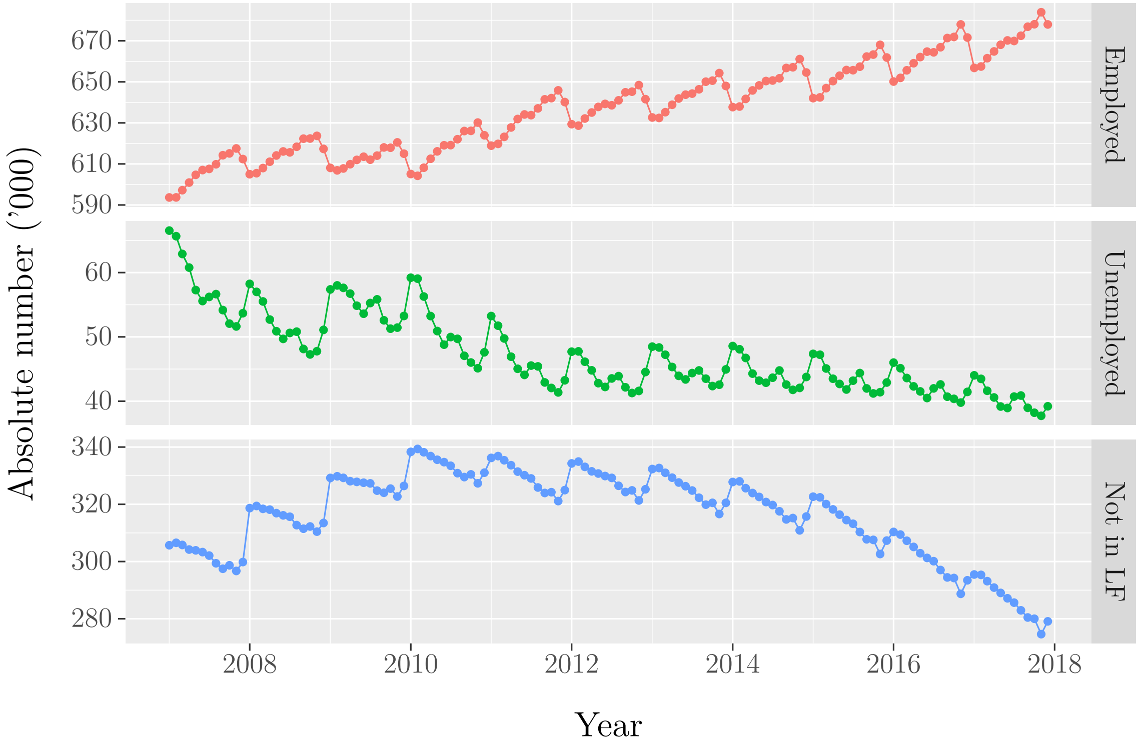

Figure 3.1: Monthly aggregates of variable `ILO`

security, including trainees and interns, in the BeH data source. Individuals which are self-employed, freelancers, soldiers, civil servants, family workers, or work, but not subject to social security, such as mini-jobbers, are not covered in the BeH data source. Therefore, those population groups are likely to be missing completely in the SIAB dataset and there is an under-coverage of the employed persons of the target population. Destatis (2018, p. 378) report that about 84% of all employed persons are either employed subject to social security or marginally part-time employed. Therefore, we assume that most of the employed persons are represented in the SIAB data and variable `ILO` still captures the large amount of employed persons.

In the SIAB dataset, we only classify persons as unemployed when they have a corresponding (X)ASU record for being registered unemployed. All other persons which are out of work, currently available for work, and seeking work are not covered in the SIAB dataset. Hence, it is plausible that there is a large under-coverage of unemployed persons in the SIAB dataset.

We only classify persons as not in LF in the SIAB dataset when they are neither considered employed nor unemployed. Large parts of the German resident population, which are categorized as not in LF, do not appear in the SIAB dataset as they are not captured by the different data sources. To be more specific, all persons who are not employed, not registered at the unemployment agency, and not receiving any benefits subject to the social code book are not included in the SIAB dataset. Hence, we expect a large under-coverage of persons not in LF in the SIAB dataset. Furthermore, in SIAB there are only persons aged $17 - 62$. All persons aged 16 and younger or 63 and over are not covered at all in the dataset.

To see whether the aggregates of variable `ILO` are realistic despite the coverage problems of the SIAB dataset, we compare them to external information. For that, we use aggregated labour market statistics, released by Destatis, the German national statistical institute. For analysis purposes, it is important to note that the SIAB dataset include only individuals aged 17-62, so age differences must be carefully considered when making comparisons. Furthermore, only relative frequencies can be used for the comparison.

Table 3.10 shows the relative frequencies of the ILO categories for the SIAB dataset and Destatis information (Destatis, 2018, Section 13.1.2), each for 2017. The relative frequencies of the categories differ substantially between the two data sources. This is largely attributable to the different age structures underlying the aggregates. The Destatis aggregates refer to the complete population in 2017, i.e. about 82 million persons. In the SIAB dataset, only persons aged $17 - 62$ are covered, wherefore the percentage of persons not in LF is expected to be substantially lower in the SIAB dataset. Persons aged 14 and younger are considered to be not in LF. Those persons are completely missing in the SIAB dataset, but appear in the Destatis numbers in Table 3.10. Similarly, also old persons, which are likely to be categorized as not in LF, are not covered in the SIAB dataset.

Table 3.10: Relative frequencies (in %) of ILO employment categories in 2017

	Employed	Unemployed	Not in LF
Destatis	50.94	1.98	47.07
SIAB	67.23	4.04	28.72

We take a look at additional statistics for specific age and person groups to get further insight into the coverage of SIAB variable `ILO`. Table 3.11 presents the relative frequencies of both sexes for the three employment categories. Destatis aggregates are from Destatis (2018, Section 13.1.3). For both Destatis and SIAB aggregates and all three employment categories, the percentage of males in employed and unemployed persons is higher than the percentage of females. The sex proportions are quite similar in the SIAB and Destatis data. Only for persons not in LF the percentages differ. In the Destatis publications, the percentage of females is about 55%, whereas only about 50% of the persons not in LF are females in the SIAB dataset.

Table 3.11: Relative frequencies (in %) of sexes for the employment categories in 2017

	Male (%)		Female (%)	
	Destatis	SIAB	Destatis	SIAB
Employed	53.49	52.17	46.51	47.83
Unemployed	59.01	57.17	40.99	42.83
Not in LF	44.94	50.28	55.06	49.73

The labour force is composed of employed and unemployed person and referred to as the active population, compare Table 3.1. Figure 3.2 shows the monthly relative frequencies of employed and unemployed persons in the labour force, for the monthly aggregated SIAB dataset and Destatis information[5] for both sexes in 2007-2017. The Destatis information on the labour force is for persons aged $15-74$. The magnitude of the relative frequencies, the seasonal patterns, and the trend in the SIAB dataset fit well to the official Destatis statistics. Although the SIAB dataset covers only part of the German resident population, on the aggregated level the monthly ILO employment categories reflect official employment statistics.

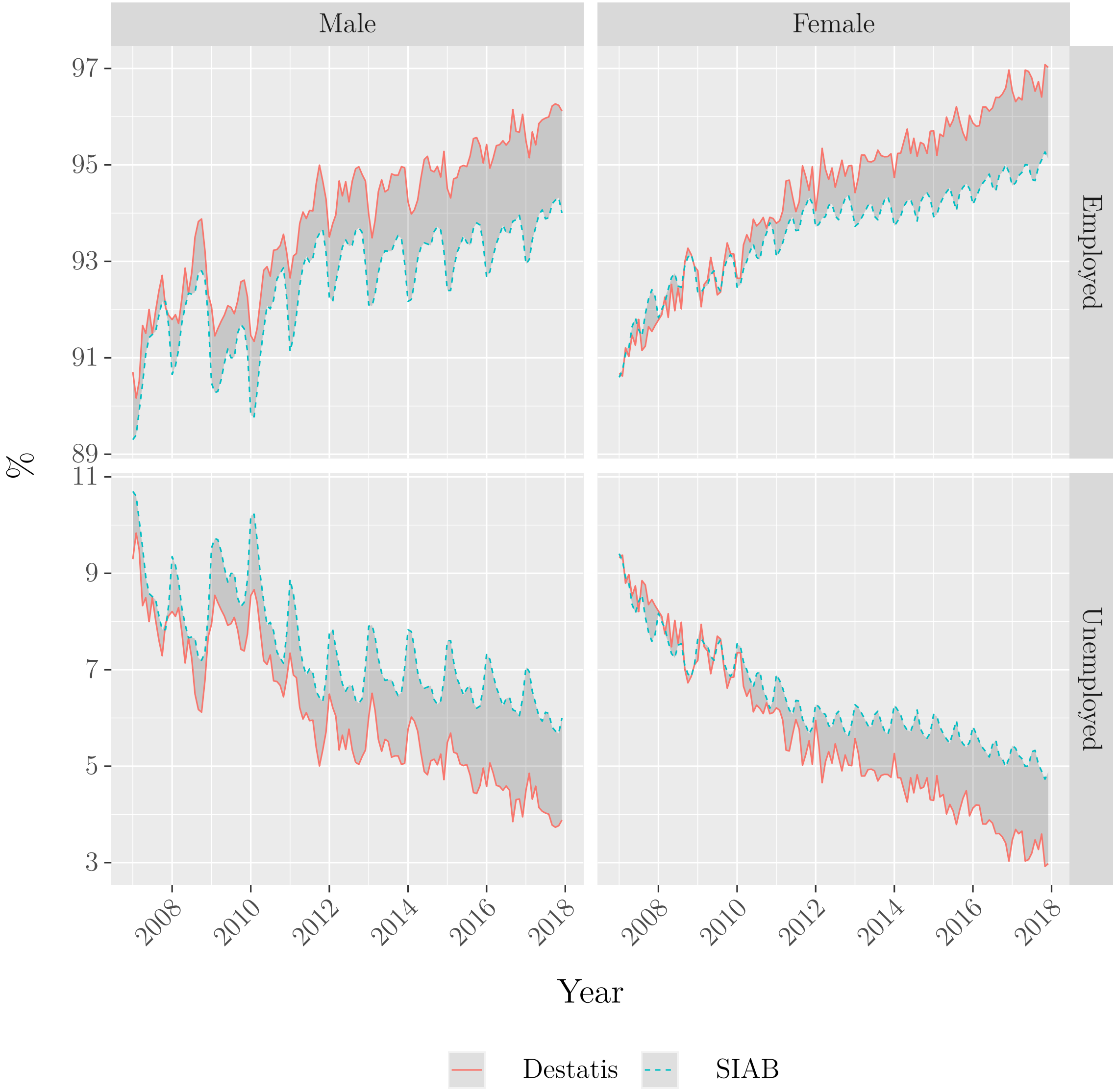

Figure 3.2: Active population: Percentages of employed and unemployed per month and sex

[5] GENESIS-Online. Code: 13231-0002. https://www-genesis.destatis.de/genesis/online.

In summary, although there are coverage problems in the SIAB dataset, particularly for persons not in the LF, the monthly SIAB employment data reflect actual seasonal patterns and trends. We therefore consider the resulting SIAB dataset with monthly employment categories as sufficient to model monthly employment transitions.

3.5 Class probability prediction

3.5.1 Introduction

In this section, we briefly introduce and discuss techniques for modelling monthly employment transitions. These techniques are later applied to the SIAB dataset in Section 3.6.

The employment status, as defined in Section 3.2, is a categorical variable with three categories. We model the monthly employment status in the SIAB dataset to generate monthly employment transitions in the RIFOSS dataset with these models. Since we want to generate synthetic data with the models, we are interested in models which give realistic probability predictions in order to then generate data from these probabilities. Low probability categories should then also be generated with frequencies according to their probabilities. We therefore also evaluate different candidate models for their probability predictions of the employment categories. In contrast, for other applications the interest is often not so much in the probability predictions, but in classifications.

There are various methods with which probability predictions can be calculated for categorical variables. For an overview, we refer to Hastie et al. (2009). For the research focus of this chapter, we identify generalised additive models (GAMs) as useful model types and briefly describe them in Section 3.5.2 for categorical dependent variables. To further improve model predictions, the ensemble methods subagging and stacking are considered in Section 3.5.3. To evaluate the probability predictions of candidate models, evaluation criteria based on Brier scores are presented in Section 3.5.4. Further, we propose an extension of the Brier score for imbalanced categorical data, which we call weighted Brier score, and show how to solve a stacking optimisation problem with the proposed weighted Brier score as a loss function.

3.5.2 Generalized additive models

Consider a dataset of n observations for which the pairs $\{y_i, \boldsymbol{x}_i\}_{i=1}^n$ are available, where y_i is an observation of a random variable Y_i with $\mathrm{E}(Y_i) = \mu_i$, and $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})^\top$ is a p-dimensional vector of auxiliary information. The following general model description is based on Hastie et al. (2009, Chapter 2). We seek for a model f with functional form

$$\mu_i = f(\boldsymbol{x}_i) = \mathrm{E}(Y_i|\boldsymbol{x}_i), \quad i = 1, \ldots, n. \tag{3.99}$$

Model f returns the expected value of the variable of interest given a set of auxiliary variables. As can be seen from (3.99), we focus on models which seek to minimize the mean squared error of the observations and predictions, compare Hastie et al. (2009, p. 18). We can choose a parametrization of f and estimate these parameters from data. The model with plugged in parameter estimates is called *fitted model* and denoted by $\hat{f}$. For any observation for which the vector of covariates of length p is available, $\hat{f}$ returns predictions of the variable of interest.

In this chapter we have decided to use GAMs as models (3.99) for the following reasons. GAMs allow for various functional relationships between the dependent variable and the auxiliary information. This allows us to take non-linear relationships such as seasonal effects into account in the model. Furthermore, by the use of GAMs we can make sure that all variables common to the RIFOSS and SIAB dataset are included in the model. Thereby, even when a variable does not have a good predictive power for employment, there will still be some correlation preserved in the generated data. For categorical dependent variables GAMs directly return probability predictions which sum up to one without the need for additional calibration. The following description of GAMs is based on Wood (2017).

In a GAM, the functional relationship of the variable of interest and the covariates (3.99) can be represented as

$$f(\boldsymbol{x}_i) = \mathrm{E}(Y_i|\boldsymbol{x}_i) = g^{-1}(\eta(\boldsymbol{x}_i)), \quad i = 1, \ldots, n, \tag{3.100}$$

where g is a *link function* and the *predictor* η specifies the functional relationship of the covariates to $g(\mu_i)$. To understand GAMs, we first describe *generalised linear models* (GLMs), i.e. we focus on the description of link function g. Then, we introduce additive models to relax the linearity assumption, i.e. we focus on the description of predictor η.

Generalized linear models

The predictor function η specifies the functional relationship of the covariates to $g(\mu_i)$. For now, we focus on the link function g and take η as a *linear predictor*. That is, η is a linear function of the covariates given by

$$\eta(\boldsymbol{x}_i) = \sum_{k=1}^{p} \beta_k x_{ik} = \boldsymbol{x}_i^\top \boldsymbol{\beta}, \quad i = 1, \ldots, n, \tag{3.101}$$

or, in matrix notation,

$$\eta(\boldsymbol{X}) = \boldsymbol{X}\boldsymbol{\beta}, \tag{3.102}$$

where $\boldsymbol{\beta} \in \mathbb{R}^p$ is the vector of fixed effects and $\boldsymbol{X} = \underset{1\leq i\leq n}{\mathrm{row}}(\boldsymbol{x}_i)$ is the model matrix which is assumed to be of full rank p. Thereby, (3.99) can be written as

$$f(\boldsymbol{x}_i) = g^{-1}(\eta(\boldsymbol{x}_i)) = g^{-1}(\boldsymbol{x}_i^\top \boldsymbol{\beta}), \quad i = 1, \ldots, n. \tag{3.103}$$

Link function g allows to model the distributions of variables Y_i, $i = 1, \ldots, n$, as any member of the *exponential family of distributions*. Note that the Y_i, $i = 1, \ldots, n$, are assumed to be independent. Examples of the exponential family of distributions are the normal distribution for continuous variables, the poisson distribution for non-negative counts, and the binomial distribution for dichotomous variables. An overview of the relationship of different data types to the parameters of the exponential family of distributions is given in Wood (2017, Table 3.1). For each member of the exponential family of distributions there exists a bijective monotonic link function g for which $\mathrm{E}(Y_i|\boldsymbol{x}_i) = g^{-1}(\eta(\boldsymbol{x}_i))$, or equivalently $g(\mathrm{E}(Y_i|\boldsymbol{x}_i)) = \eta(\boldsymbol{x}_i)$, holds. Taking the *identity link* $g(\mu_i) = \mu_i$ and η as a linear predictor (3.101) results in a *general linear model.* For a general linear model, (weighted) least squares can be applied to estimate $\boldsymbol{\beta}$. Taking the link g as any member of the exponential family of distributions and η as a linear predictor, (3.101) results in a GLM. GLMs were introduced by Nelder and Wedderburn (1972). A comprehensive overview of GLMs is given in McCullagh and Nelder (1989) and Wood (2017). In GLMs, we cannot directly solve for fixed effects $\boldsymbol{\beta}$. Instead, *iteratively re-weighted least squares* (IRLS) can be applied to estimate $\boldsymbol{\beta}$. We refer to Wood (2017, Chapter 3) for more details on IRLS.

The variable of interest in focus of this chapter is the ILO employment status, compare Table 3.1. It is a categorical variable with $J = 3$ categories. For categorical variables, we recall some basic features of the *multinomial* distribution. The description is based on Agresti (2019, p. 5). The multinomial distribution is a multi-categorical generalization of the binomial distribution, which is part of the exponential family of distributions. The probability distribution of the counts of a categorical variable follows the multinomial distribution if the category observations are independent realisations of a fixed set of category probabilities. The non-negative probabilities of the J categories are denoted by $(\pi_1, \dots, \pi_J)^\top$ with $\sum_{j=1}^J \pi_j = 1$. For n random realisations of the multinomial distribution, the absolute counts of the J categories are denoted by $n_1, \dots, n_J$, with $\sum_{j=1}^J n_j = n$. The probabilities of these counts are given by

$$\Pr(n_1, n_2, \dots, n_J) = \left(\frac{n!}{n_1! n_2! \cdots n_J!}\right) \pi_1^{n_1} \pi_2^{n_2} \cdots \pi_J^{n_J}. \tag{3.104}$$

In a GLM for multinomial data, the counts of Y_i are modelled as realisations of a multinomial distribution with probabilities $(\pi_{i1}, \dots, \pi_{iJ})^\top$. $\pi_{ij} = \Pr(Y_i = j|\boldsymbol{x}_i)$ is the non-negative probability that Y_i takes category j given covariates $\boldsymbol{x}_i$, $j = 1, \dots, J$, $i = 1, \dots, n$. For identifiability, category J is taken as the baseline category in a multinomial model and the predictor η is defined for each of the remaining $J - 1$ categories separately. For each of the $J - 1$ categories, the predictor returns the *log-odds* (Agresti, 2002, Equation 7.1)

$$\eta_j(\boldsymbol{x}_i) = \ln\left(\frac{\pi_{ij}}{\pi_{iJ}}\right), \quad j = 1, \dots, J, \quad i = 1, \dots, n, \tag{3.105}$$

such that $\eta_J(\boldsymbol{x}_i) = 0$. By re-arranging (3.105), the category-specific probabilities can be expressed as

$$\pi_{ij} = \frac{\exp(\eta_j(\boldsymbol{x}_i))}{\sum_{k=1}^J \exp(\eta_k(\boldsymbol{x}_i))}, \quad j = 1, \dots, J, \quad i = 1, \dots, n. \tag{3.106}$$

For each of the $J - 1$ categories, a different specification of predictor η_j can be chosen, e.g. a different set of auxiliary variables. For any observation i, a GLM f for multinomial data returns a vector of probability predictions for all J categories, formally

$$f : \mathbb{R}^p \to [0,1]^J, \; \boldsymbol{x}_i \mapsto (\Pr(Y_i = 1|\boldsymbol{x}_i), \dots, \Pr(Y_i = J|\boldsymbol{x}_i))^\top. \tag{3.107}$$

For further details and examples of modelling of multinomial data, we refer to Agresti (2019, Chapter 6) and Agresti (2002, Chapter 7).

In LMMs, compare Section 2.4.2, we already saw that the model matrix $\boldsymbol{X}$ usually contains an n-dimensional vector of ones as first column in order to include an intercept. Furthermore, the model matrix $\boldsymbol{X}$ allows for various transformations of potential covariates. Take the observations of a candidate variable z_i, $i = 1, \dots, n$. Then, as a covariate in a linear model, we could inter alia set the j-th variable $x_{ij} = z_i$, $x_{ij} = \log(z_i)$, or $x_{ij} = z_i^2$. Hence, it is straightforward to include polynomial relations, log-transformations, Box-Cox transformations (Box & Cox, 1964), or interaction effects between different variables into a GLM. Although these transformations, especially polynomials, may be flexible with regard to representing non-linear relationships, Wood (2017, Chapter 4.2.1) argues that they can oscillate widely, especially when they are of high degree. Additive models can solve this problem and are described in the following.

Additive models

Instead of defining η as a linear function of the covariates, it can be specified as an additive composition of smooth functions of covariates. It is then called *additive predictor* and given by

$$\eta(\boldsymbol{x}_i) = h_1(x_{i1}) + h_2(x_{i2}) + h_3(x_{i3}, x_{i4}) + \dots, \quad i = 1, \dots, n, \tag{3.108}$$

where h_1, h_2, h_3, ... are smooth functions. The following description of additive models is based on Wood (2017). Taking the *identity link* $g(\mu) = \mu$ and η as an additive predictor (3.108) results in a *general additive model*. Taking g as the link for any member of the exponential family of distributions and η as an additive predictor, (3.108) results in a GAM (Hastie & Tibshirani, 1986, 1990). Formulation (3.108) allows for arbitrary interaction smooths of covariates, see Wood (2017) for examples and further information. In this chapter, only univariate smooth functions are considered.

By further specifying some of the smooth functions in (3.108), e.g. to be linear functions in the covariates, a GAM can be written as a semi-parametric model. Without loss of generality, $\boldsymbol{x}_i$ can be reordered as $\boldsymbol{x}_i = (x_{i1}, \dots, x_{i(p_a)}, \dots, x_{i(p_a+p_b)})^\top$, $i = 1, \dots, n$, where $p = p_a + p_b$ such that p_a and p_b represent the covariates which enter the additive predictor linearly or additively (by a smooth function) respectively. A semi-parametric GAM is then given by

$$\eta(\boldsymbol{x}_i) = \underbrace{\sum_{l=1}^{p_a} \beta_l x_{il}}_{\text{parametric part}} + \underbrace{\sum_{l=(p_a+1)}^{p_a+p_b} h_l(x_{il})}_{\text{non-parametric part}}, \quad i = 1, \dots, n. \tag{3.109}$$

GAMs offer a high flexibility with respect to the choice of the functional relations while allowing for the investigation of the effects of single variables. In the modelling process, we can for example choose that some relationships are modelled as linear or cyclical functions.

The functions in (3.109) can be parametrized in order to have them in a dependence of a parameter vector which can be estimated from data. For simplicity, we demonstrate this parametrization for a single predictor variable $\boldsymbol{x} \in \mathbb{R}^n$, based on Wood (2017, Section 4.2.1). To parametrize a function h, which takes the predictor variable $\boldsymbol{x}$ as input, we choose a basis. The basis is chosen such that the assumed form of h lies within the space of functions which can be represented by that basis. Function h can then be rewritten as a sum of K *basis functions* b_k, as

$$h(x_i) = \sum_{k=1}^{K} b_k(x_i)\beta_k^h, \quad i = 1, \dots, n, \tag{3.110}$$

where β_k^h denotes the unknown parameter associated with the basis function b_k. By reformulating all smooth functions via basis functions, the predictor η (3.108) is again linear in a vector of regression coefficients which can be estimated from data by IRLS.

There are various basis functions which can be used to parametrize a function for an additive model. Wood (2017, Chapter 5) gives an overview of different smoothers. For the presented research, we consider *(cyclical) cubic regression splines* as basis functions. We briefly list some features of these splines and then justify why we chose them. The following description is based on Wood (2017, sections 5.1, 5.3.1, 5.3.2), to which we also refer for a more technical description. For a chosen covariate, a cubic regression spline is a function for which the following criteria hold. The spline is defined on the range of the covariate and is continuous to the second derivative. For the spline one has to choose a number of knots K and their placement in the range of the covariate. The space between any two neighbouring knots defines a subinterval. On each of the $K - 1$ subintervals, a cubic polynomial, i.e. a polynomial of order three, is defined. At each knot, the two neighbouring piecewise polynomials are equal in their value as well as in their first two derivatives. If furthermore, the second derivatives are zero at the first and last knot, the spline is known as a *natural cubic spline*. A cubic regression spline can also be specified as a cyclic spline. In that case, the start and end of the spline have the same value. Cyclical splines are especially useful for incorporating seasonal effects like the effect of the month of an observation. With the number and placement of the knots, the corresponding basis functions in (3.110) are fully specified. The corresponding formulas for the (cyclical) cubic splines are e.g. given in Wood (2017, Table 5.1). Natural cubic splines are a frequent choice for univariate smooth functions. They are the smoothest interpolators in the class of all those functions which are continuous on the interval of $\boldsymbol{x}$, have absolutely continuous first derivatives, and interpolate observation pairs x_i, y_i (Wood, 2017, Section 5.1.1).

A potential problem with splines is that their fit may be following the observations in the data too closely or even exactly, so-called *overfitting*. In this case, the estimated spline may be less suitable for out-of-sample predictions than a spline which is smoother and thereby follows the data points less closely. To achieve better predictions, an additional penalty term can be added, weighted by a *penalty parameter* λ. We consider penalties associated with the squared second derivative of the spline to handle its curvature. The penalty parameter can be used to set a trade-off between the data fit of the spline and its smoothness. Given a penalty parameter λ, the unknown parameters of a GAM can be estimated by *penalized IRLS*. A detailed overview of this fitting procedures is given in Wood (2017, Section 3.4.1).

The penalty parameter λ can be chosen via *cross-validation*. There are many different forms in which one can conduct cross-validation. We refer to Hastie et al. (2009) for some examples. In a simple cross-validation, the available data is randomly partitioned in two parts, one for fitting the model, the other for evaluating the predictions of the model. This procedure can be repeated several times. Cross-validation allows to approximate how well a model performs for out-of-sample prediction under different candidate values for λ.

We can see that GAMs provide a very flexible modelling tool for categorical variables of interest. They facilitate to include various functional forms in the model, e.g. cyclical non-linear effects. GAMs are implemented in standard statistical software. We use the `mgcv` package (Wood, 2003, 2004, 2011, 2017; Wood et al., 2016) in `R` (R Core Team, 2020). Function `gam` of the package facilitates to calculate GAMs for multinomial data with integrated choice of the penalty parameter, based on the theory presented in Wood et al. (2016).

3.5.3 Ensemble methods

After the editing and aggregation process of Section 3.4, the SIAB dataset contains a total of $132,084,804$ data rows. It can thus be difficult to include all observations in the estimation of a single GAM, either because of memory or time limitations. We therefore consider that a chosen GAM can only be calculated on samples of the SIAB dataset. With parallelizable infrastructure, it is, however, possible to simultaneously draw multiple samples and compute models on them, which then can be combined into a joint model. Approaches that combine several models into a new, joint model are summarized under the term *ensemble methods*. An overview of different ensemble methods is given in Hastie et al. (2009, Chapter 16), Polikar (2006), and Bühlmann (2012), on which the following description is based. Ensemble models are often found to provide a better performance for prediction than each of their input models alone and therefore considered as profitable candidate methods for the given research task. In the following, we refer to non-ensemble models as *individual models* and combinations of at least two individual models as *ensemble models*.

Let there by M fitted individual models $\hat{f}_m^*$, $m = 1, \ldots, M$. For example, each individual model is a GAM fitted to a sample of the SIAB dataset. An ensemble model $\hat{f}^E$ is built as

$$\hat{f}^E = \sum_{m=1}^{M} w_m \hat{f}_m^*, \tag{3.111}$$

where $\boldsymbol{w} = (w_1, \ldots, w_M)^\top$ are the *ensemble weights* with $w_m \geq 0$, $\sum_{m=1}^{M} w_m = 1$. An ensemble model thereby returns a weighted sum of the predictions of the individual models.

The performance gains of an ensemble model $\hat{f}^E$ over its individual input models $\hat{f}_m^*$, $m = 1, \ldots, M$, are particularly high when the individual models are diverse and relatively weak. The individual models can differ from each other especially if they are calculated on different sample data, with different weights of the same data, with different learning algorithms, or different model specifications. Popular examples of ensemble methods are *bootstrap aggregating* (bagging) (Breiman, 1996a), the *random forest* algorithm which is based on the idea of bagging (Breiman, 2001), *boosting* (Schapire, 1990), especially ADABoost (Freund, 1995; Freund & Schapire, 1997), and *stacked generalisation* (stacking) (Wolpert, 1992). Ensemble methods are mostly used when the research focus is on prediction or classification rather than on the interpretative character of the models. They are applied to various fields of interest and appear under various names like *multiple classifier systems*, *mixtures of experts*, or *composite classifier systems*, to name only a few, see Polikar (2006) for an overview. We focus on the ensemble methods bagging and stacking as they are applicable to parallel computation and allow for an easy implementation in combination with GAMs. In the following, we use $\mathcal{D}$ to denote the collection of observation tuples $\{y_i, \boldsymbol{x}_i\}_{i=1}^n$.

In bagging (Breiman, 1996a), the following steps are performed. M *bootstrap* samples are drawn from $\mathcal{D}$. For a dataset of size n, a classic bootstrap sample is a sample of size n from that data, drawn with replacement. For more information on bootstrap procedures, we refer to Efron and Tibshirani (1993). On each sample, a chosen learning algorithm, e.g. a GAM, is applied, resulting in an individual model $\hat{f}_m^*$, $m = 1, \ldots, M$. Ensemble model $\hat{f}^{bagging}$ is defined as

$$\hat{f}^{bagging} = \frac{1}{M} \sum_{m=1}^{M} \hat{f}_m^*. \tag{3.112}$$

In bagging, the ensemble weights are set to the inverse of the number of individual models and usually the same learning algorithm is used for all individual models, e.g. a specific GAM. Bagging can lead to substantial gains in accuracy when the variance between the individual models is high, as it then reduces the variance of the prediction error. As classification and regression trees often have high variance and low bias, they are frequently used as a learning algorithm for the individual models in bagging. For further information on classification and regression trees, we refer to Hastie et al. (2009, Section 9.2). Breiman (2001) extended the use of bagging with classification and regression trees by additionally sampling the covariates available to each tree, known as the random forest algorithm. Instead of bootstrap samples of size n with replacement, also subsamples, i.e. samples of size less than n, can be drawn from $\mathcal{D}$, with or without replacement. The ensemble method is then called *subsample aggregating* (subagging) (Bühlmann & Yu, 2002). Subagging can be especially useful when the available data consist of many observations as it allows to calculate the individual models on different parts of the available data. Bagging is straightforward to implement, can be used with arbitrary learning algorithms for the individual models, and the individual models can be calculated in parallel.

We consider that due to memory and computation time restrictions only a limited number of individual models, e.g. $M \leq 50$, can be calculated. With a limited number of individual models, however, the potential of variance reduction of bagging or subagging with their equally weighted individual models can be low when using GAMs. It can therefore be useful to apply them in conjunction with another ensemble method called stacking. Stacking was proposed in Wolpert (1992) and further studied in Breiman (1996b). The original idea was to apply different learning algorithms to the same data, each giving a fitted model $\hat{f}_m^*$, and building an ensemble by an optimal combination of these individual models. We define the vector of ensemble weights $\hat{\boldsymbol{w}}^{stacking}$ as the solution of optimization problem

$$\begin{aligned} &\min_{\boldsymbol{w}\in\mathbb{R}_+^M} \sum_{i=1}^{n} L(y_i, \sum_{m=1}^{M} w_m \hat{f}_m^*(\boldsymbol{x}_i)) \\ &\text{subject to} \sum_{m=1}^{M} w_m = 1. \end{aligned} \tag{3.113}$$

That is, the stacking ensemble weights $\hat{\boldsymbol{w}}^{stacking}$ are obtained as the solution of minimizing a loss function L, which takes into account the difference of observations y_i and their ensemble predictions $\sum_{m=1}^{M} w_m \hat{f}_m^*(\boldsymbol{x}_i)$ for all observation tuples in $\mathcal{D}$. We discuss suitable loss functions in the next section. The stacking ensemble model is given by

$$\hat{f}^{stacking} = \sum_{m=1}^{M} \hat{w}_m^{stacking} \hat{f}_m^*. \tag{3.114}$$

To avoid overfitting, the individual models and the ensemble weight optimisation should be calculated on different sets of data.

There is only little research about the use of GAMs in ensemble methods. De Bock et al. (2010), De Bock and Van den Poel (2012), and De Bock et al. (2019) investigated ensembles with GAMs for the task of binary classification in the context of customer churn prediction. Their studies showed that GAMs are well suited as a individual models due to their attractive properties such as flexibility regarding the relationships to be modelled. They investigated combinations of GAMs with bagging, random samples of auxiliary variables, and combinations of the two. They furthermore investigated how the interpretability of GAMs can be preserved in ensemble methods. With the presented approaches in this chapter, we add to the research of the use of GAMs with ensemble methods.

3.5.4 Evaluation of probability predictions

In the following, we discuss different criteria for evaluating the probability predictions of a model. These criteria can be used as loss function L in stacking optimisation problem (3.113). We again consider $\mathcal{D}$ as a set of observation tuples $\{y_i, \boldsymbol{x}_i\}_{i=1}^n$, where observations y_i are categorical with J categories and $\boldsymbol{x}_i$ is a vector of p auxiliary variables. Assume there is a model $\hat{f}$, which was fitted for this categorical dependent variable. The vector of probability predictions, which $\hat{f}$ returns for plugged in covariate information, is denoted by

$(\hat{p}(Y_i = 1|\boldsymbol{x}_i), \dots, \hat{p}(Y_i = J|\boldsymbol{x}_i))^\top$. The following criteria evaluate how well the probability predictions of a model fit to a set of observed categories.

A popular metric for evaluating probability predictions is the *Brier score* (Brier, 1950). It corresponds to the mean squared error between the predicted category probabilities and the actual categories and was introduced for binary variables. Based on data $\mathcal{D}$, the Brier score B of a model $\hat{f}$ is given by

$$B(\hat{f}) = \frac{1}{2n} \sum_{j=1}^{J} \sum_{i=1}^{n} (I(y_i = j) - \hat{p}(Y_i = j|\boldsymbol{x}_i))^2, \tag{3.115}$$

where I is an indicator function. The division by $2n$ ensure that $B(\hat{f}) \in [0, 1]$. Lower values of the Brier Score represent better models. From (3.115), we see that the information from the n observations enters the Brier Score with equal weights. The Brier score is frequently used to assess probability predictions for various learning algorithms, e.g. in Wood et al. (2016) and Kruppa et al. (2014a). For calculating the Brier score, a sufficient number of observations n should be considered such that the observed categories can be expected to sufficiently represent their underlying probabilities (Wilks, 2010). We note that in some publications, e.g. Bradley et al. (2008) and Kruppa et al. (2014b), authors distinguish between the Brier score and estimates of the Brier score. In the present research, we only calculate the Brier score based on some fitted models and observation data and do not make the distinction, similarly to e.g. Wood et al. (2016).

The absolute values of the Brier score as such allow for little interpretation. They can only be used to compare different models for the same data. In order to get an idea of how well a model actually works, the *Brier skill score* can be calculated, where the value of the Brier score of a model is divided by the value of the Brier score of an uninformative model such as the intercept only model (Bradley et al., 2008).

The evaluation of probability predictions for imbalanced data presents a particular challenge. A categorical variable is said to be *imbalanced*, also called *unbalanced*, if its relative frequencies differ significantly from each other. The ILO employment status in SIAB, with relative frequencies shown in Table 3.9, can be considered to be imbalanced. The category with the highest and lowest relative frequency is referred to as the *majority class* and *minority class* respectively. For classification, not probability prediction, the issues of imbalanced data are well recognized in the literature. Yang and Wu (2006) list imbalanced data as one out of ten challenging problems in data mining research, Fernández et al. (2018) and He and Garcia (2009) gave an overview of the issue of imbalanced data and different techniques to tackle it, Galar et al. (2011) focused on the combination of ensemble techniques for imbalanced data. We note that there is substantially less research concerning the evaluation of imbalanced data when the focus is on probability prediction rather than classification.

In Table 3.12, we present an illustrative example to show how different scores for probability predictions rate different models under imbalanced categorical data. Let there be three fitted models $\hat{f}^{m1}$, $\hat{f}^{m2}$, and $\hat{f}^{m3}$. The table presents some example data for which the models return probability predictions. The example data consists of 100 observations of a categorical variable with $J = 3$ categories. It is imbalanced, there are 90 data rows with category 1 and 5 data rows with category 2 and 3 each. For each model, the corresponding probability predictions are displayed in the three columns of each model. For illustration, each model returns the same probability predictions for the 90, 5, and 5 observations of each category respectively. Model $\hat{f}^{m1}$ returns probability predictions of 100% for the majority class and 0% else. Model $\hat{f}^{m2}$ is an intercept only model returning the relative frequencies as the probability predictions. Model $\hat{f}^{m1}$ and $\hat{f}^{m2}$ are considered to be uninformative, they always return the same probability predictions regardless of the covariate values. Model $\hat{f}^{m3}$ is considered to be a fairly good model. It returns probability predictions which depend on the covariate information and reflect the actually observed categories. For the observations of category 2 and 3, the probability predictions of $\hat{f}^{m3}$ even perfectly match the actually observed categories.

Table 3.12: Example 1: Brier scores for imbalanced data

Nr. of		$\hat{f}^{m1}$			$\hat{f}^{m2}$			$\hat{f}^{m3}$		
data rows	y_i	1	2	3	1	2	3	1	2	3
90	1	1	0	0	0.90	0.05	0.05	0.6	0.2	0.2
5	2	1	0	0	0.90	0.05	0.05	0.0	1.0	0.0
5	3	1	0	0	0.90	0.05	0.05	0.0	0.0	1.0
Brier score $B(.) * 100$		10.0			9.2			10.8		
Stratified Brier score $SB(.) * 100$		66.7			60.5			5.3		
Weighted Brier score $Bw(\boldsymbol{d}, .) * 100$		14.6			12.6			0.9		

Note: We take $\boldsymbol{d} = (0.027, 0.486, 0.486)^\top$, which is the normalized inverse of the relative frequencies.

Under the fitted models, the values of three different scores are displayed. For now, we only focus on the first row, the Brier score, the scores in the last two rows are explained later. The Brier score returns lower values for the uninformative models $\hat{f}^{m1}$ and $\hat{f}^{m2}$ than for model $\hat{f}^{m3}$. In the case of imbalanced data, the majority class almost entirely determines the value of the Brier score. Thus, even models that are uninformative but generally assign high probabilities to the majority class perform well according to the Brier score.

Wallace and Dahabreh (2012) recognized that the Brier score performs poorly for imbalanced data and proposed the *stratified Brier score* as an alternative. For observations $\mathcal{D}$, the stratified Brier score SB of a model $\hat{f}$ is defined as

$$SB(\hat{f}) = \frac{1}{J} \sum_{j=1}^{J} \left(\sum_{i=1}^{n} I(y_i = j) \right)^{-1} \sum_{i=1}^{n} I(y_i = j)(1 - \hat{p}(Y_i = j|\boldsymbol{x}_i))^2. \tag{3.116}$$

As with the Brier score, $SB(\hat{f}) \in [0, 1]$. The stratified Brier score is calculated as the average of category-specific scores. For the category-specific scores, only those observations y_i which are in the respective category are included in the measure. Thereby, an observation of the minority class contributes more to the stratified Brier score than an observation of the minority class. The stratified Brier score is for example applied in Collell et al. (2018) for evaluating the classification performance under imbalanced data. In Table 3.12, the values of the stratified Brier are added in the second last row. The stratified Brier score clearly favours model $\hat{f}^{m3}$ over the two uninformative models and is thus considered more suitable than the Brier score for imbalanced data.

For a categorical variable with more than two categories, we recognize a significant drawback of the stratified Brier score in (3.116). To see this, we add another illustrative example in Table 3.13. The example data presented in the table is the same as in Table 3.12, but we consider the two candidate models $\hat{f}^{m4}$ and $\hat{f}^{m5}$. Comparing the observed categories in the data and the probability predictions of the two model, we would consider $\hat{f}^{m4}$ to be a better model than $\hat{f}^{m5}$. For example, for category one, model $\hat{f}^{m4}$ assigns the highest probability predictions to that category, whereas model $\hat{f}^{m5}$ assigns the highest probability to category two. As, however, both models $\hat{f}^{m4}$ and $\hat{f}^{m5}$ assign probability 0.4 to the actually observed categories, the stratified Brier score of both models is identical. As we can see from the table and the formula of the stratified Brier score (3.116), for the category-specific measures the stratified Brier score only considers the differences between the actually observed category and the probability predictions for only this category. In the presented example, the Brier score (3.115), however, clearly favours model $\hat{f}^{m4}$ over model $\hat{f}^{m5}$. We conclude that, for our research aim, the evaluation of multi-categorical probability predictions under imbalanced data, neither the Brier nor the stratified Brier score is an optimal choice.

Table 3.13: Example 2: Brier scores for imbalanced data

Nr. of		$\hat{f}^{m4}$			$\hat{f}^{m5}$		
data rows	y_i	1	2	3	1	2	3
90	1	0.4	0.3	0.3	0.4	0.6	0.0
5	2	0.3	0.4	0.3	0.0	0.4	0.6
5	3	0.3	0.3	0.4	0.6	0.0	0.4
Brier score $B(.) * 100$			27			36	
Stratified Brier score $SB(.) * 100$			36			36	
Weighted Brier score $Bw(\boldsymbol{d}, .) * 100$			5.9			7.9	

Note: We take $\boldsymbol{d} = (0.027, 0.486, 0.486)^\top$, which is the normalized inverse of the relative frequencies.

With the examples of Tables 3.12 and 3.13, we saw that for imbalanced data with more than two categories, neither the Brier score nor the stratified Brier score is an optimal choice. A potential solution is to evaluate different models separately for each category. However, for optimisation purposes, such as the stacking problem (3.113), it is often desirable to have a scalarization, i.e. a loss function returning a single value for evaluating a model. As a solution, we propose a weighted Brier score Bw. For observations $\mathcal{D}$, the weighted Brier score Bw of a model $\hat{f}$ is given by

$$Bw(\boldsymbol{d}, \hat{f}) = \frac{J}{2n} \sum_{j=1}^{J} d_j \sum_{i=1}^{n} (I(y_i = j) - \hat{p}(Y_i = j|\boldsymbol{x}_i))^2, \tag{3.117}$$

where $\sum_{j=1}^{J} d_j = 1$, $d_j > 0$, $j = 1, \ldots, J$, and, similar to the other two Brier scores, $Bw(\boldsymbol{d}, \hat{f}) \in [0, 1]$. Compared to the original Brier score (3.115), in the weighted Brier score a vector of *category weights* $\boldsymbol{d} = (d_1, \ldots, d_J)^\top$ is added. For equal category weights, i.e. $d_j = 1/J$, $j = 1, \ldots, J$, the weighted Brier score reduces to the Brier score. The category weights allow for a category-specific weighting in the score. For example, one can choose the category-specific weights to equal the normalized inverse of the relative frequencies. Thereby, the few observations of the minority class get much more weight in the score than they would in the original Brier score.

In the last row of Tables 3.12 and 3.13, we added the values of the weighted Brier score, where category weights $\boldsymbol{d}$ are set to the normalized inverse of the relative frequencies. Unlike the stratified Brier score, the proposed weighted Brier score takes into account all probability predictions. Furthermore, it allows to account for data imbalancedness by category weights. We see that in both tables, the weighted Brier score clearly favours the prioritized models. That is, in Table 3.12, it returns the lowest value for model $\hat{f}^{m3}$ and in Table 3.13 it returns the lowest value for model $\hat{f}^{m4}$.

Proposition 3.1 shows that the stacking optimisation problem (3.113) with the weighted Brier score (3.117) as a loss function L is a convex *quadratic programming* (QP). QPs can be solved with standard statistical software such as the `R` (R Core Team, 2020) function `solve.QP` from the `quadprog` package, (Turlach & Weingessel, 2019) which uses the dual method of Goldfarb and Idnani (1982, 1983). For further information on QPs, we refer to Nocedal and Wright (2006, Chapter 16).

Proposition 3.1. The stacking optimisation problem (3.113) with the weighted Brier score (3.117) as loss function L is a convex QP of the form

$$\begin{aligned} \min_{\boldsymbol{w}} L(\boldsymbol{w}) &= (1/2)\boldsymbol{w}^\top \boldsymbol{Q} \boldsymbol{w} + \boldsymbol{v}^\top \boldsymbol{w} \\ \text{subject to } \boldsymbol{a}^\top \boldsymbol{w} &= 1 \\ \boldsymbol{w} &\geq \boldsymbol{b}, \end{aligned} \tag{3.118}$$

where $\boldsymbol{Q}$ is a symmetric positive definite $M \times M$ matrix, $\boldsymbol{v} \in \mathbb{R}^M$, $\boldsymbol{w} \in \mathbb{R}^M$, $\boldsymbol{a} = (1, \ldots, 1)^\top$ and $\boldsymbol{b} = (0, \ldots, 0)^\top$ are vectors of length M such that $\boldsymbol{a}^\top \boldsymbol{w} = \sum_{m=1}^{M} w_m$.

Proof. We introduce some additional notation. Again, the variable of interest is categorical with J categories and a dataset $\mathcal{D}$ consisting of n observations is used for finding optimal ensemble weights. The absolute frequencies of the J categories in data $\mathcal{D}$ are denoted by $n_1, \ldots, n_J$. An ensemble model $\hat{f}^E$ is given by $\hat{f}^E = \sum_{m=1}^{M} w_m \hat{f}_m^*$, where $\hat{f}_m^*$ is an individual model. An individual model $\hat{f}_m^*$ returns a vector of J predictions. Let $\hat{p}_{ij}^{m*}$ denote the j-th probability prediction of $\hat{f}_m^*(\boldsymbol{x}_i)$, $m = 1, \ldots, M$, $i = 1, \ldots, n$, $j = 1, \ldots, J$.

Set $\hat{\boldsymbol{p}}_{ij}^{M} = (\hat{p}_{ij}^{1*}, \dots, \hat{p}_{ij}^{M*})^\top$, $j = 1, \dots, J$, $i = 1, \dots, n$. The category-specific vectors are combined to form vectors of length nJ, indicated by superscript $\star$:

$$\begin{aligned} \hat{\boldsymbol{p}}^{m\star} &= (\hat{p}_{11}^{m}, \dots, \hat{p}_{n1}^{m}, \dots, \hat{p}_{1J}^{m}, \dots, \hat{p}_{nJ}^{m})^\top, && (3.119)\\ \breve{\boldsymbol{y}}^\star &= ((I(y_1 = 1))_{\times n}, \dots, (I(y_1 = J))_{\times n})^\top, && (3.120)\\ \boldsymbol{d}^\star &= ((d_1)_{\times n}, \dots, (d_J)_{\times n})^\top, && (3.121)\\ \boldsymbol{r}^\star &= ((n_1^{-1})_{\times n}, \dots, (n_J^{-1})_{\times n})^\top, && (3.122) \end{aligned}$$

where $(a)_{\times n}$ means that value a is repeated n times. Let furthermore $\hat{\boldsymbol{P}}^{M\star} = (\hat{\boldsymbol{p}}^{1\star}, \dots, \hat{\boldsymbol{p}}^{M\star})$ such that $\hat{\boldsymbol{P}}^{M\star}$ is an $n \times M$-dimensional matrix.

With the additional notation, the weighted Brier score Bw in (3.117) for an ensemble model $\hat{f}^E$ with ensemble weights $\boldsymbol{w}$ can be reformulated as

$$\begin{aligned} Bw(\boldsymbol{d}, \hat{f}^E) &= \frac{J}{2n} \sum_{j=1}^{J} d_j \sum_{i=1}^{n} (I(y_i = j) - \hat{p}(Y_i = j|\boldsymbol{x}_i))^2 \\ &\propto \sum_{j=1}^{J} d_j \sum_{i=1}^{n} (I(y_i = j) - \boldsymbol{w}^\top \hat{\boldsymbol{p}}_{ij}^{M})^2 \\ &= \sum_{j=1}^{J} \sum_{i=1}^{n} d_j I(y_i = j)^2 - 2 d_j I(y_i = j) \boldsymbol{w}^\top \hat{\boldsymbol{p}}_{ij}^{M} + d_j \boldsymbol{w}^\top \hat{\boldsymbol{p}}_{ij}^{M} \hat{\boldsymbol{p}}_{ij}^{M\top} \boldsymbol{w} \\ &= \boldsymbol{d}^{\star\top} \breve{\boldsymbol{y}}^\star - 2(\boldsymbol{d}^{\star\top} \circ \breve{\boldsymbol{y}}^{\star\top}) \hat{\boldsymbol{P}}^{M\star} \boldsymbol{w} + \boldsymbol{w}^\top \hat{\boldsymbol{P}}^{M\star\top} \operatorname{diag}(\boldsymbol{d}^\star) \hat{\boldsymbol{P}}^{M\star} \boldsymbol{w} \\ &= c + \boldsymbol{v}^\top \boldsymbol{w} + (1/2) \boldsymbol{w}^\top \boldsymbol{Q} \boldsymbol{w}, \end{aligned} \qquad (3.123)$$

where $\circ$ denotes the Hadamard product, $c = \boldsymbol{d}^{\star\top} \breve{\boldsymbol{y}}^\star$ is a constant, $\boldsymbol{v} = -2\hat{\boldsymbol{P}}^{M\star\top}(\boldsymbol{d}^\star \circ \breve{\boldsymbol{y}}^\star)$ is a vector of length M, and $\boldsymbol{Q} = \hat{\boldsymbol{P}}^{M\star\top} \operatorname{diag}(2\boldsymbol{d}^\star) \hat{\boldsymbol{P}}^{M\star}$ is a symmetric positive semi-definite $M \times M$ matrix.

The QP is strictly convex if matrix $\boldsymbol{Q}$ is positive definite. It is positive definite if and only if there exists a matrix $\mathring{\boldsymbol{Q}} \in \mathbb{R}^{\mathring{u} \times M}$, $M \leq \mathring{u}$, of full column rank M with $\boldsymbol{Q} = \mathring{\boldsymbol{Q}}^\top \mathring{\boldsymbol{Q}}$ (Gentle, 2007, Section 3.3.7).

Define nJ-dimensional vector $\boldsymbol{u}$ with

$$u_z = \begin{cases} 1, & \text{if } d_z^\star > 0 \\ 0, & \text{else} \end{cases}, \quad z = 1, \dots, nJ, \qquad (3.124)$$

and set $\mathring{u} = \sum_{z=1}^{nJ} u_z$. Let $\boldsymbol{U}$ be the diagonal matrix of $\boldsymbol{u}$, where all rows with row sum equal to zero are deleted, such that $\boldsymbol{U} \in \mathbb{R}^{\mathring{u} \times nJ}$. Then

$$\begin{aligned} \boldsymbol{Q} &= \hat{\boldsymbol{P}}^{M\star\top} \operatorname{diag}(2\boldsymbol{d}^\star) \hat{\boldsymbol{P}}^{M\star} \\ &= \hat{\boldsymbol{P}}^{M\star\top} \boldsymbol{U}^\top \operatorname{diag}(2\boldsymbol{U}\boldsymbol{d}^\star) \boldsymbol{U} \hat{\boldsymbol{P}}^{M\star} \\ &= (\hat{\boldsymbol{P}}^{M\star} \sqrt{2\boldsymbol{d}^\star})^\top \boldsymbol{U}^\top \boldsymbol{U} (\hat{\boldsymbol{P}}^{M\star} \sqrt{2\boldsymbol{d}^\star}) \\ &= \mathring{\boldsymbol{Q}}^\top \mathring{\boldsymbol{Q}}, \end{aligned} \qquad (3.125)$$

where $\mathring{\boldsymbol{Q}} = \boldsymbol{U}(\hat{\boldsymbol{P}}^{M\star}\sqrt{2\boldsymbol{d}^{\star}}) \in \mathbb{R}^{\mathring{u}\times M}$, diag($\boldsymbol{a}$) denotes the diagonal matrix of a vector $\boldsymbol{a}$. This implies that $\boldsymbol{Q}$ is positive definite as long as rank$(\mathring{\boldsymbol{Q}}) = M$. □

For an ensemble, rank$(\mathring{\boldsymbol{Q}}) = M$ holds unless at least two individual models give the same predictions. Before calculating the optimal ensemble weights, it should therefore be verified that no two individual models provide the same predictions. If two models provide the same predictions, it would not make sense to include both in a weighted ensemble anyway. Therefore, when the considered individual models lead different probability predictions, the stacking optimisation problem (3.113) with the weighted Brier score (3.117) as a loss function L is a strictly convex QP, each local optimum is then the global optimum.

3.6 Modelling and generation of monthly employment transitions

3.6.1 Choice of covariates

In this section, we take a closer look at the relationships of the potential covariates and the monthly employment status in the SIAB dataset. Based on this analysis, we then choose the concrete specification of GAMs in Section 3.6.2.

From the SIAB dataset, only those variables which are also present in the RIFOSS dataset can be used as covariates. The SIAB variables are given in Table 3.8, the RIFOSS variables are given in Table 3.2. The potential covariates are (with corresponding RIFOSS variable names in brackets): `ILO (EF29)`, `frau (EF46)`, `age (EF44)`, `schule (EF310)`, `ausbildung_gr (EF312)`, `ausbildung_imp (EF312)`. Furthermore, the `month` and `year` of an observation can be used as covariates. The simulations in Chapter 4 are based on the data generated in this chapter. For the analysis in Chapter 4, it is important that the generated employment data show realistic transitions on the person-level and realistic seasonal patterns on the aggregated level. Furthermore, it is important that the joint distributions of the monthly employment status and other variables like sex is preserved. For that, all variables common to the RIFOSS and SIAB dataset should be included in a model.

For the variables related to education, i.e. `schule`, `ausbildung_gr`, and `ausbildung_imp`, there are some discontinuities in the data in 2010-2012. Therefore, we restrict the SIAB dataset to observations in 2012-2017.

The definitions of the education variables in the SIAB dataset (`schule`, `ausbildung_gr`, `ausbildung_imp`) do not fully match those in the RIFOSS dataset (`EF310`, `EF312`), compare their descriptions in Antoni et al. (2019) and Research Data Centres of the Statistical Offices of the Federation and the Federal States (2018). The RIFOSS variable `EF310` approximately corresponds to the SIAB variable `school`. Likewise, the definitions of the RIFOSS variable `EF312` approximately corresponds to the SIAB variables `education_gr` and `education_imp`. As there are less missing values in variable `ausbildung_gr` (10.48%) than in variable ausbildung_imp (15.37%), variable ausbildung_imp is not further considered. To harmonize the SIAB and RIFOSS variables, two new dichotomous variables EF310_SIAB and EF312_SIAB are introduced in both datasets. Their definitions are shown in Table 3.14. The table shows which of the original SIAB and RIFOSS categories are assigned to the two harmonized categories for each of the two variables. After that, observations with missing values in EF310_SIAB (9.62% missings) or EF312_SIAB (10.48% missings) are deleted from the SIAB dataset. There remain 64, 024, 692 observations of 947, 897 persons in SIAB.

Table 3.14: Categories of original and harmonized SIAB and RIFOSS variables

RIFOSS variable EF310	SIAB variable schule	Harmonized variable EF310_SIAB
1-3, 6, 7	1, 4-6	1
4, 5	7-9	2

RIFOSS variable EF312	SIAB variable ausbildung_gr	Harmonized variable EF312_SIAB
1-7, 11, 12	1, 2	1
8-10	3, 4	2

For the prediction of monthly employment, it is particularly interesting to investigate the predictive power of employment lags, i.e. the employment status at previous months. Table 3.15 displays chosen employment patterns in the SIAB dataset. From the table, we can see that 80% of all observations have the same employment status in month t, $t-1$, $t-3$, $t-6$, $t-9$, and $t-12$. Furthermore, for about 97% of all observations the employment status in a month is the same as the month before. The employment status is thus relatively persistent over time and employment lags are a good predictor of the current employment status.

Table 3.16 shows the employment transitions in months t, $t-1$, and $t-12$. We use the table to see whether multiple employment lags provide more information about the employment status of a month than a single employment lag. In the table, constant employment categories in t, $t-1$, and $t-12$ are marked in bold. For example, 61.55% of observations have employment status 1 (employed) in a month t, the previous month ($t-1$), and the same month one year before ($t-12$). The table shows that not only the employment status in the previous month, but also the employment status in the previous year is an important predictor of the current employment status. For example, 5.05% of all observations are employed in months t and $t-1$, but not in LF in the previous year $t-12$. This suggests that in the models multiple employment lags should be used as predictors to capture realistic long-term employment transitions.

There are two non-categorical variables available for modelling the employment status in the SIAB dataset, the age of a person and the month of the observation. In Figures 3.3 and 3.4, the relative frequencies of the three employment categories are shown for age and

Table 3.15: SIAB: Employment patterns

Month	Pattern	Observations (in %)
t	1	69.30
	2	4.73
	3	25.96
	Σ	100.00
t-1 | t	11	67.95
	22	3.95
	33	24.71
	Σ	96.61
t-12 | t-9 | t-6 | t-3 | t-2 | t-1 | t	1111111	58.95
	2222222	1.27
	3333333	18.69
	Σ	78.91
Note: 1 employed, 2 unemployed, 3 not in LF		

Table 3.16: Employment transitions in month t, $t-1$, and $t-12$ (% of data)

		t-12		
t	t-1	1	2	3
1	1	**61.55**	1.35	5.06
	2	0.22	0.09	0.08
	3	0.41	0.04	0.51
2	1	0.26	0.07	0.05
	2	1.14	**1.85**	0.96
	3	0.11	0.11	0.18
3	1	0.57	0.04	0.25
	2	0.12	0.14	0.13
	3	3.60	1.08	**20.02**
Note: 1 employed, 2 unemployed, 3 not in LF				

`month`. The plots are shown separately for persons which were employed, unemployed, or not in LF the month before ($t-1$). Both for the `age` of a person and the `month` of the observation there is a non-linear effect on the relative frequencies of employment. The effects differ depending on the previous employment status, i.e. in column one, two, and three the non-linear relationships look different. From the figures, we decide to include `age` and `month` via non-linear effects in a GAM, with separate effects for each category of the previous employment status.

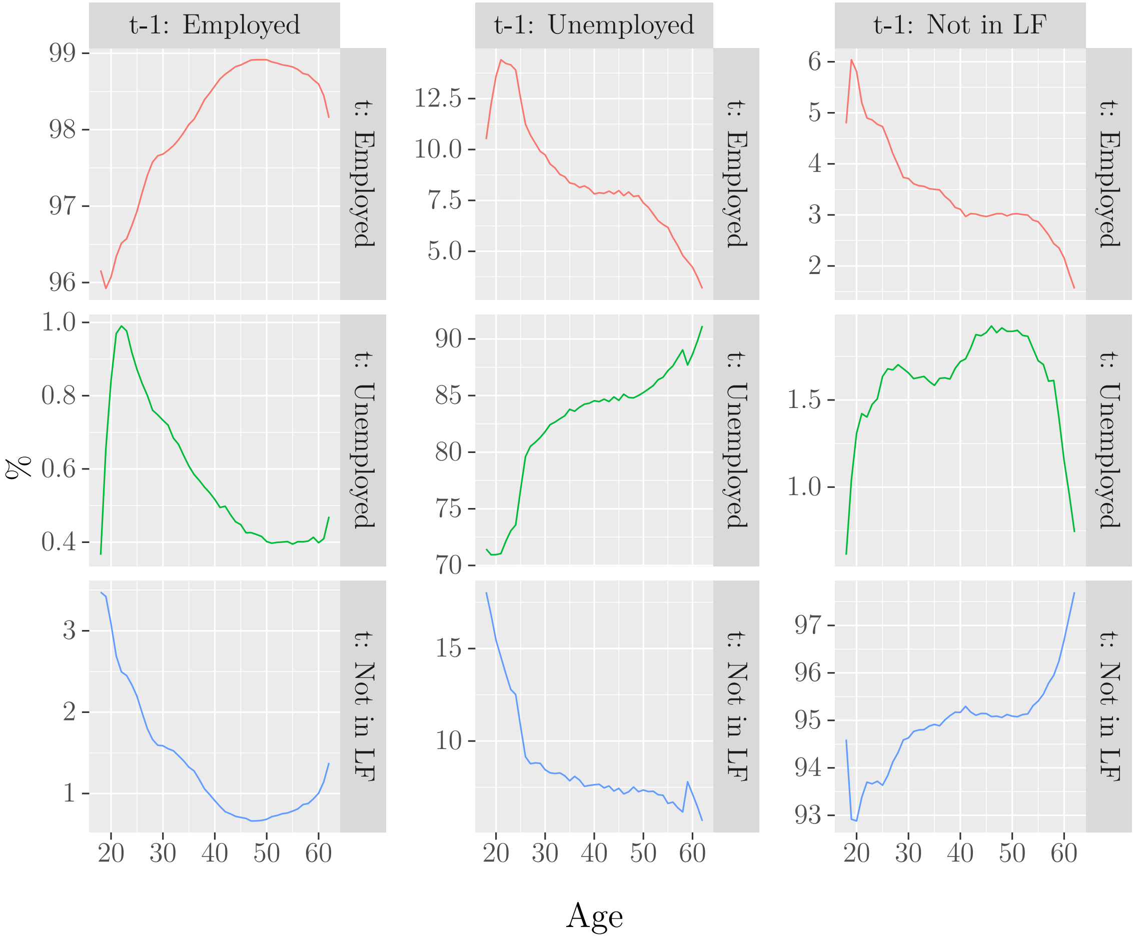

Figure 3.3: Employment aggregates per age (in %)

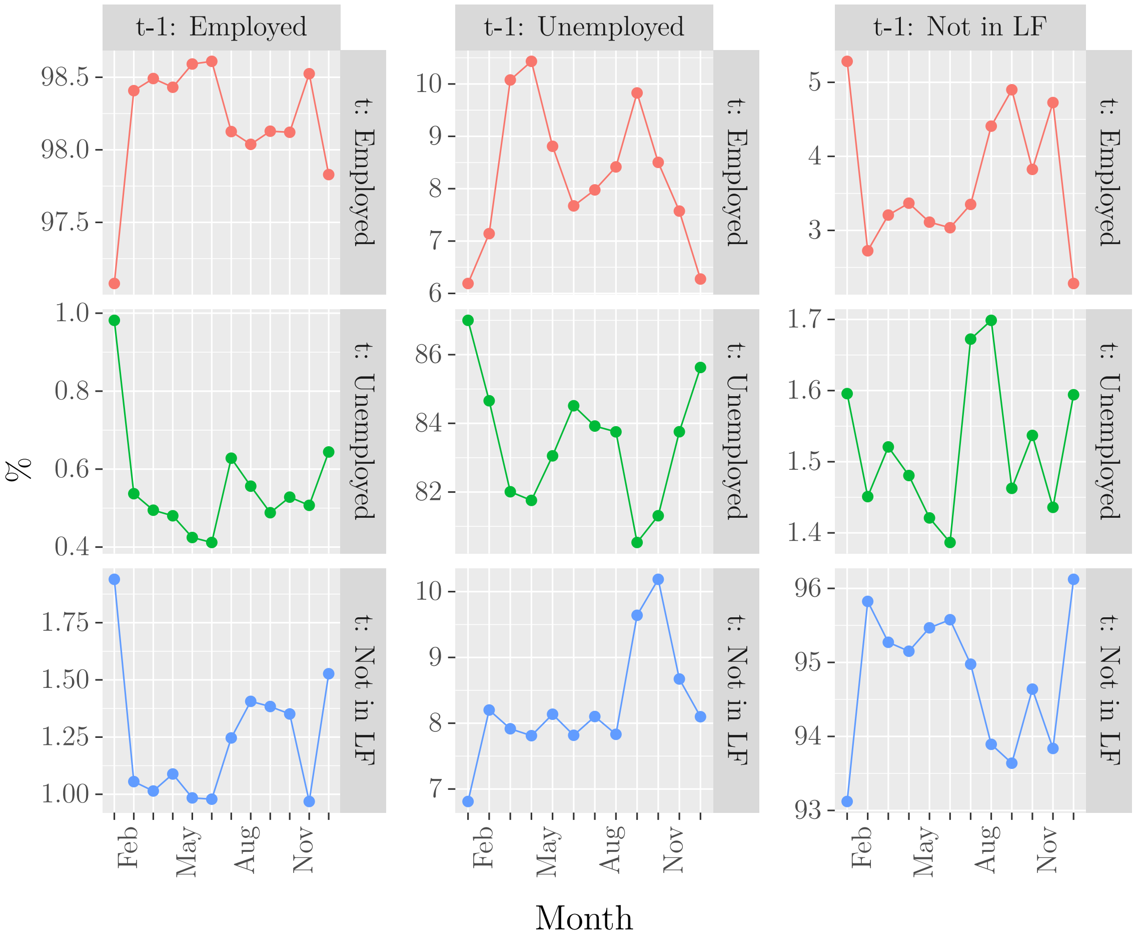

Figure 3.4: Employment aggregates per month (in %)

3.6.2 Modelling

In the following, we describe the calculation of the final models which are fitted to the SIAB dataset to model the monthly employment status. For modelling, we consider a subagging-stacking combination with GAMs as individual models and the proposed weighted Brier score (3.117) as a loss function in the stacking optimisation problem (3.113). The calculation of the individual GAMs (bagging), the calculation of the ensemble weights (stacking), and the evaluation of the individual and ensemble models should be based on different parts of the SIAB dataset. This can be done, for example, by cross-validation. However, since the SIAB dataset consists of many observations, we instead divide the SIAB dataset only once into modelling (bagging), optimisation (stacking), and evaluation data with percentages of 90%, 5%, and 5%. The data subsets are referred to as $SIAB^M$, $SIAB^O$, and $SIAB^E$. The rows of the $947,897$ persons in the SIAB dataset are randomly assigned to one of the three data subsets. Thereby, we make sure that there is no person for whom there is information in more than one data subset. The relative frequencies of employment, years of observation, and months of observation are the same in the data subsets.

The investigation of the SIAB dataset in Section 3.6.1 showed that it is useful to include several past employment categories as predictors. Auxiliary variables which correspond to past values of the dependent variable are called *lagged dependent variables*. In the context of microsimulation, the lagged employment status is often used as an auxiliary variable for modelling the employment status. For example, Leombruni and Richiardi (2006), Li and O'Donoghue (2012), and McLay et al. (2015), used lagged dependent variables for simulating employment transitions in microsimulations, Levell and Shaw (2016) used the lagged employment status of several past times.

A potential problem that arises from lagged dependent variables is *unobserved heterogeneity*. The SIAB dataset is a longitudinal dataset consisting of observations from different individuals at different time points. Therefore, it is plausible that there are cluster-specific effects such as person-specific random or fixed effects. For the definition of random effects, see the description in Section 2.4.2 in the context of LMMs. If there are person-specific effects in the data, which are not accounted for in the modelling process of a longitudinal dataset, the estimated coefficients are potentially biased (Richiardi & Poggi, 2014). If a dataset is composed of many clusters and one is interested in time-invariant effects such as sex, random instead of fixed effects should be used to account for these clusters (Honoré, 2002). However, a simultaneous consideration of a lagged dependent variable and cluster random effects is problematic as, by definition, the distribution of random effects is independent of the values of the other covariates. Furthermore, even if cluster-specific random effects are modelled, it is difficult to use these effects in out-of-sample prediction, although Richiardi and Poggi (2014) evaluated some methods for that purpose.

To circumvent the difficulties of including person-level random effects and lagged dependent variables jointly in a GAM, we apply the following procedure to the SIAB dataset. In the subagging-stacking procedure, for each individual model a sample of $SIAB^M$ is drawn and a chosen GAM is calculated on it. The M samples are drawn such that each person occurs at most once in a sample. Thereby, in the M samples there are no clusters of person observations. The existence of person-specific effects would then, in expectation, only lead to an increase of the model variance, as long as their distribution is symmetric. Also Richardson et al. (2018) argued that considering unobserved heterogeneity can be difficult and that the effect of neglecting unobserved heterogeneity is negligible when the focus is on predicting employment transitions, not analysing individual employment transitions.

We want to specify the functional form of the GAM, which we calculate in the individual models. In Section 3.6.1, we evaluated the choices and functional forms of the covariates. Additional analysis of different GAM specifications supported that including more lagged dependent variables gave better predictions. Furthermore, the consideration of several lagged dependent variables is also relevant for the investigations that follow in Chapter 4. So ideally, a GAM specification with several lags of employment is chosen. However, the RIFOSS dataset is a cross-sectional dataset and we augment it month by month by additional employment information. This means that, for example, for the first new month for which we want to generate the employment status in the RIFOSS dataset, only the employment status of the previous month is available as a predictor, no further lags of employment. When more monthly employment data is generated in the RIFOSS dataset, more employment lags are available as predictors. We therefore choose to specify six different GAMs, starting with only one lagged dependent variable, $t-1$, to all lagged dependent variables in vector $\tilde{\boldsymbol{t}} = ((t-1),(t-2),(t-3),(t-6),(t-9),(t-12))^\top$. For observation i, the additive predictor of the models is specified as

$$\begin{aligned}\eta_{jl}(\boldsymbol{x}_{ijl}) =& \beta_{j0} + \beta_{j1} I(\texttt{frau}_i = 2) + \beta_{j2} I(\texttt{EF310\_SIAB}_i = 2) \\ &+ \beta_{j3} I(\texttt{EF312\_SIAB}_i = 2) + \sum_{q=1}^{5} \beta_{j(3+q)} I(\texttt{year}_i = (2011+q)) \\ &+ \sum_{e=1}^{6} I(l \geq e) \sum_{k=1}^{2} \beta_{j(9+(2e)+k)} I(\texttt{ILO}_i^{\tilde{t}_e} = k) \\ &+ \sum_{k=1}^{3} I(\texttt{ILO}_i^{t-1} = k) s_{j1k}(\texttt{age}_i) \\ &+ \sum_{k=1}^{3} I(\texttt{ILO}_i^{t-1} = k) s_{j2k}(\texttt{month}_i), \quad j = 1,2,\ l = 1,\ldots,6.\end{aligned} \tag{3.126}$$

β_{j0} is an intercept, I is an indicator function, `frau`, `EF310_SIAB`, and `EF312_SIAB` are dichotomous variables with values in $\{1,2\}$. Index $l = 1,\ldots,6$ represents the six GAM specifications, which only differ in the availability of employment lags. We use the same model specification for all employment categories.

Variable `year` only takes six different values. As a linear effect of `year` was not supported by the data, we added dummy variables for five of the six values of `year` to capture year-specific effects. $s_{j2k}(\texttt{age})$ is a cubic regression spline with 20 evenly spaced knots. $s_{j3k}(\texttt{month})$ is a cyclical cubic regression spline to capture seasonal month effects with 10 fixed knots which are equally spaced in $[1,13]$, as proposed in Wood (2017) for monthly cyclical splines. Considering knots in $[1,13]$ instead of $[1,12]$ ensures that January and

December are not modelled as the same month. Both splines are defined separately for the three categories of employment in month $t-1$ as was suggested from the analysis of Figures 3.3 and 3.4.

We apply a subagging-stacking procedure to the SIAB subsets $SIAB^M$ and $SIAB^O$ as shown in Algorithm 3.4. On each of the $M = 50$ samples of $SIAB^M$ GAMs are calculated. The GAMs are specified according to (3.126). That is, on each sample six different GAMs are calculated. For the calculation of the GAMs, `R` (R Core Team, 2020) function `gam` from the `mgcv` package (Wood, 2003, 2004, 2011, 2017; Wood et al., 2016) is used. With the GAMs, optimal ensemble weights $\boldsymbol{w}^{opt}$ are calculated with stacking, based on $SIAB^O$. For stacking, the proposed weighted Brier score (3.117) is used as a loss function. For the weighted Brier score, category weights $\boldsymbol{d}$ have to be set. The relative frequencies of employment in SIAB are around 69.3% (employed), 4.7% (unemployed), and 26% (not in LF). To take into account the imbalance of the data and focus especially on employed and unemployed, as they are in the focus in Chapter 4, we chose $\boldsymbol{d} = (0.1, 0.6, 0.3)^\top$. For stacking, the `R` (R Core Team, 2020) function `solve.QP` from the `quadprog` package (Turlach & Weingessel, 2019) is applied, which uses the dual method of Goldfarb and Idnani (1982, 1983).

Algorithm 3.4 Calculation of prediction models based on the processed SIAB dataset

```
1. Take $SIAB^M$.
      For $m = 1, \dots, 50$ do
        i.  Draw a random sample of size $500,000$, called $s_m$. The sample is drawn
            such that it contains maximum one data row per person.
        ii. For $l = 1, \dots, 6$, calculate a multinomial GAM (3.126) based on $s_m$, called
            $\hat{f}^*_{l,m}$.
2. Take $SIAB^O$.
      For $l = 1, \dots, 6$ do
        a) Solve the stacking optimisation problem (3.113) with weighted Brier score
           $Bw(\boldsymbol{d}, \hat{f}^E_l)$ (3.117) as loss function $L$, returning weight vector $\boldsymbol{w}^{opt}_l \in \mathbb{R}^{50}$.
           The inputs of $Bw$ are $\boldsymbol{d} = (0.1, 0.6, 0.3)^\top$, $\hat{f}^E_l = \sum_{m=1}^{50} w_m \hat{f}^*_{l,m}$.
3. For $l = 1, \dots, 6$, the procedure results in the following models:
      • Individual models $\hat{f}^*_{l,m}$, $m = 1, \dots, M$,
      • Ensemble model with equal weights $\hat{f}^E_{l,\boldsymbol{w}^{equal}}$, $w^{equal}_m = 1/50$, $m = 1, \dots, 50$,
      • Ensemble with optimised weights $\hat{f}^E_{l,\boldsymbol{w}^{opt}_l}$.
```

We want to compare the predictive performance of the different models returned by Algorithm 3.4. Algorithm 3.4 returns 50 individual GAMs, an ensemble with equal weights, and an ensemble with optimised weights for each of the six scenarios of lagged dependent variables in (3.126). By comparing the performances, we can see whether the proposed subagging-stacking combination gives better probability predictions than the other models. To have a fair comparison, we calculate evaluation measures for each single category. That is, we calculate the category-specific Brier skill scores BSS_j based on $SIAB^E$, where, for

$j = 1, \dots, J$,

$$BSS_j(\hat{f}) = B_j(\hat{f})/B_j(\hat{f}^{intercept}) * 100, \tag{3.127}$$

$$B_j(\hat{f}) = \frac{1}{n}\sum_{i=1}^{n}(I(y_i = j) - pr(Y_i = j|\boldsymbol{x}_i))^2, \tag{3.128}$$

and $\hat{f}^{intercept}$ is an intercept only model.

The predictive performances of the models resulting from Algorithm 3.4 are displayed in Table 3.17, They are shown for the models with $\tilde{t}_6$ in (3.126), i.e. the GAM specification with most available employment lags. We therefore skip index l in Table 3.4. In row one, the mean scores of the $M = 50$ individual models is displayed. For all models, all category-specific values are substantially lower than 100, indicating that the models perform much better than an intercept only model. This is especially visible for categories *employed* and *not in LF*.

Among the different models, the optimally weighted ensemble $\hat{f}^E_{\boldsymbol{w}^{opt}}$ performs best for all three categories, followed by the equally weighted ensemble $\hat{f}^E_{\boldsymbol{w}^{equ}}$ and average individual model. The absolute performance differences are, however, small.

Comparing the equally weighted ensemble $\hat{f}^E_{\boldsymbol{w}^{equ}}$ and the optimally weighted ensemble $\hat{f}^E_{\boldsymbol{w}^{opt}}$, we note the following. The calculation of optimised ensemble weights $\boldsymbol{w}^{opt}$ is fast as the optimisation problem is quadratic, compare Proposition 3.1. The ensemble $\hat{f}^E_{\boldsymbol{w}^{opt}}$ performs slightly better then the ensemble $\hat{f}^E_{\boldsymbol{w}^{equ}}$, for all three employment categories. Next to the performance, there is an additional advantage of the ensemble model $\hat{f}^E_{\boldsymbol{w}^{opt}}$ over $\hat{f}^E_{\boldsymbol{w}^{equ}}$. In $\boldsymbol{w}^{opt}$ there are only 7 of the 50 values greater than 10^{-15}. To put it differently, the ensemble model $\hat{f}^E_{\boldsymbol{w}^{opt}}$ requires storing only 7 out of 50 individual models. It therefore allows for faster prediction than $\hat{f}^E_{\boldsymbol{w}^{equ}}$, which is especially useful for creating synthetic data with many observations. We therefore use ensemble models $\hat{f}^E_{l,\boldsymbol{w}^{opt}}$, $l = 1, \dots, 6$ to generate longitudinal employment in the RIFOSS dataset in the following Section 3.6.3.

Table 3.17: Prediction performance of different models based on $SIAB^E$

	Employed	Unemployed	Not in LF
	$j = 1$	$j = 2$	$j = 3$
$1/M \sum_{m=1}^{M} BSS_j(\hat{f}^*_m)$	11.477	31.357	12.628
$BSS_j(\hat{f}^E_{\boldsymbol{w}^{equal}})$	11.473	31.336	12.623
$BSS_j(\hat{f}^E_{\boldsymbol{w}^{opt}})$	11.472	31.307	12.618

3.6.3 Generation of monthly employment transitions in the RIFOSS dataset

The RIFOSS dataset is cross-sectional. The employment status in the RIOFSS dataset is treated as the employment status in January 2012. To generate additional monthly values of the employment variable `EF29` in the RIFOSS dataset, the following procedure is applied sequentially for all months in February 2012 to December 2014. All other variables, also the age of the persons, stay fixed. For each month, the optimally weighted ensemble model $\hat{f}^E_{l,\boldsymbol{w}_l^{opt}}$, $l = 1, \ldots, 6$, calculated in Section 3.6.2, with most available previous information on variable `EF29` is selected. For example, for creating employment categories for February 2012, only January 2012 is available as lagged employment information. Therefore, the ensemble model $\hat{f}^E_{1,\boldsymbol{w}_1^{opt}}$ is applied. For the next month, employment categories for January and February 2012 are available, wherefore the ensemble model $\hat{f}^E_{2,\boldsymbol{w}_2^{opt}}$ is applied.

With the chosen ensemble model, probability predictions of the employment categories are calculated in the RIFOSS dataset. For persons aged less than 15 and more than 80, the probability of being not in the LF is set to one, for persons aged over 74 the probability of being unemployed it set to zero. The cut-off at 80 years is used as the oldest employed persons in the RIFOSS dataset are 80 years old. Concrete employment categories are drawn from the probability predictions according to the multinomial distribution. In the SIAB dataset, there are persons aged 17-62, in the RIFOSS dataset the persons in the LF are aged 15-80. Therefore, for persons aged 15-17 and 62-80, the generated employment transitions should be treated with caution.

We note that in the context of microsimulation or data generation, probabilities or categories are often aligned, for example to values from official statistics. We refer to Burgard et al. (2021b), Li and O'Donoghue (2014), and Stephensen (2016) for information on different alignment approaches. We do not use any alignment in the data generation for two reasons. Firstly, for the studies in Chapter 4, it is more important to preserve the joint distributions of the variables than to return actual official aggregates. Secondly, the aggregates that we would otherwise have used for alignment can be used to validate the generated data at the aggregate level.

We take a look at the generated monthly employment data in the RIFOSS dataset. Figure 3.5 displays the absolute frequencies of the generated employment categories per month. Both seasonal patterns and a trend are visible in the generated RIFOSS employment data. This is similar to Figure 3.1 which showed the monthly aggregates in the SIAB dataset in 2007-2018.

We want to examine whether the relationships of the variables in the SIAB dataset were preserved in the RIFOSS dataset. For that, the RIFOSS aggregates are added to Figures 3.3 and 3.4, resulting in Figures 3.6 and 3.7. As there are only persons aged 17-62 in the SIAB dataset, the RIFOSS aggregates in the figures are also given for persons aged 17-62. We seen that through modelling, most of the non-linear relationships in the SIAB dataset could be replicated in the data generation, both for the age and the month of an observation.

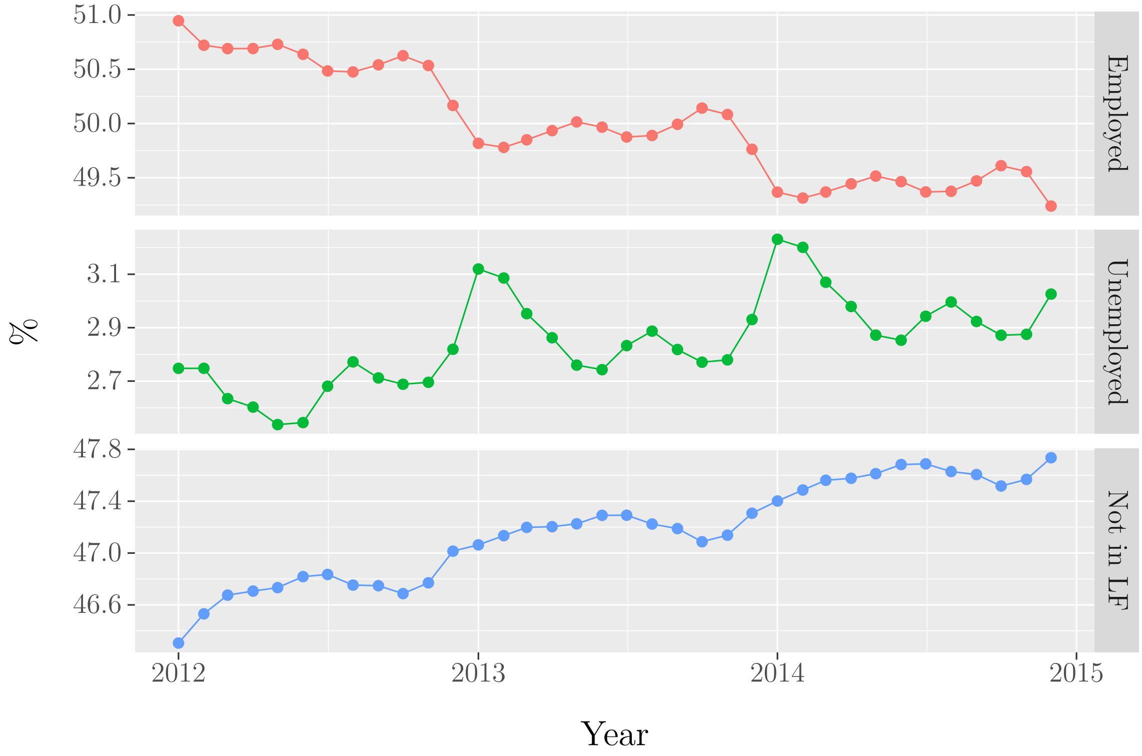

Figure 3.5: Employment aggregates in extended RIFOSS dataset

The employment patterns of SIAB and RIFOSS dataset can be compared by extending Table 3.15 by the RIFOSS aggregates of persons aged 17-62, resulting in Table 3.18. The persistence in employment at the person-level could be preserved in the generated data. About 85.04% of all observations in RIFOSS have the same employment status in months t, $t-1$, $t-2$, $t-3$, $t-6$, $t-9$, and $t-12$.

Next, we compare the aggregates of the generated data for the single years 2012-2014. This is particularly interesting as no alignment was used for the generation. Figure 3.8 displays the absolute number of persons being employed, unemployed, and not in LF for chosen age classes in the three years. In addition to the RIFOSS aggregates, official Destatis information[6] and the relative frequencies of the SIAB dataset are added. Even though no alignment was used, the generated RIFOSS employment statistics reflect the official employment statistics by age group. The quality of the generated data, however, decreases with the years. We note that the number of unemployed persons over 65 years of age increases with the years and is far off the official statistics in 2014. In the SIAB dataset there were only persons aged 17-62. Therefore, the data generation is not expected to perform well for persons aged under 17 or over 62. In Figure 3.8, we therefore see why the prediction of population groups which were not present in the modelling data should be treated with caution. As we, however, do not focus on specific age groups in the data analysis in Chapter 4, the data is considered sufficient for the three generated years.

[6] GENESIS-Online. Code: 12211-0002. https://www-genesis.destatis.de/genesis/online.

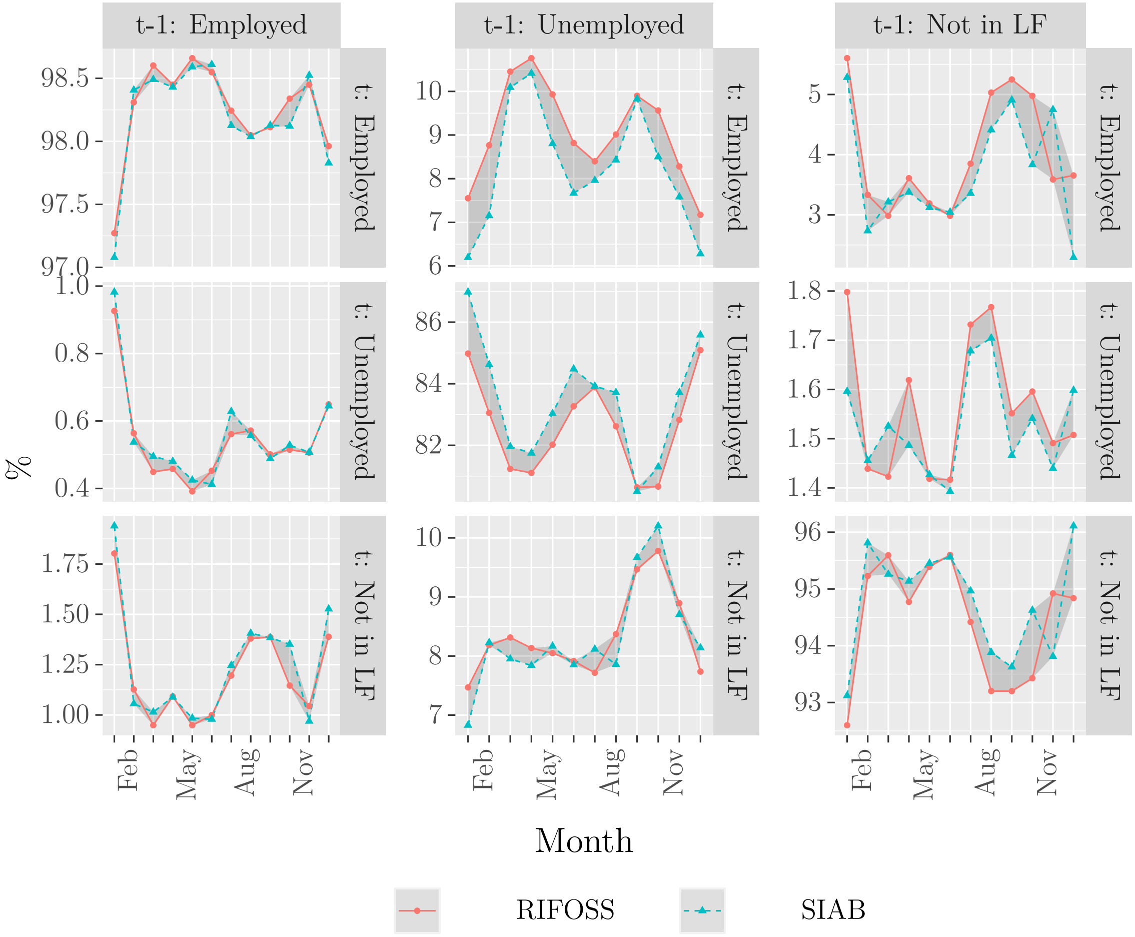

Figure 3.6: Employment aggregates per age (in %)

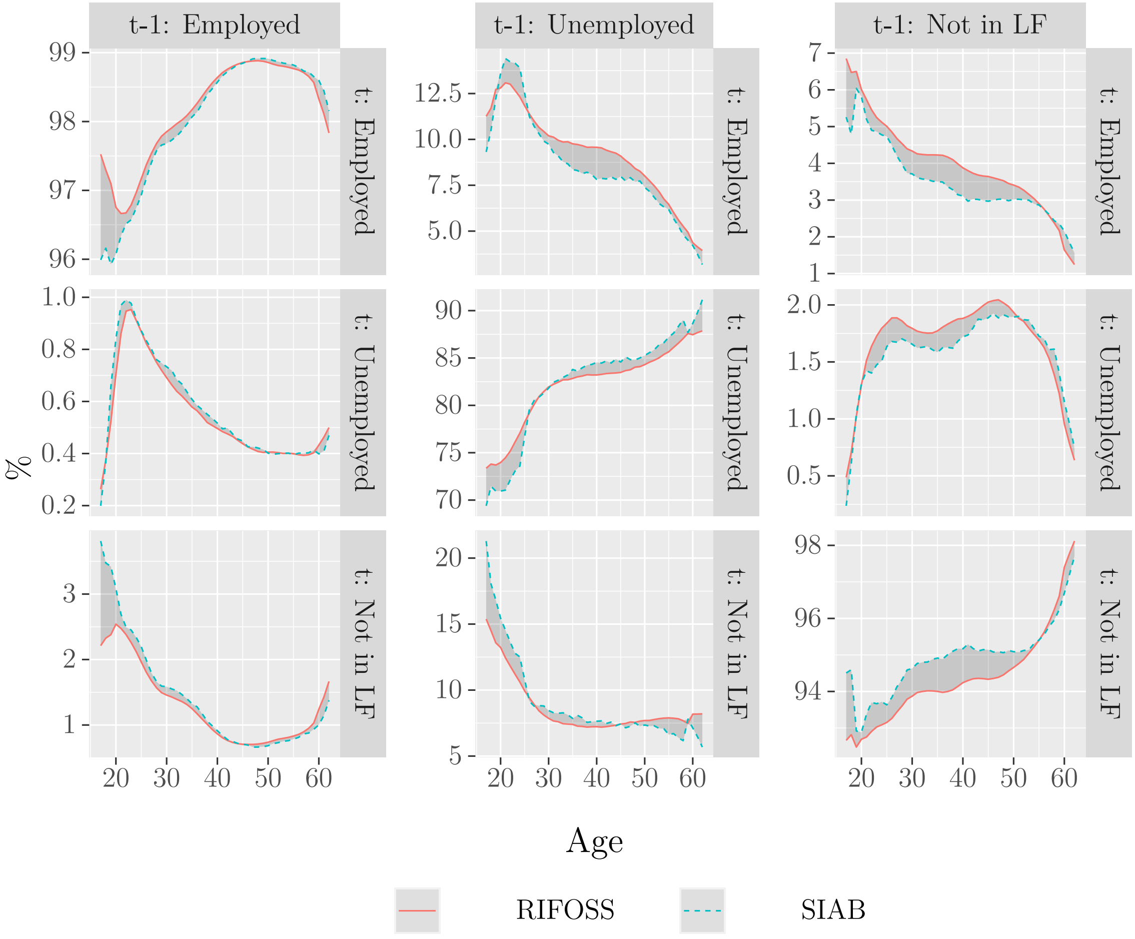

Figure 3.7: Employment aggregates per month (in %)

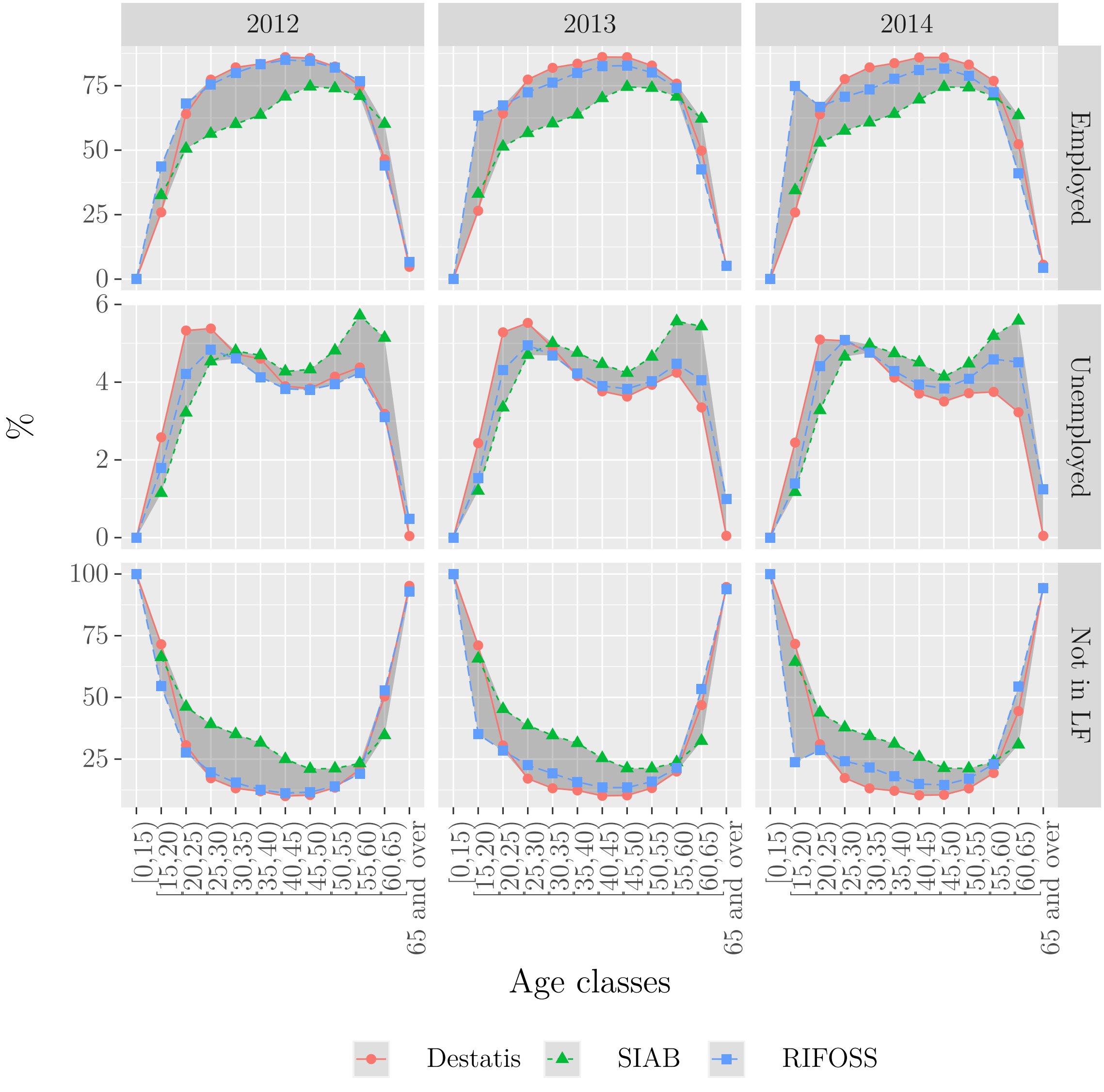

Figure 3.8: Relative frequencies (in %) of employment aggregates per age class

Table 3.18: Employment patterns

Month	Pattern	Observations (in %)	
		SIAB	RIFOSS
t	1	69.30	49.68
	2	4.73	2.93
	3	25.96	47.39
	Σ	100.00	100.00
t-1 | t	11	67.95	48.80
	22	3.95	2.43
	33	24.71	46.50
	Σ	96.61	97.73
t-12 | t-9 | t-6 | t-3 | t-2 | t-1 | t	1111111	58.95	42.67
	2222222	1.27	0.95
	3333333	18.69	41.42
	Σ	78.91	85.04

Note: 1 employed, 2 unemployed, 3 not in LF

Figure 3.9 shows the percentages of employed and unemployed persons as part of the active population, by sex. The statistics are shown for the generated RIFOSS dataset, the official Destatis data[7], and the SIAB dataset. The seasonal patterns in the aggregated data are quite similar in the three datasets although their magnitudes differ.

To summarise, the generated monthly RIFOSS employment transitions reflect the variables' relationships on the person-level as well as the person-level transitions of the SIAB dataset. On the aggregated level, both seasonal patterns and a trend of employment categories are visible, which is similar to the aggregated information from official statistics and the SIAB dataset.

[7] GENESIS-Online. Code: 13231-0002. https://www-genesis.destatis.de/genesis/online.

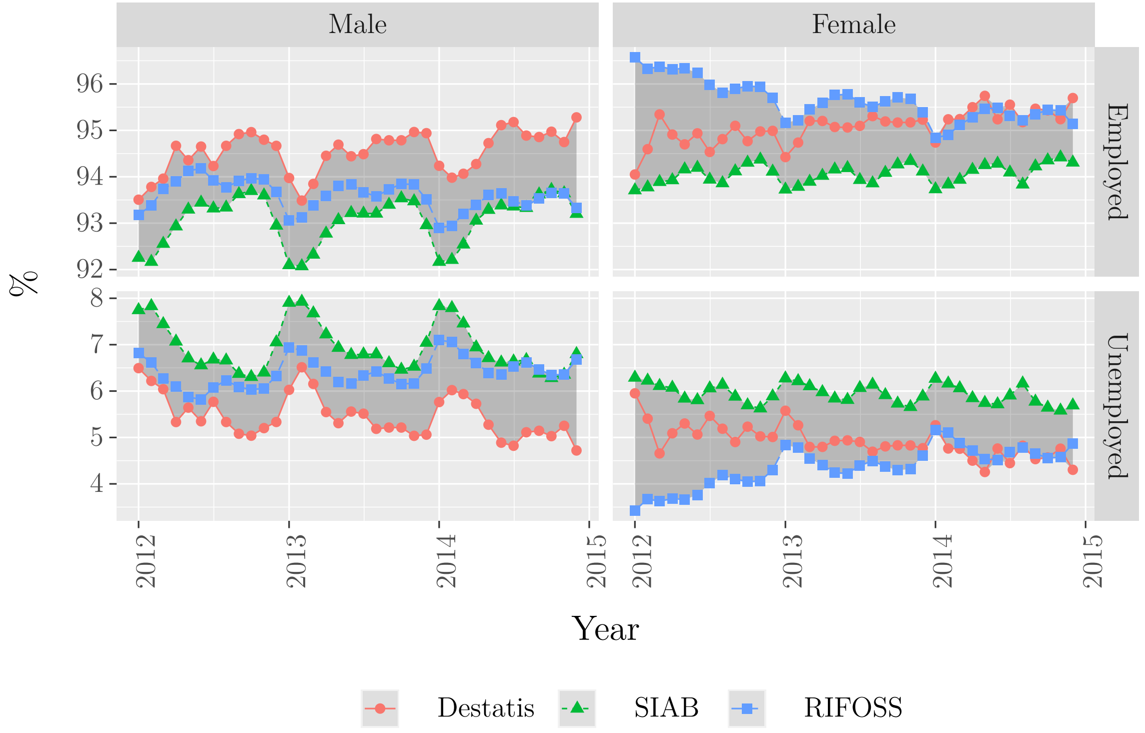

Figure 3.9: Active population: Percentages of employed and unemployed per month

3.7 Summary and outlook

In this chapter, we synthetically expanded the cross-sectional RIFOSS dataset by a monthly employment status for three years. For that, we used prediction models for monthly employment categories based on the SIAB dataset. The generated dataset serves as the simulation dataset in the design-based study in Chapter 4. Several steps were necessary to achieve the expansion of the RIFOSS dataset.

Before we could use the SIAB dataset to calculate prediction models for monthly employment information, we had to edit and aggregate the information in the dataset. We imputed missing values, removed data inconsistencies, aggregated the data to monthly data, derived employment categories in accordance to the ILO definition, and validated the final data. Based on the processed SIAB dataset, we calculated prediction models for the monthly employment transitions. Then, we used these models to generate monthly employment transitions in the RIFOSS dataset. For modelling the imbalanced employment status in the SIAB dataset, we combined GAMs with ensemble techniques subagging and stacking. For the evaluation of probability predictions under imbalanced data, we pointed out the shortcomings of the Brier score and the stratified Brier score and proposed what we call the weighted Brier score as a compromise between the two. We furthermore showed that the stacking optimisation problem with the weighted Brier score as loss function is a quadratic optimization problem.

In the application to the SIAB dataset, the proposed subagging-stacking combination showed small improvements in the quality of the probability predictions compared to an equally weighted ensemble or the individual input GAMs. We generated the monthly employment status in the RIFOSS dataset with the subagging-stacking GAMs and validated the generated data. The validation showed that the generated monthly RIFOSS employment data reflect both the patterns in the SIAB dataset on the person-level and the aggregated level and aggregate statistics published by the German statistical institute.

For future research, it would be interesting to consider additional techniques for handling imbalanced categorical data, such as balancing techniques, see e.g. Fernández et al. (2018, Section 7.3). Those could be used to adjust the frequencies of the modelling data already to the target data. In addition, a potential future research area would be to find sparse solutions to the stacking optimisation problem, e.g. by inclusion of the ℓ_q quasi-norm, $0 < q < 1$ penalty, compare Xu et al. (2012) and Zeng et al. (2014).

Chapter 4

Composite Estimation in the German Microcensus

4.1 Introduction

The design of the German Microcensus underwent major changes in 2020 (Bihler & Zimmermann, 2016; Destatis, 2021; Hochgürtel, 2013; Hundenborn & Enderer, 2019; Riede, 2013). Previously, the surveys *Microcensus* with the integrated *Labour Force Survey* (LFS), the *European Statistics on Income and Living Conditions* (EU-SILC), and the survey on *Information and Communication Technologies* (ICT) were conducted separately. As of 2020, these household surveys are conducted within an integrated system called Microcensus. Furthermore, the rotation pattern, i.e. the time sequence of a total of four interviews per sampled household, has changed. Previous to 2020, all households sampled in the Microcensus were surveyed once a year in a fixed reference week for a total of four years. Starting in 2020, there are two different rotation schemes within the Microcensus: The sampled households which are assigned to the LFS module have a different rotation scheme than all other sampled households.

Rotating panel surveys are surveys with multiple interviews, where at each time point both new population units are sampled into the survey and other population units, which already had at least one previous interview, rotate out of the survey. We refer to Kalton (2009) for a general overview of such surveys. The Microcensus, both in its old and new design, is an example of such a survey. Rotating panel surveys result in partially overlapping samples of different time points. For each time point, information from previous samples is available for a subset of the respondents of the sample. The design of the rotating panel survey determines the magnitude of the overlaps. *Composite estimators* are estimation methods which use the information of the partial overlaps with previous samples in the estimation process. Particularly for employment statistics, the inclusion of this additional information can lead to more efficient estimators since the employment status is typically rather stable over time. The estimators differ in how they handle the partially available prior information from previous interviews. In this chapter, we examine the applicability and performance of composite estimators in the new design of the Microcensus through design-based simulation studies. In the course of this, we also present adjustments of the methods to account for regionally heterogeneous sample overlaps, a particular feature of the new Microcensus design. The focus of the study is the regional (NUTS2-level) estimation of monthly and quarterly employment statistics of employed and unemployed persons such as totals and changes.

The chapter is structured as follows. In Section 4.2, we give an overview of rotating panel surveys and composite estimators. In Section 4.3, the design of the German Microcensus is described with emphasis on the design changes starting from 2020. We analyse the applicability of composite estimators in the Microcensus and present adjustments to the formulas of the composite estimators to allow for regionally heterogeneous sample overlaps. Section 4.4 presents a design-based simulation study based on the RIFOSS dataset, which was extended with longitudinal employment information in Chapter 3. In the simulation study, the adjusted composite estimators are compared for monthly and quarterly estimation of employment statistics at the NUTS2-level. The chapter closes with a summary and outlook in Section 4.5.

4.2 Estimation in rotating panel surveys

4.2.1 Rotating panel surveys

Official national household surveys are usually repeated at certain intervals, e.g. annually. Steel and McLaren (2009) gave an overview of different designs and estimation methods for surveys which are repeated over time and are taken as the main source for the following description. In repeated surveys, the sampled households are often interviewed not just once, but multiple times at fixed intervals. Monthly *rotating panel surveys* are those in which each month both new households are sampled into a survey and households leave the survey after a fixed number of interviews. The *rotation pattern* determines the timing of the interviews. In an in-for-x design, sampled household are interviewed for x successive time points and then rotate out of the survey. In an x-(y)-z design, sampled households are interviewed for x successive time points, pause for y successive time points, and are again interviewed for z successive time points. With a rotational pattern, a survey can be divided into different *rotation groups*. A rotation group contains a group of sample units that is interviewed within the same interview cycle. Rotating panel surveys result in sample overlaps. For example, in an in-for-6 months design 5/6 of the households interviewed in a month were also interviewed in the previous month. For simplicity, in the following description we assume monthly surveys, where within a month there is no sample overlap, but between different months there is. The sample overlaps lead to correlated monthly estimates. The consequences of these correlations for the HT (2.12) and the GREG (2.20) estimator are described below.

We extend the notation of the HT and GREG estimator, presented in Sections 2.3.1 and 2.3.2, for longitudinal surveys. Consider a fixed population U of size N. That is, we consider the population to stay fixed for different time points. In each month t, units are sampled from U without replacement according to a specified sampling design. The first- and second-order inclusion probabilities are denoted by π_{kt} and π_{klt}, for all $k, l \in U$, for month t. Sample s_t denotes the set of sampled elements of U in month t. The design weights are given by the inverse first-order inclusion probabilities and denoted by d_{kt}, for all $k \in s_t$. We consider a variable of interest Y which takes different values for the population U in each month t with population total $\tau_{y_t} = \sum_{k \in U} y_{kt}$. The HT estimator of population total τ_{y_t} is given by

$$\hat{\tau}_{y_t}^{\mathrm{HT}} = \sum_{k \in s_t} y_{kt} d_{kt}. \tag{4.129}$$

The GREG estimator of population total τ_{y_t} is given by

$$\hat{\tau}_{y_t}^{\mathrm{GREG}} = \sum_{k \in s_t} y_{kt} w_{kt} \tag{4.130}$$

with

$$w_{kt} = d_{kt} \left(1 + (\boldsymbol{\tau}_{x_t} - \hat{\boldsymbol{\tau}}_{x_t})^\top \left(\sum_{k \in s_t} \boldsymbol{x}_{kt} \boldsymbol{x}_{kt}^\top d_{kt} \right)^{-1} \boldsymbol{x}_{kt} \right), \tag{4.131}$$

where $\boldsymbol{x}_{kt}$ is a vector of covariate information for sampling unit $k \in s_t$ of length p with corresponding known population totals $\boldsymbol{\tau}_{x_t}$.

Rotating panel surveys are commonly applied in the context of LFSs, where the interest is in different employment statistics such as level estimates and estimates of change. The total change of a variable Y in t to t' is given by

$$\Delta_{y_{tt'}} = \tau_{y_t} - \tau_{y_{t'}}. \tag{4.132}$$

It can be estimated by plugging in the corresponding HT or GREG estimates of the totals. For example, for the HT estimator the change is estimated by $\hat{\Delta}_{y_{tt'}} = \hat{\tau}_{y_t} - \hat{\tau}_{y_{t'}}$. Although a sample overlap in t to t' does not influence the formula of the change estimators, it influences its variance. The variance of the change estimator is given by (Steel & McLaren, 2009, p. 293)

$$\begin{aligned} \operatorname{Var}(\hat{\Delta}_{y_{tt'}}) &= \operatorname{Var}(\hat{\tau}_{y_t}) + \operatorname{Var}(\hat{\tau}_{y_{t'}}) - 2\sqrt{\operatorname{Var}(\hat{\tau}_{y_t}) \operatorname{Var}(\hat{\tau}_{y_{t'}})} \operatorname{Corr}(\hat{\tau}_{y_t}, \hat{\tau}_{y_{t'}}) \\ &= \operatorname{Var}(\hat{\tau}_{y_t}) + \operatorname{Var}(\hat{\tau}_{y_{t'}}) - 2 \operatorname{Cov}(\hat{\tau}_{y_t}, \hat{\tau}_{y_{t'}}). \end{aligned} \tag{4.133}$$

The covariance of two HT estimators, $\operatorname{Cov}(\hat{\tau}_{y_t}, \hat{\tau}_{y_{t'}})$, is given in (2.17) and can similarly be given for the GREG estimator considering (2.23). Without sample overlap, i.e. with $\operatorname{Corr}(\hat{\tau}_{y_t}, \hat{\tau}_{y_{t'}}) = 0$, the variance of the change estimator is about double the variance of one of the total estimators. When the variable of interest Y is positively correlated over time and there is a sample overlap, $\operatorname{Corr}(\hat{\tau}_{y_t}, \hat{\tau}_{y_{t'}}) > 0$ holds. From (4.133), we see that a sample overlap can significantly reduce the variance of the change estimator. The exact magnitude of the variance reduction depends on the interplay of the correlation of Y over time and the magnitude of the sample overlap. The higher the sample overlap and the higher the correlation of Y in t to t', the higher the variance reduction of the change estimator. With the same argumentation, we can also see that a sample overlap between t and t' can lead to an increase in variance when estimating the sum or average of τ_{y_t} and $\tau_{y_{t'}}$, compare also Steel and McLaren (2009, pp. 294–295). For employment and unemployment the correlation over time is typically highly and moderately positive respectively, see e.g. Gambino et al. (2001). For variance estimation of the change estimator in rotating panel survey, we refer to Berger and Priam (2016).

Based on the correlations of the monthly estimates obtained from rotating panel surveys, also the *best linear unbiased estimator* (BLUE), introduced by Yansaneh and Fuller (1998), can be constructed. It is an estimator, derived to have optimal theoretical properties. Let T be the total number of time points for which estimates are available and R be the number of rotation groups. Define $\hat{\boldsymbol{\tau}}_{\boldsymbol{y}}^{TR} = (\hat{\tau}_{y11}, \ldots, \hat{\tau}_{y1R}, \ldots, \hat{\tau}_{yT1}, \ldots, \hat{\tau}_{yTR})^\top$ as the vector of rotation group specific design-unbiased estimates of true population totals $\boldsymbol{\tau}_y^T = (\tau_{y1}, \ldots, \tau_{yT})^\top$. Define the covariance matrix of the design-unbiased estimates as $\mathrm{Var}(\hat{\boldsymbol{\tau}}_{\boldsymbol{y}}^{TR}) = \dot{\boldsymbol{V}}$ and matrix $\dot{\boldsymbol{X}} = \mathbf{1}_R \otimes \boldsymbol{I}_T$, where $\mathbf{1}_R$ is a vector of 1 of length R, $\otimes$ denotes the Kronecker product, and $\boldsymbol{I}_T$ denotes the $T \times T$ identity matrix. The BLUE estimator of population totals $\boldsymbol{\tau}_y^T$ is then given by

$$\hat{\boldsymbol{\tau}}_y^{\mathrm{BLUE}} = \left(\dot{\boldsymbol{X}}^\top \dot{\boldsymbol{V}}^{-1} \dot{\boldsymbol{X}}\right)^{-1} \dot{\boldsymbol{X}}^\top \dot{\boldsymbol{V}}^{-1} \hat{\boldsymbol{\tau}}_{\boldsymbol{y}}^{TR} \tag{4.134}$$

with variance

$$\mathrm{Var}(\hat{\boldsymbol{\tau}}_y^{\mathrm{BLUE}}) = \left(\dot{\boldsymbol{X}}^\top \dot{\boldsymbol{V}}^{-1} \dot{\boldsymbol{X}}\right)^{-1}. \tag{4.135}$$

Note that (4.134) is the generalised least squares solution, similar to $\hat{\boldsymbol{\beta}}^{\mathrm{BLUE}}$ (2.31), discussed in the context of LMMs in Section 2.4.2. A multivariate version of the BLUE, i.e. for more than one variable of interest, is given in Bonnéry et al. (2020).

Despite its theoretical optimality, the BLUE is not frequently applied in rotating panel surveys. The dimensions of the matrices which need to to be stored for the BLUE increase each time the survey is conducted and require the inversion of ever-growing matrices. Furthermore, the covariance matrix $\dot{\boldsymbol{V}}$ needs to be estimated from sample data. The optimality of the BLUE, however, only holds for the exact covariance matrix $\dot{\boldsymbol{V}}$. In a design-based simulation study, Bonnéry et al. (2020) showed that when the covariance matrix $\dot{\boldsymbol{V}}$ is estimated from sample information, the performance of the BLUE estimator can be low. They argue in favour of the use of composite estimators. Furthermore, the efficiency of composite estimators can be close to the efficiency of the BLUE in practical applications (Steel & McLaren, 2009). In addition, each new time point at which the BLUE estimator is calculated results in a revision of all previous estimates. There are several extensions to the BLUE which are especially designed to circumvent growing matrices. Yansaneh and Fuller (1998) reformulated the BLUE as a recursive estimator. A BLUE based on a fixed time window is proposed in Bell (1999, 2001). Due to its disadvantages compared to other estimators and its limited application in rotating panel surveys, the BLUE is not discussed further and we focus on composite estimators instead.

In the context of rotating panel surveys and composite estimators, we note that often the presence and impact of a potential *rotation group bias* (RGB), also called *time-in-survey bias*, is examined. The RGB refers to systematic differences between different rotation groups and was first investigated by Bailar (1975). The RGB can influence certain estimators like composite estimators. It is specific to each survey design, content, and interview modes and is being studied primarily in the context of the U.S. *Current Population Survey* (CPS), for example Bonnéry et al. (2020), Cheng et al. (2013), Erkens

(2012), Halpern-Manners and Warren (2012), and Krueger et al. (2017). The U.S. CPS is a household survey with voluntary participation, which is used for the production of employment statistics (U.S. Bureau of Labor Statistics, 2006). When the response patterns are related to the employment status of the interviewed persons, the RGB can have a large influence on the quality of the employment estimates. In this chapter, the focus is on the German Microcensus, in which subjects have the obligation to provide information. We therefore expect a RGB to be less relevant than in the U.S. CPS. Moreover, under the new design of the Microcensus, see Section 4.3, there is not yet enough historical data to examine the existence and structure of a RGB. Therefore, a RGB is not considered further and remains as a possible research object for future investigations.

4.2.2 Composite estimators

Composite estimators are frequently applied in the context of rotating panel surveys. They are design-based estimators which make explicit use of the sample overlap at time t to a previous sample from t', $t' < t$, to produce estimates for time t. As they borrow strength from time, the estimators belong to the class of indirect estimators. For that, compare the definition of direct and indirect estimators in Section 2.2. Composite estimators are recursive. That is, for producing composite estimates for time t, the composite estimates of t', $t' < t$, are used as input which themselves are composite estimates using information from t'', $t'' < t'$, and so forth. When composite estimators are calculated for the first time, direct estimates such as HT or GREG estimates are used as input estimates for t' until composite estimates are available for t'. The most prominent composite estimators are different types of AK estimators, modified regression estimators, and regression composite estimators which are described in the following.

We present the estimators for a single variable of interest Y at time t and the estimation of its total $\tau_{y_t} = \sum_{k \in U} y_{kt}$. In practice, not only one, but many different statistics of different variables of a survey are of interest, for example the total number and monthly changes of employed and unemployed at the NUTS2-level. In the simulation study in Section 4.4, the estimators are evaluated for different variables of interest. For ease of description, we again consider that a time point t represents a certain month, there are no sample overlaps within a month, only between different months. Alternatively one could also consider the time points to represent quarters or years. Thereby, s_t denotes the sample in month t. For any two months t and t', where $t' < t$, sample s_t can be partitioned into $s_t \cap s_{t'}$ and $s_t \setminus s_{t'}$, the *overlapping* and *non-overlapping sample*.

AK estimator

In the *AK estimator*, the estimate of month t' is updated by the information of the sample in t to get an estimate for month t. It is given by

$$\begin{aligned}\hat{\tau}_{yt}^{\text{AK}} = &\underbrace{(1-K)\hat{\tau}_{yt}^{\text{GREG}}}_{=\text{ term 1}} \\ &+ \underbrace{K\left(\hat{\tau}_{yt'}^{\text{AK}} + \theta_{tt'}^{-1}(\hat{\tau}_{yt}^{\text{GREG, overlap}} - \hat{\tau}_{y'_t}^{\text{GREG, overlap}})\right)}_{=\text{ term 2}} \\ &+ \underbrace{A\left((1-\theta_{tt'})^{-1}\hat{\tau}_{yt}^{\text{GREG, non-overlap}} - \theta_{tt'}^{-1}\hat{\tau}_{yt}^{\text{GREG, overlap}}\right)}_{=\text{ term 3}}\end{aligned} \tag{4.136}$$

with

$$\hat{\tau}_{yt}^{\text{GREG, overlap}} = \sum_{k\in s_t\cap s_{t'}} w_{kt}y_{kt}, \quad \hat{\tau}_{yt}^{\text{GREG, non-overlap}} = \sum_{k\in s_t\setminus s_{t'}} w_{kt}y_{kt}, \tag{4.137}$$

sample overlap

$$\theta_{tt'} = \sum_{k\in s_t\cap s_{t'}} w_{kt} / \sum_{k\in s_t} w_{kt}, \tag{4.138}$$

and parameters $K \in [0,1]$ and A; A is typically chosen in the interval $[0,1]$. The estimator consists of three terms. Term 1 and 2 correspond to the original K estimator (Hansen et al., 1955). The K estimator is a convex combination of the GREG estimator in t and the K estimator in t', which is corrected for the change observed in the overlapping sample. Gurney and Daly (1965) additionally added term 3 to account for differences between the overlapping and non-overlapping sample to reduce the influence of a RGB.

For a specific variable of interest Y, the AK parameters A and K can be chosen optimally in the sense of minimizing the variance of $\hat{\tau}_{yt}^{\text{AK}}$. Variance formulas for the AK estimator are given in Cantwell (1990). For different variables, for example employment and unemployment, typically different values of A and K are variance optimal. Lent et al. (1994) and Lent et al. (1999) studied the choice of A and K for the estimation of employed and unemployed in the U.S. CPS and proposed values $K = 0.7$ and $A = 0.4$ for employed, $K = 0.4$ and $A = 0.3$ for unemployed. Consistent estimates, however, require that the same parameters are used for the production of all AK estimates. For example, the same values of A and K would have to be used for the production of the AK estimates of employment and unemployment to ensure that they add up to the labour force.

There are different extensions of the AK estimator. *AK composite weighting* (Fuller, 1990; Lent et al., 1994; Lent et al., 1999) is a two-step procedure. In the first step, for each key study variable the AK estimator is calculated with variable-specific choice of A and K. In the second step, a calibration estimator such as the GREG estimator is applied with the AK estimates of the chosen key variables as additional control totals. Thereby, the resulting calibration weights return the different AK estimates, which were calculated with different parameter choices, and consistent estimates. Next to ensuring consistency and allowing for variable-specific parameter choices, this method also has the advantage that, in contrast to the original AK estimator, it returns a single set of weights which are used to produce all estimates based on sample s_t and can be shared with micro-data users. There are further AK variants. Breau and Ernst (1983) introduced a generalised composite estimator with rotation group specific weights. Singh et al. (2001a) proposed a version of the AK estimator with micro-level matching. Ciepiela et al. (2012) suggested a dynamic K estimator for arbitrary rotation schemes. Cheng et al. (2017) presented an iterative procedure for AK estimation to calculate MSE optimal values for A and K.

Modified regression estimators

In *modified regression* (MR) estimators (Singh et al., 1997; Singh & Merkouris, 1995), the sample information from t' is used as additional auxiliary information in a GREG procedure. For a chosen set of q key variables of interest, a vector of additional auxiliary information $\boldsymbol{z}_{kt}$ of length q, for all $k \in s_t$, is defined. The corresponding control totals are denoted by $\hat{\boldsymbol{\tau}}_{z_t}^{\mathrm{MR}}$. MR estimators are given by

$$\hat{\tau}_{y_t}^{\mathrm{MR}} = \sum_{k \in s_t} w_{kt}^{\mathrm{MR}} y_{kt} \tag{4.139}$$

with

$$w_{kt}^{\mathrm{MR}} = d_{kt} \left(1 + \boldsymbol{x}_{kt}^{*\top} \left(\sum_{k \in s_t} \boldsymbol{x}_{kt}^{*} \boldsymbol{x}_{kt}^{*\top} d_{kt} \right)^{-1} \left(\boldsymbol{\tau}_{x_t}^{*} - \hat{\boldsymbol{\tau}}_{x_t}^{*} \right)^{\top} \right) \tag{4.140}$$

and

$$\boldsymbol{x}_{kt}^{*} = (\boldsymbol{x}_{kt}^{\top}, \boldsymbol{z}_{kt}^{\top})^{\top}, \quad \boldsymbol{\tau}_{x_t}^{*} = (\boldsymbol{\tau}_{x_t}^{\top}, \hat{\boldsymbol{\tau}}_{z_t}^{\mathrm{MR}\ \top})^{\top}, \quad \hat{\boldsymbol{\tau}}_{x_t}^{*} = (\hat{\boldsymbol{\tau}}_{x_t}^{\top}, \hat{\boldsymbol{\tau}}_{z_t}^{\top})^{\top}. \tag{4.141}$$

The formula of MR estimators thus corresponds to the formula of the GREG estimator (4.130) with q additional variables. For the q additional variables the true population values are unknown and estimates $\hat{\boldsymbol{\tau}}_{z_t}^{\mathrm{MR}}$ are used instead. With weights w_{kt}^{MR}, all statistics of interest can be calculated based on the sample s_t. The choice of the q key variables, for which additional auxiliary variables are defined, is typically limited as too many additional auxiliary variables could lead to a distortion of the final weights w_{kt}^{MR} (Gambino et al., 2001). To make an example, in the Canadian LFS the key study variables chosen in the MR procedure are ILO statistics by age and sex groups and provincial levels as well as specific industries (Singh et al., 2001b).

The use of the q additional auxiliary variables is expected to give more precise estimators than the corresponding GREG estimator if the additional auxiliary variables are highly correlated to the variable of interest. With high correlation, they assist in further reducing the residuals in the variance formula, compare the variance formula of the GREG estimator (2.23). That is, the efficiency gains from using an MR instead of the GREG estimator are expected to be high for those key variables which are used as additional auxiliary information and those which are highly correlated to the key variables. This is empirically shown, for example in Bell (2001), Gambino et al. (2001), and Salonen (2014).

There are different ways of defining the additional auxiliary variables and corresponding estimated control totals in MR estimators. The information of the selected key study variables is only partially available for previous time points as there is only a partial overlap between different samples. The different types of MR estimators differ essentially in how they deal with this partially missing information. From the class of MR estimators, we present the so-called MR1, MR2, MR3, MRR, and RC estimator below. They are explained for a single key variable of interest Y with additional auxiliary variable z_{kt} with control total $\hat{\tau}_{z_t}^{\mathrm{MR}}$.

In all MR estimators, the additional covariate and control total are defined in such a way that the design weighted sample values, in expectation, match the corresponding control total, i.e. that

$$\mathrm{E}\Big[\sum_{k\in s_t} d_{kt} z_{kt}\Big] = \hat{\tau}_{z_t}^{\mathrm{MR}} \tag{4.142}$$

holds (Preston, 2015). For the MR estimators, we define the overlap of samples s_t and $s_{t'}$, $\theta_{tt'}$, as

$$\theta_{tt'} = \sum_{k\in s_t\cap s_{t'}} d_{kt} / \sum_{k\in s_t} d_{kt}. \tag{4.143}$$

The *MR1 estimator*, proposed by Singh (1996), emphasises the values of Y in t' observed from the overlapping sample. In the MR1 estimator, the additional auxiliary variable z_{kt}^{MR} and its control total $\hat{\tau}_{z_t}^{\mathrm{MR}}$ are defined as

$$z_{kt}^{\mathrm{MR}} = z_{kt}^{\mathrm{MR1}} = \begin{cases} y_{kt'} & k \in s_t \cap s_{t'} \\ \hat{\tau}_{y_{t'}}^{\mathrm{MR1}}/N_{t'} & k \in s_t \setminus s_{t'} \end{cases}, \quad \hat{\tau}_{z_t}^{\mathrm{MR}} = \hat{\tau}_{y_{t'}}^{\mathrm{MR1}}. \tag{4.144}$$

For the overlapping sample, the observations of t' are used. For the non-overlapping sample, the values are filled via mean imputation such that (4.142) holds.

Instead of emphasising the observed values themselves, the *MR2 estimator*, proposed by Singh (1996) and Singh et al. (1997), emphasises the change observed from the overlapping sample. In the MR2 estimator, the additional auxiliary variable z_{kt}^{MR} and its control total $\hat{\tau}_{z_t}^{\mathrm{MR}}$ are defined as

$$z_{kt}^{\mathrm{MR}} = z_{kt}^{\mathrm{MR2}} = \begin{cases} y_{kt} + \theta_{tt'}^{-1}(y_{kt'} - y_{kt}) & k \in s_t \cap s_{t'} \\ y_{kt} & k \in s_t \setminus s_{t'} \end{cases}, \quad \hat{\tau}_{z_t}^{\mathrm{MR}} = \hat{\tau}_{y_{t'}}^{\mathrm{MR2}}. \tag{4.145}$$

The values of the overlapping sample are set to the observed values in t adjusted for the observed change from t' to t. To ensure that (4.142) holds, in the MR2 estimator carry-backward imputation is used for the values of the non-overlapping sample. Fuller and Rao (2001) recognized a potential *drift problem* for the MR2 estimator. The drift problem describes a situation where the MR2 estimates deviate substantially from the direct estimates over a potentially long period of time. The authors expect the drift problem to appear for variables which are highly correlated over time, i.e. for employment rather than for unemployment.

An additional MR approach was proposed by Gatto et al. (2009) and Loriga (2014) and is henceforth called *MR3 estimator*. The idea here is to avoid any imputation in the additional auxiliary variable and instead adjust its control total. In the MR3 estimator, the additional auxiliary variable z_{kt}^{MR} and its control total $\hat{\tau}_{z_t}^{\mathrm{MR}}$ are defined as

$$z_{kt}^{\mathrm{MR}} = z_{kt}^{\mathrm{MR3}} = \begin{cases} y_{kt'} & k \in s_t \cap s_{t'} \\ 0 & k \in s_t \setminus s_{t'} \end{cases}, \quad \hat{\tau}_{z_t}^{\mathrm{MR}} = \theta_{tt'} \hat{\tau}_{y_t'}^{\mathrm{MR3}}. \tag{4.146}$$

Similar to the MR1 estimator, for the overlapping sample the additional auxiliary variable takes the observed values in t'. For the non-overlapping sample, there is no information on the values in t' and z_{kt}^{MR3} is set to zero. The control total $\theta_{tt'}\hat{\tau}_{y_t'}^{\mathrm{MR3}}$ adjusts the MR3 estimate of t' for the sampling fraction of the overlapping sample to ensure that (4.142) holds.

Beaumont and Bocci (2005), based on Beaumont (2005), proposed another MR variant, called *MRR* estimator. The values for the overlapping sample and the control totals are defined similar to the MR1 estimator (4.144). For the non-overlapping sample a correction factor is calculated based on the observed change in the overlapping sample. In the MRR estimator, the additional auxiliary variable z_{kt}^{MR} and its control total $\hat{\tau}_{z_t}^{\mathrm{MR}}$ are defined as

$$z_{kt}^{\mathrm{MR}} = z_{kt}^{\mathrm{MRR}} = \begin{cases} y_{kt'} & k \in s_t \cap s_{t'} \\ y_{kt} + \frac{1-\theta_{tt'}}{\theta_{tt'}} \frac{\sum_{k \in s_t \cap s_{t'}} d_{kt}(y_{kt'} - y_{kt})}{\sum_{k \in s_t \setminus s_{t'}} d_{kt}} & k \in s_t \setminus s_{t'} \end{cases}, \quad \hat{\tau}_{z_t}^{\mathrm{MR}} = \hat{\tau}_{y_{t'}}^{\mathrm{MRR}}. \tag{4.147}$$

In different empirical applications, the MR1 estimator showed good results for level and the MR2 estimator for change estimation. For the design of the Canadian LFS, Fuller and Rao (2001) gave a theoretical illustration of this finding based on limit variances and a first-order autoregressive process. In theory, one could use both, additional auxiliary variables according to the MR1 and MR2 definition in a MR estimator to ensure efficient level and change estimation. A large number of auxiliary variables, however, may lead to a distortion of the final weights (Gambino et al., 2001, p. 67). As a compromise, designed to work well for level and change estimation while avoiding a large number of additional auxiliary variables, Fuller and Rao (2001) proposed the *regression composite* (RC) estimator. In the RC estimator, the additional auxiliary variable z_{kt}^{MR} and its control total $\hat{\tau}_{z_t}^{\mathrm{MR}}$ are defined as

$$z_{kt}^{\mathrm{MR}} = z_{kt}^{\mathrm{RC}} = (1-\alpha) z_{kt}^{\mathrm{MR1}} + \alpha z_{kt}^{\mathrm{MR2}}, \quad \hat{\tau}_{z_t}^{\mathrm{MR}} = \hat{\tau}_{y_{t'}}^{\mathrm{RC}} \tag{4.148}$$

with $\alpha \in [0, 1]$.

The RC estimator requires the choice of parameter α. It reduces to the MR1 and MR2 estimator for $\alpha = 0$ and $\alpha = 1$ respectively. An optimal choice of α, in the sense of minimizing the variance of the RC estimator, depends on the sampling design, the variable under study, and the importance of level versus change estimation. For the Canadian LFS, the performance of the RC estimator with different α was evaluated in Chen and Liu (2002), Fuller and Rao (2001), and Gambino et al. (2001). In the survey, the RC estimator is applied with $\alpha = 2/3$ (Gambino et al., 2001, p. 67). Fuller and Rao (2001) argued that the choice of α in the RC estimator can also be used to reduce the probability of the drift problem of the MR2 estimator. In a design-based simulation study, Preston (2015), however, found indications of a potential drift problem also for the RC estimator.

Gambino et al. (2001, Section 3) and Singh et al. (2001a, pp. 33–34) summarized the features of MR estimators. MR estimators allow for a simple implementation with standard statistical software, use the information of the overlapping and non-overlapping sample on the level of the sampling units, require only a single step to produce final weights, ensure consistency between different estimates, and facilitate to increase the efficiency of level and change estimators compared to the corresponding GREG estimator. When the population size changes considerably in t to t', an adjusted version of the MR estimators can be used, as proposed in Preston (2015).

For estimating the variance of MR estimators, it not only has to be considered that the additional control totals are themselves estimates, but also that the additional auxiliary information is partly imputed or altered. The variance estimation of composite estimators is out of the scope of this thesis and we refer to Berger et al. (2009) and Dever and Valliant (2010) for further information.

4.2.3 Composite estimators in selected Labour Force Surveys

The underlying sampling design is critical to the performance and choice of composite estimators. Various combinations of rotating panel designs and composite estimators are used in official LFSs. We briefly review some selected applications and related research.

The U.S. CPS is conducted with a 4-8-4 months rotation scheme and AK composite weighting is applied for the estimation of employment statistics (U.S. Bureau of Labor Statistics, 2006). The parameters for AK composite weighting are set to $K = 0.4$ and $A = 0.3/4 = 0.075$ for the estimation of unemployed and $K = 0.7$ and $A = 0.4/4 = 0.1$ for the estimation of employed. We note that due to slight differences in the formulas, the values of A in U.S. Bureau of Labor Statistics (2006, Section 10-10) need to be divided by 4 to fit formula (4.136).

Bonnéry et al. (2020) evaluated the performance of different composite estimators for the U.S. CPS in a design-based simulation study. In their study, both the AK and BLUE estimator performed poorer than the GREG or RC estimator once the parameters and input quantities of the estimators were estimated from sample data. Compared to the corresponding GREG estimator, the RC estimator showed a good performance in their study, for different choices of α.

The Canadian LFS rotates with an in-for-6 months design. The RC estimator is used with $\alpha = 2/3$ as a compromise between level and change estimation (Statistics Canada, 2017). In their theoretical studies, Fuller and Rao (2001) found that for the Canadian LFS the MR2 estimator performs similar to the GREG estimator for level and significantly better than the GREG estimator for change estimation. Additionally, they found that the MR1 estimator performed well for estimation of levels, but less well than the MR2 estimator for change. Gambino et al. (2001) empirically compared the GREG and three-month moving averages with the $\text{RC}_{\alpha=0.67}$ estimator showing efficiency gains of the RC estimator over the two others for the key variables which were included as additional calibration constraints. Also Beaumont and Bocci (2005) empirically investigated different composite estimators for the Canadian LFS using historical survey data, comparing the MR1, MR2, $\text{RC}_{\alpha=0.7}$, and MRR estimator. For the estimation of employment levels, the RC estimator performed best followed by the MR2 and MR1 estimator. For unemployment, the performance of the estimators was more close and the RC estimator performed slightly better than the others. For monthly changes of employed and unemployed the order of best to worst performance was MRR, MR2, RC, and MR1. The findings are somewhat surprising as for the same survey Fuller and Rao (2001) argued that MR1 works best for level and MR2 for change estimation. Consistent with Fuller and Rao (2001), other works, such as Steel and McLaren (2009, p. 301) and Gambino et al. (2001, pp. 66–67), also treated MR1 as the go-to estimator for level and MR2 for change estimation.

The Australian LFS rotates with an in-for-8 months design. Bell (2001) empirically evaluated different composite estimators for the application to the Australian LFS, including the BLUE, BLUE with a fixed window, the AK, MR2, and $\text{RC}_{\alpha=0.7}$ estimator. They recognized a potential drift problem of the MR2 and RC estimator as their estimates differed considerably from the GREG estimates over time. Considering the standard errors of the estimates, both MR2 and RC estimator performed significantly better than the AK and BLUE estimator, both for level and change estimation. This was especially true for the estimation of employment. For unemployment, the results of the different estimators were quite close. The performance superiority of the MR2 and RC estimator was, however, only visible for the key study variables, i.e. for those variables used as additional calibration constraints. Due to the potential drift problem in the MR2 and RC estimator, a variant of the fixed window BLUE is applied to the Australian LFS since 2007 (Australian Bureau of Statistics, 2018; Pink, 2007).

Salonen (2014, 2016) investigated the use of the RC and MR3 estimator for the Finish LFS which rotates with a 1-(2)-1-(2)-1-(5)-1-(2)-1 months pattern. The empirical investigation showed high efficiency gains for level and change estimation for both estimators. The efficiency gains were high for the estimation of employed and modest for unemployed.

To summarise, the different applications show that the performance of composite estimators is highly dependent on the underlying sampling design and the statistics of interest. For employment estimation in LFSs, different composite estimators are applied. What most of the applications and investigations have in common is the following. Most composite estimators perform better than the corresponding GREG estimator for both level and change estimation of employment statistics. The efficiency gains are especially high for change as opposed to level estimation. The efficiency gains are higher for statistics of employed than of unemployed as employment is higher correlated over time. Some results also slightly contradict each other, such as the performance of the MR1 and MR2 estimator for level and change estimation in the Canadian LFS in Fuller and Rao (2001) and Gambino et al. (2001) versus Beaumont and Bocci (2005).

The results emphasise that the evaluation of composite estimators should be tailored to a specific application and that a careful investigation with realistic survey data is necessary.

In the simulation study in Section 4.4, therefore the simulation population and procedure of the design-based study are tailored to the design and estimation procedure in the German Microcensus.

4.3 German Microcensus

4.3.1 Overview

The German Microcensus is a yearly conducted 1% household survey by the German national statistical office Destatis. General information about the Microcensus is given in Destatis (2021). The survey covers the entire resident population in Germany, that is all persons in private households and shared accommodations at their main and secondary residence.

It is conducted as a one-stage stratified cluster sampling. The strata are the 38 German NUTS2 regions. In the strata, the sampling units are clusters of households, formed by street segments, among other criteria. In the sampled clusters, each household is interviewed. The participation in the survey is mandatory, on the basis of §13 of the German Microcensus Law. For more detailed information, e.g. on the stratification and the formation of selection districts, we refer to Destatis (2021) and the literature cited there.

The German Microcensus gives information on the socio-economic, demographic, and household structure of the German population including working and living conditions, health, migration, education, and employment, among others. The estimates of the Microcensus have particular political relevance and scope. Among other things, they are the basis for many studies of the labour market and occupational research and the federal government's annual pension insurance report.

The LFS is conducted as part of the Microcensus. It is in the focus of this chapter. The LFS questionnaire is harmonized to fulfil the standards of the ILO and the European Statistical Office (EUROSTAT). The standardization aims to make the results of the LFS comparable between different European countries. The main aim of the LFS is to provide population statistics of employment according to the ILO definition, compare Section 3.2. Example of such estimates are the number of persons aged 15 and older being employed, unemployed, and not in labour force. The results of the LFS are inter alia used for the distribution of regional and social funds of the European Union.

4.3.2 Design changes since 2020

There were major changes to the design of the German Microcensus in 2020. Overview articles of the changes are given in Bihler and Zimmermann (2016), Hochgürtel (2013), Hundenborn and Enderer (2019), and Riede (2013), which are the main sources for the following description of the design.

Until 2020, the official household surveys Microcensus, with the integrated LFS, EU-SILC, and the survey on ICT were conducted separately. Since 2020, they have been conducted as an integrated system of household surveys called Microcensus. In the integrated system, each sampled household receives a core programme of questions. The core programme contains those questions for which high quality estimates are needed such as the ILO employment status. Next to the core programme, each sampled household can receive maximum one additional module of questions. There are the LFS, SILC, and ICT module, and a potential extra module. Over the whole year the sampling fraction of the Microcensus at NUTS0, i.e. federal level, is 1%. Over the year on NUTS0 about 45%, 12%, and 3.5% of the Microcensus sampling units receive modules LFS, SILC, and ICT respectively (Hundenborn & Enderer, 2019). At the NUTS2-level, i.e. for the 38 government districts, both the overall sampling fraction as well as the relative frequencies of the modules systematically differ from the above values to meet certain quality criteria.

The rotation design of the Microcensus has also changed. Until 2020, households were interviewed a total of four times, once per year for four consecutive years. Since 2020, household are still interviewed a total of four times, but the rotation pattern of the four interviews changed, as described in the following. For the description, we distinguish two parts of the Microcensus which correspond to two different rotation schemes: (1) The part of the Microcensus which is assigned to the LFS module, i.e. the core plus LFS module, and (2) The part of the Microcensus which is not assigned to the LFS module. Households which are assigned to the LFS module are interviewed with a 2-(2)-2 quarters pattern, i.e. they are interviewed in a quarter, in the next quarter, pause for two quarters, and are again interviewed for two consecutive quarters. They remain in the survey for six quarters. All households which are not assigned to the LFS module are interviewed once a year, thus remaining in the survey for four years.

4.3.3 Sample overlaps

Composite estimators take advantage of sample overlaps to previous time points which result from a rotating panel survey. Therefore, to assess the applicability of composite estimators for the production of employment statistics from the Microcensus, we analyse which sample overlaps arise from the Microcensus design in the following. For this purpose, Table 4.1 shows the Microcensus interviews of the rotation groups in the two rotation schemes; the Microcensus with the LFS module and the Microcensus without the LFS module. By design, there are no overlapping sampling units within a quarter. Therefore, the rotation groups are defined by the quarter of the first interview. The absolute sizes of the rotation groups within th two rotation schemes are equal in expectation. In Table 4.1, we added the quarters of the previous interview to see sample overlaps resulting from the design.

From Table 4.1, we see that in the Microcensus plus LFS module, the sample overlap of a quarter to the previous quarter (Q-1) is 50%, the overlap of a quarter to the same quarter in the previous year (Q-4) is 50%. In the Microcensus without the LFS module, the sample overlap of a quarter to the previous quarter (Q-1) is 0%, the overlap of a quarter to the

Table 4.1: Microcensus rotation groups (RGs) for the quarters of one year

		Quarters (Q) in previous year				Quarters (Q) in current year			
	RG	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Microcensus with LFS module	1	2			3	4			
	2	1	2			3	4		
	3		1	2			3	4	
	4			1	2			3	4
	5				1	2			3
	6					1	2		
	7						1	2	
	8							1	2
	9								1
Microcensus without LFS module	1					1			
	2	1				2			
	3	2				3			
	4	3				4			
	5						1		
	6		1				2		
	7		2				3		
	8		3				4		
	9							1	
	10			1				2	
	11			2				3	
	12			3				4	
	13								1
	14				1				2
	15				2				3
	16				3				4

Table 4.2: Microcensus sample overlaps for composite estimators

Sample overlap Q to	Time proximity	Magnitude of overlap
Q-1	Higher	$\approx 27\%$
Q-4	Lower	$\approx 61\%$

same quarter in the previous year (Q-4) is 75%. As over the year on NUTS0 about 45% of the Microcensus sample receive the LFS module, the approximative overall sample overlaps of Table 4.2 result. Note that the overlaps in Table 4.2 are only on NUT0-level; at the NUTS2-level they differ.

From Section 4.2.1, we saw that for composite estimators two factors are important: The magnitude of the sample overlap and the correlation over time of those variables, which are used as key study variables. In most applications of composite estimators that we presented in Section 4.2.3, there was only one sample overlap that could be considered for use in composite estimators. For example, in the Canadian LFS the monthly overlap is 5/6. The information from the previous month in the Canadian LFS has both, high magnitude (5/6) and high time proximity to the current month. Similarly, also in the other applications of composite estimators covered in Section 4.2.3 the previous time point is the most promising sample overlap to use for the composite estimators. This is not the case for the German Microcensus as can be seen from Table 4.2. In the Microcensus, there are two candidate overlaps to be used for composite estimators. Taking the overlap to the previous quarter (Q-1) ensures higher time proximity to the current quarter, but means low sample overlap ($\approx$ 27%). Taking the overlap to the quarter one year before (Q-4) has lower time proximity, but ensures higher sample overlap ($\approx$ 61%). A central question for the simulation study in Section 4.4 is therefore not only how the different composite estimators perform, but also whether they should be applied using the sample overlap information from Q-1 or Q-4.

To simulate the Microcensus sampling and estimation process as in the original design in the studies of Section 4.4, we note some additional features of the Microcensus. In the German Microcensus, population units are sampled via *stratified one-stage cluster sampling* (Destatis, 2021). On the first stage, strata are formed by the cross-combination of address size categories and regional domains. In each strata, clusters are sampled randomly with stratum-specific inclusion probabilities. The clusters are for example formed by streets; all households located in a street belong to the same cluster. All households belonging to a sampled cluster are interviewed. In Sections 2.3.1 and 2.3.2, the HT and GREG estimator were described for general sampling designs. As shown in Särndal et al. (1992, Chapter 4), it is straightforward to apply the formulas of the HT and GREG estimator under one-stage cluster sampling; the inclusion probabilities of all units in a cluster are equal to the inclusion probability of the corresponding cluster. We note that for the variance estimation, however, the cluster sampling has to be explicitly accounted for. The focus of this chapter is on point estimation, not on variance estimation. We therefore refer to Särndal et al. (1992, Chapter 4) for the variance estimation formulas of the HT and GREG estimator under cluster sampling. For the German Microcensus in particular, a description of the variance estimation can be found in Destatis (2020b, Appendix B), Afentakis and Bihler (2005) and, for the scientific use file of the Microcensus, in Schimpl-Neimanns (2011).

For the production of estimates from the Microcensus, the GREG estimator is applied to quarterly sample data, after non-response adjustment (Destatis, 2021). The resulting GREG weights are used for producing quarterly and monthly estimates of the Microcensus.

For the monthly estimate the number of weeks per month relative to the total number of weeks per quarter is used to adjust the quarterly GREG weights. A detailed overview of the GREG application in the Microcensus can be found in Afentakis and Bihler (2005, Table 2). For the quarterly application of the GREG estimator the auxiliary variables are: (1) Age classes (under 15, 15-44, 45 and over) crossed by sex on NUTS0, (2) Regular and professional soldiers including federal and riot police, basic military conscripts, civilian population on NUTS0, (3) Total population on NUTS0, (4) Nationality (German, not German) crossed by sex on NUTS1, i.e. the level of the 16 federal states, (5) Total population on NUTS2, i.e. the level of the 38 government districts.

4.3.4 Adjustments of composite estimators to account for regionally heterogeneous sample overlaps

For the Microcensus, Table 4.2 displays the resulting sample overlaps of a quarter to Q-1 (one quarter before) and Q-4 (the same quarter a year before) on NUTS0. These overlaps differ in the different NUTS2 regions. Applying the composite estimators with the formulas of Section 4.2.2 right away would lead to biased NUTS2-level estimates of employment statistics. In the following, we therefore present adjustments to the formulas of the estimators to account for regionally heterogeneous sample overlaps. The adjustments are similar to the presentation of composite estimators in Preston (2015) who adjusted the formulas composite estimators for stratified designs and survey frame changes in business surveys.

Consider that the target population can be partitioned into D strata, the NUTS2 regions. We can rewrite all population-specific quantities as stratum-specific quantities and denotes these by adding subscripts d, $d = 1, \ldots, D$. For example, with s_{td} we denote the part of sample s_t which corresponds to domain d. For simplicity, we do not distinguish between the names of the original and adjusted estimators as we only use the adjusted estimators hereafter.

The AK estimator, adjusted to stratum-specific sample overlaps, is given by

$$\begin{aligned}\hat{\tau}_{y_t}^{\mathrm{AK}} =& (1-K)\hat{\tau}_{y_t}^{\mathrm{GREG}} \\ &+ K\left(\hat{\tau}_{y_{t'}}^{\mathrm{AK}} + \sum_{d=1}^{D} \theta_{tt'd}^{-1}(\hat{\tau}_{y_{td}}^{\text{GREG, overlap}} - \hat{\tau}_{y_{t'd}}^{\text{GREG, overlap}})\right) \\ &+ A\left(\sum_{d=1}^{D}(1-\theta_{tt'd})^{-1}\hat{\tau}_{y_{td}}^{\text{GREG, non-overlap}} - \theta_{tt'd}^{-1}\hat{\tau}_{y_{td}}^{\text{GREG, overlap}}\right)\end{aligned} \tag{4.149}$$

with

$$\hat{\tau}_{y_{td}}^{\text{GREG, overlap}} = \sum_{k \in s_{td} \cap s_{t'd}} w_{kt} y_{kt}, \quad \hat{\tau}_{y_{td}}^{\text{GREG, non-overlap}} = \sum_{k \in s_{td} \setminus s_{t'd}} w_{kt} y_{kt}, \tag{4.150}$$

$K \in [0,1]$, and stratum-specific sample overlap

$$\theta_{tt'd} = \sum_{k \in s_{td} \cap s_{t'd}} w_{kt} / \sum_{k \in s_{td}} w_{kt}, \; d = 1, \dots, D. \tag{4.151}$$

Similarly, with $d = 1, \dots, D$, an additional auxiliary variable for the adjusted MR estimators is defined as

$$z_{kt}^{\text{MR1}} = \begin{cases} y_{kt'} & k \in s_{td} \cap s_{t'd} \\ \hat{\tau}_{y_{t'd}}^{\text{MR1}} / N_d & k \in s_{td} \setminus s_{t'd} \end{cases}, \tag{4.152}$$

$$z_{kt}^{\text{MR2}} = \begin{cases} y_{kt} + \theta_{tt'd}^{-1}(y_{kt'} - y_{kt}) & k \in s_{td} \cap s_{t'd} \\ y_{kt} & k \in s_{td} \setminus s_{t'd} \end{cases}, \tag{4.153}$$

$$z_{kt}^{\text{MR3}} = \begin{cases} y_{kt'} & k \in s_{td} \cap s_{t'd} \\ 0 & k \in s_{td} \setminus s_{t'd} \end{cases}, \tag{4.154}$$

$$z_{kt}^{\text{MRR}} = \begin{cases} y_{kt'} & k \in s_{td} \cap s_{t'd} \\ y_{kt} + \frac{1-\theta_{tt'd}}{\theta_{tt'd}} \frac{\sum_{k \in s_{td} \cap s_{t'd}} d_{kt}(y_{kt'} - y_{kt})}{\sum_{k \in s_{td} \setminus s_{t'd}} d_{kt}} & k \in s_{td} \setminus s_{t'd} \end{cases}, \tag{4.155}$$

with

$$\hat{\tau}_{y_{t'd}}^{\text{MR1}} = \sum_{k \in s_{t'd}} w_{kt'}^{\text{MR 1}} y_{kt'}, \tag{4.156}$$

$$\theta_{tt'd} = \sum_{k \in s_{td} \cap s_{t'd}} d_{kt} / \sum_{k \in s_{td}} d_{kt}, \tag{4.157}$$

and N_d as the size of stratum d. For the MR3 estimator the control total of the additional variable has to be adjusted and is given by $\hat{\tau}_{z_t}^{\text{MR}} = \sum_{d=1}^{D} \hat{\tau}_{y_{t'd}}^{\text{MR3}} \theta_{tt'd}^{-1}$. The formula of the RC estimator remains as (4.148), only that now the adjusted formulas of the MR1 and MR2 auxiliaries are used.

4.4 Simulation

To evaluate the composite estimators for employment estimation under the Microcensus design, we apply a design-based simulation study. The study aims at replicating the target population, design, and estimation procedure of the German Microcensus as closely as possible. We therefore use the RIFOSS dataset as the simulation population; the extension of the RIFOSS dataset to include longitudinal employment categories is discussed in Chapter 3. The performance evaluation of the estimators is conducted for both quarterly and monthly estimates of employed and unemployed on the NUTS2-level. Three different statistics are considered: Level, change to previous time point (change_t1), and change to previous year (change_ty). In Section 4.3, we saw that there are two candidate overlaps for composite estimators resulting from the Microcensus design. That is, the overlap of a quarter to the previous quarter (Q-1) with higher time proximity and lower overlap magnitude and the overlap of a quarter to the same quarter the year before (Q-4) with lower time proximity but higher overlap magnitude. We therefore not only compare different composite estimators, but also evaluate whether the sample overlap to the previous quarter or previous year is better suited in the Microcensus. For composite estimators, we not only analyse which temporal overlap is best to use, but also at which regional level. To this end, we examine the estimators with both auxiliary information at NUTS1 and NUTS2.

4.4.1 Simulation population

The RIFOSS dataset with longitudinal employment information is considered as the simulation population. Chapter 3 gives a detailed description of how the data were enriched with monthly employment categories according to the ILO definition. For the simulation, we only use those observations from the RIFOSS dataset referring to persons in their main residence. The RIFOSS dataset consists of about 82.5 million persons in their main residence which are grouped in about 38 million households. The dataset enables simulations of the Microcensus design in all its detail and regional depth. As can be seen in Chapter 3, the monthly employment information in the RIFOSS dataset were generated replicating the months of the years 2012-2014. Since the new Microcensus design has only been used since 2020, we treat the data as if they represent years 2022-2024. We chose 2022-2024 as the number of weeks per quarter is more regularly distributed in 2022-2024 than e.g. in 2020-2022. In 2022-2024, all quarters contain 13 weeks. In 2020-2022, there are quarters with 12, 13, and 14 weeks, which would require additional adjustments in the sampling and estimation process.

The monthly and quarterly aggregates of employed and unemployed persons in the RIFOSS dataset are given in Figure 4.1. On the aggregated level, the data show both seasonal effects and a trend for employed and unemployed. There is a negative and positive trend for the number of employed and unemployed persons respectively. On average, the percentages of employment, unemployment, and not in labour force in the RIFOSS dataset, i.e. for persons all ages, are 49.56%, 2.82%, and 47.61%. For persons aged 15 to 64 the percentages are 73.6%, 4%, and 22.4%.

As discussed previously, the performance of the composite estimators for the estimation of employed and unemployed is significantly influenced by the correlation of employment or unemployment over time and the magnitude of the sample overlap used. From Section 4.3.3, we saw that the sample overlaps to one quarter before (Q-1) and one year before (Q-4) both have interesting features for composite estimators in terms of time proximity and magnitude of sample overlap. Table 4.3 shows the average employment transitions from one quarter to others in the RIFOSS dataset. For example, 96% means that in the RIFOSS dataset 96% of the persons which are employed in a quarter were also employed the quarter before. The correlation of employment is high to Q-1, even to Q-4. For unemployment, the correlation to Q-1 is moderate and decreasing significantly in time, it is only at 40% for Q-4. The correlations shown in Table 4.3 are in line with other empirical LFS data, e.g. Gambino et al. (2001).

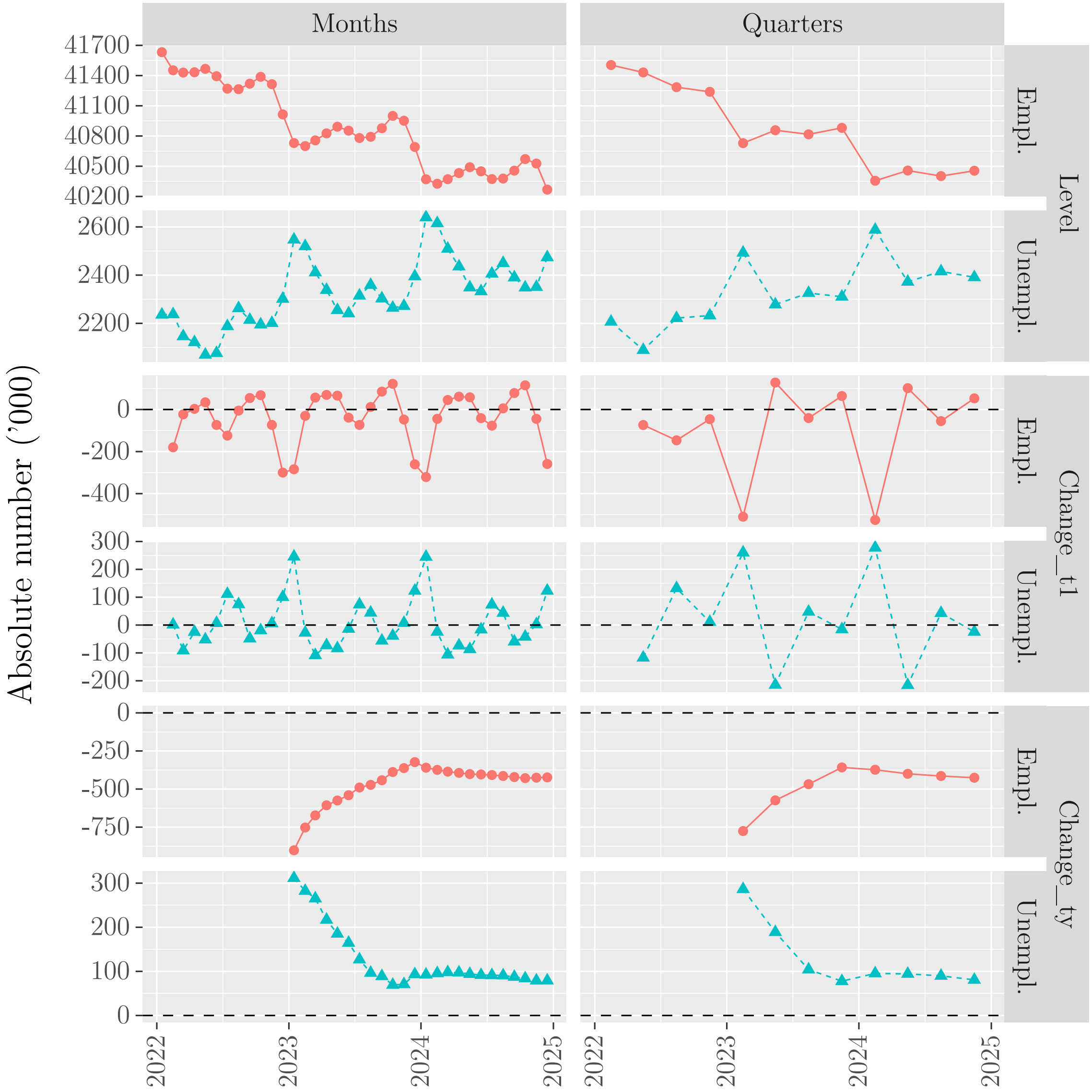

Figure 4.1: Employment aggregates in the simulation population

Table 4.3: Observations in RIFOSS population having the same employment status in quarter Q and quarter Q-1 or Q-4 (in %)

Same status in Q and	Employed	Unemployed
Q-1	96%	65%
Q-4	91%	40%

4.4.2 Simulation setup

The steps for the simulation are given in Algorithm 4.1. We will go through them step by step.

Algorithm 4.1 Simulation steps

```
1. For $r = 1, \dots, 4.000$ do
     a) Draw monthly samples from the RIFOSS population for 2022-2024 according
        to the design of the German Microcensus.
     b) For quarter $q = 2022.q1, \dots, 2024.q4$ do
          i. Calculate the GREG estimator resulting in a set of weights.
         ii. Calculate all composite estimators for all combinations of additional auxiliary
             information in Table 4.4. Each combination results in a set of weights.
        iii. Calculate estimates of all quantities of interest, i.e. the cross-combinations
             of the variables in Table 4.5 with the HT, the GREG, and the composite
             estimators.
2. For all estimators $\zeta$ and all different statistics of interest $\gamma$ (Table 4.5) calculate
     $\mathrm{RBias}(\hat{\gamma}^{\zeta}) = 100R^{-1}\sum_{r=1}^{R}\frac{\hat{\gamma}^{\zeta(r)} - \gamma}{\gamma}, \quad \mathrm{MSE}(\hat{\gamma}^{\zeta}) = R^{-1}\sum_{r=1}^{R}(\hat{\gamma}^{\zeta(r)} - \gamma)^2$
     $\mathrm{MSE.rel}(\hat{\gamma}^{\zeta}) = 100\frac{\mathrm{MSE}(\hat{\gamma}^{\zeta})}{\mathrm{MSE}(\hat{\gamma}^{\mathrm{GREG}})}.$
```

Table 4.4: Additional auxiliary information for composite estimators

Auxiliary information		Details
Regional	- Aux. NUTS1	16 NUTS1 regions resulting in 32 additional auxiliary variables
	- Aux. NUTS2	38 NUTS2 regions resulting in 76 additional auxiliary variables
Time	- Aux. Q-1	Using the overlapping sample of a quarter to the previous quarter
	- Aux. Q-4	Using the overlapping sample of a quarter to the same quarter a year before

In each simulation run a sample is drawn in accordance with the Microcensus sampling design. For each sample, sequentially for each quarter in the data in 2022-2024, the HT and GREG estimator are applied as baseline estimators. The estimators are applied similar to the actual procedure in the German Microcensus, described e.g. in Afentakis and Bihler (2005). Although we aim at reproducing the Microcensus procedure as closely as possible, there are some deviations from the GREG calculation in the simulation study to its Microcensus implementation. Auxiliary variables related to nationality and soldiers are

Table 4.5: Quantities of interest

Level	Values
Regional	- 38 NUTS2 regions
Time	- 12 quarters in 2022-2024 - 36 months in 2022-2024
Variable	- Employed - Unemployed
Measure	- Level: Quarterly and monthly totals - Change_t1: Change to previous time point (quarters: change from one quarter to next quarter, months: change from one month to next month) - Change_ty: Change to previous year (quarters: change from one quarter to the same quarter the year after, months: change from one month to the same month the year after)

not available in the RIFOSS dataset and therefore not included as auxiliary information in the GREG estimator. Since the employment status in the RIFOSS dataset is available at a monthly, not a weekly, basis the simulation study is set up as if all months contained the same number of weeks. Otherwise, with the interviews distributed evenly over the weeks of a year, biased estimates would occur with monthly data. We do not simulate any non-response and therefore do not apply any non-response adjustment. We calculate the final weights of the GREG estimator using formula (4.131). In the Microcensus, an additional iterative procedure is used to ensure non-negative weights (Afentakis & Bihler, 2005).

The composite estimators are all based on the baseline GREG estimator, i.e. the AK estimator uses the GREG estimates as input and the MR estimators use the GREG auxiliary information. We apply composite estimators MR1, MR2, MR3, MRR, RC ($\alpha \in \{0.25, 0.5, 0.75\}$), and the AK estimator. Three different versions of the AK composite weighting are applied with the same parameters K and A for the estimation of employment and unemployment, AK1 ($K = 0.7$, $A = 0$), AK2 ($K = 0.4$, $A = 0$), and AK3 ($K = 0.4$, $A = 0.05$), similar to the parameters used in the U.S. CPS (U.S. Bureau of Labor Statistics, 2006, Chapter 10). For the composite estimators, the formulas of Section 4.3.4 are applied, i.e. the formulas adjusted for regionally heterogeneous sample overlaps.

The composite estimators are calculated sequentially for the quarters of 2022-2024 as they are recursive, compare Section 4.2.2. Table 4.4 displays the different versions of auxiliary data each composite estimator is calculated with. The composite estimators are applied using the overlap to the previous quarter Q-1 and the quarter in the previous year Q-4 as these overlaps were shown to be the most promising ones, see Section 4.3.3. For composite estimators, it is necessary to specify not only which sample overlap to use, but also the regional level of the information. Using information on employed and unemployed on NUTS1 (16 regions) results in 32 additional auxiliary variables. Using information on employed and unemployed on NUTS2 (38 regions) is more detailed, but results in 76 additional auxiliary variables which could lead to a distortion of the final weights. Therefore, both NUTS1- and NUTS2-level information is investigated for both Q-1 and Q-4 auxiliary information for composite estimators. Note that due to their recursive nature, the composite estimators with sample overlap Q-1 can be calculated beginning in the second quarter of 2022. With sample overlap Q-4, the composite estimators can be calculated beginning in the first quarter of 2023. Before that, only HT/GREG estimates can be calculated.

With each combination of auxiliary information in Table 4.4, we get a different vector of quarterly calibration weights per estimator. With these final weights, all quantities of interest are calculated for all estimators. As quantities of interest we consider all cross-combinations of Table 4.5. Note that change_t1 is the month-to-month change and quarter-to-quarter change when considering monthly and quarterly estimates respectively. Equivalently, change_ty is the month to month in year before and quarter to quarter in year before change when considering monthly and quarterly estimates respectively.

As evaluation metrics, we calculate the relative bias, RBias, and the relative MSE, MSE.rel. MSE.rel is frequently used to assess the efficiency gains of a composite estimator over the corresponding GREG estimator.

4.4.3 Overall performance

We first evaluate whether the proposed formulas of the composite estimators, which were adjusted to account for regionally heterogeneous sample overlaps, give unbiased estimates for quarterly employment statistics on NUTS2. Table 4.6 displays the average RBias of the estimates for the levels, i.e. the totals, of employed and unemployed. The values are the average values over the 38 NUTS2 regions and 12 quarters in 2022-2024. The different columns refer to the different sets of additional auxiliary information used in the estimators, i.e. sample overlap information from Q-1 or Q-4, on NUTS1 or NUTS2. The RBias of the HT and GREG estimator resulting from the simulation lie within $[-0.01, 0.00]$. For unemployment the RBias values are larger than for employment as there are substantially fewer unemployed persons in the simulation population. Thus, the MC convergence rate of the RBias is lower for unemployed and we expect that they become more similar with more iterations. Only for the AK estimators the values of the RBias notably differ from zero, especially for sample overlap Q-1. For the other estimators, the RBias is close to zero. Therefore, we see no indication off a bias for the adjusted composite estimators (except the AK). Although not shown here, we note that the original formulas of the composite estimators gave significantly biased NUTS2-level estimates.

To get an overview of the performance of the different composite estimators, the mean values of MSE.rel over the 38 NUTS2 regions and 12 quarters are shown in Tables 4.7, 4.8, and 4.9 for estimates of level, change_t1, and change_ty, for quarterly NUTS2-level statistics. Values of MSE.rel smaller than 100 indicate that an estimator is more efficient than the GREG estimator, values greater than 100 indicate it is less efficient.

Table 4.6: Mean RBias (in %) of quarterly NUTS2-level estimates

(a) Employed

	Additional auxiliaries			
	NUTS1		NUTS2	
	Q-1	Q-4	Q-1	Q-4
MR1	0.00	-0.01	0.00	-0.01
MR2	0.00	0.00	0.01	0.00
MR3	-0.01	0.00	0.00	0.00
MRR	-0.04	-0.11	-0.02	0.00
$\mathrm{RC}_{\alpha=0.25}$	0.00	-0.01	0.01	0.00
$\mathrm{RC}_{\alpha=0.5}$	0.00	0.00	0.02	0.00
$\mathrm{RC}_{\alpha=0.75}$	0.00	0.00	0.01	0.00
AK1	-0.30	-0.07	-0.32	-0.07
AK2	-0.11	-0.04	-0.11	-0.04
AK3	-0.10	-0.04	-0.11	-0.04

(b) Unemployed

	Additional auxiliaries			
	NUTS1		NUTS2	
	Q-1	Q-4	Q-1	Q-4
MR1	0.00	-0.01	-0.01	-0.01
MR2	-0.09	-0.05	-0.23	-0.14
MR3	-0.01	0.00	-0.01	0.00
MRR	-0.11	0.12	0.46	0.00
$\mathrm{RC}_{\alpha=0.25}$	0.03	-0.01	0.07	-0.01
$\mathrm{RC}_{\alpha=0.5}$	-0.04	-0.02	-0.09	-0.05
$\mathrm{RC}_{\alpha=0.75}$	-0.07	-0.04	-0.18	-0.10
AK1	-0.31	-0.07	-0.33	-0.07
AK2	-0.11	-0.03	-0.12	-0.04
AK3	-0.11	-0.04	-0.12	-0.04

Table 4.7: Mean MSE.rel (in %) of quarterly NUTS2-level estimates

Employed

	Additional auxiliaries			
	NUTS1		NUTS2	
	Q-1	Q-4	Q-1	Q-4
MR1	95	94	83	81
MR2	87	95	64	86
MR3	96	95	87	84
MRR	185	112	259	109
$\mathrm{RC}_{\alpha=0.25}$	97	92	165	76
$\mathrm{RC}_{\alpha=0.5}$	94	92	107	78
$\mathrm{RC}_{\alpha=0.75}$	88	93	70	82
AK1	248	196	409	311
AK2	140	121	184	146
AK3	141	122	186	148

Unemployed

	Additional auxiliaries			
	NUTS1		NUTS2	
	Q-1	Q-4	Q-1	Q-4
MR1	96	98	91	96
MR2	96	99	92	98
MR3	96	98	91	96
MRR	407	126	719	134
$\mathrm{RC}_{\alpha=0.25}$	96	98	90	96
$\mathrm{RC}_{\alpha=0.5}$	96	98	90	96
$\mathrm{RC}_{\alpha=0.75}$	96	99	91	97
AK1	173	123	274	154
AK2	103	101	108	103
AK3	106	103	115	108

Table 4.8: Mean MSE.rel (in %) of quarterly NUTS2 change_t1 estimates

Employed					Unemployed				
	Additional auxiliaries					Additional auxiliaries			
	NUTS1		NUTS2			NUTS1		NUTS2	
	Q-1	Q-4	Q-1	Q-4		Q-1	Q-4	Q-1	Q-4
MR1	95	94	82	79	MR1	96	98	91	96
MR2	72	97	24	91	MR2	91	100	80	100
MR3	95	94	84	82	MR3	96	98	91	96
MRR	74	111	25	116	MRR	120	127	133	142
RC$_{\alpha=0.25}$	83	92	54	76	RC$_{\alpha=0.25}$	92	99	81	97
RC$_{\alpha=0.5}$	76	93	32	80	RC$_{\alpha=0.5}$	91	99	79	98
RC$_{\alpha=0.75}$	73	95	24	86	RC$_{\alpha=0.75}$	91	100	79	99
AK1	264	201	449	320	AK1	110	127	123	162
AK2	161	122	228	147	AK2	95	102	87	106
AK3	160	124	226	151	AK3	95	105	89	112

Table 4.9: Mean MSE.rel (in %) of quarterly NUTS2 change_ty estimates

Employed					Unemployed				
	Additional auxiliaries					Additional auxiliaries			
	NUTS1		NUTS2			NUTS1		NUTS2	
	Q-1	Q-4	Q-1	Q-4		Q-1	Q-4	Q-1	Q-4
MR1	93	94	79	79	MR1	97	98	93	96
MR2	93	80	76	43	MR2	99	97	98	93
MR3	95	94	84	79	MR3	97	98	93	96
MRR	135	83	147	39	MRR	278	106	466	100
RC$_{\alpha=0.25}$	89	87	110	61	RC$_{\alpha=0.25}$	98	97	95	94
RC$_{\alpha=0.5}$	88	83	69	48	RC$_{\alpha=0.5}$	99	97	97	93
RC$_{\alpha=0.75}$	91	80	70	43	RC$_{\alpha=0.75}$	99	97	98	93
AK1	229	200	382	324	AK1	173	104	275	110
AK2	130	137	164	183	AK2	108	98	120	96
AK3	132	137	169	181	AK3	112	99	129	98

We start analysing the results by looking at the overall performance of the composite estimators in Tables 4.7, 4.8, and 4.9, i.e. for all different sets of sample overlap information. Except for the MRR and AK estimators, the composite estimators show efficiency gains over the GREG estimator for the different statistics of employment and unemployment and different sets of sample overlap information. As expected, efficiency gains are generally higher for employment than for unemployment estimation, expect for the AK estimators. Especially for the estimation of change, the efficiency gains are significantly high, see the MR2 and RC estimator with sample overlap Q-1. This is in line with other research on composite estimators, e.g. Bell (2001) and Beaumont and Bocci (2005).

The performance of the AK estimators stands out. Only for certain sets of additional auxiliary information they give efficiency gains over the GREG estimator, and that only for the estimation of unemployment change. The AK3 ($K = 0.4$, $A = 0.05$) estimator performs slightly worse than AK2 ($K = 0.4$, $A = 0$) estimator for most scenarios, which can be attributed to the fact that there is no RGB in the simulation which would justify choosing A different from zero. The AK1 ($K = 0.7$, $A = 0$) estimator performs much worse than the AK2 ($K = 0.4$, $A = 0$) estimator, for both employment and unemployment. This highlights the sensitivity of the performance of the estimator with respect to the choice of K. In the present simulation study, for all scenarios and statistics where an AK estimator yields efficiency gains over the GREG estimator, the efficiency gains from the MR2 estimator are higher. The results are similar to those in Bell (2001, Section 6.1), where the AK estimator ($K = 0.7$, $A = 0.06$) had higher standard errors than the MR estimators. Also the design-based study in Bonnéry et al. (2020) showed that the performance of the AK estimator is sensitive to the parameter choices. Since the other composite estimators have better performance than the AK estimators, the AK estimators are not further considered.

Both MR2 and MRR estimator emphasise the change observed from the overlapping sample. The MR2 estimator performs similar to the MRR estimator for cases where the MRR estimator is highly efficient, but performs significantly better for cases where the MRR estimator is less efficient than the GREG estimator. It seems that the MRR estimator is sensitive to the magnitude of the sample overlap as the performance varies to a great extend when using additional auxiliary information from Q-1 versus Q-4. By comparing the MR2 and MRR formulas (4.145) and (4.147), we see that in the MRR estimator the auxiliary values are adjusted for the non-overlapping sample whereas in the MR2 estimator they are adjusted for the overlapping sample. With small sample overlaps, i.e. Q-1, there is more adjustment in the MRR than in the MR2 estimator. In Beaumont and Bocci (2005) and Preston (2015), the MRR estimator performed well for change estimation of employment and unemployment as well as level estimation of employment. Also, the performance was more close to that of the MR2 estimator. The results of the simulation study presented here emphasise that performance results from specific surveys cannot be expected to be similar under another sampling design and that each application requires an evaluation tailored to the specific design. The MRR estimator is not further considered.

The MR1 and MR3 estimator show a similar performance. They are also close in their definition, only that in the MR1 estimator mean imputation is used for the non-overlapping sample and in the MR3 estimator the control total is adjusted by the value of the sample overlap. In both estimators the definition of the additional auxiliary information is based on the observed levels in the previous sample. For all quantities of interest and combinations of auxiliary information shown in Tables 4.7, 4.8, and 4.9, the MR1 performs better than the MR3 estimator. Therefore, the MR3 estimator is not further considered.

4.4.4 Sample overlap information on NUTS1 versus NUTS2

The remaining estimators are the MR1, MR2, and the RC estimator. Apart from the RC estimator with $\alpha \in \{0.25, 0.5\}$, the estimators give efficiency gains for employment and unemployment for both level and change estimation and all sets of additional auxiliary information. The efficiency gains are generally higher when considering additional auxiliary information from NUTS2 instead of from NUTS1. As we evaluate the estimators for NUTS2-level statistics, we also expect NUTS2-level auxiliary information to give more valuable information than NUTS1-level auxiliary information. The results indicate that using the many NUTS2 auxiliary variables (76 additional auxiliary variables) does not lead to a distortion of the resulting weights and should be preferred over NUTS1 auxiliary variables. In the following, therefore only the results of the MR1, MR2, and the RC estimator with NUTS2 auxiliary variables are considered.

4.4.5 Performance for single NUTS2 regions and quarters

The performance of the RC estimator with $\alpha \in \{0.25, 0.5\}$ in Table 4.7 for the level estimates of employment are particularly striking. There, with additional auxiliary information from Q-1, the RC estimator with $\alpha \in \{0.25, 0.5\}$ performs better with NUTS1 than with NUTS2 auxiliaries. To analyse these results, we display the performance for quarterly NUTS2-level estimation for the single NUTS2 regions (average values over the 12 quarters) in Figure 4.2 and single time points (average values over 38 NUTS2 regions) in Figure 4.3. The values in Table 4.7 correspond to the mean values over the quarters and NUTS2 regions in Figures 4.2 and 4.3. Note that from Figure 4.2, we can also see that the earliest possible calculation of the composite estimators with sample overlap Q-4 is the first quarter of 2023. With sample overlap Q-1, composite estimators can already be calculated for the second quarter of 2022.

From Figures 4.2 and 4.3 we see that the performance of the MR1, MR2, and $\text{RC}_{\alpha=0.75}$ estimator is relatively constant across the quarters of 2022-2024 and the single NUTS2 regions. For the RC ($\alpha \in \{0.25, 0.5\}$) estimator, however, with additional auxiliaries on NUTS2 from Q-1, the performance of the employment level estimation is very noticeable. The efficiency gains of the RC ($\alpha \in \{0.25, 0.5\}$) estimator over the GREG estimator drift away for proceeding quarters in Figure 4.2. Similar to the drift problem described in Fuller and Rao (2001) for the MR2 estimator, the efficiency loss only applies to the estimation of employment, not for unemployment. This indicates that it only appears when the additional auxiliary information is highly correlated to the variable of interest, compare

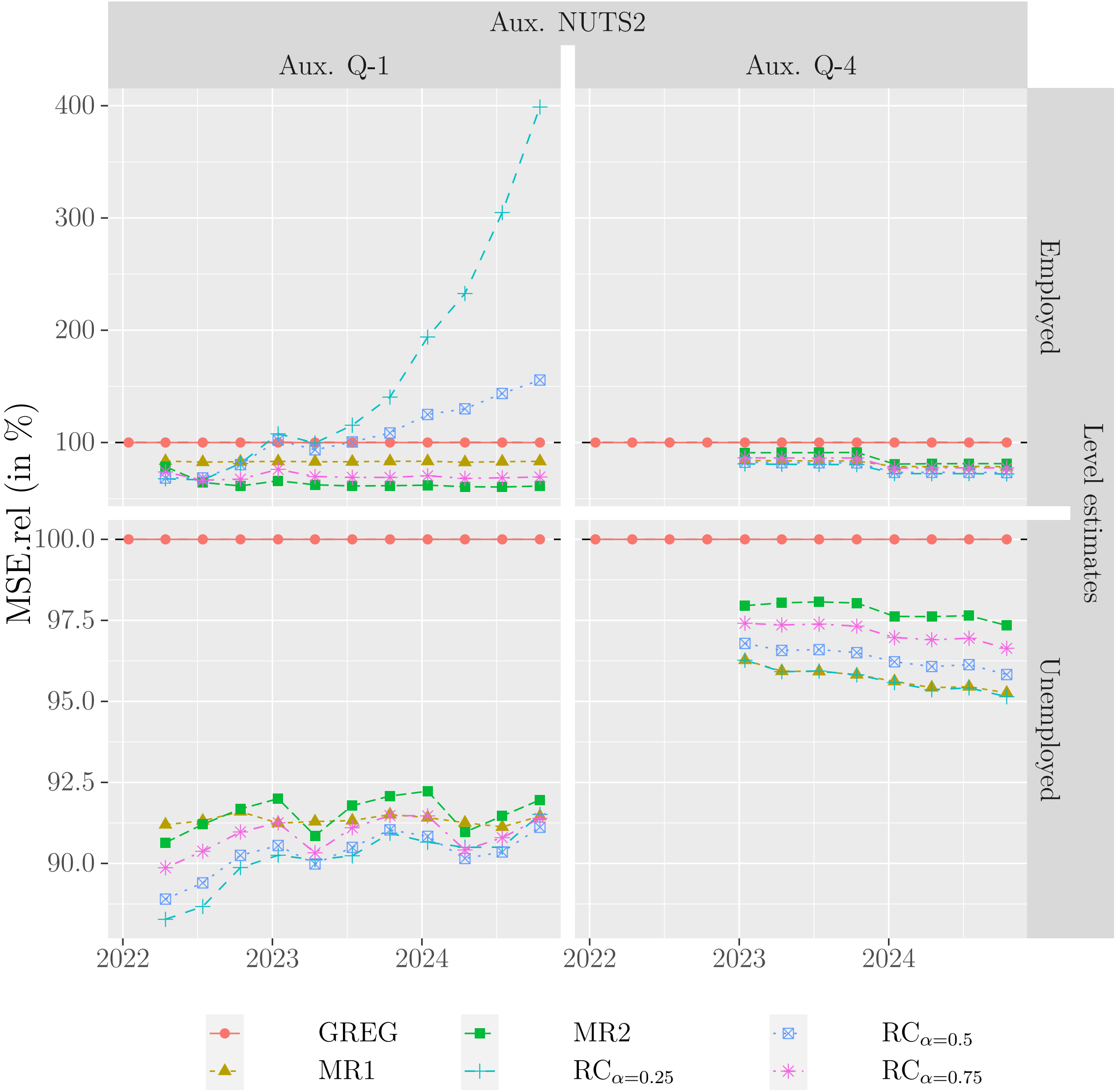

Figure 4.2: MSE.rel of quarterly NUTS2-level estimates, mean values over 38 NUTS2 regions

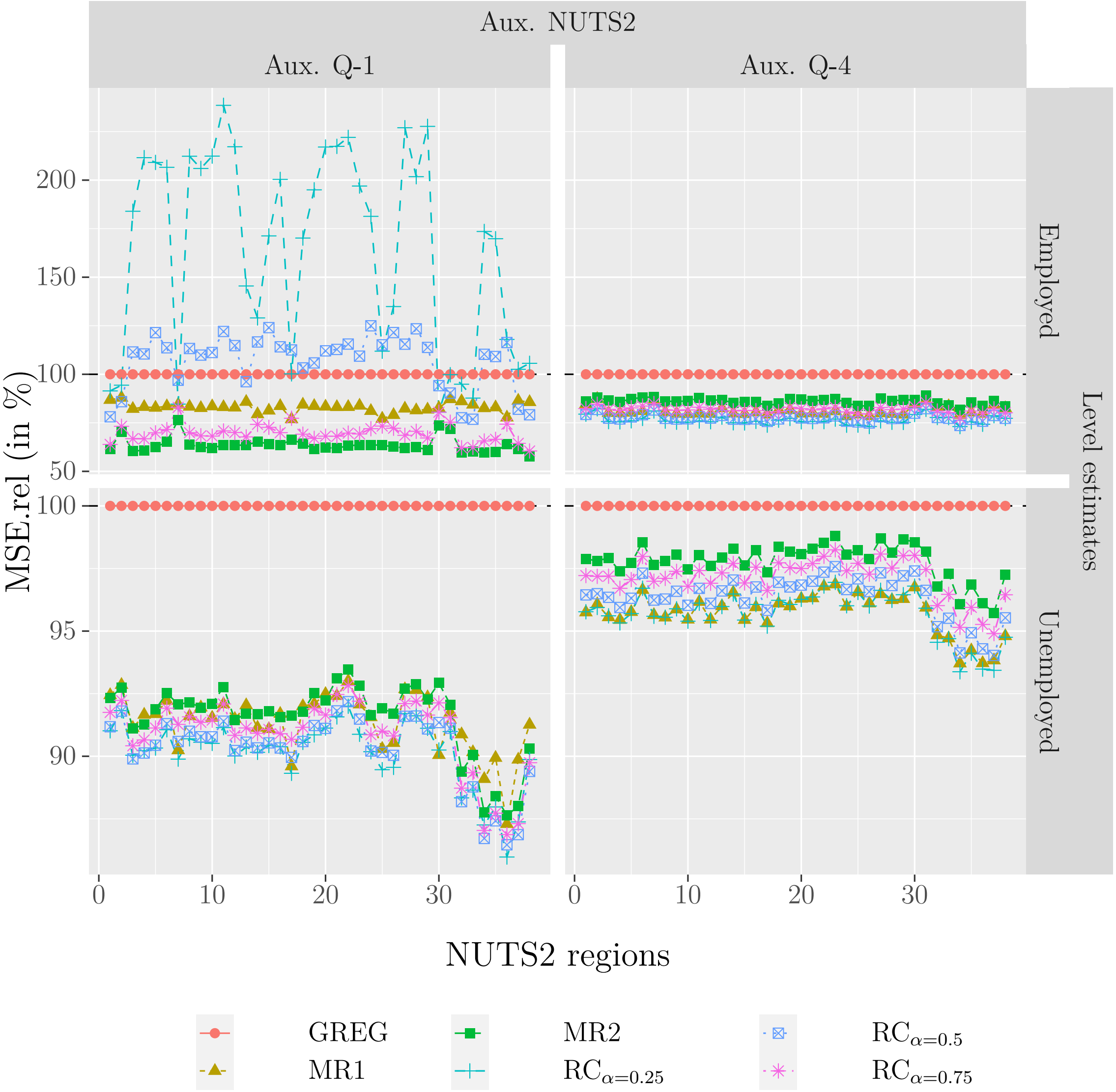

Figure 4.3: MSE.rel of quarterly NUTS2-level estimates, mean values over 12 quarters

Table 4.3. Also Preston (2015) found indications of a potential drift problem for the RC estimator. Contrary to the results in Fuller and Rao (2001) or Preston (2015), in the presented figures only the RC ($\alpha \in \{0.25, 0.5\}$) estimator shows this pattern, not the MR1, MR2, or RC$_{\alpha=0.75}$ estimator. The results imply that the performance of the RC estimator can be sensitive to the choice of α.

The performance of the RC ($\alpha \in \{0.25, 0.5\}$) estimator also differs to a great extent for the different NUTS2 regions as can be seen in Figure 4.3. In some NUTS2 regions, the values of MSE.rel of the RC ($\alpha \in \{0.25, 0.5\}$) estimator are significantly lower than in others. The NUTS2 regions for which the performance is better are those where the percentage of the LFS module is higher, leading to higher sample overlap to Q-1 than in the other regions. This implies that not only the correlation of the variable of interest to the additional auxiliary information, but also the magnitude of the sample overlap can play an important role for the efficiency and sensitivity of the RC estimator and optimal choice of α. Estimators RC$_{\alpha=0.25}$ and RC$_{\alpha=0.5}$ are not further considered.

4.4.6 Sample overlap information from Q-1 versus Q-4

Estimators MR1, MR2, and RC$_{\alpha=0.75}$ show a good performance for the different level and change estimations with NUTS2 auxiliary information compared to the corresponding GREG estimator. It remains to be analysed whether they perform better with sample overlap information from Q-1 versus Q-4. Table 4.2 showed that the overlap to Q-1 gives higher time proximity, but has a lower magnitude of overlap ($\approx 27\%$), whereas Q-4 gives lower time proximity, but has a higher magnitude of overlap ($\approx 61\%$).

The performance of the MR1, MR2, and RC$_{\alpha=0.75}$ estimator with Q-1 versus Q-4 NUTS2-level auxiliaries can be compared using the two right columns of Tables 4.7, 4.8, and 4.9. For the estimation of levels and quarterly changes, the MR2 and RC$_{\alpha=0.75}$ estimator perform best with sample overlap information from Q-1, both for employment and unemployment estimation. The efficiency gains over the GREG estimator can be high for quarterly change estimation. The values of MSE.rel of the MR2 estimator is only 24% for the estimation of quarterly employment change. When, however, the focus is on yearly change, both estimators perform better with Q-4 auxiliaries as then the additional auxiliaries are more highly correlated to the statistic of interest.

For the MR1 estimator the picture is a bit different. With the MR1 estimator, the estimation of employment statistics is more efficient using auxiliary information from Q-4, the estimation of unemployment statistics is more efficient using auxiliary information from Q-1. The MR1 estimator slightly outperforms the MR2 and RC$_{\alpha=0.75}$ estimator when NUTS2-level sample overlap information from Q-4 is used, for level and quarterly change estimation of employed and unemployed. This indicates that the magnitude of the sample overlap effects the performance of the MR1 and MR2 estimator and that for the choice of MR1 versus MR2 or RC the underlying sampling design and magnitude of sample overlap should be considered.

Overall, for level and quarterly change estimation of employed and unemployed on NUTS2, we recommend the use of the MR2 and $\mathrm{RC}_{\alpha=0.75}$ estimator with auxiliaries from Q-1 on NUTS2. For the estimation of yearly change of employed and unemployed on NUTS2, we recommend the use of the MR2 and $\mathrm{RC}_{\alpha=0.75}$ estimator with auxiliaries from Q-4 on NUTS2.

4.4.7 Performance of monthly estimation

As in the Microcensus, compare Afentakis and Bihler (2005), we applied the GREG estimator quarter-wise in the simulation study. Similarly, also the composite estimators were applied to quarterly sample data. The quarterly weights returned from the GREG and the composite estimators can be used to calculate monthly estimates by adjusting them for the relative number of weeks that each month represents in a given quarter. Table 4.10 shows the relative MSE of the monthly NUTS2-level estimates of the MR1, MR2, and $\mathrm{RC}_{\alpha=0.75}$ estimator. The estimators are shown for additional auxiliary information on NUTS2 from Q-1 as this combination showed the best performance gains over the GREG estimator in the previous analyses.

Although the estimators are always at least as efficient as the corresponding GREG estimator, we see that the efficiency gains from using composite estimators are significantly lower for monthly than for quarterly statistics. This suggests that the additional quarterly auxiliary information in MR estimators is less suited for monthly than for quarterly estimation. It also confirms the findings of the other presented applications of composite estimators: They give efficiency gains especially for those variables which as used as the key variables in composite estimators, compare e.g. Bell (2001). For other variables, in this case the monthly quantities, the efficiency gains are small. Therefore, thorough consideration should be given to which variables are important enough to be used as key study variables.

Table 4.10: Mean MSE.rel (in %) of monthly NUTS2 estimates with additional auxiliaries on NUTS2 from Q-1

Employed			
	Level	Change_t1	Change_ty
MR1	99	100	99
MR2	98	99	99
$\mathrm{RC}_{\alpha=0.75}$	98	99	98

Unemployed			
	Level	Change_t1	Change_ty
MR1	98	100	98
MR2	98	98	99
$\mathrm{RC}_{\alpha=0.75}$	97	99	99

4.5 Summary and outlook

We evaluated the use of composite estimators for the production of employment statistics under the new design of the German Microcensus. We started by presenting the formulas of different composite estimators, the design and design changes in the Microcensus in 2020, and an analyses which sample overlap information available from the Microcensus sampling design should be used in the composite estimators. After that, we presented adjustments of the formulas of the composite estimators to account for the regionally heterogeneous sample overlaps arising from the Microcensus design.

In a design-based simulation study, we compared the performance of the adjusted composite estimators in a setting replicating the population and design od the Microcensus as well as the production of employment estimates used in the survey. For that, the RIFOSS dataset was used as the simulation population, see Chapter 3 for further details on the generation of longitudinal employment information in this dataset. In the Microcensus there are two candidate overlaps which can be used in composite estimators, the overlap to the previous quarter (Q-1) and previous year (Q-4). Both options were studied in the simulation study, on NUTS1- and NUTS2-level. The estimators were evaluated for quarterly and monthly estimation of the levels, change to the previous time point (change_t1), and change to the previous year (change_ty) of employed and unemployed at the NUTS2-level.

The simulation study showed that there is no systematic bias in the adjusted composite estimators, which indicates that the presented adjustments of the formulas of the composite estimators work as expected. Among the composite estimators, the MR1, MR2, and RC ($\alpha = 0.75$) estimator showed the best performance, in terms of MSE, for the different estimation goals. The efficiency gains over the corresponding GREG estimator were generally higher for the estimation of employment than for unemployment, higher for the estimation of change_t1 than levels, and higher with NUTS2- instead of NUTS1-level sample overlap information. Overall, the interest of LFSs is mostly in the estimation of levels and recent changes. Based on the results of the simulation study, we recommend the use of the MR2 and RC$_{\alpha=0.75}$ estimator for the Microcensus, using the NUTS2-level auxiliaries based on the sample overlap to Q-1. The simulation study also showed that we can only expect performance gains from the estimators for those variables that are explicitly included as key study variables in the composite estimators. In the study, we achieve significant performance gains in the quarterly estimates but only small performance gains in the monthly estimates from composite estimators using quarterly sample overlap information.

The simulation study revealed further interesting results. The performance of the RC estimator was very sensitive with respect to the choice of α, indicating a kind of drift problem which was not apparent for the MR1 or MR2 estimator. The analysis showed that the sensitivity of the method is related to the magnitude of the sample overlap. Also the different versions of the AK estimator were sensitive to the magnitude of the sample overlap and performed worse than the corresponding GREG estimator for most quantities of interest. Overall, the simulation study sheds light on the interplay of sampling design and sample overlap for the performance of composite estimators. This was achieved in particular by examining not only their performance under a particular design, but also under different sets of sample overlaps as auxiliary information. In other studies, generally only a single sample overlap is studied with composite estimators.

In the design-based study we neither considered non-response nor potential differences between rotation groups as there was no historic information available on the two under the new German Microcensus design. In future research it would be interesting to see how a potential RGB influences the performance of the estimators. Potential future research also involves the investigation of variance estimation procedures for the composite estimators.

Chapter 5

Empirical Best Prediction in Multivariate Fay-Herriot Models

5.1 Introduction

In Chapter 4, we considered different design-based estimators for domain-specific employment estimation in the German Microcensus. The sample sizes in the considered domains were large enough such that the designed-based methods provided accurate estimates. However, if we were to consider finer and finer domains, e.g. estimates for finer and finer territorial units, we would find that the variances of the design-based methods become too large to call them accurate. That is, for ever smaller domains we run into small area problems. Among others, area-level model-based small area methods like the FH model, presented in Section 2.4.3, are designed for such problems and can give more accurate domain estimates than classic design-based estimators at the cost of being model-dependent.

In this chapter and the following Chapter 6 we present methodological developments of multivariate versions of the FH model. With the methodological extensions, we address two challenges that researchers face in practical applications of multivariate FH (MFH) models. Namely, the prediction of arbitrary multi-variable domain indicators (in this chapter) and dealing with partially missing direct estimates (in Chapter 6). The theory of the two chapter can be combined to predict multi-variable domain indicators with MFH models even for domains for which the corresponding multi-variable survey estimates are partially missing.

With MFH models, several dependent variables can be modelled jointly. For example, one can take the domain-specific totals of employed and unemployed persons as the dependent variables in a MFH model. As the model explicitly considers the joint distribution of the dependent variables, for many situation it leads to more precise predictions than the corresponding univariate FH models.

In practical applications, frequently the interest is not in the domain-specific values of different variables alone, but in arbitrary (potentially non-linear) indicators of these. Again, consider the totals of employed and unemployed. The interest can be in the totals themselves or non-linear combinations of them like the unemployment rate, which is defined as the total number of unemployed divided by the sum of employed and unemployed, or the relative difference of the two variables.

To compute survey estimates of such indicators, often plug-in estimators are used. That is, estimates for the components of an indicator are plugged into its formula to get an estimate. Plug-in estimators, however, do not consider the joint distribution of the input estimates, which can be inefficient. Furthermore, plug-in estimators are biased for non-linear functions and asymptotic unbiasedness of plug-in estimators cannot be assumed for small area problems with small sample sizes.

For small area problems, as an alternative to plug-in predictors, we examine the theory of best predictors (BPs) of multi-variable domain indicators under MFH models. By definition, the BPs are the predictors with minimum MSE in the class of all model-unbiased predictors and therefore theoretically advantageous to the plug-in predictors. For general indicators, the BPs of the MFH models are given by multi-dimensional integrals. As these integrals do not have a closed-form solution, we consider different integral approximations and compare the performance of the approximations to the performance of the corresponding plug-in predictors.

The chapter is structured as follows. We briefly cover different techniques for approximating multivariate integrals over the normal distribution in Section 5.2. The MFH model, together with additional remarks and a literature overview, is presented in Section 5.3. Based on the integration techniques of Section 5.2, we discuss different approaches for approximating the BPs of general multi-variable domain indicators in MFH models in Section 5.5. Furthermore, we present bootstrap approaches for their MSE estimation in Section 5.6. The different approaches are evaluated in model-based simulation studies in Section 5.7. In the studies, we also compare the performance of the BP approximations to the performance of the corresponding plug-in predictors. In an illustrative application to publicly available data from the Spanish LFS, we use the presented methodology to predict unemployment rate in Spanish provinces crossed by sex and age classes in Section 5.8. The chapter closes with a summary and outlook in Section 5.9.

5.2 Integral approximation

In Section 5.5, we are interested in m-dimensional integrals over the m-variate normal distribution. Formally, we are interested in m-dimensional integrals of the form

$$\int_{\mathbb{R}^m} g(\boldsymbol{x})f(\boldsymbol{x})d\boldsymbol{x}, \tag{5.158}$$

where $g : \mathbb{R}^m \to \mathbb{R}$ is a continuous and bounded function, and $f : \mathbb{R}^m \to \mathbb{R}$ is the m-variate normal density function with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Then, integral (5.158) exists and is finite, i.e. $|\int_{\mathbb{R}^m} g(\boldsymbol{x})f(\boldsymbol{x})d\boldsymbol{x}| < \infty$ holds. To name an example, we will later consider the case of $m = 2$ and g as the unemployment rate which takes the total number of employed and unemployed as input.

We assume that it is not analytically feasible to calculate (5.158) exactly. In the following, we therefore briefly cover some basics of numerical integration, so-called *quadrature* techniques, for approximating (5.158). The numerical integration techniques which we cover approximate the integral by a weighted sum of T function evaluations, i.e.

$$\int_{\mathbb{R}^m} g(\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x} \approx \sum_{t=1}^{T} w_t g(\boldsymbol{x}_t). \tag{5.159}$$

The points where the function is evaluated, $\boldsymbol{x}_t \in \mathbb{R}^m$, are called *nodes*, with corresponding *weights* $w_t \in \mathbb{R}$, $t = 1 \ldots, T$.

We consider the integral approximation techniques Gauss-Hermite, Monte Carlo (MC), and Quasi MC. In each method, the nodes and weight of the integral approximation (5.159) are chosen differently. Gauss-Hermite quadrature can especially be used for low-dimensional problems, provided that the function g can be well approximated polynomially. For higher-dimensional problems and situations where little is known about the functional form of g, MC and Quasi MC integration are more commonly used. As we provide a general approach for approximating BPs for m-dimensional problems, we briefly cover all listed methods. For a concrete application, choosing the appropriate approximation method depends on the form of the concrete non-linear indicator of interest and its dimension. We refer to Gentle (2003, p. 233) or Press et al. (2007, Section 4.8) for a general discussion of the different multi-dimensional integration methods. González et al. (2006) compared the methods under different dimensionalities in the context of logistic-normal models.

5.2.1 Gauss-Hermite quadrature

General form

Gauss-Hermite quadrature is a specific Gaussian quadrature method. We refer to Press et al. (2007, Section 4.6) for general information on Gaussian quadrature methods. We will see why Gauss-Hermite quadrature is particularly useful for integrating over the normal distribution shortly. We start with one-dimensional integration, i.e. $m = 1$. In Gauss-Hermite quadrature, we have (Davis & Rabinowitz, 1984, p. 222)

$$\int_{\mathbb{R}} \tilde{g}(x) \exp\Big(-x^2\Big) dx = \sum_{t=1}^{T} w_t \tilde{g}(x_t), \tag{5.160}$$

where $\tilde{g}$ is a polynomial of degree $2T - 1$. The nodes in (5.160) correspond to the roots of the Hermite polynomial $H_T(x)$. Hermite polynomials $H_T(x)$ have the recurrence relation (Davis & Rabinowitz, 1984, p. 41)

$$H_0(x) = 1,\ H_1(x) = 2x,\ H_{T+1}(x) = 2xH_T(x) - 2T\ H_{T-1}(x). \tag{5.161}$$

A detailed description of Hermite polynomials is given in Davis and Rabinowitz (1984, Sections 1.12, 1.13) and Hochstrasser (1968). The weights w_t of (5.160) are given by (Davis & Rabinowitz, 1984, p. 224)

$$w_t = \frac{2^{T+1} T! \sqrt{\pi}}{(H_{T+1}(x_t))^2}. \tag{5.162}$$

Press et al. (2007, pp. 185–186) present an algorithm for calculating the Gauss-Hermite nodes and weights which exploits recurrence relations of Hermite polynomials and is especially useful for large T. From the above formulas we see that the Gauss-Hermite nodes and weights can be calculated independently from function g. They can therefore be found in standard tables for various T, e.g. Davis and Polonsky (1968, Table 25.10). The better the function of interest g can be approximated by polynomials of degree T, the better the Gauss-Hermite quadrature approximates the integral (5.158).

Gauss-Hermite quadrature for specific normal distribution

Let us consider the univariate case, $m = 1$, and f as the density of a normal distribution with mean μ and standard deviation σ. The integral which we are interested in is given by

$$\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) g(y) dy. \tag{5.163}$$

To apply Gauss-Hermite quadrature, we can transform (5.163) in accordance with the left hand side of (5.160) . By applying the substitution rule, similar to e.g. Liu and Pierce (1994), taking $x = (y-\mu)/\sqrt{2}\sigma \Leftrightarrow y = \sqrt{2}\sigma x + \mu$, and $dy = \sqrt{2}\sigma dx$, we get

$$\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-x^2\right) g(\sqrt{2}\sigma x + \mu)\sqrt{2}\sigma dx = \frac{1}{\sqrt{\pi}} \int_{\mathbb{R}} \exp\left(-x^2\right) g(\sqrt{2}\sigma x + \mu) dx. \tag{5.164}$$

Integral (5.163) can thus be approximated by Gauss-Hermite quadrature via

$$\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) g(y) dy \approx \sum_{t=1}^{T} \tilde{w}_t g(\tilde{x}_t), \tag{5.165}$$

with $\tilde{x}_t = \sqrt{2}\sigma x_t + \mu$, and $\tilde{w}_t = \pi^{-0.5} w_t$, $t = 1, \ldots, T$. The x_t and w_t are the standard Gauss-Hermite nodes and weights for T function evaluations. $\tilde{w}_t = \pi^{-0.5} w_t$ are also known as *normalized weights.*

Multi-dimensional Gauss-Hermite quadrature

For multi-dimensional integrals such as (5.158), we can repeat the one-dimensional quadrature over the different dimensions, according to the so-called *product rule.* We then have, compare e.g. Cools (1997, p. 10),

$$\int_{\mathbb{R}^m} \exp\left(-\boldsymbol{x}^\top \boldsymbol{x}\right) g(\boldsymbol{x}) d\boldsymbol{x} \approx \sum_{t_1=1}^{T} \cdots \sum_{t_m=1}^{T} \prod_{k=1}^{m} w_{t_k} g(x_{t_1}, \ldots, x_{t_m}). \tag{5.166}$$

Next to the product rule, there exist various other techniques for multivariate numerical quadrature like sparse grids. We only consider the product rule and refer to Cools (1997, 2002) and Stroud (1971) for alternative methods. When we consider T evaluations for each dimension, we end up with a total of T^m evaluations in multivariate quadrature with product rule. Therefore, for higher-dimensional integrals, the use of (Quasi) MC methods is often preferred. These are described in Sections 5.2.2 and 5.2.3.

Multi-dimensional Gauss-Hermite quadrature for m-variate normal distribution

For a concrete m-dimensional multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, we are interested in approximating an integral

$$\int_{\mathbb{R}^m} \frac{1}{(2\pi)^{m/2}\det(\boldsymbol{\Sigma})^{1/2}} \exp\Big(-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\Big) g(\boldsymbol{y}) d\boldsymbol{y}. \tag{5.167}$$

We can again apply a substitution of the integration variable to get to the Gauss-Hermite quadrature form. For that we need a decomposition of the covariance matrix $\boldsymbol{\Sigma}$ in the form $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^\top$, where $\boldsymbol{L}$ is an $m \times m$ matrix, like the Cholesky decomposition, e.g. described in Press et al. (2007, p. 378), or the eigen-decomposition, e.g. described in Wood (2017, B.9).

We take $\boldsymbol{x} = 2^{-0.5}\boldsymbol{L}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}) \Leftrightarrow \boldsymbol{y} = \sqrt{2}\boldsymbol{L}\boldsymbol{x} + \boldsymbol{\mu}$, note that $\sqrt{\det(\boldsymbol{\Sigma})} = |\det(\boldsymbol{L})|$ and $d\boldsymbol{y} = \sqrt{2}|\det(\boldsymbol{L})|d\boldsymbol{x}$ and get

$$\begin{aligned} &\int_{\mathbb{R}^m} \frac{1}{(2\pi)^{m/2}\det(\boldsymbol{\Sigma})^{1/2}} \exp\Big(-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\Big) g(\boldsymbol{y}) d\boldsymbol{y} \\ =&\pi^{-m/2}\int_{\mathbb{R}^m} \exp\Big(-\boldsymbol{x}^\top\boldsymbol{x}\Big) g(\sqrt{2}\boldsymbol{L}\boldsymbol{x}+\boldsymbol{\mu}) d\boldsymbol{x}. \end{aligned} \tag{5.168}$$

Integral (5.158) can thus be approximated by Gauss-Hermite quadrature with product rule via

$$\begin{aligned} &\int_{\mathbb{R}^m} \frac{1}{(2\pi)^{m/2}\det(\boldsymbol{\Sigma})^{1/2}} \exp\Big(-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\Big) g(\boldsymbol{y}) d\boldsymbol{y} \\ \approx&\sum_{t_1=1}^{T} \cdots \sum_{t_m=1}^{T} \prod_{k=1}^{m} \tilde{w}_{t_k} g(\tilde{x}_{t_1,\dots,t_m}) \end{aligned} \tag{5.169}$$

with $\tilde{w}_{t_k} = \pi^{-m/2} w_{t_k}$, $\tilde{x}_{t_1,\dots,t_m} = \sqrt{2}\boldsymbol{L}(x_{t_1},\dots,x_{t_m}) + \boldsymbol{\mu}$, weights w_{t_k} and nodes x_{t_k} from the univariate Gauss-Hermite quadrature, $t = 1,\dots,T$, $k = 1,\dots,m$. This product formula for Gauss-Hermite quadrature for a specific multivariate normal distribution is e.g. used in Wu et al. (2006) and Judd et al. (2011).

5.2.2 Monte Carlo integration

MC integration is a numerical integration technique which gives unbiased estimates of integrals, independent from the concrete functional form of g. It is especially useful for higher-dimensional integrals and when the function g cannot be well approximated by polynomials. Another advantage of MC integration is that it is always possible draw more random numbers until the convergence of the approximation is satisfactory. This is not possible for the Gauss-Hermite quadrature as the weights and nodes depend on the chosen number of function evaluations T. In Gauss-Hermite quadrature, the nodes for a chosen number of function evaluations $\tilde{T}$ cannot be reused for any other $\tilde{T} \neq T$.

MC integration is based on random numbers. The integral in (5.158) over the m-variate normal density f with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ can be approximated via MC integration as

$$T^{-1}\sum_{t=1}^{T} g(\boldsymbol{x}_t), \quad \boldsymbol{x}_t \sim N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad t = 1, \ldots, T. \tag{5.170}$$

The formula corresponds to (5.159) with weights $w_t = 1/T$, $t = 1, \ldots, T$, and random draws from the multivariate normal distribution f as nodes $\boldsymbol{x}_t$. The approximation error is of order $\mathcal{O}(1/\sqrt{T})$. That is, if we want to half the MC approximation error, we have to multiply the number of function evaluations T by four (Gentle, 2003, p. 233).

We refer to Gentle (2007) and Press et al. (2007, Chapter 7) for information on random number generation for specific distributions and the features of the algorithms. In this thesis, we use `R` package `baseR` with functions `runif` and `rnorm` whenever we use random numbers from the uniform and normal distribution respectively. To generate multivariate normal draws from univariate normal draws, Algorithm 5.1 can be used. The algorithm is based on Press et al. (2007, Section 7.8), who also present a proof.

Algorithm 5.1 Generation of realisations from multivariate normal distribution

```
The aim is to have $n$ random numbers from a specific $m$-multivariate normal distribution
with mean vector $\boldsymbol{\mu} \in \mathbb{R}^m$ and $m \times m$ covariance matrix $\boldsymbol{\Sigma}$.
    1. Take a vector of random numbers $\tilde{\boldsymbol{x}} \overset{iid}{\sim} N(0,1)$ of length $m \times n$.
    2. With $\tilde{\boldsymbol{x}}$, form a $n \times m$ matrix $\tilde{\boldsymbol{X}}$ of arbitrary order.
    3. Calculate (Cholesky or eigenvalue-) decomposition $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^\top$.
    4. Take $\boldsymbol{X} = \boldsymbol{L}\tilde{\boldsymbol{X}} + \boldsymbol{\mu}$.
The resulting $\boldsymbol{X}$ is a matrix of $n$ random realisations from the desired $m$-variate normal
distribution.
```

Antithetic variates

In the standard MC integration, we aim for independently and identically distributed random numbers. To reduce the variance of the approximation, there exist several approaches which use correlated values as integration nodes. An overview of these variance reduction techniques is for example given in Gentle (2003, section 7.5). Here, we only consider *antithetic variates* as a variance reduction technique. The following description is based on Gentle (2003, p. 245).

When integrating over a symmetric distribution like the normal or uniform distribution, we can use the symmetry to reduce the variance of the MC integral approximation. That is, for each random draw, we define an antithetic variate corresponding to the symmetric counterpart of that random draw. With antithetic variates, only $T/2$ random draws have to be generated to get T function evaluations. For T function evaluations, the approximation of (5.158) with MC and antithetic variates is calculated as

$$T^{-1}\sum_{i=1}^{T/2}(g(\boldsymbol{x}_t) + g(\tilde{\boldsymbol{x}}_t)), \quad \boldsymbol{x}_t \sim N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \ \tilde{\boldsymbol{x}}_t = 2\boldsymbol{\mu} - \boldsymbol{x}_t, \quad t = 1, \ldots, T/2. \tag{5.171}$$

Following the description in Gentle (2003, p. 246), the variance of approximation (5.171) is the variance of a sum. The variance of a sum of two elements equals the variance of the summands plus two times the covariance, compare e.g. (4.133). The covariance of the nodes and their antithetic variates is negative. Therefore, antithetic variates lead to a variance reduction of MC quadrature.

5.2.3 Quasi Monte Carlo integration

MC integration is based on random numbers, which simulate random processes. Especially for a small to medium number of function evaluations T it can happen that, by chance, the generated values are not evenly spaced. For example, samples from two independent uniform distributions can by chance be clustered on a unit-cube. We already saw that we can reduce the MC variance by taking into account the symmetry of the uniform or normal distribution with the use of antithetic variates. With the standard normal distribution, using antithetic variates ensures that the mean of the random numbers plus antithetic variates is equal to zero.

The idea of correlated numbers is taken even further with *quasi-random numbers*. We base our description of quasi-random numbers on Gentle (2003, Chapter 3). The idea of quasi-random numbers is to define algorithms which produce sequences of values which, on the unit cube, are maximally apart from each other and thereby span the unit cube more evenly than random numbers would. The resulting sequences are referred to as *low-discrepancy sequences*. A particular advantage of these sequences compared to Gauss-Hermite is that we can always produce additional values of a quasi-random sequence when we want to increase the number of function evaluations T. That is, when T is increased, the already produced quasi-random numbers can be re-used. There is no stochastic element in quasi-random numbers, they are deterministic. For general overview of quasi-random number theory and different quasi-random sequences and their algorithms, we refer to Press et al. (2007, Section 7.8), Gentle (2003, Chapter 3), and the references therein.

In this chapter, we consider the *Halton sequence* (Halton, 1960) and the *Sobol sequence* (Sobol', 1967, 1976) for quasi-random number generation. The algorithms for generating these sequences are e.g. given in Press et al. (2007, Section 7.8). The sequences are implemented in standard statistical software, e.g. the `R` (R Core Team, 2020) package `fOptions` (Wuertz et al., 2017).

Figure 5.1 displays 200 random numbers and numbers from the Halton and Sobol sequence, all for both the bivariate uniform distribution in interval $[0, 1]$ and the standard bivariate normal distribution. For the random numbers with antithetic variates, the first 100 values are the random numbers, the rest 100 values are their antithetic variates. Especially when we focus on the unit cube, we see that the random numbers, by chance, tend to cluster and the quasi-random numbers span the grid more evenly. Focusing on the standard bivariate normal, we see that the antithetic variates ensue that the values are placed symmetrically around the origin, compared to the completely random numbers. We can furthermore see that both the Halton and Sobol sequence result in different sets of quasi-random numbers.

For higher-dimensional problems, we note that Gentle (2003, p. 95) recommended not to use the Halton sequence. Furthermore, we refer to the studies in Jäckel (2002, Chapter 8), who compared different quasi-random sequences for high-dimensional finance applications. We only consider low-dimensional integrals in this chapter. The unemployment rate as a function of employment and unemployment, for example, requires the approximation of a two-dimensional integral only.

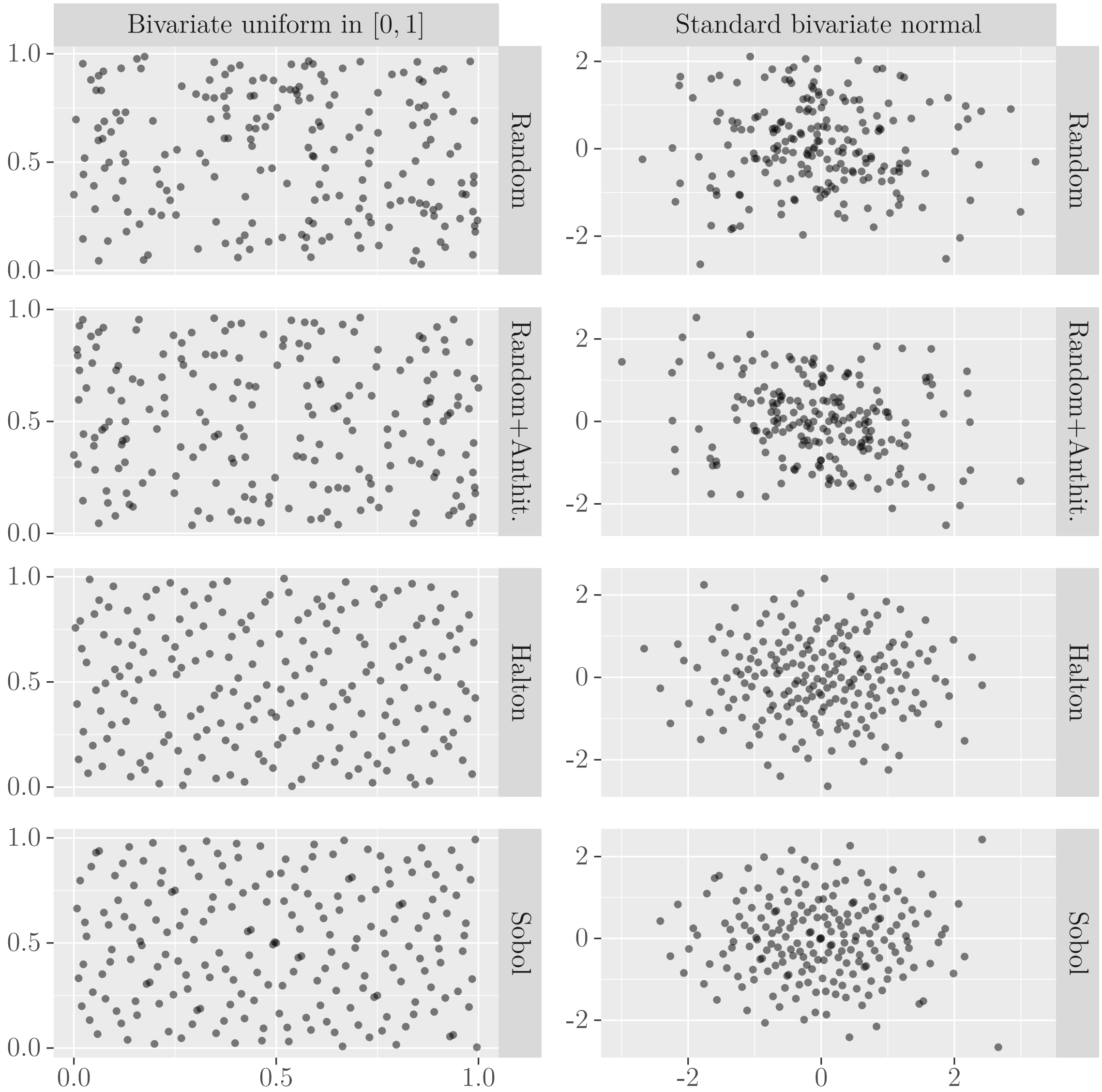

Figure 5.1: 200 (quasi-)random numbers

5.3 Multivariate Fay-Herriot model

5.3.1 Model

Multivariate models are those in which not one, but several dependent variables are modelled simultaneously. Instead of a joint multivariate model, one can also calculate separate univariate models for each dependent variable. However, a joint modelling approach comes with several advantages, e.g. listed in Snijders and Bosker (2012, Section 16.1). If the dependent variables are sufficiently correlated, a joint modelling approach can result in more precise parameter estimates and predictions than the corresponding univariate models. A multivariate analysis allows to make inferences about the covariance structure of the dependent variables such as in-group variances. There are also research questions where the focus is on the simultaneous effect of some auxiliary variables on several dependent variables; only a multivariate analysis is suitable then. In Sections 5.4 and 5.5, we use the joint distribution of the dependent variables to give best predictions of multi-variable indicators.

We presented the (univariate) FH model as a special kind of a LMM with block-diagonal covariance structure in Section 2.4.3. A *multivariate FH* (MFH) model is a FH model which takes into account multiple dependent variables. The MFH model was introduced by Datta et al. (1991) and Fay (1987) and further investigated by Benavent and Morales (2016). A detailed description of the model for the case of two dependent variables, also called *bivariate FH model*, was given in Morales et al. (2021, Chapter 19). In the following, we describe the MFH model, based on Benavent and Morales (2016).

Let U be a finite population which can be partitioned into D domains $U_1, \ldots, U_D$, $\boldsymbol{\mu}_d = (\mu_{d1}, \ldots, \mu_{dm})^\top$ be a vector of m characteristics of interest in domain d, and $\boldsymbol{y}_d = (y_{d1}, \ldots, y_{dm})^\top$ be a vector of the corresponding m direct estimates of $\boldsymbol{\mu}_d$, calculated by using the data of the target survey sample, $d = 1, \ldots, D$.

The multivariate Fay-Herriot model is defined in two stages. The first stage indicates that direct estimators $\boldsymbol{y}_d$ are unbiased and follow the sampling model

$$\boldsymbol{y}_d = \boldsymbol{\mu}_d + \boldsymbol{e}_d, \quad d = 1, \ldots, D, \tag{5.172}$$

where the vectors $\boldsymbol{e}_d = (e_{d1}, \ldots, e_{dm})^\top \sim N_m(0, \boldsymbol{V}_{ed})$ are independent with known covariance matrices $\boldsymbol{V}_{ed} \in \mathbb{R}^{m \times m}$. The covariance matrices $\boldsymbol{V}_{ed}$ of direct estimators $\boldsymbol{y}_d$ are given by

$$\boldsymbol{V}_{ed} = \begin{pmatrix} \sigma^2_{ed1} & \sigma_{ed12} & \cdots & \sigma_{ed1m} \\ \sigma_{ed12} & \sigma^2_{ed2} & \cdots & \sigma_{ed2m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{ed1m} & \sigma_{ed2m} & \cdots & \sigma^2_{edm} \end{pmatrix}, \quad d = 1, \ldots, D. \tag{5.173}$$

The diagonal elements of $\boldsymbol{V}_{ed}$ correspond to the variances of the m direct estimators, $\sigma_{edab} = \rho_{edab}\sqrt{\sigma^2_{eda}\sigma^2_{edb}}$ is the sampling error covariance of variables a and b with sampling error correlation ρ_{edab}, $a, b = 1, \ldots, m$, $a \neq b$.

In the second stage, for domain $d = 1, \ldots, D$, the true domain characteristic μ_{dk} is assumed to be linearly related to p_k explanatory variables, $k = 1, \ldots, m$. We define $p = \sum_{k=1}^{m} p_k$ as the total number of explanatory variables for all m target variables. The domain-specific aggregates of the p_k explanatory variables for μ_{dk} are given by $\boldsymbol{x}_{dk} = (x_{dk1}, \ldots, x_{dkp_k})^\top$, $k = 1, \ldots, m$. For every domain, we can combine the auxiliary information of the m dependent variables into a $m \times p$ block-diagonal auxiliary matrix $\boldsymbol{X}_d = \operatorname{diag}\left(\boldsymbol{x}_{d1}^\top, \ldots, \boldsymbol{x}_{dm}^\top\right)$, which is assumed to be of full rank. Let $\boldsymbol{\beta}_k = (\beta_{k1}, \ldots, \beta_{kp_k})^\top \in \mathbb{R}^{p_k}$ contain the regression parameters for μ_{dk}, $k = 1, \ldots, m$, and let $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_m^\top)^\top \in \mathbb{R}^p$ be the vector of fixed effects for all m characteristics.

The linking model is

$$\boldsymbol{\mu}_d = \boldsymbol{X}_d\boldsymbol{\beta} + \boldsymbol{u}_d, \quad \boldsymbol{u}_d = (u_{d1}, \ldots, u_{dm})^\top \sim N_m(0, \boldsymbol{V}_{ud}), \quad d = 1, \ldots, D, \tag{5.174}$$

where random effects $\boldsymbol{u}_d$ and sampling errors $\boldsymbol{e}_d$ are independent and vectors $\boldsymbol{u}_a$ and $\boldsymbol{u}_b$ are independent, $a, b = 1, \ldots, D$, $a \neq b$.

The covariance matrix of the random effects $\boldsymbol{V}_{ud} \in \mathbb{R}^{m \times m}$ depends on $q = m\,(m+1)\,/2$ variance parameters, consisting of m variances and $m\,(m-1)\,/2$ covariances. It is given by

$$\boldsymbol{V}_{ud} = \begin{pmatrix} \sigma_{u1}^2 & \rho_{12}\sigma_{u1}\sigma_{u2} & \cdots & \rho_{1m}\sigma_{u1}\sigma_{um} \\ \rho_{12}\sigma_{u1}\sigma_{u2} & \sigma_{u2}^2 & \cdots & \rho_{2m}\sigma_{u2}\sigma_{um} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1m}\sigma_{u1}\sigma_{um} & \rho_{2m}\sigma_{u2}\sigma_{um} & \cdots & \sigma_{um}^2 \end{pmatrix}, \tag{5.175}$$

where σ_{ua}^2 denotes the random effects variance of variable a and ρ_{ab} denotes the random effect correlation of variables a and b, $a, b = 1, \ldots, m$, $a \neq b$. The vector of variance parameters is denoted by

$$\boldsymbol{\theta} = (\sigma_{u1}^2, \sigma_{u2}^2, \ldots, \sigma_{um}^2, \rho_{12}, \rho_{13}, \ldots, \rho_{23}, \rho_{24}, \ldots, \rho_{m-1,m})^\top \in \mathbb{R}^q. \tag{5.176}$$

The first m elements of $\boldsymbol{\theta}$ correspond to the random effect variances, the $q - m$ last elements correspond to the random effect correlations.

The covariance matrix $\boldsymbol{V}_{ud}$ given by formula (5.175) is an *unspecified* covariance matrix with $q = m\,(m+1)\,/2$ unknown parameters. Depending on the data, it can make sense to further specify the matrix and thereby restrict the number of unknown parameters. For example, in the context of MFH models, Benavent and Morales (2016) specified $\boldsymbol{V}_{ud}$ according to an auto-regressive process. Models with a specified covariance matrix are nested within the model with unspecified $\boldsymbol{V}_{ud}$. Therefore, they can be compared via *likelihood ratio tests* (Littell, 2002, p. 482). We refer to Littell et al. (2004) for some examples of specifications of $\boldsymbol{V}_{ud}$. In this thesis we only consider unspecified covariance matrices $\boldsymbol{V}_{ud}$.

Taking together the sampling and linking model, the MFH model can be expressed as a single model in the form

$$\boldsymbol{y}_d = \boldsymbol{X}_d\boldsymbol{\beta} + \boldsymbol{u}_d + \boldsymbol{e}_d, \quad d = 1, \ldots, D, \tag{5.177}$$

or in matrix form

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u} + \boldsymbol{e}, \tag{5.178}$$

with

$$\begin{aligned} \boldsymbol{y} &= \underset{1\le d\le D}{\operatorname{col}}(\boldsymbol{y}_d) \in \mathbb{R}^{mD}, \quad \boldsymbol{u} = \underset{1\le d\le D}{\operatorname{col}}(\boldsymbol{u}_d) \in \mathbb{R}^{mD}, \\ \boldsymbol{e} &= \underset{1\le d\le D}{\operatorname{col}}(\boldsymbol{e}_d) \in \mathbb{R}^{mD}, \quad \boldsymbol{X} = \underset{1\le d\le D}{\operatorname{col}}(\boldsymbol{X}_d) \in \mathbb{R}^{mD\times p}. \end{aligned} \tag{5.179}$$

The MFH model (5.177) is a generalisation of the bivariate, trivariate, and multivariate Fay-Herriot models studied by Burgard et al. (2021c), Esteban et al. (2020), and Benavent and Morales (2016) respectively. Under model (5.177), it holds that $\boldsymbol{y} \sim N_m(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{V})$ with covariance matrix $\boldsymbol{V} = \boldsymbol{V}_u + \boldsymbol{V}_e = \underset{1\le d\le D}{\operatorname{diag}}(\boldsymbol{V}_d) \in \mathbb{R}^{mD\times mD}$, where

$$\boldsymbol{V}_u = \underset{1\le d\le D}{\operatorname{diag}}(\boldsymbol{V}_{ud}), \quad \boldsymbol{V}_e = \underset{1\le d\le D}{\operatorname{diag}}(\boldsymbol{V}_{ed}), \quad \boldsymbol{V}_d = \boldsymbol{V}_{ud} + \boldsymbol{V}_{ed}, \quad d = 1, \dots, D. \tag{5.180}$$

For the variance components, Lemma 5.1 gives a useful relationship.

Lemma 5.1. It holds that

$$\boldsymbol{V}_{ud}\boldsymbol{V}_d^{-1} = (\boldsymbol{V}_{ed}^{-1} + \boldsymbol{V}_{ud}^{-1})^{-1}\boldsymbol{V}_{ed}^{-1}, \quad d = 1, \dots, D \tag{5.181}$$

Proof. By applying the inversion formula

$$(\boldsymbol{A} + \boldsymbol{B}\boldsymbol{C}\boldsymbol{D})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{C}^{-1} + \boldsymbol{D}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{D}\boldsymbol{A}^{-1}, \tag{5.182}$$

with $\boldsymbol{A} = \boldsymbol{V}_{ed}^{-1}$, $\boldsymbol{C} = \boldsymbol{V}_{ud}^{-1}$, $\boldsymbol{B} = \boldsymbol{D} = \boldsymbol{I}_m$, where $\boldsymbol{I}_m$ is the $m \times m$ identity matrix, we have

$$\left(\boldsymbol{V}_{ed}^{-1} + \boldsymbol{V}_{ud}^{-1}\right)^{-1} = \boldsymbol{V}_{ed} - \boldsymbol{V}_{ed}(\boldsymbol{V}_{ud} + \boldsymbol{V}_{ed})^{-1}\boldsymbol{V}_{ed} = \boldsymbol{V}_{ed} - \boldsymbol{V}_{ed}\boldsymbol{V}_d^{-1}\boldsymbol{V}_{ed}. \tag{5.183}$$

Therefore

$$\begin{aligned} (\boldsymbol{V}_{ed}^{-1} + \boldsymbol{V}_{ud}^{-1})^{-1}\boldsymbol{V}_{ed}^{-1} &= \left(\boldsymbol{V}_{ed} - \boldsymbol{V}_{ed}\boldsymbol{V}_d^{-1}\boldsymbol{V}_{ed}\right)\boldsymbol{V}_{ed}^{-1} = \boldsymbol{I}_m - \boldsymbol{V}_{ed}\boldsymbol{V}_d^{-1} && (5.184) \\ &= (\boldsymbol{V}_d - \boldsymbol{V}_{ed})\boldsymbol{V}_d^{-1} = \boldsymbol{V}_{ud}\boldsymbol{V}_d^{-1}, \quad d = 1, \dots, D. && (5.185) \end{aligned}$$

□

5.3.2 Parameter estimation

The vector of model parameters is $\boldsymbol{\psi} = (\boldsymbol{\beta}^\top, \boldsymbol{\theta}^\top)^\top \in \mathbb{R}^{p+q}$. For estimating $\boldsymbol{\psi}$, we can use the ML or REML approach with a Fisher-Scoring algorithm as described in Section 2.4.2.

In Chapter 6, we present the Fisher-Scoring algorithm for the MFH model under partially missing direct estimates (MMFH). When there are no missing direct estimates, the MMFH model reduces to the MFH model. Therefore, the algorithms presented in Section 6.5 can also be used for the MFH model and we refrain from repeating the formulas here.

5.3.3 Prediction

Proposition 5.1. The conditional distribution of $\boldsymbol{u}_d$ given $\boldsymbol{y}_d$ is multivariate normal with with mean vector and variance matrix

$$\mathrm{E}\left[\boldsymbol{u}_d|\boldsymbol{y}_d;\boldsymbol{\beta},\boldsymbol{\theta}\right] = \boldsymbol{\Phi}_d\boldsymbol{V}_{ed}^{-1}(\boldsymbol{y}_d-\boldsymbol{X}_d\boldsymbol{\beta}) = \boldsymbol{V}_{ud}\boldsymbol{V}_d^{-1}(\boldsymbol{y}_d-\boldsymbol{X}_d\boldsymbol{\beta}), \tag{5.186}$$

$$\mathrm{Var}(\boldsymbol{u}_d|\boldsymbol{y}_d;\boldsymbol{\beta},\boldsymbol{\theta}) = \boldsymbol{\Phi}_d = \left(\boldsymbol{V}_{ed}^{-1}+\boldsymbol{V}_{ud}^{-1}\right)^{-1} = \boldsymbol{V}_{ud}\boldsymbol{V}_d^{-1}\boldsymbol{V}_{ed}, \tag{5.187}$$

$$d=1,\ldots,D.$$

Proof. We recall that the kernel of the m-variate normal probability density function for variables $\tilde{Y}_1,\ldots,\tilde{Y}_m$ with mean $\tilde{\boldsymbol{\mu}}$ and covariance matrix $\tilde{\boldsymbol{\Sigma}}$ is

$$\begin{aligned} f(\tilde{\boldsymbol{y}}|\tilde{\boldsymbol{\mu}},\tilde{\boldsymbol{\Sigma}}) &= \frac{1}{(2\pi)^{m/2}\det\left(\tilde{\boldsymbol{\Sigma}}\right)^{1/2}}\exp\Big\{-\frac{1}{2}(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{\mu}})^\top\tilde{\boldsymbol{\Sigma}}^{-1}(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{\mu}})\Big\} \\ &\propto \exp\Big\{-\frac{1}{2}\tilde{\boldsymbol{y}}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\boldsymbol{y}}+\tilde{\boldsymbol{\mu}}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\boldsymbol{y}}\Big\}. \end{aligned} \tag{5.188}$$

The conditional distribution of $\boldsymbol{u}_d$ given $\boldsymbol{y}_d$, for $d=1,\ldots,D$, is then given by

$$\begin{aligned} &f(\boldsymbol{u}_d|\boldsymbol{y}_d) \\ &\quad\propto f(\boldsymbol{y}_d|\boldsymbol{u}_d)f(\boldsymbol{u}_d) \\ &\quad= \frac{1}{(2\pi)^{m/2}\det(\boldsymbol{V}_{ed})^{1/2}}\exp\Big\{-\frac{1}{2}(\boldsymbol{y}_d-\boldsymbol{X}_d\boldsymbol{\beta}-\boldsymbol{u}_d)^\top\boldsymbol{V}_{ed}^{-1}(\boldsymbol{y}_d-\boldsymbol{X}_d\boldsymbol{\beta}-\boldsymbol{u}_d)\Big\} \\ &\qquad\cdot\frac{1}{(2\pi)^{m/2}\det(\boldsymbol{V}_{ud})^{1/2}}\exp\Big\{-\frac{1}{2}\boldsymbol{u}_d^\top\boldsymbol{V}_{ud}^{-1}\boldsymbol{u}_d\Big\} \\ &\quad\propto \exp\Big\{-\frac{1}{2}\boldsymbol{u}_d^\top\boldsymbol{V}_{ed}^{-1}\boldsymbol{u}_d+\boldsymbol{u}_d^\top\boldsymbol{V}_{ed}^{-1}(\boldsymbol{y}_d-\boldsymbol{X}_d\boldsymbol{\beta})\Big\}\exp\Big\{-\frac{1}{2}\boldsymbol{u}_d^\top\boldsymbol{V}_{ud}^{-1}\boldsymbol{u}_d\Big\} \\ &\quad= \exp\Big\{-\frac{1}{2}\boldsymbol{u}_d^\top\left(\boldsymbol{V}_{ed}^{-1}+\boldsymbol{V}_{ud}^{-1}\right)\boldsymbol{u}_d+\boldsymbol{u}_d^\top\boldsymbol{\Phi}_d^{-1}\left(\boldsymbol{\Phi}_d\boldsymbol{V}_{ed}^{-1}(\boldsymbol{y}_d-\boldsymbol{X}_d\boldsymbol{\beta})\right)\Big\}. \end{aligned} \tag{5.189}$$

Therefore, $f(\boldsymbol{u}_d|\boldsymbol{y}_d)$ is a multivariate normal distribution with parameters

$$\mathrm{E}\left[\boldsymbol{u}_d|\boldsymbol{y}_d;\boldsymbol{\beta},\boldsymbol{\theta}\right] = \boldsymbol{\Phi}_d\boldsymbol{V}_{ed}^{-1}(\boldsymbol{y}_d-\boldsymbol{X}_d\boldsymbol{\beta}) \tag{5.190}$$

$$\mathrm{Var}(\boldsymbol{u}_d|\boldsymbol{y}_d;\boldsymbol{\beta},\boldsymbol{\theta}) = \left(\boldsymbol{V}_{ed}^{-1}+\boldsymbol{V}_{ud}^{-1}\right)^{-1} = \boldsymbol{\Phi}_d. \tag{5.191}$$

This and Lemma 5.1 complete the proof. □

If $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are known, the best predictors (BP) of $\boldsymbol{u}$ and $\boldsymbol{\mu}$ under the MFH model are

$$\hat{\boldsymbol{u}}^{\text{BP}} = \text{E}\left[\boldsymbol{u}_d | \boldsymbol{y}_d; \boldsymbol{\beta}, \boldsymbol{\theta}\right] = \boldsymbol{V}_u \boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}), \tag{5.192}$$

$$\hat{\boldsymbol{\mu}}^{\text{BP}} = \boldsymbol{X}\boldsymbol{\beta} + \hat{\boldsymbol{u}}^{\text{BP}}. \tag{5.193}$$

If $\boldsymbol{\theta}$ is known and $\boldsymbol{\beta}$ is unknown, the best linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$ and the best linear unbiased predictor (BLUP) of $\boldsymbol{u}$ and $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$ are

$$\hat{\boldsymbol{\beta}}^{\text{BLUE}} = (\boldsymbol{X}^\top \boldsymbol{V}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{V}^{-1} \boldsymbol{y}, \tag{5.194}$$

$$\hat{\boldsymbol{u}}^{\text{BLUP}} = \boldsymbol{V}_u \boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}^{\text{BLUE}}), \tag{5.195}$$

$$\hat{\boldsymbol{\mu}}^{\text{BLUP}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}^{\text{BLUE}} + \hat{\boldsymbol{u}}^{\text{BLUP}}. \tag{5.196}$$

By substituting $\boldsymbol{\theta}$ by a consistent estimator $\hat{\boldsymbol{\theta}}$, we obtain the empirical BLUE (EBLUE) of $\boldsymbol{\beta}$ and the empirical BLUP (EBLUP) of $\boldsymbol{u}$ and $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$, i.e.

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \hat{\boldsymbol{V}}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \hat{\boldsymbol{V}}^{-1} \boldsymbol{y}, \tag{5.197}$$

$$\hat{\boldsymbol{u}}^{\text{EBLUP}} = \hat{\boldsymbol{V}}_u \hat{\boldsymbol{V}}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}), \tag{5.198}$$

$$\hat{\boldsymbol{\mu}}^{\text{EBLUP}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{u}}^{\text{EBLUP}}, \tag{5.199}$$

where $\hat{\boldsymbol{V}}_u = \boldsymbol{V}_u(\hat{\boldsymbol{\theta}})$ and $\hat{\boldsymbol{V}} = \hat{\boldsymbol{V}}_u + \boldsymbol{V}_e$ are obtained by plugging $\hat{\boldsymbol{\theta}}$ in the place of $\boldsymbol{\theta}$.

5.3.4 Mean squared error

In the MFH model, we consider a vector of characteristics of interest $\boldsymbol{\mu}_d$ of length m in each domain $d = 1, \dots, D$. The MSE of the EBLUPs of $\boldsymbol{\mu}_d$ are therefore given by an $m \times m$ MSE matrix, which is often referred to as the *matrix of mean squared crossed errors*. The diagonal of the matrix corresponds to the MSEs of the single characteristics. The derivation of the matrix for some classes of MFH models is given in Benavent and Morales (2016). Morales et al. (2021, Chapter 19) gave a detailed description of the MSEs of the bivariate FH model. We note that the MSE matrix can also be derived from the formulation of LMMs with block-diagonal covariance matrices presented in Section 2.4.2, by rewriting (2.27) for $m \geq 2$ characteristics of interest.

For the MSE matrix, we recall that the vector of variance parameters $\boldsymbol{\theta}$ is of length q. The formulas of the MSE matrix in Morales et al. (2021, Chapter 19) are given for bivariate FH models, i.e. $m = 2$. For general $m \geq 2$, MSE approximation (2.72) becomes

$$\text{MSE}\left(\hat{\boldsymbol{\mu}}^{\text{EBLUP}}\right) \approx \boldsymbol{G}_1(\boldsymbol{\theta}) + \boldsymbol{G}_2(\boldsymbol{\theta}) + \boldsymbol{G}_3(\boldsymbol{\theta}) \tag{5.200}$$

with

$$\boldsymbol{G}_1\left(\boldsymbol{\theta}\right) = \boldsymbol{T}, \tag{5.201}$$

$$\boldsymbol{G}_2\left(\boldsymbol{\theta}\right) = (\boldsymbol{X} - \boldsymbol{T}\boldsymbol{V}_e^{-1}\boldsymbol{X})\boldsymbol{Q}(\boldsymbol{X}^\top - \boldsymbol{X}^\top\boldsymbol{V}_e^{-1}\boldsymbol{T}), \tag{5.202}$$

$$\boldsymbol{G}_3\left(\boldsymbol{\theta}\right) = \sum_{a=1}^{q}\sum_{b=1}^{q}\operatorname{Cov}(\hat{\theta}_a, \hat{\theta}_b)\boldsymbol{L}^{(a)}\boldsymbol{V}\boldsymbol{L}^{(b)\top}, \tag{5.203}$$

$$\boldsymbol{T} = \boldsymbol{V}_u - \boldsymbol{V}_u\boldsymbol{V}^{-1}\boldsymbol{V}_u, \quad \boldsymbol{Q} = (\boldsymbol{X}^\top\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}, \tag{5.204}$$

$$\boldsymbol{L} = \frac{\partial\boldsymbol{V}_u\boldsymbol{V}^{-1}}{\partial\sigma_u^2}, \quad \boldsymbol{L}^{(a)} = (\boldsymbol{I}_{mD} - \boldsymbol{V}_u\boldsymbol{V}^{-1})\frac{\partial\boldsymbol{V}_u}{\partial\theta_a}\boldsymbol{V}^{-1}, \quad a = 1, \dots, q. \tag{5.205}$$

$\boldsymbol{I}_{mD}$ is the $mD \times mD$ identity matrix. The MSE approximation (5.200) is of order $\mathcal{O}(D^{-1})_{mD\times mD}$. The covariances $\operatorname{Cov}(\hat{\theta}_a, \hat{\theta}_b)$, $a, b = 1, \dots, q$, depend on the estimation method, in Section 2.4.2 they are displayed for ML and REML estimators.

Similar to formula (2.78), based on Datta and Lahiri (2000) and Prasad and Rao (1990), an estimator of the MSE matrix is given by

$$\widehat{\text{MSE}}\left(\hat{\boldsymbol{\mu}}^{\text{EBLUP}}\right) = \boldsymbol{G}_1\left(\hat{\boldsymbol{\theta}}\right) + \boldsymbol{G}_2\left(\hat{\boldsymbol{\theta}}\right) + 2\boldsymbol{G}_3\left(\hat{\boldsymbol{\theta}}\right), \tag{5.206}$$

where $\hat{\boldsymbol{\theta}}$ is the vector of the q likelihood-based variance parameter estimates. In this chapter, we focus on parameter and MSE estimation via REML. For parameter estimation via ML, an additional bias term has to be considered, similar to (2.79). We refer to Datta and Lahiri (2000) for more information on the bias correction for ML.

5.3.5 Literature review and remarks

In the class of area-level small area models, multivariate Fay-Herriot models have received more attention in recent years. The MFH model was introduced in Fay (1987) and further studied by Benavent and Morales (2016), Datta et al. (1996), Datta et al. (1991), and Ghosh et al. (1996). The studies showed that the multivariate model can give predictions with lower MSE than the corresponding univariate models when the dependent variables are sufficiently correlated. Since then, there were many different contributions to MFH models of which we will list some, without being exhaustive.

González-Manteiga et al. (2008) evaluated different analytical and bootstrap MSE approximations for a special case of MFH models. Benavent and Morales (2016) proposed MSE estimators for general MFH models and compared different MFH model specifications. A detailed description of the theory of bivariate FH models was given in Morales et al. (2021, Chapter 19). Esteban et al. (2020) introduced area-level compositional mixed models by applying transformations to a MFH model.

MFH models are commonly used for time-series data, i.e. the dependent variables in the model correspond to one variable at different time points. Esteban et al. (2012) considered

MFH time models for poverty indicators. Benavent and Morales (2021) studied a bivariate FH model with independent time effects. Furthermore, Marchetti and Secondi (2017) and Ubaidillah et al. (2019) used MFH models for the prediction of household consumption expenditures. Porter et al. (2015) considered MFH models with latent spatial dependencies. Huang and Bell (2004, 2012) empirically investigated bivariate FH models for SAIPE poverty estimation.

Many research contributions were also made to consider different kinds of measurement errors in the covariates of FH models. Ybarra and Lohr (2008) investigated measurement errors in (univariate) FH models. In Lohr and Ybarra (2002) and Ybarra and Lohr (2008), the authors showed the connection of the resulting *measurement error model* to MFH models. Their model was further extended by Burgard et al. (2020b) for the univariate and Burgard et al. (2021a) for the bivariate case. Arima et al. (2017) considered measurement errors in MFH models using a Bayesian approach. Krause et al. (2022) proposed robust estimation in the presence of measurement errors for generalised versions of MFH models.

In the following, we give two additional remarks for the MFH model which help to understand certain results of the simulation studies which are presented in Section 5.7.

Remark 5.1. Special case: $\hat{\boldsymbol{\beta}}^{\mathrm{EBLUE}}$ and $\hat{\boldsymbol{\beta}}^{\mathrm{BLUE}}$ coincide in the FH model.

Consider the (univariate) FH model. It corresponds to the MFH model with $m = 1$ dependent variable and is described in Section 2.4.3. We consider the case $\sigma_{ed}^2 = \sigma_e^2$, $d = 1, \dots, D$, i.e. the sampling error variance is the same in all domains. In that case, $\hat{\boldsymbol{\beta}}^{\mathrm{EBLUE}}$ and $\hat{\boldsymbol{\beta}}^{\mathrm{BLUE}}$ coincide in the FH model. To see that, compare formula (2.84), where term $1/(\sigma_{ed}^2 + \sigma_u^2)$ cancels out when $\sigma_{ed}^2 = \sigma_e^2$, $d = 1, \dots, D$. Note that this remark only holds for the FH, not the MFH model.

Remark 5.2. Special case: FH and MFH model give the same predictions of random effects $\boldsymbol{u}_d$.

Consider $m = 2$ dependent variables and the special case $\sigma_{u1}^2 = \sigma_{u2}^2$, $\sigma_{ed1}^2 = \sigma_{ed2}^2$, $\sigma_{ed1}^2 = \sigma_{e1}^2$, $d = 1, \dots, D$, $\rho_u = \rho_e = \rho$. That is, the random effects variances of the two variables of interest coincide, the sampling error variances are the same in all areas and for both variables of interest, the random effects and sampling error correlation coincide. In that case, we can rewrite the variance matrices as follows

$$\boldsymbol{V}_{ud} = \begin{pmatrix} \sigma_{u1}^2 & \rho\sigma_{u1}\sigma_{u1} \\ \rho\sigma_{u1}\sigma_{u1} & \sigma_{u1}^2 \end{pmatrix} = \begin{pmatrix} \sigma_{u1}^2 & \rho\sigma_{u1}^2 \\ \rho\sigma_{u1}^2 & \sigma_{u1}^2 \end{pmatrix}, \tag{5.207}$$

$$\boldsymbol{V}_{ed} = \begin{pmatrix} \sigma_{e1}^2 & \rho\sigma_{e1}\sigma_{e1} \\ \rho\sigma_{e1}\sigma_{e1} & \sigma_{e1}^2 \end{pmatrix} = \begin{pmatrix} \sigma_{e1}^2 & \rho\sigma_{e1}^2 \\ \rho\sigma_{e1}^2 & \sigma_{e1}^2 \end{pmatrix}, \tag{5.208}$$

$$\boldsymbol{V}_d = \boldsymbol{V}_{ud} + \boldsymbol{V}_{ed} = \begin{pmatrix} \sigma_{u1}^2 + \sigma_{e1}^2 & \rho\left(\sigma_{u1}^2 + \sigma_{e1}^2\right) \\ \rho\left(\sigma_{u1}^2 + \sigma_{e1}^2\right) & \sigma_{u1}^2 + \sigma_{e1}^2 \end{pmatrix}, \tag{5.209}$$

$$\boldsymbol{V}_d^{-1} = \frac{1}{\det(\boldsymbol{V}_d)} \begin{pmatrix} \sigma_{u1}^2 + \sigma_{e1}^2 & -\rho\left(\sigma_{u1}^2 + \sigma_{e1}^2\right) \\ -\rho\left(\sigma_{u1}^2 + \sigma_{e1}^2\right) & \sigma_{u1}^2 + \sigma_{e1}^2 \end{pmatrix}, \quad d = 1, \dots, D. \tag{5.210}$$

In this special case, the off-diagonal elements of $\boldsymbol{V}_{ud}\boldsymbol{V}_d^{-1}$ are zero. For example, take the upper-right element of $\boldsymbol{V}_{ud}\boldsymbol{V}_d^{-1}$,

$$(\boldsymbol{V}_{ud}\boldsymbol{V}_d^{-1})_{(1,2)} = \frac{\sigma_{u1}^2\left(-\rho\left(\sigma_{u1}^2+\sigma_{e1}^2\right)\right)+\rho\sigma_{u1}^2\left(\sigma_{u1}^2+\sigma_{e1}^2\right)}{\det(\boldsymbol{V}_d)} = 0. \tag{5.211}$$

In this special case, the BLUP formulas of the random effects in the FH (2.85) and MFH model (5.195) give the same results.

The remark shows that also for non-zero random effects and sampling error correlations, there are special cases in which the MFH model does not give efficiency gains over the corresponding FH models. For illustration, we have shown the remark for the case of $m = 2$ variables. It can, however, be readily extended to the general multivariate case with $m \geq 2$ variables.

5.4 Empirical best prediction of linear domain indicators

In Section 5.3, we considered the prediction of the $\boldsymbol{\mu}_d \in \mathbb{R}^m$. We now take into account a more general concept and consider that the domain parameters of interest take the form

$$G_d = G_d(\boldsymbol{u}_d, \boldsymbol{\beta}) = g(\boldsymbol{\mu}_d), \quad d = 1, \ldots, D, \tag{5.212}$$

where $g : \mathbb{R}^m \mapsto \mathbb{R}$ is a function.

We first consider the case where G_d corresponds to a linear function in $\boldsymbol{u}_d$ and $\boldsymbol{\beta}$ such that g is a linear function in $\boldsymbol{\mu}_d$. For example, take a MFH model where the dependent variables are the domain totals of employed and unemployed persons. Assume our domain parameter of interest is the sum of employed and unemployed and $g : \mathbb{R}^2 \mapsto \mathbb{R}$. We can then simply plug-in the output of the MFH model to calculate (empirical) BLUPs of G_d and their MSE estimates as shown in Remark 5.3.

Remark 5.3. When G_d is a linear function in $\boldsymbol{u}_d$ and $\boldsymbol{\beta}$, the domain parameters of interest take the form

$$G_d = G_d(\boldsymbol{u}_d, \boldsymbol{\beta}) = g(\boldsymbol{\mu}_d) = \boldsymbol{\lambda}^\top \boldsymbol{\mu}_d + \alpha, \quad d = 1, \ldots, D, \tag{5.213}$$

with $\boldsymbol{\lambda} \in \mathbb{R}^m$ and $\alpha \in \mathbb{R}$ such that $g : \mathbb{R}^m \mapsto \mathbb{R}$ is linear function in $\boldsymbol{\mu}_d$, $d = 1, \ldots, D$. By the *linearity of expectation*, see e.g. Dekking et al. (2005, p. 137), it holds that

$$\mathrm{E}[\boldsymbol{\lambda}^\top \boldsymbol{\mu}_d + \alpha] = \boldsymbol{\lambda}^\top \, \mathrm{E}[\boldsymbol{\mu}_d] + \alpha, \quad d = 1, \ldots, D. \tag{5.214}$$

The BLUP of G_d is given by

$$\hat{G}_d^{\mathrm{BLUP}} = g(\hat{\boldsymbol{\mu}}_d^{\mathrm{BLUP}}) = \boldsymbol{\lambda}^\top \hat{\boldsymbol{\mu}}_d^{\mathrm{BLUP}} + \alpha, \quad d = 1, \ldots, D. \tag{5.215}$$

The EBLUP of G_d is obtained by substituting $\hat{\boldsymbol{\mu}}^{\text{BLUP}}$ in (5.199) for $\hat{\boldsymbol{\mu}}^{\text{EBLUP}}$ (5.193) and given by

$$\hat{G}_d^{\text{EBLUP}} = g(\hat{\boldsymbol{\mu}}_d^{\text{EBLUP}}) = \boldsymbol{\lambda}^\top \hat{\boldsymbol{\mu}}_d^{\text{EBLUP}} + \alpha, \quad d = 1, \dots, D. \tag{5.216}$$

Likewise, an estimator of $\text{MSE}(\hat{G}_d^{\text{EBLUP}}) = \text{E}\left[(\hat{G}_d^{\text{EBLUP}} - G_d)^2\right]$ is given by

$$\widehat{\text{MSE}}(\hat{G}_d^{\text{EBLUP}}) = \boldsymbol{\lambda}^\top \widehat{\text{MSE}}(\hat{\boldsymbol{\mu}}_d^{\text{EBLUP}}) \boldsymbol{\lambda}, \tag{5.217}$$

where $\widehat{\text{MSE}}(\hat{\boldsymbol{\mu}}_d^{\text{EBLUP}})$ is given by (5.206).

Thus, when the target domain parameters are given by a linear function of $\boldsymbol{\mu}_d$, the predictions $\hat{G}_d^{\text{EBLUP}}$ and MSE estimates $\widehat{\text{MSE}}(\hat{G}_d^{\text{EBLUP}})$, $d = 1, \dots, D$, can be directly calculated from the output of the MFH model.

5.5 Empirical best prediction of non-linear domain indicators

Consider again that the domain parameters of interest take the form

$$G_d = G_d(\boldsymbol{u}_d, \boldsymbol{\beta}) = g(\boldsymbol{\mu}_d), \quad d = 1, \dots, D, \tag{5.218}$$

where $g : \mathbb{R}^m \mapsto \mathbb{R}^m$ is a function. In contrast to Section 5.4, we now assume that we cannot simplify G_d any further and that it corresponds to a general, potentially non-linear, function in $\boldsymbol{u}_d$ and $\boldsymbol{\beta}$. Consequently, g is a general, potentially non-linear, function in $\boldsymbol{\mu}_d$.

The best predictor (BP) of G_d is

$$\hat{G}_d^{\text{BP}} = \text{E}_{\boldsymbol{\psi}}[G_d | \boldsymbol{y}_d] = \int_{\mathbb{R}^m} G_d(\boldsymbol{u}_d, \boldsymbol{\beta}) f_{\boldsymbol{\psi}}(\boldsymbol{u}_d | \boldsymbol{y}_d)\, d\boldsymbol{u}_d, \quad d = 1, \dots, D. \tag{5.219}$$

The EBP of G_d is obtained by substituting $\boldsymbol{\psi} = (\boldsymbol{\beta}^\top, \boldsymbol{\theta}^\top)^\top$ by a consistent estimator in (5.219), like the ML/REML estimator. The EBP of G_d is

$$\hat{G}_d^{\text{EBP}} = \text{E}_{\hat{\boldsymbol{\psi}}}[G_d | \boldsymbol{y}_d] = \int_{\mathbb{R}^m} G_d(\boldsymbol{u}_d, \hat{\boldsymbol{\beta}}) f_{\hat{\boldsymbol{\psi}}}(\boldsymbol{u}_d | \boldsymbol{y}_d)\, d\boldsymbol{u}_d, \quad d = 1, \dots, D. \tag{5.220}$$

In practice, instead of approximating the integral form of G_d, often plug-in predictors are used. We denote the best plug-in predictor (BI) as

$$\hat{G}_d^{\text{BI}} = G_d(\hat{\boldsymbol{u}}_d^{\text{BLUP}}, \boldsymbol{\beta}) = g(\hat{\boldsymbol{\mu}}_d^{\text{BLUP}}), \quad d = 1, \dots, D. \tag{5.221}$$

The empirical BI (EBI) of G_d is obtained by substituting $\hat{\boldsymbol{\mu}}_d^{\text{BLUP}}$ and $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\mu}}_d^{\text{EBLUP}}$ and consistent estimator $\hat{\boldsymbol{\beta}}$ and given by

$$\hat{G}_d^{\text{EBI}} = G_d(\hat{\boldsymbol{u}}_d^{\text{EBLUP}}, \hat{\boldsymbol{\beta}}) = g(\hat{\boldsymbol{\mu}}_d^{\text{EBLUP}}), \quad d = 1, \dots, D. \tag{5.222}$$

Note that although we use the term *best*, the quantities $\hat{G}_d^{\mathrm{BI}}$ and $\hat{G}_d^{\mathrm{EBI}}$ are not defined according to the LMM theory from Section 2.4.2, where term *best* is used to denote MSE optimality. Rather, we chose the terms to allow for easier comparison of plug-ins to BPs. From (5.221) and (5.222), we see that the plug-in predictors do not take the joint distribution of the domain random effects into account and, in general, we have $\hat{G}_d^{\mathrm{BI}} \neq \hat{G}_d^{\mathrm{BP}}$, $d = 1, \ldots, D$.

The m-dimensional integral (5.219) of the BPs is, in general, not analytically calculable. We therefore approximate it with the techniques presented in Section 5.2. To simplify the description of the integral approximations, we use the same notation for the actual BPs, i.e. with their integral form, and the approximated BPs. Formally, we denote both by $\hat{G}_d^{\mathrm{BP}}$, $d = 1, \ldots, D$. We proceed in the same way for the EBPs and approximated EBPs. Recall that in practice, we only calculate the approximated (E)BPs.

The integral $\hat{G}_d^{\mathrm{BP}}$ (5.219) can be presented in different forms. Table 5.1 shows the different variants of (5.219) which we consider in the following, namely $\boldsymbol{I_1}$, $\boldsymbol{I_2}$, and $\boldsymbol{I_3}$. For the EBP, the same formulas can be used, substituting $\boldsymbol{\psi} = (\boldsymbol{\beta}^\top, \boldsymbol{\theta}^\top)^\top$ by a consistent estimator. For the integral forms $\boldsymbol{I_1}$, $\boldsymbol{I_2}$, and $\boldsymbol{I_3}$, we can apply the approximation methods presented in Section 5.2. They are listed in Table 5.2, together with the abbreviations which we hereafter use for them. In Simulation 1, we investigate the performance of the cross-combinations of the approximation methods in Table 5.2 and integral forms in Table 5.1.

The application of the integral approximation methods of Table 5.2 is very similar for integral forms $\boldsymbol{I_1}$, $\boldsymbol{I_2}$, and $\boldsymbol{I_3}$ of Table 5.1. We therefore illustrate them only for integral form $\boldsymbol{I_1}$ in Algorithms 5.2, 5.3, 5.4, and 5.5. For the algorithms, note that for $f_{\boldsymbol{\psi}}(\boldsymbol{u}_d|\boldsymbol{y}_d)$ we have

$$\boldsymbol{\mu}_{\boldsymbol{u}_d|\boldsymbol{y}_d}(\boldsymbol{\beta}, \boldsymbol{\theta}) = \boldsymbol{V}_{ud}(\boldsymbol{\theta})\boldsymbol{V}_d^{-1}(\boldsymbol{\theta})(\boldsymbol{y}_d - \boldsymbol{X}_d\boldsymbol{\beta}), \tag{5.227}$$

$$\boldsymbol{V}_{\boldsymbol{u}_d|\boldsymbol{y}_d}(\boldsymbol{\theta}) = \boldsymbol{V}_{ud}(\boldsymbol{\theta})\boldsymbol{V}_d^{-1}(\boldsymbol{\theta})\boldsymbol{V}_{ed}, \quad d = 1, \ldots, D. \tag{5.228}$$

Table 5.1: Integral forms of $\hat{G}_d^{\mathrm{BP}}$

$\boldsymbol{I_1}$	We consider the integral form in (5.219) $\hat{G}_d^{\mathrm{BP}} = \int_{\mathbb{R}^m} G_d(\boldsymbol{u}_d, \boldsymbol{\beta}) f_{\boldsymbol{\psi}}(\boldsymbol{u}_d \mid \boldsymbol{y}_d)\, d\boldsymbol{u}_d, \quad d = 1, \ldots, D. \qquad (5.223)$
$\boldsymbol{I_2}$	For the conditional distribution it holds that $f_{\boldsymbol{\psi}}(\boldsymbol{u}_d \mid \boldsymbol{y}_d) = \frac{f_{\boldsymbol{\psi}}(\boldsymbol{u}_d, \boldsymbol{y}_d)}{f_{\boldsymbol{\psi}}(\boldsymbol{y}_d)} = \frac{f_{\boldsymbol{\psi}}(\boldsymbol{y}_d \mid \boldsymbol{u}_d) f_{\boldsymbol{\theta}}(\boldsymbol{u}_d)}{f_{\boldsymbol{\psi}}(\boldsymbol{y}_d)}, \quad d = 1, \ldots, D. \qquad (5.224)$ We can therefore rewrite the BP of G_d as $\hat{G}_d^{\mathrm{BP}} = \int_{\mathbb{R}^m} G_d(\boldsymbol{u}_d, \boldsymbol{\beta}) \frac{f_{\boldsymbol{\psi}}(\boldsymbol{y}_d \mid \boldsymbol{u}_d) f_{\boldsymbol{\theta}}(\boldsymbol{u}_d)}{f_{\boldsymbol{\psi}}(\boldsymbol{y}_d)}\, d\boldsymbol{u}_d$ $= \frac{\int_{\mathbb{R}^m} G_d(\boldsymbol{u}_d, \boldsymbol{\beta}) f_{\boldsymbol{\psi}}(\boldsymbol{y}_d \mid \boldsymbol{u}_d) f_{\boldsymbol{\theta}}(\boldsymbol{u}_d)\, d\boldsymbol{u}_d}{f_{\boldsymbol{\psi}}(\boldsymbol{y}_d)}$ $= \frac{\int_{\mathbb{R}^m} G_d(\boldsymbol{u}_d, \boldsymbol{\beta}) f_{\boldsymbol{\psi}}(\boldsymbol{y}_d \mid \boldsymbol{u}_d) f_{\boldsymbol{\theta}}(\boldsymbol{u}_d)\, d\boldsymbol{u}_d}{\int_{\mathbb{R}^m} f_{\boldsymbol{\psi}}(\boldsymbol{y}_d \mid \boldsymbol{u}_d) f_{\boldsymbol{\theta}}(\boldsymbol{u}_d)\, d\boldsymbol{u}_d}$ $= \frac{A_d(\boldsymbol{y}_d, \boldsymbol{\beta}, \boldsymbol{\theta})}{B_d(\boldsymbol{y}_d, \boldsymbol{\beta}, \boldsymbol{\theta})}, \quad d = 1, \ldots, D.$
$\boldsymbol{I_3}$	The denominator in formula (5.1) is $f_{\boldsymbol{\psi}}(\boldsymbol{y}_d)$ and we have $\boldsymbol{y}_d \sim N_m(\boldsymbol{X}_d \boldsymbol{\beta}, \boldsymbol{V}_d)$. Therefore, we have additional integral form $\hat{G}_d^{\mathrm{BP}} = \frac{\det(\boldsymbol{V}_d)^{1/2}}{\det(\boldsymbol{V}_{ed})^{1/2}} \frac{A_d(\boldsymbol{y}_d, \boldsymbol{\beta}, \boldsymbol{\theta})}{C_d(\boldsymbol{y}_d, \boldsymbol{\beta}, \boldsymbol{\theta})}, \quad d = 1, \ldots, D, \qquad (5.225)$ where $C_d(\boldsymbol{y}_d, \boldsymbol{\beta}, \boldsymbol{\theta}) = (2\pi)^{m/2} \det(\boldsymbol{V}_d)^{1/2} f_{\boldsymbol{\psi}}(\boldsymbol{y}_d) = \exp\Big(-\frac{1}{2} \boldsymbol{r}_d^\top \boldsymbol{V}_d^{-1} \boldsymbol{r}_d\Big). \qquad (5.226)$ Compared to $\boldsymbol{I_2}$, in integral form $\boldsymbol{I_3}$, we only have to approximate an integral in the nominator, not the denominator.

Table 5.2: Integral approximation techniques

Name	Approximation technique
GH	Gauss-Hermite quadrature (Section 5.2.1, Algorithm 5.2)
MC	Monte Carlo integration (Section 5.2.2, Algorithm 5.3)
MCA	Monte Carlo integration with antithetic variates (Section 5.2.2, Algorithm 5.4)
QMCH	Quasi Monte Carlo integration with Halton sequence (Section 5.2.3, Algorithm 5.5)
QMCS	Quasi Monte Carlo integration with Sobol sequence (Section 5.2.3, Algorithm 5.5)

Algorithm 5.2 Gauss-Hermite quadrature (GH) for $\boldsymbol{I_1}$

```
Choose $T \in \mathbb{N}_+$.
For $d = 1, \dots, D$ do
    1. Take decomposition $\boldsymbol{V}_{\boldsymbol{u}_d|\boldsymbol{y}_d}(\boldsymbol{\theta}) = \boldsymbol{L}_d \boldsymbol{L}_d^\top$ and
       $\boldsymbol{z}_d = 2^{-0.5} \boldsymbol{L}_d^{-1} (\boldsymbol{u}_d - \boldsymbol{\mu}_{\boldsymbol{u}_d|\boldsymbol{y}_d}) \iff \boldsymbol{u}_d = \sqrt{2} \boldsymbol{L}_d \boldsymbol{z}_d + \boldsymbol{\mu}_{\boldsymbol{u}_d|\boldsymbol{y}_d}$, such that

       $\hat{G}_d^{\mathrm{BP}} = \pi^{-m/2} \int_{\mathbb{R}^m} G_d(\sqrt{2} \boldsymbol{L}_d \boldsymbol{z}_d + \boldsymbol{\mu}_{\boldsymbol{u}_d|\boldsymbol{y}_d}, \boldsymbol{\beta}) \exp\Big( - \boldsymbol{z}_d^\top \boldsymbol{z}_d \Big) \, d\boldsymbol{z}_d.$

    2. Approximate

       $\hat{G}_d^{\mathrm{BP}} \approx \sum_{t_1=1}^{T} \cdots \sum_{t_m=1}^{T} \pi^{-m/2} G_d(\tilde{z}_{t_1}, \dots, \tilde{z}_{t_m}) \prod_{k=1}^{m} w_{t_k},$

       with $\tilde{z}_{t_1}, \dots, \tilde{z}_{t_m} = \sqrt{2} \boldsymbol{L}_d (z_{t_1}, \dots, z_{t_m})^\top + \boldsymbol{\mu}_{\boldsymbol{u}_d|\boldsymbol{y}_d}$, where $z_{t_k}$ and $w_{t_k}$ are the standard
       Gauss-Hermite nodes and weights for chosen $T$, $k = 1, \dots, m$, $t = 1, \dots, T$.
```

Algorithm 5.3 Monte Carlo integration (MC) for $\boldsymbol{I_1}$

```
Choose $T \in \mathbb{N}_+$.
For $d = 1, \dots, D$ do
    1. For $t = 1, \dots, T$ do
        a) Draw random numbers $\boldsymbol{u}_d^{(t)} \overset{\text{iid}}{\sim} N_m(\boldsymbol{\mu}_{\boldsymbol{u}_d|\boldsymbol{y}_d}(\boldsymbol{\theta}), \boldsymbol{V}_{\boldsymbol{u}_d|\boldsymbol{y}_d}(\boldsymbol{\theta}))$.
        b) Calculate $G_d^{(t)} = G_d(\boldsymbol{u}_d^{(t)}, \boldsymbol{\beta})$.
    2. Approximate $\hat{G}_d^{\mathrm{BP}} \approx T^{-1} \sum_{t=1}^{T} G_d^{(t)}$.
```

Algorithm 5.4 Monte Carlo integration with antithetic variates (MCA) for $\boldsymbol{I_1}$

```
Choose $T \in \mathbb{N}_+$.
For $d = 1, \dots, D$ do
    1. For $t = 1, \dots, T$ do
        a) Draw random numbers $\boldsymbol{u}_d^{(t)} \overset{\text{iid}}{\sim} N_m(\boldsymbol{\mu}_{\boldsymbol{u}_d|\boldsymbol{y}_d}(\boldsymbol{\theta}), \boldsymbol{V}_{\boldsymbol{u}_d|\boldsymbol{y}_d}(\boldsymbol{\theta}))$.
        b) Set antithetic variates $\boldsymbol{u}_d^{(T+t)} = 2\boldsymbol{\mu}_{\boldsymbol{u}_d|\boldsymbol{y}_d}(\boldsymbol{\beta}, \boldsymbol{\theta}) - \boldsymbol{u}_d^{(t)}$.
        c) Calculate $G_d^{(t)} = G_d(\boldsymbol{u}_d^{(t)}, \boldsymbol{\beta})$, $G_d^{(T+t)} = G_d(\boldsymbol{u}_d^{(T+t)}, \boldsymbol{\beta})$.
    2. Approximate $\hat{G}_d^{\mathrm{BP}} \approx 2T^{-1} \sum_{t=1}^{2T} G_d^{(t)}$.
```

Algorithm 5.5 Quasi Monte Carlo integration (QMC) for $\boldsymbol{I_1}$

```
Choose $T \in \mathbb{N}_+$.
For $d = 1, \dots, D$ do
  1. For $t = 1, \dots, T$ do
     a) Take the values $\boldsymbol{u}_d^{(t)}$ of the Halton/Sobol sequence which correspond to
        $N_m(\boldsymbol{\mu}_{\boldsymbol{u}_d|\boldsymbol{y}_d}(\boldsymbol{\theta}), \boldsymbol{V}_{\boldsymbol{u}_d|\boldsymbol{y}_d}(\boldsymbol{\theta}))$.
        For that, we use R (R Core Team, 2020) functions rnorm.sobol and
        rnorm.halton from the fOptions package (Wuertz et al., 2017). The sequences
        are adjusted to a specific $m$-variate normal distribution using Algorithm 5.1.
     b) Calculate $G_d^{(t)} = G_d(\boldsymbol{u}_d^{(t)}, \boldsymbol{\beta})$.
  2. Approximate $\hat{G}_d^{\mathrm{BP}} \approx T^{-1} \sum_{t=1}^{T} G_d^{(t)}$.
```

5.6 Mean squared error of non-linear domain indicators

5.6.1 Parametric bootstrap for MSE

The MSE of the approximations of $\hat{G}_d^{\mathrm{EBP}}$ cannot be given analytically for general functional forms of G_d. Instead, we propose a *parametric bootstrap estimator* for estimating the MSE of $\hat{G}_d^{\mathrm{EBP}}$, $\mathrm{MSE}(\hat{G}_d^{\mathrm{EBP}}) = \mathrm{E}\left[(\hat{G}_d^{\mathrm{EBP}} - G_d)^2\right]$. For estimating the MSE, we calculate the ML/REML estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$ by using the observed data $(\boldsymbol{y}_d, \boldsymbol{X}_d)$, $d = 1, \dots, D$, and run the following Algorithm 5.6.

Algorithm 5.6 Parametric bootstrap for estimating the MSE of $\hat{G}_d^{\mathrm{EBP}}$

```
1. For $b = 1, \dots, B$, $B \in \mathbb{N}_+$, do
   a) For $d = 1, \dots, D$ do
      i. Generate $\boldsymbol{u}_d^{*(b)} \sim N_m(0, \hat{\boldsymbol{V}}_u)$, $\boldsymbol{e}_d^{*(b)} \sim N_m(0, \boldsymbol{V}_{ed})$.
      ii. Calculate
          $$\begin{aligned} \boldsymbol{y}_d^{*(b)} &= \boldsymbol{\mu}_d^{*(b)} + \boldsymbol{e}_d^{*(b)}, \quad \boldsymbol{\mu}_d^{*(b)} = \boldsymbol{X}_d\hat{\boldsymbol{\beta}} + \boldsymbol{u}_d^{*(b)}, \\ G_d^{*(b)} &= g(\boldsymbol{\mu}_{d1}^{*(b)}, \dots, \boldsymbol{\mu}_{dm}^{*(b)}). \end{aligned} \qquad (5.229)$$
   b) By using the bootstrap data $(\boldsymbol{y}_d^{*(b)}, \boldsymbol{X}_d)$, $d = 1, \dots, D$, calculate
      i. ML/REML estimates $\hat{\boldsymbol{\beta}}^{*(b)}$ and $\hat{\boldsymbol{\theta}}^{*(b)}$,
      ii. EBPs $\hat{G}_d^{\mathrm{EBP}\,*(b)}$. For the approximation, use one of the techniques of Table
          5.2 with one of the integral forms of Table 5.1 under the MFH model
          (5.229).
2. For $d = 1, \dots, D$, calculate $\widehat{\mathrm{MSE}}_{1d}^{*} = B^{-1} \sum_{b=1}^{B} \left(\hat{G}_d^{\mathrm{EBP}\,*(b)} - G_d^{*(b)}\right)^2$.
```

5.6.2 Parametric bootstrap for a component of the MSE

We note that the MSE of $\hat{G}_d^{\mathrm{EBP}}$ can be divided into two parts. The MSE contains integrals with respect to the joint probability density function $f(\boldsymbol{y}_d, \boldsymbol{u}_d) = f(\boldsymbol{u}_d|\boldsymbol{y}_d)f(\boldsymbol{y}_d)$. Note that we have

$$\begin{aligned} G_d &= G_d(\boldsymbol{u}_d, \boldsymbol{\beta}), && (5.230)\\ \hat{G}_d^{\mathrm{BP}} &= \hat{G}_d^{\mathrm{BP}}(\boldsymbol{y}_d, \boldsymbol{\beta}, \boldsymbol{\theta}), && (5.231)\\ \hat{G}_d^{\mathrm{EBP}} &= \hat{G}_d^{\mathrm{BP}}(\boldsymbol{y}_d, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) = \hat{G}_d^{\mathrm{EBP}}(\boldsymbol{y}_d), \quad d = 1, \dots, D. && (5.232) \end{aligned}$$

We can rewrite the MSE of $\hat{G}_d^{\mathrm{EBP}}$ as

$$\begin{aligned} \mathrm{E}\Big[(\hat{G}_d^{\mathrm{EBP}} - G_d)^2\Big] &= \mathrm{E}\Big[\Big((\hat{G}_d^{\mathrm{EBP}} - \hat{G}_d^{\mathrm{BP}}) + (\hat{G}_d^{\mathrm{BP}} - G_d)\Big)^2\Big]\\ &= \underbrace{\mathrm{E}\Big[\Big(\hat{G}_d^{\mathrm{EBP}} - \hat{G}_d^{\mathrm{BP}}\Big)^2\Big]}_{=S_{1d}} + \underbrace{\mathrm{E}\Big[\Big(\hat{G}_d^{\mathrm{BP}} - G_d\Big)^2\Big]}_{=S_{2d}}, \quad d = 1, \dots, D. \end{aligned} \tag{5.233}$$

As $\mathrm{E}\Big[G_d|\boldsymbol{y}_d\Big] = \hat{G}_d^{\mathrm{BP}}$, in the above formulas the cross-moment term is

$$\mathrm{E}\Big[(\hat{G}_d^{\mathrm{EBP}} - \hat{G}_d^{\mathrm{BP}}) + (\hat{G}_d^{\mathrm{BP}} - G_d)\Big] = \mathrm{E}_{\boldsymbol{y}_d}\Big[\Big(\hat{G}_d^{\mathrm{EBP}} - \hat{G}_d^{\mathrm{BP}}\Big)\,\mathrm{E}\Big(\hat{G}_d^{\mathrm{BP}} - G_d|\boldsymbol{y}_d\Big)\Big] = 0. \tag{5.234}$$

As an alternative to Algorithm 5.6, we propose an independent estimation of the terms S_{1d} and S_{2d} in (5.233). We calculate the ML/REML estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$ by using the observed data $(\boldsymbol{y}_d, \boldsymbol{X}_d)$, $d = 1, \dots, D$, and run Algorithm 5.7 for the estimation of S_{1d} in (5.233).

Algorithm 5.7 Parametric bootstrap for estimating S_{1d} in (5.233)

```
1. For $b = 1, \dots, B$, $B \in \mathbb{N}_+$, do
    a) For $d = 1, \dots, D$ do
        i. Generate $\boldsymbol{u}_d^{*(b)} \sim N_m(0, \hat{\boldsymbol{V}}_u)$, $\boldsymbol{e}_d^{*(b)} \sim N_m(0, \boldsymbol{V}_{ed})$.
        ii. Calculate
            $\boldsymbol{y}_d^{*(b)} = \boldsymbol{\mu}_d^{*(b)} + \boldsymbol{e}_d^{*(b)}, \quad \boldsymbol{\mu}_d^{*(b)} = \boldsymbol{X}_d\hat{\boldsymbol{\beta}} + \boldsymbol{u}_d^{*(b)}.$   (5.235)
    b) By using the bootstrap data $(\boldsymbol{y}_d^{*(b)}, \boldsymbol{X}_d)$, $d = 1, \dots, D$, calculate
        i. ML/REML estimates $\hat{\boldsymbol{\beta}}^{*(b)}$ and $\hat{\boldsymbol{\theta}}^{*(b)}$,
        ii. BPs $\hat{G}_d^{\mathrm{BP}\,*(b)}$ and EBPs $\hat{G}_d^{\mathrm{EBP}\,*(b)}$. For the approximation, use one of the
            techniques of Table 5.2 with one of the integral forms of Table 5.1 under
            model (5.235).
2. For $d = 1, \dots, D$ calculate $\hat{S}_{1d}^{*} = B^{-1} \sum_{b=1}^{B} \Big(\hat{G}_d^{\mathrm{EBP}\,*(b)} - \hat{G}_d^{\mathrm{BP}\,*(b)}\Big)^2$.
```

The second summand in (5.233) is

$$S_{2d} = \mathrm{E}\left[\left(\hat{G}_d^{\mathrm{BP}} - G_d\right)^2\right] = \mathrm{E}_{\boldsymbol{y}_d}\left[\mathrm{E}\left(\left(G_d - \hat{G}_d^{\mathrm{BP}}\right)^2 |\boldsymbol{y}_d\right)\right] = \mathrm{E}_{\boldsymbol{y}_d}\left[\mathrm{Var}\left(G_d|\boldsymbol{y}_d\right)\right]. \tag{5.236}$$

In integral form, we have

$$S_{2d} = \int_{\mathbb{R}^m} \underbrace{\left(\int_{\mathbb{R}^m}\left(\hat{G}_d^{\mathrm{BP}} - G_d\right)^2 f(\boldsymbol{u}_d|\boldsymbol{y}_d)\,d\boldsymbol{u}_d\right)}_{=\,S_{2d}^{\mathrm{inner}}} f(\boldsymbol{y}_d)\,d\boldsymbol{y}_d. \tag{5.237}$$

We calculate the ML/REML estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$ by using the observed data $(\boldsymbol{y}_d, \boldsymbol{X}_d)$, $d = 1, \ldots, D$, and run the following Algorithm 5.8 to approximate S_{2d}. In Algorithm 5.8, we use a nested GH technique to approximate the nested integral of (5.237). Alternatively, also (Quasi) MC integration could be used to approximate S_{2d}.

Finally, an estimator of the MSE of $\hat{G}_d^{\mathrm{EBP}}$ is

$$\widehat{\mathrm{MSE}}_{2d}^{*} = \hat{S}_{1d}^{*} + \hat{S}_{2d}. \tag{5.238}$$

For higher-dimensional integrals, Algorithm 5.8, might get too demanding because of the double integral. For the estimation of S_{2d}, we therefore propose an additional estimator which only approximates the inner integral in (5.237). To put it differently, we avoid approximating the double integral by estimating an expected value (expectation with respect to $f(\boldsymbol{y}_d)$) with observed values. The inner integral is

$$S_{2d}^{\mathrm{inner}} = \mathrm{Var}\left(G_d|\boldsymbol{y}_d\right) = \int_{\mathbb{R}^m}\left(G_d - \hat{G}_d^{\mathrm{BP}}\right)^2 f(\boldsymbol{u}_d|\boldsymbol{y}_d)\,d\boldsymbol{u}_d, \quad \hat{G}_d^{\mathrm{BP}} = \mathrm{E}\left[G_d|\boldsymbol{y}_d\right]. \tag{5.239}$$

We use the procedure for calculating approximation $\hat{S}_{2d}^{\mathrm{inner}}$ in Algorithm 5.8 with observed $\boldsymbol{y}_d$ and set $\hat{S}_{2d}^{\mathrm{inner}} = \hat{S}_{2d}^{\mathrm{inner}}(\boldsymbol{y}_d)$. To put it differently, we only apply the inner GH approximation of Algorithm 5.8 with observed $\boldsymbol{y}_d$. Finally, an estimator of the MSE of $\hat{G}_d^{\mathrm{EBP}}$ is

$$\widehat{\mathrm{MSE}}_{3d}^{*} = \hat{S}_{1d}^{*} + \hat{S}_{2d}^{\mathrm{inner}}. \tag{5.240}$$

Algorithm 5.8 Nested GH quadrature for estimating S_{2d} in (5.233)

For $d = 1, \dots, D$ do

1. Take decomposition $\hat{\boldsymbol{V}}_d(\hat{\boldsymbol{\theta}}) = \boldsymbol{P}_d \boldsymbol{P}_d^\top$ and $\boldsymbol{z}_d = 2^{-0.5}\boldsymbol{P}_d^{-1}(\boldsymbol{y}_d - \boldsymbol{X}_d\hat{\boldsymbol{\beta}}) \iff \boldsymbol{y}_d = \sqrt{2}\boldsymbol{P}_d\boldsymbol{z}_d + \boldsymbol{X}_d\hat{\boldsymbol{\beta}}$, such that

$$\begin{aligned} S_{2d} &= \int_{\mathbb{R}^m} \underbrace{\left(\int_{\mathbb{R}^m}\left(\hat{G}_d^{\mathrm{BP}} - G_d\right)^2 f(\boldsymbol{u}_d|\boldsymbol{y}_d)\,d\boldsymbol{u}_d\right)}_{=\,S_{2d}^{\mathrm{inner}}} f(\boldsymbol{y}_d)\,d\boldsymbol{y}_d \\ &= \pi^{-m/2}\int_{\mathbb{R}^m} S_{2d}^{\mathrm{inner}}(\sqrt{2}\boldsymbol{P}_d\boldsymbol{z}_d + \boldsymbol{X}_d\hat{\boldsymbol{\beta}})\exp\Big(-\boldsymbol{z}_d^\top\boldsymbol{z}_d\Big)\,d\boldsymbol{z}_d \end{aligned}$$

2. Calculate approximation

$$\hat{S}_{2d} = \sum_{t_1=1}^{T}\cdots\sum_{t_m=1}^{T}\pi^{-m/2}\hat{S}_{2d}^{\mathrm{inner}}(\tilde{z}_{t_1},\dots,\tilde{z}_{t_m})\prod_{k=1}^{m} w_{t_k},$$

with $\tilde{z}_{t_1},\dots,\tilde{z}_{t_m} = \sqrt{2}\boldsymbol{P}_d(z_{t_1},\dots,z_{t_m})^\top + \boldsymbol{X}_d\hat{\boldsymbol{\beta}}$, where z_{t_k} and w_{t_k} and are the standard GH nodes and weights for chosen $T \in \mathbb{N}_+$, $k = 1,\dots,m$, $t = 1,\dots,T$. $\hat{S}_{2d}^{\mathrm{inner}} = \hat{S}_{2d}^{\mathrm{inner}}(\boldsymbol{y}_d)$ is approximated as follows, based on (5.227) and (5.228).

   a) Take decomposition $\hat{\boldsymbol{V}}_{\boldsymbol{u}_d|\boldsymbol{y}_d}(\hat{\boldsymbol{\theta}}) = \boldsymbol{L}_d\boldsymbol{L}_d^\top$ and $\boldsymbol{c}_d = 2^{-0.5}\boldsymbol{L}_d^{-1}(\boldsymbol{u}_d - \boldsymbol{\mu}_{\boldsymbol{u}_d|\boldsymbol{y}_d}) \iff \boldsymbol{u}_d = \sqrt{2}\boldsymbol{L}_d\boldsymbol{c}_d + \boldsymbol{\mu}_{\boldsymbol{u}_d|\boldsymbol{y}_d}$, such that

$$\hat{G}_d^{\mathrm{BP}} = \pi^{-m/2}\int_{\mathbb{R}^m} G_d(\sqrt{2}\boldsymbol{L}_d\boldsymbol{c}_d + \boldsymbol{\mu}_{\boldsymbol{u}_d|\boldsymbol{y}_d}, \hat{\boldsymbol{\beta}})\exp\Big(-\boldsymbol{c}_d^\top\boldsymbol{c}_d\Big)\,d\boldsymbol{c}_d.$$

   b) Calculate approximation

$$\hat{S}_{2d}^{\mathrm{inner}} = \sum_{r_1=1}^{R}\cdots\sum_{r_m=1}^{R}\pi^{-m/2}(G_d(\tilde{c}_{r_1},\dots,\tilde{c}_{r_m}) - \hat{G}_d^{\mathrm{EBP}})^2\prod_{k=1}^{m} v_{r_k}$$

   with $\tilde{c}_{r_1},\dots,\tilde{c}_{r_m} = \sqrt{2}\boldsymbol{L}_d(c_{r_1},\dots,c_{r_m})^\top + \boldsymbol{\mu}_{\boldsymbol{u}_d|\boldsymbol{y}_d}$, where c_{r_k} and v_{r_k} and are the standard GH nodes and weights for chosen $R \in \mathbb{N}_+$, $k = 1,\dots,m$, $r = 1,\dots,R$. The EBPs $\hat{G}_d^{\mathrm{EBP}}$ in $\hat{S}_{2d}^{\mathrm{inner}}$ are the EBP predictions from the MFH model, calculated using one of the techniques of Table 5.2 with one of the integral forms of Table 5.1.

5.7 Simulation

We proposed different approximations of $\hat{G}_d^{\text{BP}}$ and $\hat{G}_d^{\text{EBP}}$ and the MSE of $\hat{G}_d^{\text{EBP}}$. In the following, we evaluate the different approaches in simulation studies and compare the performance of the proposed approximations of $\hat{G}_d^{\text{BP}}$ and $\hat{G}_d^{\text{EBP}}$ to the performance of the corresponding plug-in predictors $\hat{G}_d^{\text{BI}}$ and $\hat{G}_d^{\text{EBI}}$.

The parameter settings of the simulation studies are inspired by the illustrative application which we present in Section 5.8. We consider a bivariate FH (BFH) model, i.e. a MFH model with $m = 2$ dependent variables, and take the percentages of employed and unemployed as the dependent variables in the model. Our domain parameter of interest is the unemployment rate which is given by the percentage of unemployed divided by the sum of the percentages of employed and unemployed. The unemployment rate is an example of a multi-variable non-linear domain indicator. Formally, under the BFH model, the domain-specific unemployment rates are given by

$$R_d = R_d(\boldsymbol{u}_d, \boldsymbol{\beta}) = \frac{\mu_{d1}}{\mu_{d1} + \mu_{d2}} = \frac{\boldsymbol{x}_{d1}^\top \boldsymbol{\beta}_1 + u_{d1}}{\boldsymbol{x}_{d1}^\top \boldsymbol{\beta}_1 + u_{d1} + \boldsymbol{x}_{d2}^\top \boldsymbol{\beta}_2 + u_{d2}}, \quad d = 1, \dots, D, \tag{5.241}$$

where μ_{d1} and μ_{d2} are the percentages of unemployed and employed respectively.

5.7.1 Simulation setup

We use the following simulation setup for all of the following simulation studies 1, 2, 3, and 4. Consider the particular bivariate Fay-Herriot (BFH) model

$$\boldsymbol{y}_d = \boldsymbol{X}_d \boldsymbol{\beta} + \boldsymbol{u}_d + \boldsymbol{e}_d, \quad d = 1, \dots, D, \tag{5.242}$$

the parameters of which we set in the following.

We consider one auxiliary variable plus intercept for both dependent variables in the model. Therefore, $p_1 = p_2 = 2$, $p = p_1 + p_2 = 4$. Set $\boldsymbol{\beta}_1 = (\boldsymbol{\beta}_{11}, \boldsymbol{\beta}_{12})^\top = (0, 0.2)^\top$, $\boldsymbol{\beta}_2 = (\boldsymbol{\beta}_{21}, \boldsymbol{\beta}_{22})^\top = (0.6, 0.2)^\top$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$.

The auxiliary information is generated once and remains fixed throughout the iterations of the simulations. For $d = 1, \dots, D$, we take $\boldsymbol{X}_d = \text{diag}(\boldsymbol{x}_{d1}^\top, \boldsymbol{x}_{d2}^\top)_{2\times 4}$, $\boldsymbol{x}_{d1} = (x_{d11}, x_{d12})^\top$, $\boldsymbol{x}_{d2} = (x_{d21}, x_{d22})^\top$, $x_{d11} = x_{d21} = 1$. Generate $x_{d12} = U_{d1}$, $U_{d1} \overset{\text{ind}}{\sim} Unif(0.49, 0.51)$, $x_{d22} = U_{d2}$, $U_{d2} \overset{\text{ind}}{\sim} Unif(0.49, 0.51)$, where $Unif$ is the uniform distribution.

The random effects $\boldsymbol{u}_d$ and the sampling errors $\boldsymbol{e}_d$, $d = 1, \dots, D$, of the model are generated in each iteration of the simulations. We take $\boldsymbol{u}_d \sim N_2(0, \boldsymbol{V}_{ud})$ and $\boldsymbol{e}_d \sim N_2(0, \boldsymbol{V}_{ed})$ with

$$\boldsymbol{V}_{ud} = \begin{pmatrix} \sigma_{u1}^2 & \rho_u \sigma_{u1} \sigma_{u2} \\ \rho_u \sigma_{u1} \sigma_{u2} & \sigma_{u2}^2 \end{pmatrix}, \quad \boldsymbol{V}_{ed} = \begin{pmatrix} \sigma_{e1}^2 & \rho_e \sigma_{e1} \sigma_{e2} \\ \rho_e \sigma_{e1} \sigma_{e2} & \sigma_{e2}^2 \end{pmatrix}, \quad d = 1, \dots, D, \tag{5.243}$$

where $\sigma_{u1} = \sigma_{u2} = 0.012$, $\sigma_{e1} = \sigma_{e2} = 0.014$. The variance components are given by vector $\boldsymbol{\theta}$ of length 3 with $\theta_1 = \sigma_{u1}^2$, $\theta_2 = \sigma_{u2}^2$, and $\theta_3 = \rho_{12} = \rho_u$.

Correlations are crucial when working with multivariate models. In the simulations, we therefore consider different values of random effects correlation ρ_u and sampling error correlation ρ_e. For the proportions of unemployed and employed, we would expect both random effects and sampling errors to be negatively correlated. Therefore, especially scenario $\rho_e = \rho_u = -0.5$ is of interest.

The above parameter configurations generate y_{1d} and y_{2d} as the proportions of unemployed and employed people in domain d. The proportions of unemployed, employed, and inactive people resulting from the simulation setting are around 0.1, 0.7 and 0.2, respectively. This implies unemployment rates around 0.125 (12.5%).

5.7.2 Simulation 1: Integral approximations of $\hat{R}_d^{\text{BP}}$

Research question
We proposed different approximations of $\hat{G}_d^{\text{BP}}$ in Section 5.5. In this simulation, we want to see which integral approximation, given by the cross-combinations of the integral form (Table 5.1) and approximation technique (Table 5.2), works best for the unemployment rate R_d. The best working combination is then used for the follow-up simulations.

To have a fair comparison between GH, MC, MCA, QMCH, and QMCS, we compare their performance for similar numbers of function evaluations e. For GH, we can only have squared numbers of function evaluations because of the product formula. For MCA, we can only have even numbers of function evaluations because of the antithetic variates. To give an example, with $e = 100$ function evaluations, we calculate the following. GH is calculated with 10 nodes and weights for each of the two dimensions. MC, QMCH, and QMCS are calculated with 100 (quasi-)random numbers. MCA is calculated with 50 random numbers and the corresponding 50 antithetic variates.

Simulation settings
The focus of this simulation is purely on the approximation of R_d as defined in (5.241). We therefore choose $D = 1$ to focus on a single domain and use the true values $\boldsymbol{\psi} = (\boldsymbol{\beta}^\top, \boldsymbol{\theta}^\top)^\top$ in the predictions. We conduct the simulation via Algorithm 5.9. The only randomness considered in this simulation is from generating random effects $\boldsymbol{u}_d$ and sampling errors $\boldsymbol{e}_d$ in each iteration. We simulate the data with random effects correlation $\rho_u = -0.5$ and sampling error correlation $\rho_e \in \{-0.5, 0, 0.25\}$.

Results
We evaluate the results for scenario $\rho_u = \rho_e = -0.5$. In the Appendix A.1, the results are displayed for $\rho_u = -0.5$ and sampling error correlation $\rho_e \in \{0, 0.25\}$. The conclusions are quite similar to the conclusions under $\rho_e = -0.5$.

For the plug-in predictors $\hat{R}_d^{\text{FH,BI}}$ and $\hat{R}_d^{\text{BI}}$ the ARBias (in %) is 0.03 and 0.03 and the RRMSE (in %) is 8.83 and 8.83 respectively. With Remark 5.2, we see why $\hat{R}_d^{\text{FH,BI}}$ and

Algorithm 5.9 Steps of Simulation 1

```
The steps of Simulation 1 are
  1. Generate $\boldsymbol{x}_{dk}$, $d = 1$, $k = 1, 2$.
  2. For $i = 1, \dots, I$, $I = 1,000$, do
      a) Generate $\boldsymbol{u}_d^{(i)} \sim N_2(\mathbf{0}, \boldsymbol{V}_{ud})$, $\boldsymbol{e}_d^{(i)} \sim N_2(\mathbf{0}, \boldsymbol{V}_{ed})$.
      b) For $d = 1, \dots, D$, calculate
           $\boldsymbol{\mu}_d^{(i)} = \boldsymbol{X}_d\boldsymbol{\beta} + \boldsymbol{u}_d^{(i)}, \quad R_d^{(i)} = \frac{\boldsymbol{\mu}_{d1}^{(i)}}{\boldsymbol{\mu}_{d1}^{(i)} + \boldsymbol{\mu}_{d2}^{(i)}}, \quad \boldsymbol{y}_d^{(i)} = \boldsymbol{X}_d\boldsymbol{\beta} + \boldsymbol{u}_d^{(i)} + \boldsymbol{e}_d^{(i)}.$
      c) With two marginal FH models, known $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$, calculate, $k = 1, 2$,
           $\hat{u}_{dk}^{\text{FH,BP}(i)} = \frac{\sigma_{uk}^2}{\sigma_{uk}^2 + \sigma_{ekd}^2}\Big(y_{dk}^{(i)} - \boldsymbol{x}_{dk}\boldsymbol{\beta}_k\Big), \ \hat{\mu}_{dk}^{\text{FH,BP}(i)} = \boldsymbol{x}_{dk}\boldsymbol{\beta}_k + \hat{u}_{dk}^{\text{FH,BP}(i)},$
           $\hat{R}_d^{\text{FH,BI}(i)} = \frac{\hat{\mu}_{d1}^{\text{FH,BP}(i)}}{\hat{\mu}_{d1}^{\text{FH,BP}(i)} + \hat{\mu}_{d2}^{\text{FH,BP}(i)}}.$
      d) With the BFH model, known $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$, calculate
           $\hat{\boldsymbol{u}}_d^{\text{BP}(i)} = \boldsymbol{V}_{ud}\boldsymbol{V}_d^{-1}\Big(\boldsymbol{y}_d^{(i)} - \boldsymbol{X}_d\boldsymbol{\beta}\Big), \quad \hat{\boldsymbol{\mu}}_d^{\text{BP}(i)} = \boldsymbol{X}_d\boldsymbol{\beta} + \hat{\boldsymbol{u}}_d^{\text{BP}(i)},$
           $\hat{R}_d^{\text{BI}(i)} = \frac{\hat{\mu}_{d1}^{\text{BP}(i)}}{\hat{\mu}_{d1}^{\text{BP}(i)} + \hat{\mu}_{d2}^{\text{BP}(i)}}.$
      e) Approximate $\hat{R}_d^{\text{BP}(i)}$ by the integral forms of Table 5.1 and the approximation
         techniques given in Table 5.2.
  3. For $d = 1, \dots, D$, $\hat{R}_d^{(i)} \in \{\hat{R}_d^{\text{FH,BI}(i)}, \hat{R}_d^{\text{BI}(i)}, \hat{R}_d^{\text{BP}(i)}\}$, calculate
       $\text{ARBias}_d = 100\frac{|I^{-1}\sum_{i=1}^I(\hat{R}_d^{(i)} - R_d^{(i)})|}{|I^{-1}\sum_{i=1}^I R_d^{(i)}|}, \ \text{RRMSE}_d = 100\frac{\Big(I^{-1}\sum_{i=1}^I(\hat{R}_d^{(i)} - R_d^{(i)})^2\Big)^{1/2}}{|I^{-1}\sum_{i=1}^I R_d^{(i)}|}.$
```

$\hat{R}_d^{\mathrm{BI}}$ give the same predictions in this simulation scenario and therefore also the same values of ARBias and RRMSE.

The ARBias (in %) and RRMSE (in %) of the different approximation techniques and integral forms are given in Tables 5.3 and 5.4. Note the case of $e = 24/25$ function evaluations. In this case, we use 25 function evaluations for GH (product formula) and 24 function evaluations for the other approximation techniques.

For $40,000$ function evaluations, the different approximation techniques and integral forms give very similar results. Comparing the different integral forms, the approximations work best for $\boldsymbol{I}_1$, both in terms of ARBias and RRMSE. Comparing the approximation techniques, GH gives the best and most stable results for different integral forms and a small number of function evaluations. Comparing MC and MCA, the use of the antithetic variates reduces the variance of the Monte Carlo integration for most considered cases. Comparing QMCH and QMCS, the performance difference between the Halton and Sobol sequences are rather small. Comparing the integration using quasi-random numbers in QMCH and QMCS with random numbers in MC and MCA, the RRMSE under quasi-random numbers is often smaller, see e.g. the RRMSE for $\boldsymbol{I}_3$ and small e.

Comparing the plug-in predictors with GH under $\boldsymbol{I}_1$, the ARBias of GH is lower, already for $e = 16$ function evaluations, and we see no difference in the RRMSE. This is consistent with the theory as the BPs are derived to be best (minimal MSE) in the class of model-unbiased predictors. The plug-in predictors are not model-unbiased, but theoretically can exhibit a smaller RRMSE. In the following simulations, we use GH with $\boldsymbol{I}_1$ and $e = 25$ function evaluations for approximating $\hat{R}_d^{\mathrm{BP}}$ and $\hat{R}_d^{\mathrm{EBP}}$. Especially for the MSE approximation by bootstrap, it is advantageous to have only few function evaluations.

Table 5.3: ARBias (in %), $\rho_e = -0.5$

		Number of function evaluations e						
Int.	Approx.	16	24/25	100	400	2,500	10,000	40,000
$\boldsymbol{I_1}$	MC	0.01	0.03	0.02	0.04	0.02	0.02	0.01
	MCA	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	GH	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	QMCH	0.51	0.49	0.20	0.07	0.03	0.02	0.01
	QMCS	0.65	0.10	0.06	0.06	0.02	0.01	0.01
$\boldsymbol{I_2}$	MC	0.03	0.10	0.02	0.02	0.01	0.01	0.02
	MCA	0.10	0.10	0.08	0.01	0.01	0.01	0.01
	GH	0.02	0.01	0.01	0.01	0.01	0.01	0.01
	QMCH	0.40	0.40	0.16	0.06	0.02	0.02	0.01
	QMCS	0.44	0.13	0.05	0.04	0.02	0.01	0.01
$\boldsymbol{I_3}$	MC	0.91	1.12	0.28	0.14	0.08	0.04	0.04
	MCA	0.50	0.95	0.29	0.39	0.04	0.00	0.01
	GH	0.03	0.00	0.01	0.01	0.01	0.01	0.01
	QMCH	1.21	1.08	0.36	0.09	0.03	0.01	0.01
	QMCS	1.26	0.08	0.33	0.13	0.03	0.01	0.01

Table 5.4: RRMSE (in %), $\rho_e = -0.5$

		Number of function evaluations e						
Int.	Approx.	16	24/25	100	400	2,500	10,000	40,000
$\boldsymbol{I_1}$	MC	9.16	9.04	8.88	8.84	8.84	8.83	8.83
	MCA	8.83	8.83	8.83	8.83	8.83	8.83	8.83
	GH	8.83	8.83	8.83	8.83	8.83	8.83	8.83
	QMCH	8.85	8.85	8.84	8.83	8.83	8.83	8.83
	QMCS	8.86	8.83	8.83	8.83	8.83	8.83	8.83
$\boldsymbol{I_2}$	MC	9.35	9.18	8.99	8.90	8.83	8.84	8.84
	MCA	9.39	9.17	8.95	8.82	8.84	8.84	8.84
	GH	8.85	8.83	8.83	8.83	8.83	8.83	8.83
	QMCH	8.95	9.00	8.86	8.84	8.84	8.83	8.83
	QMCS	9.02	8.94	8.89	8.85	8.83	8.83	8.83
$\boldsymbol{I_3}$	MC	32.93	27.00	15.51	10.92	9.14	8.87	8.84
	MCA	38.93	29.71	13.98	10.99	9.16	8.89	8.86
	GH	8.88	8.83	8.83	8.83	8.83	8.83	8.83
	QMCH	20.37	17.07	10.72	9.07	8.87	8.84	8.84
	QMCS	19.48	14.16	10.75	9.12	8.85	8.84	8.84

5.7.3 Simulation 2: EBPs and plug-in predictors

Research question
Simulation 2 evaluates the performance of the plug-in predictors from two marginal FH models and the BFH model against the performance of the proposed BFH EBP approximations of the unemployment rates R_d. For approximating the EBPs of the unemployment rates, we use GH quadrature with a total of 25 function evaluations and integral form $\boldsymbol{I}_1$ because of its good performance in Simulation 1. The different predictors and approximations are compared for different combinations of sampling error and random effect correlations ρ_e and ρ_u.

Simulation settings
We take $D = 200$ domains, correlations $\rho_e \in \{-0.5, 0, 0.25\}$, $\rho_u \in \{-0.75$, -0.5, -0.25, 0, 0.25, 0.5, $0.75\}$, and conduct the simulation via Algorithm 5.10.

Results
Tables 5.5, 5.6, and 5.7 present the simulation results, i.e. the RBIAS, ARBias, and RRMSE of the different predictions of the unemployment rates for different combinations of sampling error and random effect correlation.

Let us first compare the plug-in predictors based on the FH and BFH model. The plug-ins of the BFH model give better results than the plug-ins of the FH models, for most scenarios, especially in terms of RRMSE. There are two reasons for that. First, the BFH model takes into account the joint distribution of the two dependent variables which, in most applications, leads to more efficient estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. Second, it gives more efficient predictions of $\boldsymbol{\mu}_d$, $d = 1, \ldots, D$. For that, consider also the additional material to Simulation 1 in Appendix A.1, where the true parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ were used. There, with $\rho_u = -0.5$ and $\rho_e \in \{0, 0.25\}$, the BFH plug-in has lower RRMSE than the FH plug-in.

We compare the BFH plug-in and the approximation of the EBP of R_d by GH quadrature in Tables 5.5, 5.6, and 5.7. Approximating the EBP results in lower RBias and ARBias and very similar RRMSE compared to the BFH plug-in, for most scenarios. Because of the lower bias, approximating R_d^{EBP} is to be preferred over the BFH plug-in predictor. The approximation also needs only $e = 25$ function evaluations and is therefore computationally not much more complex than the plug-in predictor.

Comparing the different correlation scenarios, mainly we see differences between the FH and BFH plug-in for varying correlations. When sampling error and random effect correlation are of high magnitude and opposite sign, the gains of using the BFH instead of the FH plug-in are highest.

Algorithm 5.10 Steps of Simulation 2

```
The steps of Simulation 2 are
  1. Generate $\boldsymbol{x}_{dk}$, $d = 1$, $k = 1, 2$.
  2. For $i = 1, \dots, I$, $I = 2,000$, do
     a) Generate $\boldsymbol{u}_d^{(i)} \sim N_2(\boldsymbol{0}, \boldsymbol{V}_{ud})$, $\boldsymbol{e}_d^{(i)} \sim N_2(\boldsymbol{0}, \boldsymbol{V}_{ed})$.
     b) For $d = 1, \dots, D$, calculate
          $\boldsymbol{\mu}_d^{(i)} = \boldsymbol{X}_d \boldsymbol{\beta} + \boldsymbol{u}_d^{(i)}, \quad R_d^{(i)} = \frac{\mu_{d1}^{(i)}}{\mu_{d1}^{(i)} + \mu_{d2}^{(i)}}, \quad \boldsymbol{y}_d^{(i)} = \boldsymbol{X}_d \boldsymbol{\beta} + \boldsymbol{u}_d^{(i)} + \boldsymbol{e}_d^{(i)}.$
     c) With two marginal FH models, calculate FH-REML estimators $\hat{\sigma}_{uk}^2$ and $\hat{\boldsymbol{\beta}}$,
        $k = 1, 2$, and
          $\hat{u}_{dk}^{\text{FH,EBLUP}(i)} = \frac{\hat{\sigma}_{uk}^2}{\hat{\sigma}_{uk}^2 + \sigma_{ekd}^2} \Big( y_{dk}^{(i)} - x_{dk} \hat{\boldsymbol{\beta}}_k(\hat{\sigma}_{uk}^2) \Big),$
          $\hat{\mu}_{dk}^{\text{FH,EBLUP}(i)} = x_{dk} \hat{\boldsymbol{\beta}}_k(\hat{\sigma}_{uk}^2) + \hat{u}_{dk}^{\text{FH,EBLUP}(i)},$
          $\hat{R}_d^{\text{FH,EBI}(i)} = \frac{\hat{\mu}_{d1}^{\text{FH,EBLUP}(i)}}{\hat{\mu}_{d1}^{\text{FH,EBLUP}(i)} + \hat{\mu}_{d2}^{\text{FH,EBLUP}(i)}}.$
     d) With the BFH model, calculate BFH-REML estimators $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\beta}}$, and
          $\hat{\boldsymbol{u}}_d^{\text{EBLUP}(i)} = \hat{\boldsymbol{V}}_u \boldsymbol{V}_d^{-1} \Big( \boldsymbol{y}_d^{(i)} - \boldsymbol{X}_d \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}) \Big), \hat{\boldsymbol{\mu}}_d^{\text{EBLUP}(i)} = \boldsymbol{X}_d \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}) + \hat{\boldsymbol{u}}_d^{\text{EBLUP}(i)},$
          $\hat{R}_d^{\text{EBI}(i)} = \frac{\hat{\mu}_{d1}^{\text{EBLUP}(i)}}{\hat{\mu}_{d1}^{\text{EBLUP}(i)} + \hat{\mu}_{d2}^{\text{EBLUP}(i)}}.$
     e) Apply GH, with BFH-REML estimators $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\beta}}$, for obtaining $\hat{R}_d^{\text{EBP}(i)}$.
  3. For $d = 1, \dots, D$, $\hat{R}_d^{(i)} \in \{\hat{R}_d^{\text{FH,EBI}(i)}, \hat{R}_d^{\text{EBI}(i)}, \hat{R}_d^{\text{EBP}(i)}\}$, calculate
       $\text{RBias}_d = 100 \frac{I^{-1} \sum_{i=1}^I (\hat{R}_d^{(i)} - R_d^{(i)})}{|I^{-1} \sum_{i=1}^I R_d^{(i)}|}, \ \text{RRMSE}_d = 100 \frac{\Big( I^{-1} \sum_{i=1}^I (\hat{R}_d^{(i)} - R_d^{(i)})^2 \Big)^{1/2}}{|I^{-1} \sum_{i=1}^I R_d^{(i)}|},$
       $\text{ARBias} = D^{-1} \sum_{d=1}^D |\, \text{RBias}_d \,|, \quad \text{RRMSE} = D^{-1} \sum_{d=1}^D \text{RRMSE}_d \,.$
```

Table 5.5: RBias (in %)

		ρ_u						
ρ_e	Method	-0.75	-0.5	-0.25	0	0.25	0.5	0.75
-0.5	Plug-in FH	0.02	0.03	0.08	0.08	0.11	0.16	0.19
	Plug-in BFH	0.04	0.03	0.07	0.04	0.05	0.07	0.08
	EBP GH	0.01	−0.01	0.02	−0.01	−0.01	0.01	0.01
0	Plug-in FH	−0.03	0.01	0.04	0.09	0.12	0.13	0.16
	Plug-in BFH	0	0.03	0.05	0.09	0.10	0.09	0.10
	EBP GH	−0.03	−0.01	−0.01	0.01	0.02	0	0
0.25	Plug-in FH	−0.03	−0.01	0.04	0.05	0.08	0.12	0.14
	Plug-in BFH	0.01	0.02	0.07	0.06	0.08	0.10	0.11
	EBP GH	−0.02	−0.03	0	−0.02	−0.01	−0.01	−0.01

Table 5.6: ARBias (in %)

		ρ_u						
ρ_e	Method	-0.75	-0.5	-0.25	0	0.25	0.5	0.75
-0.5	Plug-in FH	0.16	0.15	0.17	0.16	0.17	0.20	0.21
	Plug-in BFH	0.16	0.15	0.16	0.15	0.13	0.14	0.12
	EBP GH	0.16	0.14	0.15	0.14	0.13	0.13	0.10
0	Plug-in FH	0.15	0.16	0.15	0.16	0.17	0.18	0.19
	Plug-in BFH	0.14	0.16	0.16	0.16	0.16	0.16	0.15
	EBP GH	0.14	0.15	0.15	0.14	0.13	0.14	0.12
0.25	Plug-in FH	0.14	0.16	0.15	0.14	0.16	0.17	0.19
	Plug-in BFH	0.13	0.15	0.15	0.14	0.16	0.16	0.16
	EBP GH	0.13	0.15	0.13	0.14	0.14	0.13	0.14

Table 5.7: RRMSE (in %)

		ρ_u						
ρ_e	Method	-0.75	-0.5	-0.25	0	0.25	0.5	0.75
-0.5	Plug-in FH	8.88	8.71	8.57	8.43	8.27	8.12	7.96
	Plug-in BFH	8.83	8.74	8.54	8.24	7.78	7.17	6.28
	EBP GH	8.83	8.74	8.54	8.24	7.78	7.17	6.28
0	Plug-in FH	8.62	8.48	8.34	8.18	8.03	7.88	7.70
	Plug-in BFH	8.11	8.28	8.30	8.21	8.00	7.63	7.02
	EBP GH	8.11	8.28	8.30	8.21	8.00	7.63	7.02
0.25	Plug-in FH	8.52	8.37	8.21	8.05	7.92	7.73	7.59
	Plug-in BFH	7.59	7.87	7.98	8.01	7.94	7.68	7.24
	EBP GH	7.59	7.87	7.98	8.01	7.94	7.68	7.24

5.7.4 Simulation 3: Number of bootstrap samples B

Research question

Simulations 3 and 4 evaluate the proposed MSE estimators of the approximated EBPs of the unemployment rates, which were presented in Section 5.6. Simulation 3 is a preliminary study for the two bootstrap procedures in Algorithms 5.6 and 5.7. With the simulation, we want to evaluate how many bootstrap samples B are needed such that the convergence of the bootstrap algorithms is acceptable.

Simulation settings

For the simulation, we again take $D = 200$ areas and $\rho_u = \rho_e = -0.5$. As in Simulation 2, we apply GH quadrature with integral form $\boldsymbol{I}_1$ and 25 function evaluations to approximate the EBPs of the unemployment rates. We simulate the data once, estimate the coefficients of the BFH model, and apply $B = 1,000$ bootstrap replicates. The steps of this preliminary simulation are given in Algorithm 5.11.

Algorithm 5.11 Steps of Simulation 3

```
The steps of Simulation 3 are
  1. Generate data $\boldsymbol{x}_{dk}$, $k = 1, 2$, $\boldsymbol{u}_d \sim N_2(\boldsymbol{0}, \boldsymbol{V}_{ud})$, $\boldsymbol{e}_d \sim N_2(\boldsymbol{0}, \boldsymbol{V}_{ed})$, $\boldsymbol{y}_d = \boldsymbol{X}_d\boldsymbol{\beta} + \boldsymbol{u}_d + \boldsymbol{e}_d$
     $d = 1, \dots, D$.
  2. Calculate the BFH-REML estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$ by using the observed data $(\boldsymbol{y}_d, \boldsymbol{X}_d)$,
     $d = 1, \dots, D$,
  3. For $b = 1, \dots, B$, $B = 1,000$, do
       a) For $d = 1, \dots, D$, generate $\boldsymbol{u}_d^{*(b)} \sim N_m(0, \hat{\boldsymbol{V}}_u)$, $\boldsymbol{e}_d^{*(b)} \sim N_m(0, \boldsymbol{V}_{ed})$ and calculate
          $\boldsymbol{y}_d^{*(b)} = \boldsymbol{\mu}_d^{*(b)} + \boldsymbol{e}_d^{*(b)}$, $\boldsymbol{\mu}_d^{*(b)} = \boldsymbol{X}_d\hat{\boldsymbol{\beta}} + \boldsymbol{u}_d^{*(b)}$, $R_d^{*(b)} = g(\boldsymbol{\mu}_{d1}^{*(b)}, \dots, \boldsymbol{\mu}_{dm}^{*(b)})$.
       b) By using the bootstrap data $(\boldsymbol{y}_d^{*(b)}, \boldsymbol{X}_d)$, $d = 1, \dots, D$, calculate BFH REML
          estimators $\hat{\boldsymbol{\beta}}^{*(b)}$ and $\hat{\boldsymbol{\theta}}^{*(b)}$, EBPs $\hat{R}_d^{\mathrm{EBP}\,*(b)}$ by GH.
  4. For $d = 1, \dots, D$, $\tilde{b} = 1, \dots, B$, $B = 1,000$, calculate
```

$$\widehat{\mathrm{MSE}}_{1d}^{*,\tilde{b}} = \tilde{b}^{-1}\sum_{b=1}^{\tilde{b}}\left(\hat{R}_d^{\mathrm{EBP}\,*(b)} - R_d^{*(b)}\right)^2, \qquad \mathrm{RDiff}(\widehat{\mathrm{MSE}}_{1d}^{*,\tilde{b}}) = 100\frac{(\widehat{\mathrm{MSE}}_{1d}^{*,\tilde{b}} - \widehat{\mathrm{MSE}}_{1d}^{*,B})}{\widehat{\mathrm{MSE}}_{1d}^{*,B}},$$

$$\hat{S}_{1d}^{*,\tilde{b}} = \tilde{b}^{-1}\sum_{b=1}^{\tilde{b}}\left(\hat{R}_d^{\mathrm{EBP}\,*(b)} - R_d^{\mathrm{BP}\,*(b)}\right)^2, \qquad \mathrm{RDiff}(\hat{S}_{1d}^{*,\tilde{b}}) = 100\frac{(\hat{S}_{1d}^{*,\tilde{b}} - \hat{S}_{1d}^{*,B})}{\hat{S}_{1d}^{*,B}}.$$

Results

Figures 5.2 and 5.3 show the relative differences $\mathrm{RDiff}(\widehat{\mathrm{MSE}}_{1d}^{*,\tilde{b}})$ and $\mathrm{RDiff}(\hat{S}_{1d}^{*,\tilde{b}})$. They are the relative difference of the bootstrap values in dependence of the number of iterations $\tilde{b}$ with respect to $\widehat{\mathrm{MSE}}_{1d}^{*,B}$ and $\hat{S}_{1d}^{*,B}$ for $\tilde{b} = 1, \dots, B$, $B = 1,000$. The plots show the mean values over the $D = 200$ domains and the inner 90% and 70% percent of domains for each $\tilde{b} = 1, \dots, 1000$.

For $\widehat{\text{MSE}}^{*}_{1d}$, we recommend to implement the bootstrap algorithm with $B = 400$ to obtain a similar precision to $B = 1,000$ replicates. For $\hat{S}^{*}_{1d}$, we recommend to implement the bootstrap algorithm with $B = 500$ replicates to obtain a similar precision to $B = 1,000$ replicates.

Note that $\hat{S}^{*}_{1d}$ is a very small quantity compared to $\widehat{\text{MSE}}^{*}_{1d}$. The area-mean of $\hat{S}^{*,1,000}_{1d}$ is only about 1/74 of that of $\widehat{\text{MSE}}^{*,1,000}_{1d}$. Therefore, in computing $\text{RDiff}(\hat{S}^{*,\tilde{b}}_{1d})$, the denominator is close to zero which typically results in volatile results as we can see in Figure 5.3. We want to recall that the value of $\hat{S}_{1d}$ plays only a minor role for the total MSE. Therefore, we do not consider the convergence behaviour visible in Figure 5.3 to be problematic and recommend $B = 500$ replicates.

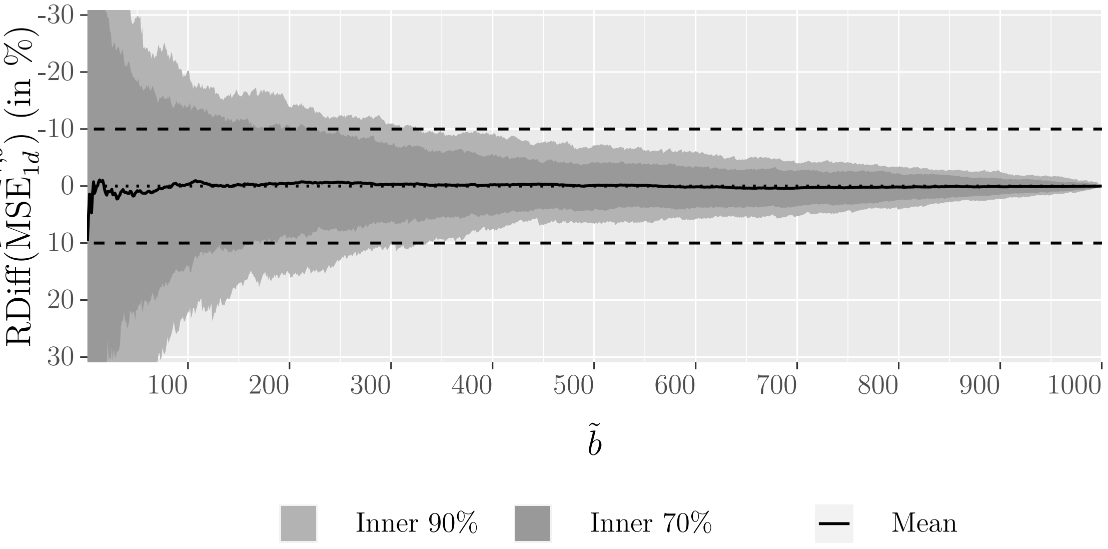

Figure 5.2: Convergence of $\widehat{\text{MSE}}^{*,\tilde{b}}_{1d}$ relative to $\widehat{\text{MSE}}^{*,B}_{1d}$, $d = 1, \ldots, 200$

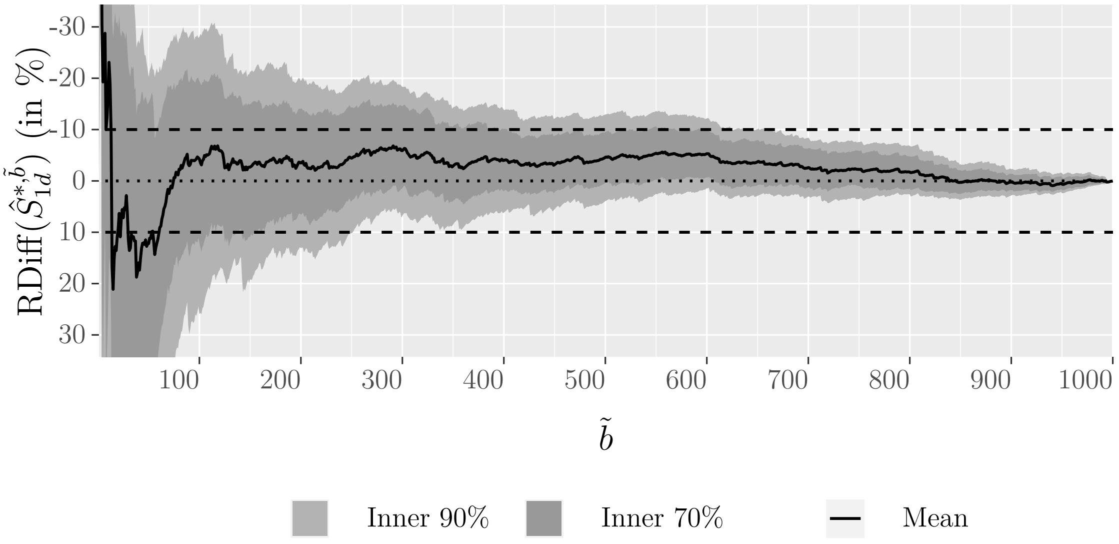

Figure 5.3: Convergence of $\hat{S}^{*,\tilde{b}}_{1d}$ relative to $\hat{S}^{*,B}_{1d}$, $d = 1, \ldots, 200$

5.7.5 Simulation 4: MSE estimators

Research question

Simulation 4 evaluates the MSE estimators presented in Section 5.6. That is, $\widehat{\text{MSE}}^{*}_{1d}$, calculated with Algorithm 5.6, $\widehat{\text{MSE}}^{*}_{2d}$ (5.238), calculated with Algorithms 5.7 and 5.8, and $\widehat{\text{MSE}}^{*}_{3d}$ (5.240), calculated with Algorithm 5.7 and Algorithm 5.8 for only the inner integral.

Simulation settings

Based on the results of Simulation 3, we choose $B = 500$ bootstrap replicates for the two bootstrap procedures in Algorithms 5.6 and 5.7. For the simulation, we again take $D = 200$ areas and $\rho_u = \rho_e = -0.5$. As in Simulation 2 and 3, we apply GH quadrature with integral form $\boldsymbol{I}_1$ and 25 function evaluations to approximate the EBPs of the unemployment rates. The steps of Simulation 3 are given in Algorithm 5.12.

Results

Figure 5.4 shows the RBias and RRMSE (in %) of the three different MSE estimators. The mean values are presented in Table 5.8.

For all three estimators, there is a slight negative relative bias of 2.39%. As the bias is small, we still consider the estimators suitable. There are no differences visible between the RBias of the different estimators. In terms of RRMSE, the proposed separate estimation of the two MSE summands, as done in $\widehat{\text{MSE}}^{*}_{2d}$ and $\widehat{\text{MSE}}^{*}_{3d}$, is superior to the solely parametric bootstrap estimation of the MSE given by $\widehat{\text{MSE}}^{*}_{1d}$. Also, the measures for $\widehat{\text{MSE}}^{*}_{2d}$ and $\widehat{\text{MSE}}^{*}_{3d}$ are close, implying that S^{inner}_{2d} is a good estimator of S_{2d}. The use of $\widehat{\text{MSE}}^{*}_{3d}$ with

Algorithm 5.12 Steps of Simulation 4

```
The steps of Simulation 4 are
  1. Generate data $\boldsymbol{x}_{dk}$, $k = 1, 2$, $\boldsymbol{u}_d \sim N_2(\mathbf{0}, \boldsymbol{V}_{ud})$, $\boldsymbol{e}_d \sim N_2(\mathbf{0}, \boldsymbol{V}_{ed})$, $\boldsymbol{y}_d = \boldsymbol{X}_d\boldsymbol{\beta} + \boldsymbol{u}_d + \boldsymbol{e}_d$ $d = 1, \dots, D$.
  2. For $i = 1, \dots, I$, $I = 1,000$, do
     a) Generate $\boldsymbol{u}_d^{(i)} \sim N_2(\mathbf{0}, \boldsymbol{V}_{ud})$, $\boldsymbol{e}_d^{(i)} \sim N_2(\mathbf{0}, \boldsymbol{V}_{ed})$.
     b) Calculate the true means, the true unemployment rate, and the direct estimates, $\boldsymbol{\mu}_d^{(i)} = \boldsymbol{X}_d\boldsymbol{\beta} + u_d^{(i)}$, $R_d^{(i)} = \frac{\boldsymbol{\mu}_{d1}^{(i)}}{\boldsymbol{\mu}_{d1}^{(i)} + \boldsymbol{\mu}_{d2}^{(i)}}$, $\boldsymbol{y}_d^{(i)} = \boldsymbol{X}_d\boldsymbol{\beta} + \boldsymbol{u}_d^{(i)} + \boldsymbol{e}_d^{(i)}$, $d = 1, \dots, D$.
     c) Calculate the BFH REML estimators $\hat{\boldsymbol{\beta}}^{(i)}$ and $\hat{\boldsymbol{\theta}}^{(i)}$ by using the observed data $(\boldsymbol{y}_d^{(i)}, \boldsymbol{X}_d)$, $d = 1, \dots, D$. Calculate BFH EBPs $\hat{R}_d^{\mathrm{EBP}(i)}$, $d = 1, \dots, D$ by applying GH quadrature. For the MSE estimation run the following algorithm.
        i. For $b = 1, \dots, B$ $(B = 500)$ do
           A. For $d = 1, \dots, D$, generate $\boldsymbol{u}_d^{*(b)} \sim N_m(0, \hat{\boldsymbol{V}}_u)$, $\boldsymbol{e}_d^{*(b)} \sim N_m(0, \boldsymbol{V}_{ed})$ and calculate $\boldsymbol{y}_d^{*(b)} = \boldsymbol{\mu}_d^{*(b)} + \boldsymbol{e}_d^{*(b)}$, $\boldsymbol{\mu}_d^{*(b)} = \boldsymbol{X}_d\hat{\boldsymbol{\beta}}^{(i)} + \boldsymbol{u}_d^{*(b)}$, $R_d^{*(b)} = g(\boldsymbol{\mu}_{d1}^{*(b)}, \dots, \boldsymbol{\mu}_{dm}^{*(b)})$.
           B. By using the bootstrap data $(\boldsymbol{y}_d^{*(b)}, \boldsymbol{X}_d)$, $d = 1, \dots, D$, calculate BFH REML estimators $\hat{\boldsymbol{\beta}}^{*(b)}$ and $\hat{\boldsymbol{\theta}}^{*(b)}$. Calculate BFH EBPs $\hat{R}_d^{\mathrm{EBP}\,*(b)}$ by applying GH quadrature.
        ii. For $d = 1, \dots, D$, calculate $\widehat{\mathrm{MSE}}_{1d}^{*(i)} = B^{-1}\sum_{b=1}^{B}\left(\hat{R}_d^{\mathrm{EBP}\,*(b)} - R_d^{*(b)}\right)^2$,
            $\hat{S}_{1d}^{*(i)} = B^{-1}\sum_{b=1}^{B}\left(\hat{R}_d^{\mathrm{EBP}\,*(b)} - R_d^{\mathrm{BP}\,*(b)}\right)^2$.
     d) For $d = 1, \dots, D$ calculate
        $\widehat{\mathrm{MSE}}_{2d}^{*(i)} = \hat{S}_{1d}^{*(i)} + \hat{S}_{2d}^{(i)}$, $\widehat{\mathrm{MSE}}_{3d}^{*(i)} = \hat{S}_{1d}^{*(i)} + \hat{S}_{2d}^{\mathrm{inner}(i)}$, where $\hat{S}_{2d}^{(i)}$ and $\hat{S}_{2d}^{\mathrm{inner}(i)}$ are calculated using Algorithms 5.8 and Algorithm 5.8 for only the inner integral with $T = R = 5$, the total number of nodes in each dimension for the outer and inner integral. For Algorithms 5.8 and Algorithm 5.8 for only the inner integral, this results in $5^4 = 625$ (due to the double integral) and $5^2 = 25$ function evaluations respectively.
  3. For $d = 1, \dots, D$, $\widehat{\mathrm{MSE}}_d^{*(i)} \in \{\widehat{\mathrm{MSE}}_{1d}^{*(i)}, \widehat{\mathrm{MSE}}_{2d}^{*(i)}, \widehat{\mathrm{MSE}}_{3d}^{*(i)}\}$ calculate

$$\mathrm{MSE}_d = I^{-1}\sum_{i=1}^{I}(\hat{R}_d^{\mathrm{EBP}(i)} - R_d^{(i)})^2, \qquad \widehat{\mathrm{MSE}}_d^{*} = I^{-1}\sum_{i=1}^{I}\widehat{\mathrm{MSE}}_d^{*(i)},$$

$$\mathrm{RBias}_d = 100\frac{\widehat{\mathrm{MSE}}_d^{*} - \mathrm{MSE}_d}{\mathrm{MSE}_d}, \qquad \mathrm{RRMSE}_d = 100\frac{\left(I^{-1}\sum_{i=1}^{I}(\widehat{\mathrm{MSE}}_d^{*(i)} - \mathrm{MSE}_d)^2\right)^{1/2}}{\mathrm{MSE}_d},$$

$$\mathrm{RBias} = D^{-1}\sum_{d=1}^{D}|\,\mathrm{RBias}_d\,|, \qquad \mathrm{RRMSE} = D^{-1}\sum_{d=1}^{D}\mathrm{RRMSE}_d\,.$$
```

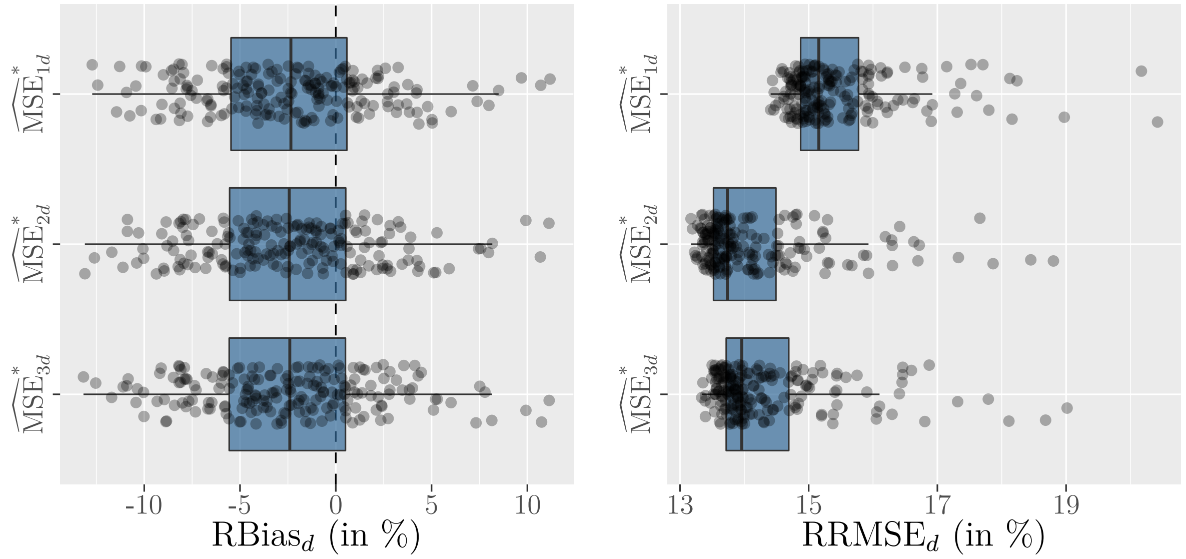

Figure 5.4: Performance of MSE estimators

S_{2d}^{inner} instead of $\widehat{\text{MSE}}^*_{2d}$ with S_{2d} can be especially useful for higher-dimensional problems as it significantly reduces the number of function evaluations.

Table 5.8: Performance of MSE estimators

MSE estimator	RBias (in %)	RRMSE (in %)
$\widehat{\text{MSE}}^*_{1d}$	-2.39	15.47
$\widehat{\text{MSE}}^*_{2d}$	-2.39	14.14
$\widehat{\text{MSE}}^*_{3d}$	-2.39	14.34

5.8 Application

5.8.1 Data description

As an illustrative example of the proposed methodology, we estimate unemployment rates for small domains using publicly available data from the *Spanish Labour Force Survey* (SLFS) for the first quarter of 2021.

Similar to the German LFS, which is conducted as part of the German Microcensus, compare Section 4.3, the questions of the Spanish LFS are also harmonized to the standards of the ILO and EUROSTAT. The focus of the SLFS is to provide different statistics on the number of persons being employed, unemployed, and not in labour force, among others. Note that, different from the ILO definition applied by most European countries including

Germany, the minimum age for persons to be considered employed/unemployed in the SLFS is 16, not 15[1].

The SLFS is conducted quarterly by the *Spanish Statistical Office* (INE). The samples are drawn via two-stage stratified random sampling for each Spanish province. At the first stage, municipality areas are drawn as the primary sampling units. At the second stage, dwellings are drawn from strata of each municipality. In the dwellings, all households are interviewed.

On the INE website[2], SLFS micro-data for the first quarter of 2021 are publicly availably. The data contain around $140,000$ observations at the person-level. Geographically, the most detailed information available in the micro-data is an indicator for the Spanish provinces. Next to geographical information, the sample micro-data contain information on persons demography, education, studies, labour activity during the reference week, and employment characteristics, among others. Furthermore, the data contain elevation factors. The elevation factors correspond to the inverses of the inclusion probabilities corrected for non-response and calibrated to known province quantities. With the evaluation factors, we can calculate direct estimators by province from the micro-level SLFS data.

5.8.2 Domains of interest and direct estimates

From the SLFS age variable, which differentiates between five age categories, we define the broader categories `AGE1` (16-24 years), `AGE2` (25-54 years), and `AGE3` (55-64 years). As domains of interest, we take the cross-combinations of the 50 Spanish provinces plus the autonomous cities Ceuta and Melilla, persons sex, and the three age classes, resulting in a total of $D = 52 \times 2 \times 3 = 312$ domains. As dependent variables for a bivariate FH (BFH) model, we calculate the proportions of employed and unemployed in the domains, similar to the simulation studies in Section 5.7.

In the micro-data, dummy variables y_{di1} and y_{di2} indicates whether person i in domain d is employed or unemployed respectively. For person i in domain d, the non-negative elevation factor is given by w_{di}. With s_d, we denote the sample in domain d, $d = 1, \dots, D$. The domain sizes are unknown and estimated by

$$\hat{N}_d = \sum_{i \in s_d} w_{di}, \quad d = 1, \dots, D. \tag{5.244}$$

The proportions of employed and unemployed in the domains of interest are calculated as

$$\hat{\mu}_{dk}^{\mathrm{Dir}} = \hat{N}_d^{-1} \sum_{i \in s_d} w_{di} y_{dik}, \quad d = 1, \dots, D, \quad k = 1, 2. \tag{5.245}$$

Direct estimates $\hat{\mu}_{dk}^{\mathrm{Dir}}$ are used as dependent variables y_{dk} in BFH model 5.177.

[1] https://ec.europa.eu/eurostat/cache/metadata/en/lfsi_esms.htm.

[2] https://www.ine.es/en/index.htm.

There is one domain (province 49 - Zamora, male, `AGE1`) with no unemployed person. Hence, the direct estimate of the proportion of unemployed is zero and no variance estimate can be calculated. We exclude this domain and thereby consider the remaining $D = 311$ domains as domains interest.

For the BFH model, sampling covariance matrices $\boldsymbol{V}_{ed}$, $d = 1, \ldots, D$, are needed. From the INE, no second-order inclusion probabilities are provided. We therefore approximate the design-based variance of $\hat{\mu}_{dk}^{\text{Dir}}$ by using the Hajek-approximation which only uses the first-order inclusion probabilities (Hájek, 1964), giving

$$\sigma_{edk}^2 = \widehat{\text{Var}}(\hat{\mu}_{dk}^{\text{Dir}}), \quad d = 1, \ldots, D, \quad k = 1, 2. \tag{5.246}$$

Elements σ_{ed1}^2 and σ_{ed2}^2 are the diagonal elements of $\boldsymbol{V}_{ed}$, $d = 1, \ldots, D$. We further need the off-diagonal elements of $\boldsymbol{V}_{ed}$, given by the covariances of the direct estimates. We approximate the covariances by (Wood, 2008)

$$\widehat{\text{Cov}}(\hat{\mu}_{d1}^{\text{Dir}}, \hat{\mu}_{d2}^{\text{Dir}}) = \frac{1}{2}(\widehat{\text{Var}}(\hat{\mu}_{d1}^{\text{Dir}}) + \widehat{\text{Var}}(\hat{\mu}_{d2}^{\text{Dir}}) - \widehat{\text{Var}}(\hat{\mu}_{d1}^{\text{Dir}} - \hat{\mu}_{d2}^{\text{Dir}})), \quad d = 1, \ldots, D. \tag{5.247}$$

The sampling error correlation of $\hat{\mu}_{d1}^{\text{Dir}}$ and $\hat{\mu}_{d2}^{\text{Dir}}$, $d = 1, \ldots, D$, is calculated by

$$\hat{\rho}_{ed} = \widehat{\text{Cov}}(\hat{\mu}_{d1}^{\text{Dir}}, \hat{\mu}_{d2}^{\text{Dir}}) \Big/ \Big(\widehat{\text{Var}}(\hat{\mu}_{d1}^{\text{Dir}})\widehat{\text{Var}}(\hat{\mu}_{d2}^{\text{Dir}})\Big)^{1/2}, \quad d = 1, \ldots, D, \tag{5.248}$$

Table 5.9 displays the quantiles of $\hat{\rho}_{ed}$, $d = 1, \ldots, D$. The sampling errors of the dependent variables have a negative correlation of moderate to small magnitude.

Table 5.9: Quantiles of $\hat{\rho}_{ed}$, $d = 1, \ldots, D$

0%	25%	50%	75%	100%
-0.58	-0.38	-0.24	-0.15	-0.04

5.8.3 Model choice

We have calculated direct estimates of the domain-specific proportions of employed and unemployed $\hat{\mu}_{dk}^{\text{Dir}}$, $k = 1, 2$, including estimates of their sampling error covariance matrices $\boldsymbol{V}_{ed}$, $d = 1, \ldots, D$. For a BFH model, we additionally need auxiliary information on the domain level.

The INE does not publish aggregate data for our domains of interest, which are provinces crossed by sex and three age classes. For this illustration, we therefore take four waves of SLFS sample data from 2020, estimate domain aggregates with these data, and take these estimated aggregates as known auxiliary information. The auxiliary estimates are then based on four times as many observations as the estimates of the dependent variables, wherefore we expect their sampling variances to be considerably lower than those of the dependent variables.

We note that the auxiliary information as defined above are subject to measurement errors, violating the assumptions of the MFH model, as we introduced it in Section 5.3. When working with estimated auxiliary information which are subject to sampling errors, measurement error models as proposed by Arima et al. (2017), Burgard et al. (2020b, 2021a), and Burgard et al. (2022), Ybarra and Lohr (2008) should be applied to account for the extra uncertainty in the model. A combination of EBP approximations in MFH models and measurement errors in the covariates, however, is out of the scope of the presented chapter and left as future research area. For an illustration of the proposed method, we find the estimated auxiliary data suitable.

As auxiliary information from the four SLFS waves of 2020, we consider person-level dummy variables `Edu.Pri` (indicator for primary education of less), `Edu.Sec` (indicator for secondary education), and `Nat.Spa` (indicator for Spanish nationality). The reference category `Edu.Sup` (indicator for superior education) is excluded.

A BFH model is fitted. As input, we use the domain-specific proportions of employed and unemployed $\hat{\bar{\mu}}_{dk}^{\text{Dir}}$ as the dependent variables y_{dk}, $k = 1, 2$, with sampling error covariance matrices $\boldsymbol{V}_{ed}$, $d = 1, \ldots, D$, estimated from the SLFS data. As covariates, we consider SLFS estimates (from the four SLFS waves of 2020) of the domain-specific proportions of dummy variables `Edu.Pri`, `Edu.Sec`, and `Nat.Spa`. In addition, we consider domain-specific dummy variables `AGE2`, `AGE3`, and `sex.male` as covariates. After considering different models, we chose the proportions of variables `Edu.Pri` and `Edu.Sec` and the dummy variables `AGE2`, `AGE3`, and `sex.male` as covariates, for both dependent variables. With the data, we estimate a BFH model and apply the REML Fisher-Scoring algorithm to estimate the model parameters as we did in the simulation studies.

5.8.4 Parameter estimates

We first discuss the estimated variances components and fixed effects of the BFH model. Table 5.10 displays the estimated variance components with 95% confidence intervals. The confidence intervals of the random effect variances do not contain zero, indicating that the model including the domain-specific random effects explains part of the variability of the target variables. The estimated correlation of the random effects is considerably negative, indicating that when the domain-specific random effect of the employment proportion is high in a domain, the unemployment proportion is low and vice versa.

The estimated fixed effects of the model are displayed in Table 5.11. All parameter estimates are significant at the 5% level. For the estimated fixed effects, we recall that the reference category for education variables `Edu.Pri` and `Edu.Sec` is superior education. A higher proportion of persons in primary and secondary education (`Edu.Pri` and `Edu.Sec`) is associated with a higher proportion of unemployed and a lower proportion of employed. Domains for the male population are associated with higher employment and lower unemployment proportions than domains for the female population. As this is an

Table 5.10: Variance component estimates and asymptotic 95% confidence intervals

	$\hat{\theta}$	Lower limit	Upper limit
$\hat{\sigma}_{u1}$	0.028	0.024	0.031
$\hat{\sigma}_{u2}$	0.052	0.046	0.058
$\hat{\rho}_{12}$	-0.731	-0.832	-0.629

illustrative example based on estimated covariate records, the coefficients values, however, should be treated with caution.

Table 5.11: Estimated fixed effects for proportion of unemployed and employed

	Unemployed				Employed			
	$\hat{\beta}$	std.err.	t-value	p-value	$\hat{\beta}$	std.err.	t-value	p-value
(Intercept)	0.025	0.028	0.915	0.360	0.404	0.051	7.903	0.000
Edu.Pri	0.231	0.037	6.308	0.000	-0.611	0.067	-9.094	0.000
Edu.Sec	0.102	0.034	3.019	0.003	-0.319	0.062	-5.121	0.000
AGE2	0.038	0.011	3.425	0.001	0.474	0.020	24.078	0.000
AGE3	-0.035	0.009	-3.978	0.000	0.362	0.016	22.534	0.000
sex.male	-0.015	0.004	-3.428	0.001	0.122	0.008	15.469	0.000

5.8.5 Model diagnostics

In Figure 5.5, we present different diagnostics to check the validity of the calculated BFH model. The diagnostics are separately shown for the dependent variables proportion of unemployed (left) and proportion of employed (right).

In the first row, as proposed in Brown et al. (2001), the direct estimates are plotted against the EBLUPs and the diagonal line $y = x$ is added to the plots. From theory, the direct estimates are design-unbiased, although they can exhibit large variances for small domains. Therefore, plotting the direct estimates versus the EBLUPs of a FH model can indicate a potential model bias. From the figure, we see no systematic deviations from direct estimates to EBLUPs for the proportion of employment. For the proportion of unemployed, the model EBLUPs are systematically lower than the direct estimates for higher unemployment proportions. We note that in the application high direct estimates of unemployment proportions are associated with high sampling variances of the direct estimators. Therefore, for large unemployment proportions, the BFH model smooths the predictions more towards the model-based part than the direct estimates. We therefore do not see indications of a model bias from the figures.

In row 2 of Figure 5.5, we plot the BFH EBLUPs versus standardized residuals. With the figures, we want to see whether the assumption of normally distributed residuals is valid. The residuals are calculated as $r_{dk}^{\text{EBLUP}} = \hat{\mu}_{dk}^{\text{Dir}} - \hat{\mu}_{dk}^{\text{EBLUP}}$, $k = 1, 2$, $d = 1, \ldots, D$. The standardized residuals are calculated as $(r_{dk}^{\text{EBLUP}} - D^{-1}\sum_{d=1}^{D} r_{dk}^{\text{EBLUP}})/std(r_{dk}^{\text{EBLUP}})$, where $std(r_{dk}^{\text{EBLUP}})$ is the standard deviation of the set of residuals, $k = 1, 2$, $d = 1, \dots, D$. By the definition of the BFH model, the residuals should be normally distributed with mean zero. Furthermore, no systematic differences should be visible in the residual distribution for changing magnitudes of EBLUPs. Both for employment and unemployment most residuals are within range. 1.3% (4/311) of the absolute residual values are larger than three for each, the proportion of unemployed and employed. We consider this as acceptable and hence see no model violation in it. For unemployment, there is a noticeably large standardized residual. The underlying sample size of the corresponding direct estimate is very small (41 persons) such that the BFH EBLUP is smoothed more towards the model-based estimator.

To see whether there are efficiency gains in applying the BFH model instead of direct estimators, the standard errors of the direct estimators are plotted against the RMSEs of the BFH EBLUPs in the last row of Figure 5.5 with the additional diagonal line $y = x$. The BFH EBPLUPs are at least as efficient as the direct estimators as all points lie along or below the diagonal line. For large standard errors of direct estimates, the BFH EBLUPs give high efficiency gains over the direct estimators. For these cases, the BFH model puts more weight on the estimated model than on the direct estimates. For future research it would be interesting to see the efficiency gains when the measurement errors in the auxiliaries are fully accounted for.

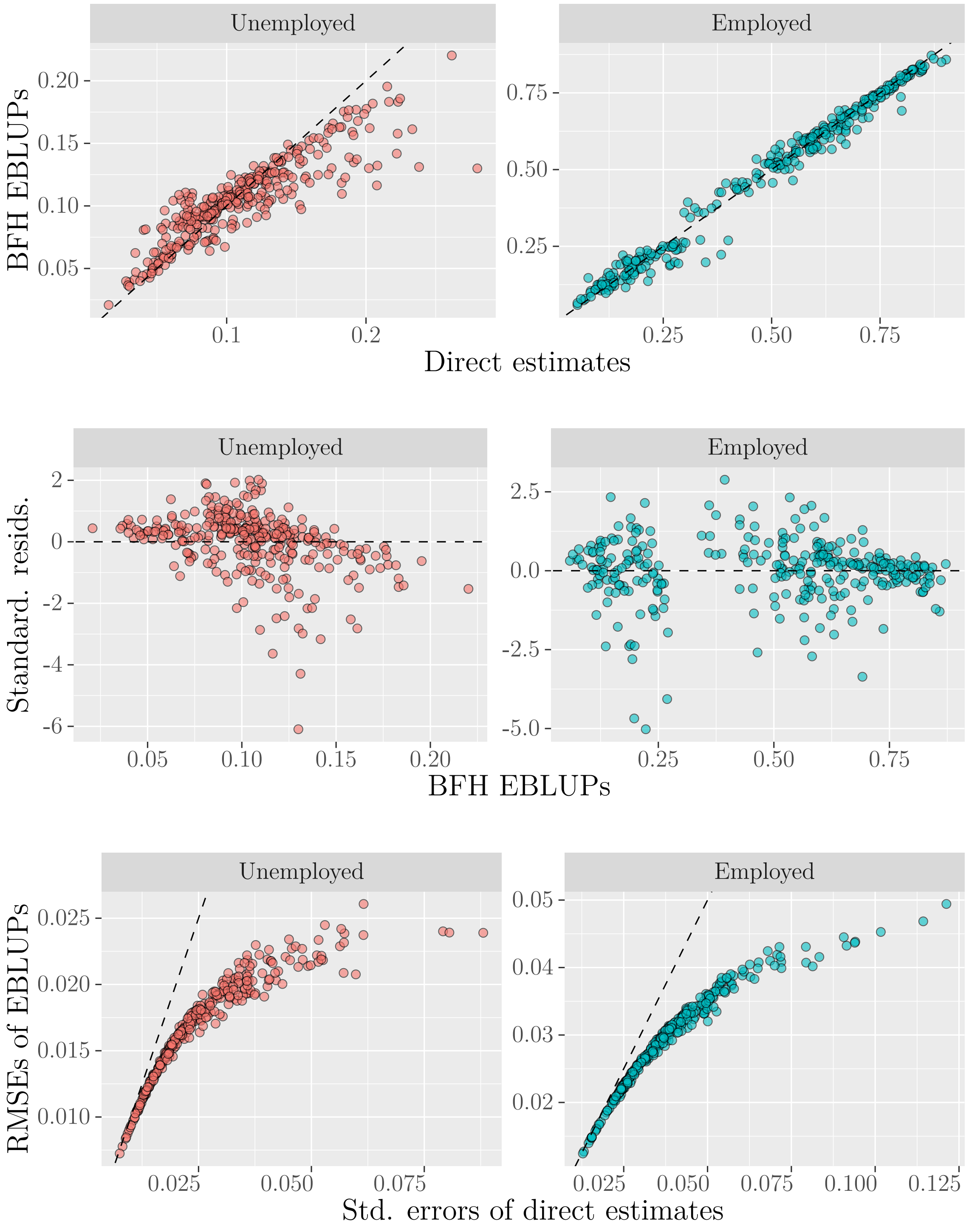

Figure 5.5: Model diagnostics for the proportions of unemployed (left) and employed (right)

5.8.6 Unemployment rates

From the calculated model, we can calculate plug-in EBIs and BFH EBPs of the unemployment rate as shown in Section 5.5 and Simulation 2. Similar to Simulation 2, the plug-in EBIs are calculated as $\hat{R}_d^{\mathrm{EBI}} = \hat{\mu}_{d1}^{\mathrm{EBLUP}} / (\hat{\mu}_{d1}^{\mathrm{EBLUP}} + \hat{\mu}_{d2}^{\mathrm{EBLUP}})$, where $\hat{\mu}_{d1}^{\mathrm{EBLUP}}$ and $\hat{\mu}_{d2}^{\mathrm{EBLUP}}$ are the BFH EBLUPs of the proportions of unemployed and employed, and the EBPs $\hat{R}_d^{\mathrm{EBP}}$ are calculated with GH quadrature with a total of 25 function evaluations, $d = 1, \dots, D$.

To see whether the resulting predictions of the unemployment rates are realistic, we additionally calculate the direct estimates of the unemployment rates from the SLFS data and compare them to the model EBIs and EBPs. Figure 5.6 shows the direct estimates of the unemployment rates versus EBIs (left) and EBPs (right). The diagonal line $y = x$ is added to the plot. Neither EBPs nor EBIs show systematic differences to the direct estimates. Furthermore, we see that in this particular application the differences between the EBPs and EBIs are small. From the figure, we barely see any differences between the two plots, the correlation of the values is close to one, so they are almost identical. The results from Simulation 2 show that the performance difference of EBIs and EBPs is small when the correlations of sampling errors and random effects have same sign and similar magnitude. In this application, both correlations are estimated to be moderately negative, compare Tables 5.9 and 5.10. From the correlations and the results of Simulation 2, we would therefore also not expect the predictions to differ to a great extend.

For the BFH EBPs, we can calculate MSE estimates based on the theory presented in Section 5.6 and Simulations 3 and 4. Following the results of the simulation studies, we use estimator $\widehat{\mathrm{MSE}}_{2d}^{*} = \hat{S}_{1d}^{*} + \hat{S}_{2d}$ (5.238) with $B = 500$ bootstrap samples for $\hat{S}_{1d}^{*}$ and nested GH with $T = 5$ evaluations for each dimension, resulting in a total of 625 function evaluations, for $\hat{S}_{2d}$.

We are especially interested in the EBPs of the unemployment rate for `AGE1` (16-24 years) for the 52 regional domains by sex. On the INE website[3], the national unemployment rates of persons under 25 years of age are estimated at 38.18% (males) and 41.18% (females) respectively. To put the numbers in perspective, we take the labour force statistics for persons aged 15-24 in the first quarter of 2021 for the European countries provided by Eurostat[4]. The overall unemployment rates for the European Union 27 countries are 18.2% (males) and 19.0% (females). For the comparison, recall that, unlike most European countries, the INE sets the minimum age for unemployed persons at 16, whereas other European countries, e.g. Germany, set it to 15. Despite this small difference in definition, we see that the Spanish unemployment rates for persons aged 24 and younger is one of the highest among all European countries and therefore the unemployment rates for `AGE1` (16-24 years) are of special political interest.

Figure 5.7 shows the BFH EBPs (row one) and corresponding root MSEs (RMSEs) (row two) for the 52 provinces by sex for `AGE1`. The corresponding figures for `AGE2` and `AGE3` are given in Appendix A.2. Note that we use different colour scales per age class. The

[3] https://www.ine.es/jaxiT3/Datos.htm?t=4247.

[4] Eurostat. Unemployment by sex and age quarterly data [UNE_RT_Q___custom_2134929].

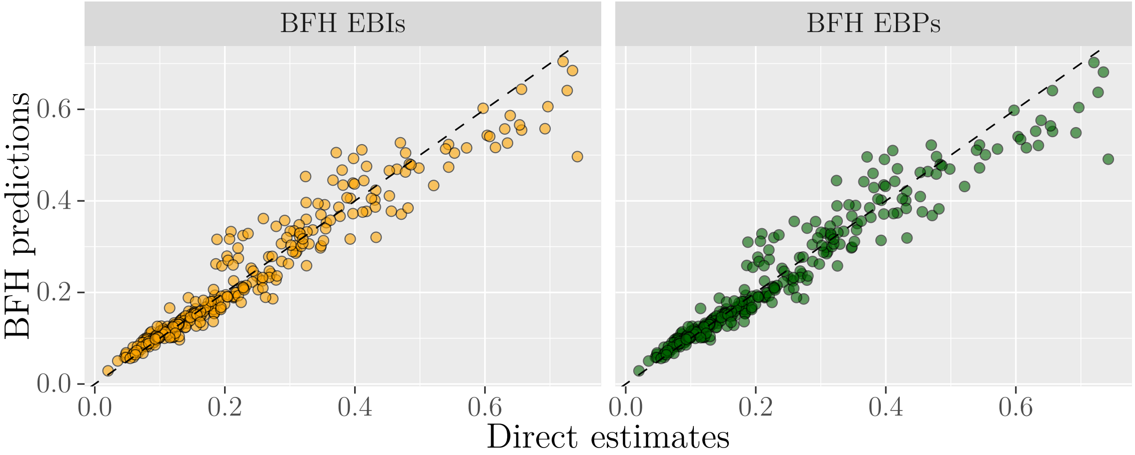

Figure 5.6: Direct estimates versus model-based predictions of unemployment rates

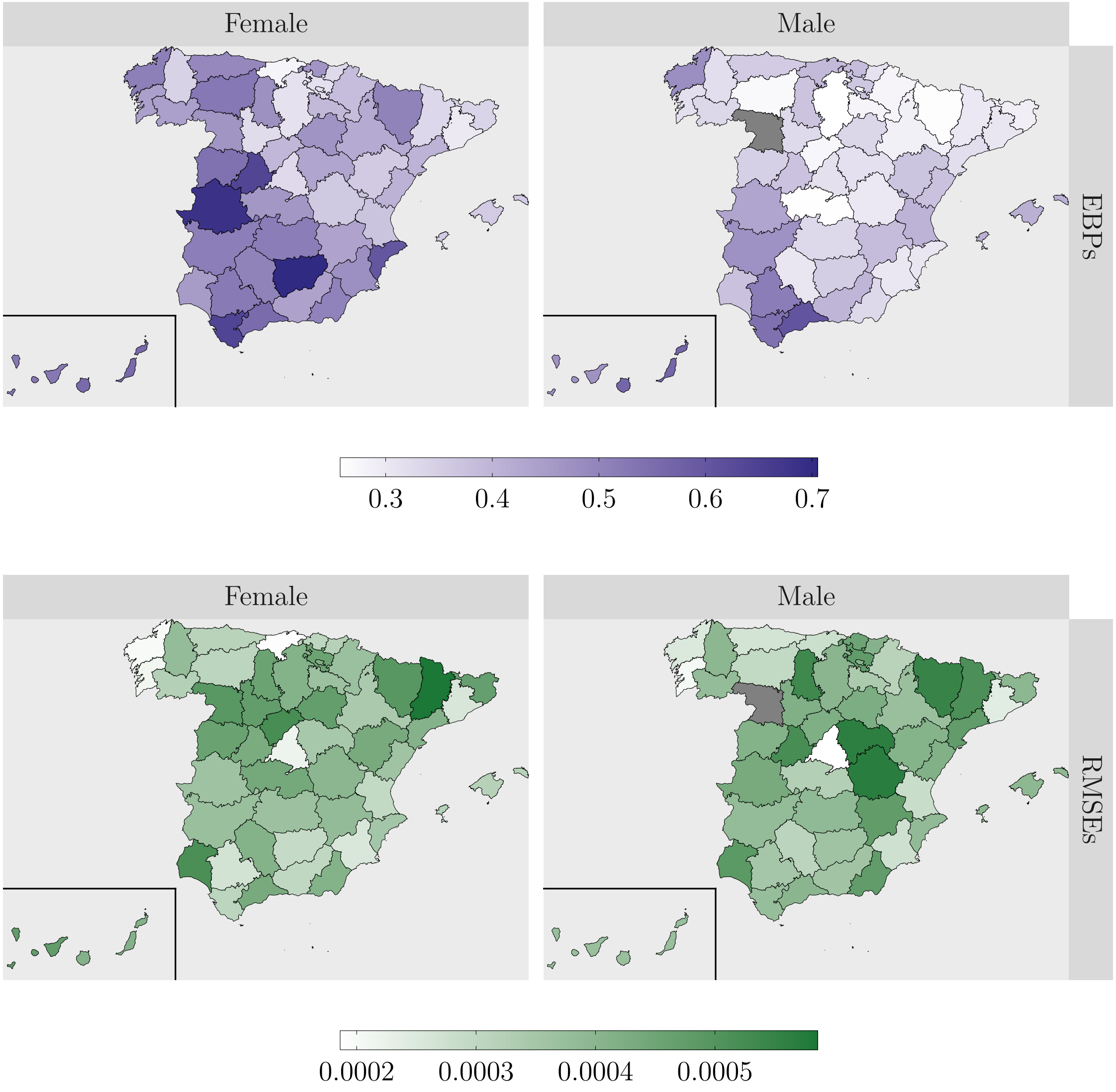

Figure 5.7: BFH EBPs and their RMSEs for `AGE1` (16-24 years)

unemployment rates for AGE1 are significantly higher than for AGE2 and AGE3. From Figure 5.7, we can see that the unemployment rates are higher for females than for males and that they differ to a great extend between the 52 geographical areas. For the EBPs of males, there is a grey shaded province in Figure 5.7, which corresponds to the province which was excluded from the model as there were no unemployed persons in the sample and we could not calculate a variance estimate. From the figure, we see that the magnitude of the MSEs is quite heterogeneous, indicating that for some domains the model-based predictions worked well whereas for others the predicted unemployment rates should be interpreted with caution. Recall that in the model and MSE estimation we did not consider the measurement error in the covariates, but left it as a future research area.

5.9 Summary and outlook

In this chapter, we introduced approximations of BPs of general, potentially non-linear, multi-variable domain indicators like the unemployment rate in MFH models. Under the linear mixed model theory, the introduced BPs of these indicators are unbiased with minimum MSE in the class of model-unbiased predictors. The BPs, however, are given in integral forms and can only be approximated. We therefore presented different integral approximation techniques aimed at different dimensions of the non-linear indicator considered.

With several simulation studies, we empirically investigated the applicability of the proposed BP approximations by simulating employment and unemployment proportions as dependent variables for a BFH model and approximating the (E)BPs of the unemployment rates from the model.

Simulation 1 investigated the choice of the integral approximation for the BPs. For the approximation, we considered Gauss-Hermite quadrature, MC integration plus antithetic variates, and Quasi MC integration with the Sobol and Halton sequence in combination with different integral forms. Especially Gauss-Hermite quadrature showed a good performance in approximating the BPs of the unemployment rates, especially with only few function evaluations.

Simulation 2 compared the performance of the plug-in predictors to the EBP approximations of the unemployment rates under different correlation scenarios. The simulation revealed that approximating the EBPs of the unemployment rates mostly gave lower relative bias than the corresponding plug-in predictors, while inhibiting very similar RRMSEs, which is in line with the theory.

Simulation 3 investigated the number of samples B for the parametric bootstrap procedures for estimating the MSE of the approximated EBPs and Simulation 4 analysed the performance of the different MSE estimators. We saw that estimating the two components of the MSE separately gave the best results. Furthermore, another proposed MSE estimator which only approximates the expectation of the inner integral of the second MSE component and therefore needs significantly fewer function evaluations, showed a good performance.

In an illustrative application, we applied the proposed procedure to publicly available Spanish LFS data. Similar to the simulation studies, we applied a BFH model using the proportions of employed and unemployed as dependent variables. The domains of interest were constituted by the cross-combinations of 52 regions, three age classes, and sex. The indicators of interest were the unemployment rates, which we predicted with the proposed EBP approximation and the BFH plug-in predictor. We evaluated the model fit and the estimated unemployment rates.

The application gave rise to potential future research. As there were not suitable, publicly available auxiliary information available, we estimated auxiliary variables from a larger sample of micro-data from the SLFS. In the application, we took this auxiliary information as given and estimated without error. For future application, it would be interesting to combine the theory of multi-variable area-level domain indicators and measurement errors in the covariates. Furthermore, we illustrated the approach for a rather simple non-linear indicator, the unemployment rate. For future research it would be interesting to investigate the behaviour of the proposed approach for different non-linear indicators of different dimensions.

Chapter 6

Multivariate Fay-Herriot Models under Missing Direct Estimates

6.1 Introduction

The standard multivariate Fay-Herriot (MFH) model presented in Section 5.3 is based on the assumption that all direct estimates $\boldsymbol{y}_d \in \mathbb{R}^m$ of parameters $\boldsymbol{\mu}_d \in \mathbb{R}^m$ are known for all domains $d = 1, \ldots, D$. In practical application, this assumption is usually not fulfilled. In particular, the direct estimates of one of the m variables may be missing for some domains while direct estimates of another variable may be missing in other domains.

As an example, consider a MFH model where the m dependent variables correspond to the values of one variable at m different time points. When the domains of interest are not explicitly considered in the sampling process, e.g. as strata, it can happen that for small domains the sample sizes are, by chance, zero at specific time points. Then no direct estimates can be calculated for some domain-time combinations, resulting in partially missing direct estimates for MFH models.

In a MFH model, based on the theory presented in Section 5.3, only those domains can be used for which the direct estimates of the m variables are complete. Already a domain where only one of the m direct estimates is missing, would have to be excluded from the modelling process. This has two effects. One is that not all available information is used for parameter estimation of the model. The other is that for the excluded domain, i.e. the domain for which at least one direct estimate was missing, only synthetic predictions can be computed under the MFH model.

In this chapter, we introduce the *multivariate Fay-Herriot model under partially missing direct estimates* (MMFH) to solve these problems. The MMFH model is capable of including all observations of direct estimates into the model for efficient parameter estimation. Furthermore, we derive best predictions for multiple domain variables with missing direct estimates. They draw from the information of the available direct estimates of a domain to predict the domain random effects of all m variables of that domain.

We note that part of the work presented in this chapter is already published. Preliminary work on this chapter dealing with the case of bivariate FH models under partial missing direct estimators is published in Burgard et al. (2019b, 2021c). The application shown in this chapter is the same application which is published in Burgard et al. (2021c).

The chapter proceeds as follows. Section 6.2 gives additional information on MFH models under missing direct estimates. We discuss situations in which partially missing direct estimates occur to stress the practical necessity of the proposed approach. Furthermore, we propose an MSE estimator for MFH synthetic predictions. The multivariate FH model under partially missing direct estimates (MMFH) is presented in Section 6.3. We derive best predictions under the MMFH model in Section 6.4, and algorithms for the ML/REML parameter estimation under the model in Section 6.5. Formulas for approximating the MSE of the predictions are derived in Section 6.6. In Section 6.7, we use simulation studies to analyse the behaviour of the proposed parameter estimation, prediction of the characteristics of interest, and MSE estimation. Section 6.8 presents an illustrative application of the proposed approach to publicly available data from the U.S. Census Bureau.

6.2 Missing direct estimates in MFH models

6.2.1 Reasons for missing direct estimates

While we have already briefly touched on the possibility of missing direct estimates, we would like to go into a little more detail about the problem. We consider two main reasons for missing direct estimates: Domain-specific sample sizes smaller than two and publication restrictions.

We first take a look at the case of sample sizes smaller than two. When small domains are not explicitly accounted for in a sampling design, so-called *unplanned domains*, the randomisation process of the sampling design can result in domain-specific sample sizes smaller than two, compare Lehtonen and Veijanen (2009, Section 2.2.1). When a domain-specific sample size is zero, the domain is referred to as an *unsampled domain* and no direct estimate can be calculated. When a domain-specific sample size is one, we can calculate a simple direct estimator like the domain mean or total, but we cannot calculate a variance estimator unless there are at least two sampling observations. Sampling techniques like stratification can assure a fixed and high sample size in chosen domains. However, a survey is usually constructed to have a variance-optimal allocation for key variables while considering cost restrictions and response burden of the survey objects. These considerations naturally prohibit the possibility of designing surveys which give direct estimates with low design-variance for various variable and domain combinations. Therefore, depending on the domains of interest, there may be unsampled domains and domains with small sample sizes also in large national surveys.

Another reason for missing direct estimates are publication restrictions. In official statistics direct estimates are typically only published when they meet certain requirements. Publications in official statistics are usually bound to estimators where the *coefficient of variation* is not higher than 20% (Molina & Marhuenda, 2015, p. 85). For example, in the German Microcensus, yearly estimates of population counts with values under 5.000 are not published as the anticipated standard error of the estimates is considered too high (Destatis, 2020a, Section 4.2). That is, also when domain-specific sample sizes are high enough to compute direct estimates and estimates of their variances, the estimates are not necessarily publicly available.

As an example of data with partially missing direct estimates, we include an illustrative application of the proposed MMFH model to publicly available data in Section 6.8. There, we focus on publicly available U.S. county-level data from a variable in two consecutive years t and t'. For both t and t', some of the county direct estimates are missing. In particular, many of the direct estimates which are not available in t are available in t' and vice versa. Hence, we have partially missing direct estimates and this is exactly the data situation for which we present the MFH model with partially missing values (MMFH), develop the algorithms for the parameter estimation, and derive the best prediction and MSE formulas in Sections 6.3, 6.4, 6.5, and 6.6.

We would like to point out that with the MMFH model, auxiliary information that is partially missing and subject to measurement error can also be included in the model as additional dependent variables.

6.2.2 MFH synthetic predictors

When we encounter partially missing direct estimates in MFH models, for the domains with missing direct estimates only synthetic predictions can be calculated. In the following, we therefore briefly present synthetic predictors and an estimator to their MSEs under MFH models. In Section 5.3, we presented the MFH model (5.177) as

$$\boldsymbol{y}_d = \boldsymbol{X}_d\boldsymbol{\beta} + \boldsymbol{u}_d + \boldsymbol{e}_d, \quad d = 1, \dots, D, \tag{5.177}$$

with $\boldsymbol{u}_d \sim N_m(\boldsymbol{0}, \boldsymbol{V}_{ud})$, $\boldsymbol{e}_d \sim N_m(\boldsymbol{0}, \boldsymbol{V}_{ed})$, and independent $\boldsymbol{e}_d$ and $\boldsymbol{u}_d$. As in Section 5.3, the parameters of $\boldsymbol{V}_{ud}$ are denoted by $\boldsymbol{\theta} \in \mathbb{R}^q$, with $q = m(m+1)/2$, and the $\boldsymbol{V}_{ed}$ are assumed known, $d = 1, \dots, D$.

Let us assume that some of the $\boldsymbol{y}_d$ are partially or fully missing, while all the $\boldsymbol{X}_d$ are known, $d = 1, \dots, D$. Without loss of generality, we reorder the domains such that we can partition the set of domains $\mathcal{D} = \{1, \dots, D\}$ into the two subsets

$\mathcal{D}^{\text{mis}} = \{1, \dots, D^{\text{mis}}\}$ with $D^{\text{mis}} \le D$ and

$\mathcal{D}^{\text{obs}} = \{D^{\text{mis}} + 1, \dots, D\}$.

We assume that the vector of direct estimates $\boldsymbol{y}_d$ of length m is fully observed if $d \in \mathcal{D}^{\text{obs}}$ and that at least one of the m direct estimates $\boldsymbol{y}_d$ is missing if $d \in \mathcal{D}^{\text{mis}}$. Further, $\boldsymbol{V}_{ed}$, with diagonal elements $\sigma_{edk} > 0$, $k = 1, \dots, m$, is assumed to be known for $d \in \mathcal{D}^{\text{obs}}$, but is missing for $d \in \mathcal{D}^{\text{mis}}$. For the subset $\mathcal{D}^{\text{obs}}$, we consider the vectors and matrices

$$\boldsymbol{X}^{\text{obs}} = \underset{1 \le d \le D^{\text{obs}}}{\text{col}}(\boldsymbol{X}_d), \quad \boldsymbol{V}^{\text{obs}} = \underset{1 \le d \le D^{\text{obs}}}{\text{diag}}(\boldsymbol{V}_{ud} + \boldsymbol{V}_{ed}). \tag{6.249}$$

The MFH synthetic predictor is given by

$$\hat{\boldsymbol{\mu}}_d^{\text{syn}} = \boldsymbol{X}_d\hat{\boldsymbol{\beta}}, \quad d = 1, \dots, D, \tag{6.250}$$

where $\hat{\boldsymbol{\beta}}$ is estimated by using the data from $\mathcal{D}^{\text{obs}}$.

So far, there exist only vague MSE estimators for the MFH synthetic predictor. In Burgard et al. (2021c, Appendix 5), we presented an MSE approximation of the (univariate) FH synthetic predictor, Morales et al. (2021, pp. 441–442) extended the approximation to the bivariate FH model. Here, we extend the proposed approximation to the multivariate case.

If $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$, both estimated using the data from $\mathcal{D}^{\text{obs}}$, are asymptotically consistent and independent estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, as the ML and REML estimators are, then the mean and variance of $\hat{\boldsymbol{\mu}}_d^{\text{syn}}$ are

$$\begin{aligned} \mathrm{E}[\hat{\boldsymbol{\mu}}_d^{\text{syn}}] &\approx \boldsymbol{X}_d\boldsymbol{\beta}, && (6.251)\\ \mathrm{Var}(\hat{\boldsymbol{\mu}}_d^{\text{syn}}) &= \boldsymbol{X}_d \,\mathrm{Var}(\hat{\boldsymbol{\beta}})\boldsymbol{X}_d^\top \approx \boldsymbol{X}_d \left[(\boldsymbol{X}^{\text{obs}})^\top(\boldsymbol{V}^{\text{obs}})^{-1}\boldsymbol{X}^{\text{obs}}\right]^{-1}\boldsymbol{X}_d^\top && (6.252)\\ & d = 1,\dots,D. \end{aligned}$$

As $\boldsymbol{\mu}_d = \boldsymbol{X}_d\boldsymbol{\beta} + \boldsymbol{u}_d$, the MSE of $\hat{\boldsymbol{\mu}}_d^{\text{syn}} = \boldsymbol{X}_d\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}})$, is

$$\begin{aligned} \mathrm{MSE}(\hat{\boldsymbol{\mu}}_d^{\text{syn}}) &= \mathrm{E}\left[(\hat{\boldsymbol{\mu}}_d^{\text{syn}} - \boldsymbol{\mu}_d)^2\right] = \mathrm{E}\left[(\boldsymbol{X}_d\hat{\boldsymbol{\beta}} - \boldsymbol{X}_d\boldsymbol{\beta} - \boldsymbol{u}_d)^2\right]\\ &= \mathrm{E}\left[(\boldsymbol{X}_d(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \boldsymbol{u}_d)^2\right]\\ &= \boldsymbol{X}_d\,\mathrm{E}\left[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top\right]\boldsymbol{X}_d^\top + \mathrm{E}[\boldsymbol{u}_d^2] - 2\,\mathrm{E}\left[\boldsymbol{X}_d(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\boldsymbol{u}_d\right] \qquad (6.253)\\ &\approx \boldsymbol{X}_d\left[(\boldsymbol{X}^{\text{obs}})^\top(\boldsymbol{V}^{\text{obs}}(\hat{\boldsymbol{\theta}}))^{-1}\boldsymbol{X}^{\text{obs}}\right]^{-1}\boldsymbol{X}_d^\top + \boldsymbol{V}_{ud} - 2\,\mathrm{E}\left[\boldsymbol{X}_d\hat{\boldsymbol{\beta}}\boldsymbol{u}_d\right]\\ & d = 1,\dots,D. \end{aligned}$$

If $d \in \mathcal{D}^{\text{mis}}$, then $\boldsymbol{u}_d$ and $\hat{\boldsymbol{\beta}}$ are independent, so that $\mathrm{E}[\boldsymbol{X}_d\hat{\boldsymbol{\beta}}\boldsymbol{u}_d] = \boldsymbol{X}_d\,\mathrm{E}[\hat{\boldsymbol{\beta}}]\,\mathrm{E}[\boldsymbol{u}_d] = 0$ and

$$\mathrm{MSE}(\hat{\boldsymbol{\mu}}_d^{\text{syn}}) \approx \boldsymbol{X}_d\left[(\boldsymbol{X}^{\text{obs}})^\top(\boldsymbol{V}^{\text{obs}}(\hat{\boldsymbol{\theta}}))^{-1}\boldsymbol{X}^{\text{obs}}\right]^{-1}\boldsymbol{X}_d^\top + \boldsymbol{V}_{ud}, \quad \forall d \in \mathcal{D}^{\text{mis}}. \qquad (6.254)$$

An estimator is given by substituting $\boldsymbol{V}_{ud}$ by $\hat{\boldsymbol{V}}_{ud} = \boldsymbol{V}_{ud}(\hat{\boldsymbol{\theta}})$ resulting in

$$\widehat{\mathrm{MSE}}(\hat{\boldsymbol{\mu}}_d^{\text{syn}}) = \boldsymbol{X}_d\left[(\boldsymbol{X}^{\text{obs}})^\top(\boldsymbol{V}^{\text{obs}}(\hat{\boldsymbol{\theta}}))^{-1}\boldsymbol{X}^{\text{obs}}\right]^{-1}\boldsymbol{X}_d^\top + \hat{\boldsymbol{V}}_{ud}, \quad \forall d \in \mathcal{D}^{\text{mis}}. \qquad (6.255)$$

In the simulation presented in Section 6.7, we evaluate the proposed MSE estimator.

6.3 Model

MFH model
In the following, we introduce the MFH model with partial missing direct estimates (MMFH) based on the MFH model. For this purpose, we briefly repeat the different quantities of the MFH model, which is described in more detail in Section 5.3.1.

Let U be a finite population which can be partitioned into D domains $U_1, \dots, U_D$, $\boldsymbol{\mu}_d = (\mu_{d1}, \dots, \mu_{dm})^\top$ be a vector of m characteristics of interest in domain d, and $\boldsymbol{y}_d = (y_{d1}, \dots, y_{dm})^\top$ be a vector of the corresponding m direct estimates of $\boldsymbol{\mu}_d$, calculated by using the data of the target survey sample, $d = 1, \dots, D$.

The MFH model can be expressed as a single model in the form

$$\boldsymbol{y}_d = \boldsymbol{X}_d\boldsymbol{\beta} + \boldsymbol{u}_d + \boldsymbol{e}_d, \quad d = 1, \dots, D, \tag{5.177}$$

or in the matrix form

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u} + \boldsymbol{e}, \tag{5.178}$$

with

$$\begin{aligned} \boldsymbol{y} &= \underset{1\le d\le D}{\operatorname{col}}(\boldsymbol{y}_d) \in \mathbb{R}^{mD}, \quad \boldsymbol{u} = \underset{1\le d\le D}{\operatorname{col}}(\boldsymbol{u}_d) \in \mathbb{R}^{mD}, \\ \boldsymbol{e} &= \underset{1\le d\le D}{\operatorname{col}}(\boldsymbol{e}_d) \in \mathbb{R}^{mD}, \quad \boldsymbol{X} = \underset{1\le d\le D}{\operatorname{col}}(\boldsymbol{X}_d) \in \mathbb{R}^{mD\times p}. \end{aligned} \tag{6.256}$$

The quantities appearing in 5.177 and 5.178 are described in the following for domains $d = 1, \dots, D$ and variables of interest $k = 1, \dots, m$.

Sampling errors $\boldsymbol{e}_d = (e_{d1}, \dots, e_{dm})^\top \sim N_m(\boldsymbol{0}, \boldsymbol{V}_{ed})$ are independent with known covariance matrices $\boldsymbol{V}_{ed} \in \mathbb{R}^{m\times m}$, given by

$$\boldsymbol{V}_{ed} = \begin{pmatrix} \sigma^2_{ed1} & \sigma_{ed12} & \cdots & \sigma_{ed1m} \\ \sigma_{ed12} & \sigma^2_{ed2} & \cdots & \sigma_{ed2m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{ed1m} & \sigma_{ed2m} & \cdots & \sigma^2_{edm} \end{pmatrix}. \tag{6.257}$$

There are p_k explanatory variables for variable k, $k = 1, \dots, m$. The total number of explanatory variables for all m target variables is given by $p = \sum_{k=1}^m p_k$. The domain-specific aggregates of the p_k explanatory variables of μ_{dk} are given by $\boldsymbol{x}_{dk} = (x_{dk1}, \dots, x_{dkp_k})^\top$. For every domain d, we can combine the auxiliary information of the m dependent variables into a $m \times p$ block-diagonal auxiliary matrix $\boldsymbol{X}_d = \operatorname{diag}\left(\boldsymbol{x}_{d1}^\top, \dots, \boldsymbol{x}_{dm}^\top\right)$, which is assumed to be of full rank. Let $\boldsymbol{\beta}_k = (\beta_{k1}, \dots, \beta_{kp_k})^\top \in \mathbb{R}^{p_k}$ contain the regression parameters for μ_{dk} and let $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \dots, \boldsymbol{\beta}_m^\top)^\top \in \mathbb{R}^p$ be the vector of fixed effects for all m characteristics.

Random effects $\boldsymbol{u}_d$ are given by $\boldsymbol{u}_d = (u_{d1}, \dots, u_{dm})^\top \sim N_m(\boldsymbol{0}, \boldsymbol{V}_{ud})$. Random effects $\boldsymbol{u}_d$ and sampling errors $\boldsymbol{e}_d$ are assumed to be independent. The covariance matrix of the random effects $\boldsymbol{V}_{ud} \in \mathbb{R}^{m\times m}$ depends on $q = m\,(m+1)\,/2$ variance parameters, consisting of m variances and $m\,(m-1)\,/2$ covariances. It is given by

$$\boldsymbol{V}_{ud} = \begin{pmatrix} \sigma^2_{u1} & \rho_{12}\sigma_{u1}\sigma_{u2} & \cdots & \rho_{1m}\sigma_{u1}\sigma_{um} \\ \rho_{12}\sigma_{u1}\sigma_{u2} & \sigma^2_{u2} & \cdots & \rho_{2m}\sigma_{u2}\sigma_{um} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1m}\sigma_{u1}\sigma_{um} & \rho_{2m}\sigma_{u2}\sigma_{um} & \cdots & \sigma^2_{um} \end{pmatrix}, \tag{5.175}$$

where σ^2_{ua} denotes the random effects variance of variable a and ρ_{ab} denotes the random effects correlation of variables a and b, $a, b = 1, \ldots, m$, $a \neq b$. The vector of variance parameters is denoted by

$$\boldsymbol{\theta} = (\sigma^2_{u1}, \sigma^2_{u2}, \ldots, \sigma^2_{um}, \rho_{12}, \rho_{13}, \ldots, \rho_{23}, \rho_{24}, \ldots, \rho_{m-1,m})^\top \in \mathbb{R}^q. \tag{6.258}$$

The first m elements of $\boldsymbol{\theta}$ refer to the random effect variances, the $q - m$ last elements refer to the random effect correlations.

Under model (5.177), it holds that $\boldsymbol{y} \sim N_{mD}(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{V})$ with covariance matrix $\boldsymbol{V} = \boldsymbol{V}_u + \boldsymbol{V}_e = \underset{1 \leq d \leq D}{\operatorname{diag}}(\boldsymbol{V}_d) \in \mathbb{R}^{mD \times mD}$, where

$$\boldsymbol{V}_u = \underset{1 \leq d \leq D}{\operatorname{diag}}(\boldsymbol{V}_{ud}), \quad \boldsymbol{V}_e = \underset{1 \leq d \leq D}{\operatorname{diag}}(\boldsymbol{V}_{ed}), \quad \boldsymbol{V}_d = \boldsymbol{V}_{ud} + \boldsymbol{V}_{ed}, \quad d = 1, \ldots, D. \tag{6.259}$$

Additional quantities for missing direct estimates

Let us assume that some entries of the $\boldsymbol{y}_d$ are missing. For $d = 1, \ldots, D$, $k = 1, \ldots, m$, we introduce the additional quantities

$$\lambda_{dk} = \begin{cases} 1, & \text{if } y_{dk} \text{ is observed,} \\ 0, & \text{otherwise,} \end{cases} \qquad \boldsymbol{\lambda}_d = (\lambda_{d1}, \ldots, \lambda_{dm})^\top, \tag{6.260}$$

and sets

$$\mathcal{A}_d = \Big\{k \in \mathbb{N} : 1 \leq k \leq m, \lambda_{dk} = 1\Big\}, \tag{6.261}$$

to indicate which of the m variables are observed in domain d.

To account for arbitrary missing structures in the domains, we define domain-specific sub-vectors and sub-matrices of certain quantities of the MFH model representing the structure of the observed variables of interest in domain $d = 1, \ldots, D$. For any quantity $\boldsymbol{r}_d$, for example a vector, we define quantity $\breve{\boldsymbol{r}}_d$ which corresponds to $\boldsymbol{r}_d$ reduced to the observed variables of interest in domain d.

Define $\breve{m}_d = \sum_{k=1}^m \lambda_{dk} = \#\big(\mathcal{A}_d\big)$, representing the number of observed variables of interest in each domain. Let $\boldsymbol{\Lambda}_d$ be the diagonal matrix of $\boldsymbol{\lambda}_d$ where all rows with row sum equal to zero are deleted, such that $\boldsymbol{\Lambda}_d \in \mathbb{R}^{\breve{m}_d \times m}$. Let $\breve{\boldsymbol{y}}_d = \boldsymbol{\Lambda}_d \boldsymbol{y}_d \in \mathbb{R}^{\breve{m}_d}$ containing the observed variables of interest in domain d. We also consider domain-specific auxiliary variables referring to the observed variables of interest. For this sake, we define $\breve{p}_d = \sum_{k=1}^m p_k \lambda_{dk}$ and $\breve{\boldsymbol{\beta}}_d = \underset{k \in \mathcal{A}_d}{\operatorname{col}}(\boldsymbol{\beta}_k)$, such that $\breve{\boldsymbol{\beta}}_d \in \mathbb{R}^{\breve{p}_d}$. Further, let $\breve{\boldsymbol{X}}_d$ be a block-diagonal matrix of those $\boldsymbol{x}_{dk}$ for which $\lambda_{dk} = 1$, such that $\breve{\boldsymbol{X}}_d \in \mathbb{R}^{\breve{m}_d \times \breve{p}_d}$.

In a similar manner, let the domain-specific random effects of the observed variables be denoted by $\breve{\boldsymbol{u}}_d = \boldsymbol{\Lambda}_d \boldsymbol{u}_d \in \mathbb{R}^{\breve{m}_d}$. The corresponding domain-specific covariance matrices of the random terms are defined by $\breve{\boldsymbol{V}}_{ed} = \boldsymbol{\Lambda}_d \boldsymbol{V}_{ed} \boldsymbol{\Lambda}_d^\top \in \mathbb{R}^{\breve{m}_d \times \breve{m}_d}$ and $\breve{\boldsymbol{V}}_{ud} = \boldsymbol{\Lambda}_d \boldsymbol{V}_{ud} \boldsymbol{\Lambda}_d^\top \in \mathbb{R}^{\breve{m}_d \times \breve{m}_d}$, respectively, such that $\breve{\boldsymbol{V}}_d = \breve{\boldsymbol{V}}_{ud} + \breve{\boldsymbol{V}}_{ed}$. We furthermore define

$$\mathcal{Q}_d = \Big\{a \in \mathbb{N} : 1 \leq a \leq q, \, \theta_a \text{ is a parameter of } \breve{\boldsymbol{V}}_{ud}\Big\} \tag{6.262}$$

and $\breve{\boldsymbol{\theta}}_d = \operatorname*{col}_{a\in\mathcal{Q}_d}(\theta_a)$ as the vector of parameters $\boldsymbol{\theta}$ appearing in $\breve{\boldsymbol{V}}_{ud}$.

MMFH model

If the MFH model (5.177) holds for $d \in \{1,\dots,D\}$ with arbitrary missing data of interest in each domain, we say that target vectors $\boldsymbol{y}_d$ obey a *missing data MFH* (MMFH) model.

The MMFH model can be expressed as a single model in the form

$$\breve{\boldsymbol{y}}_d = \breve{\boldsymbol{X}}_d\breve{\boldsymbol{\beta}}_d + \breve{\boldsymbol{u}}_d + \breve{\boldsymbol{e}}_d, \quad d = 1,\dots,D, \tag{6.263}$$

or in the matrix form

$$\breve{\boldsymbol{y}} = \boldsymbol{\Lambda}(\boldsymbol{X}\boldsymbol{\beta}) + \breve{\boldsymbol{u}} + \breve{\boldsymbol{e}}, \tag{6.264}$$

with

$$\begin{aligned}\breve{\boldsymbol{y}} &= \operatorname*{col}_{1\le d\le D}(\breve{\boldsymbol{y}}_d) \in \mathbb{R}^{\breve{M}}, \quad & \breve{\boldsymbol{u}} &= \operatorname*{col}_{1\le d\le D}(\breve{\boldsymbol{u}}_d) \in \mathbb{R}^{\breve{M}},\\ \breve{\boldsymbol{e}} &= \operatorname*{col}_{1\le d\le D}(\breve{\boldsymbol{e}}_d) \in \mathbb{R}^{\breve{M}}, \quad & \boldsymbol{\Lambda} &= \operatorname*{col}_{1\le d\le D}\boldsymbol{\Lambda}_d \in \mathbb{R}^{\breve{M}\times m},\end{aligned} \tag{6.265}$$

$\breve{M} = \sum_{d=1}^{D}\breve{m}_d$ and $\boldsymbol{\Lambda}$ is the diagonal matrix of $(\boldsymbol{\lambda}_1^\top,\dots,\boldsymbol{\lambda}_D^\top)^\top \in \mathbb{R}^{mD}$ where all rows with row sum equal to zero are deleted, such that $\boldsymbol{\Lambda} \in \mathbb{R}^{\breve{M}\times(mD)}$.

If the MMFH model holds, then

$$\breve{\boldsymbol{y}}_d \sim N_{\breve{m}_d}(\breve{\boldsymbol{X}}_d\breve{\boldsymbol{\beta}}_d, \breve{\boldsymbol{V}}_{ud} + \breve{\boldsymbol{V}}_{ed}) \tag{6.266}$$

$$\breve{\boldsymbol{y}}_d|_{\breve{\boldsymbol{u}}_d} \sim N_{\breve{m}_d}(\breve{\boldsymbol{X}}_d\breve{\boldsymbol{\beta}}_d + \breve{\boldsymbol{u}}_d, \breve{\boldsymbol{V}}_{ed}), \quad d = 1,\dots,D. \tag{6.267}$$

Note that the (univariate) FH model (2.83), the MFH model (5.177), and the bivariate FH model under missing direct estimates, proposed in Burgard et al. (2021c), are special cases of the above formulation.

6.4 Prediction

Let $\mathcal{D} = \{1,\dots,D\}$ contain all domains of interest. Without loss of generality, we reorder the domains such that we can partition the set of domains $\mathcal{D}$ into the two subsets:

> $\mathcal{D}^{\text{mis}} = \{1,\dots,D^{\text{mis}}\}$ contains the $D^{\text{mis}} \le D$ domains where $1 \le \breve{m}_d < m$ and thus at least one variable of interest is not observed. We assume that at least one of the m variables is observed in each domain.
>
> $\mathcal{D}^{\text{obs}} = \{D^{\text{mis}}+1,\dots,D\}$ contains the $D^{\text{obs}} = D - D^{\text{mis}}$ domains where $\breve{m}_d = m$ and thus all m variables of interest are observed.

In a situation where the target data follows a MMFH model, the MFH model is strictly applicable only to $\mathcal{D}^{\text{obs}}$, but not to $\mathcal{D}^{\text{mis}}$. For example, under the MFH model we can only calculate EBLUPs of $\boldsymbol{\mu}_d$ or $\boldsymbol{u}_d$ for $d \in \mathcal{D}^{\text{obs}}$. However, in what follows we show that it is possible calculate EBLUPs for $d \in \mathcal{D}^{\text{mis}}$ under the MMFH model. The only requirements for that are: At least one of the m direct estimates has to be observed in each domain $d \in \mathcal{D}^{\text{mis}}$ and there have to be enough domains $\mathcal{D}$ with combinations of the m direct estimates available such that their random effect correlations can be estimated.

In order to derive EBLUPs for domains with arbitrary missing structures of the target variables, we introduce additional domain-specific quantities for $d = 1, \dots, D$ which are of same size for all domains, but where certain entries are set to zero depending on the observed variables of interest. For any quantity $\boldsymbol{r}_d$, for example a vector, we define quantity $\acute{\boldsymbol{r}}_d$ corresponding to $\boldsymbol{r}_d$, where all elements referring to missing direct estimates are set to zero.

We define $\acute{\boldsymbol{y}}_d = \boldsymbol{\Lambda}_d^\top \boldsymbol{\Lambda}_d \boldsymbol{y}_d = \boldsymbol{\Lambda}_d^\top \breve{\boldsymbol{y}} \in \mathbb{R}^m$ such that $\acute{\boldsymbol{y}}_d$ is a vector containing zeros where direct estimates $\boldsymbol{y}_d$ are missing. Let $\acute{\boldsymbol{V}}_{ed} = \boldsymbol{\Lambda}_d^\top \breve{\boldsymbol{V}}_{ed} \boldsymbol{\Lambda}_d \in \mathbb{R}^{m \times m}$ and $\acute{\boldsymbol{V}}_{ed}^{\text{inv}} = \boldsymbol{\Lambda}_d^\top \breve{\boldsymbol{V}}_{ed}^{-1} \boldsymbol{\Lambda}_d \in \mathbb{R}^{m \times m}$, be matrices containing the entries of $\breve{\boldsymbol{V}}_{ed}$ and $\breve{\boldsymbol{V}}_{ed}^{-1}$ for the variables observed in domain d and zeros else.

Proposition 6.1. The conditional distribution of $\boldsymbol{u}_d$ given $\boldsymbol{y}_d$ under the MMFH model is multivariate normal with with mean vector and variance matrix

$$\begin{aligned} \mathrm{E}[\boldsymbol{u}_d | \breve{\boldsymbol{y}}_d] &= \mathrm{Var}(\boldsymbol{u}_d | \breve{\boldsymbol{y}}_d; \boldsymbol{\beta}, \boldsymbol{\theta}) = \boldsymbol{\Phi}_d \acute{\boldsymbol{V}}_{ed}^{\text{inv}} (\acute{\boldsymbol{y}}_d - \boldsymbol{X}_d \boldsymbol{\beta}) && (6.268) \\ &= (\acute{\boldsymbol{V}}_{ed}^{\text{inv}} + \boldsymbol{V}_{ud}^{-1})^{-1} \acute{\boldsymbol{V}}_{ed}^{\text{inv}} (\acute{\boldsymbol{y}}_d - \boldsymbol{X}_d \boldsymbol{\beta}), \\ \boldsymbol{\Phi}_d &= \mathrm{Var}(\boldsymbol{u}_d | \breve{\boldsymbol{y}}_d; \boldsymbol{\beta}, \boldsymbol{\theta}) = (\acute{\boldsymbol{V}}_{ed}^{\text{inv}} + \boldsymbol{V}_{ud}^{-1})^{-1}, \quad d = 1, \dots, D. && (6.269) \end{aligned}$$

Proof. We recall that the kernel of the m-variate normal probability density function for variables $\tilde{Y}_1, \dots, \tilde{Y}_m$ with mean $\tilde{\boldsymbol{\mu}}$ and covariance matrix $\tilde{\boldsymbol{\Sigma}}$ is

$$\begin{aligned} f(\tilde{\boldsymbol{y}} | \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}) &= \frac{1}{(2\pi)^{m/2} \det\left(\tilde{\boldsymbol{\Sigma}}\right)^{1/2}} \exp\Big\{ -\frac{1}{2} (\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{\mu}})^\top \tilde{\boldsymbol{\Sigma}}^{-1} (\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{\mu}}) \Big\} \\ &\propto \exp\Big\{ -\frac{1}{2} \tilde{\boldsymbol{y}}^\top \tilde{\boldsymbol{\Sigma}}^{-1} \tilde{\boldsymbol{y}} + \tilde{\boldsymbol{\mu}}^\top \tilde{\boldsymbol{\Sigma}}^{-1} \tilde{\boldsymbol{y}} \Big\}. \end{aligned} \qquad (6.270)$$

The conditional distribution of $\boldsymbol{u}_d$ given $\breve{\boldsymbol{y}}_d$ is then given by

$$\begin{aligned} & f(\boldsymbol{u}_d | \breve{\boldsymbol{y}}_d) \\ &\quad \propto f(\breve{\boldsymbol{y}}_d | \boldsymbol{u}_d) f(\boldsymbol{u}_d) = f(\breve{\boldsymbol{y}}_d | \breve{\boldsymbol{u}}_d) f(\boldsymbol{u}_d) \\ &\quad = \frac{1}{(2\pi)^{\breve{m}_d/2} \det\left(\breve{\boldsymbol{V}}_{ed}\right)^{1/2}} \exp\Big\{ -\frac{1}{2} (\breve{\boldsymbol{y}}_d - \breve{\boldsymbol{X}}_d \breve{\boldsymbol{\beta}}_d - \breve{\boldsymbol{u}}_d)^\top \breve{\boldsymbol{V}}_{ed}^{-1} (\breve{\boldsymbol{y}}_d - \breve{\boldsymbol{X}}_d \breve{\boldsymbol{\beta}}_d - \breve{\boldsymbol{u}}_d) \Big\} \\ &\qquad \cdot \frac{1}{(2\pi)^{m_d/2} \det(\boldsymbol{V}_{ud})^{1/2}} \exp\Big\{ -\frac{1}{2} \boldsymbol{u}_d^\top \boldsymbol{V}_{ud}^{-1} \boldsymbol{u}_d \Big\} \\ &\quad \propto \exp\Big\{ -\frac{1}{2} \breve{\boldsymbol{u}}_d^\top \breve{\boldsymbol{V}}_{ed}^{-1} \breve{\boldsymbol{u}}_d + \breve{\boldsymbol{u}}_d^\top \breve{\boldsymbol{V}}_{ed}^{-1} (\breve{\boldsymbol{y}}_d - \breve{\boldsymbol{X}}_d \breve{\boldsymbol{\beta}}_d) - \frac{1}{2} \boldsymbol{u}_d^\top \boldsymbol{V}_{ud}^{-1} \boldsymbol{u}_d \Big\} \\ &\quad = \exp\Big\{ -\frac{1}{2} \boldsymbol{u}_d^\top \left(\boldsymbol{\Lambda}_d^\top \breve{\boldsymbol{V}}_{ed}^{-1} \boldsymbol{\Lambda}_d + \boldsymbol{V}_{ud}^{-1} \right) \boldsymbol{u}_d + \boldsymbol{u}_d^\top \boldsymbol{\Lambda}_d^\top \breve{\boldsymbol{V}}_{ed}^{-1} \boldsymbol{\Lambda}_d (\acute{\boldsymbol{y}}_d - \boldsymbol{X}_d \boldsymbol{\beta}) \Big\} \\ &\quad = \exp\Big\{ -\frac{1}{2} \boldsymbol{u}_d^\top (\acute{\boldsymbol{V}}_{ed}^{\text{inv}} + \boldsymbol{V}_{ud}^{-1}) \boldsymbol{u}_d + \boldsymbol{u}_d^\top \boldsymbol{\Phi}_d^{-1} \left(\boldsymbol{\Phi}_d \acute{\boldsymbol{V}}_{ed}^{\text{inv}} (\acute{\boldsymbol{y}}_d - \boldsymbol{X}_d \boldsymbol{\beta}) \right) \Big\}. \end{aligned} \qquad (6.271)$$

Therefore, $f(\boldsymbol{u}_d|\breve{\boldsymbol{y}}_d)$ is a multivariate normal distribution with parameters

$$\mathrm{E}[\boldsymbol{u}_d|\breve{\boldsymbol{y}}_d] = \boldsymbol{\Phi}_d \acute{\boldsymbol{V}}_{ed}^{\mathrm{inv}}(\acute{\boldsymbol{y}}_d - \boldsymbol{X}_d\boldsymbol{\beta}) \tag{6.272}$$

$$\mathrm{Var}(\boldsymbol{u}_d|\breve{\boldsymbol{y}}_d) = \left(\acute{\boldsymbol{V}}_{ed}^{\mathrm{inv}} + \boldsymbol{V}_{ud}^{-1}\right)^{-1} = \boldsymbol{\Phi}_d, \quad d = 1, \dots, D. \tag{6.273}$$

□

If $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are known, the best predictors (BPs) of $\boldsymbol{u}_d$ and $\boldsymbol{\mu}_d$ are

$$\hat{\boldsymbol{u}}_d^{\mathrm{BP}} = \mathrm{E}\left[\boldsymbol{u}_d|\breve{\boldsymbol{y}}_d; \boldsymbol{\beta}, \boldsymbol{\theta}\right] = \boldsymbol{\Phi}_d \acute{\boldsymbol{V}}_{ed}^{\mathrm{inv}}(\acute{\boldsymbol{y}}_d - \boldsymbol{X}_d\boldsymbol{\beta}), \tag{6.274}$$

$$\hat{\boldsymbol{\mu}}_d^{\mathrm{BP}} = \boldsymbol{X}\boldsymbol{\beta} + \hat{\boldsymbol{u}}^{\mathrm{BP}}, \quad d = 1, \dots, D. \tag{6.275}$$

If $\boldsymbol{\theta}$ is known and $\boldsymbol{\beta}$ is unknown, the best linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$ and the best linear unbiased predictor (BLUP) of $\boldsymbol{u}$ and $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$ are

$$\hat{\boldsymbol{\beta}}^{\mathrm{BLUE}} = (\breve{\boldsymbol{X}}^\top \breve{\boldsymbol{V}}^{-1}(\boldsymbol{\theta})\breve{\boldsymbol{X}})^{-1}\breve{\boldsymbol{X}}^\top \breve{\boldsymbol{V}}^{-1}(\boldsymbol{\theta})\breve{\boldsymbol{y}}, \tag{6.276}$$

$$\hat{\boldsymbol{u}}_d^{\mathrm{BLUP}} = \boldsymbol{\Phi}_d\acute{\boldsymbol{V}}_{ed}^{\mathrm{inv}}(\acute{\boldsymbol{y}}_d - \boldsymbol{X}_d\hat{\boldsymbol{\beta}}^{\mathrm{BLUE}}) = \boldsymbol{\Phi}_d\acute{\boldsymbol{V}}_{ed}^{\mathrm{inv}}(\boldsymbol{\Lambda}_d^\top\breve{\boldsymbol{y}}_d - \boldsymbol{X}_d\hat{\boldsymbol{\beta}}^{\mathrm{BLUE}}), \tag{6.277}$$

$$\hat{\boldsymbol{\mu}}_d^{\mathrm{BLUP}} = \boldsymbol{X}_d\hat{\boldsymbol{\beta}}^{\mathrm{BLUE}} + \hat{\boldsymbol{u}}_d^{\mathrm{BLUP}}, \quad d = 1, \dots, D, \tag{6.278}$$

where $\breve{\boldsymbol{y}} = (\breve{\boldsymbol{y}}_1^\top, \dots, \breve{\boldsymbol{y}}_D^\top)^\top \in \mathbb{R}^{\breve{M}}$, $\breve{\boldsymbol{V}} \in \mathbb{R}^{\breve{M}\times\breve{M}}$ is the block-diagonal covariance matrix of $\breve{\boldsymbol{y}}$ with blocks of size $\breve{m}_d \times \breve{m}_d$, and $\breve{\boldsymbol{X}} = (\breve{\boldsymbol{X}}_1, \dots, \breve{\boldsymbol{X}}_D)^\top \in \mathbb{R}^{\breve{M}\times p}$, $d = 1, \dots, D$.

Note that the formulas are very close to the formulas for the MFH model. For the MFH model, the derivations of the BLUPs and BLUE are listed in detail for example in Morales et al. (2021, Section 16.2). Compared to the MFH model, what changes in the MMFH model is simply that different quantities of the model are adjusted to represent only those entries which refer to be observed values of $\boldsymbol{y}$. Furthermore, the linearity in the BLUE and BLUP formulas in the MMFH model is with respect to the observed values $\breve{\boldsymbol{y}}$.

By substituting $\boldsymbol{\theta}$ by an estimator $\hat{\boldsymbol{\theta}}$, we obtain the empirical BLUE (EBLUE) of $\boldsymbol{\beta}$ and the empirical BLUP (EBLUP) of $\boldsymbol{u}_d$ and $\boldsymbol{\mu}_d = \boldsymbol{X}_d\boldsymbol{\beta}_d + \boldsymbol{u}_d$, i.e.

$$\hat{\boldsymbol{\beta}} = (\breve{\boldsymbol{X}}^\top \breve{\boldsymbol{V}}^{-1}(\hat{\boldsymbol{\theta}})\breve{\boldsymbol{X}})^{-1}\breve{\boldsymbol{X}}^\top \breve{\boldsymbol{V}}^{-1}(\hat{\boldsymbol{\theta}})\breve{\boldsymbol{y}}, \tag{6.279}$$

$$\hat{\boldsymbol{u}}_d^{\mathrm{EBLUP}} = \boldsymbol{\Phi}_d\acute{\boldsymbol{V}}_{ed}^{\mathrm{inv}}(\acute{\boldsymbol{y}}_d - \boldsymbol{X}_d\hat{\boldsymbol{\beta}}) = \boldsymbol{\Phi}_d\acute{\boldsymbol{V}}_{ed}^{\mathrm{inv}}(\boldsymbol{\Lambda}_d^\top\breve{\boldsymbol{y}}_d - \boldsymbol{X}_d\hat{\boldsymbol{\beta}}), \tag{6.280}$$

$$\hat{\boldsymbol{\mu}}_d^{\mathrm{EBLUP}} = \boldsymbol{X}_d\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{u}}_d^{\mathrm{EBLUP}}, \quad d = 1, \dots, D. \tag{6.281}$$

Examples

To illustrate Proposition 6.1 and the BLUP of $\boldsymbol{u}_d$ (6.277) under the MMFH model, we consider three data examples. For that, we need some additional notation. For an arbitrary matrix $\tilde{\boldsymbol{M}} = \left(\tilde{m}_{ij}\right) \in \mathbb{R}^{\tilde{a}\times\tilde{b}}$, we use $(\tilde{\boldsymbol{M}})_{i,j}$, $(\tilde{\boldsymbol{M}})_{i,*}$, and $(\tilde{\boldsymbol{M}})_{*,j}$ to indicate entry $\tilde{m}_{ij}$, the i-th row vector $(\tilde{m}_{i1}, \dots, \tilde{m}_{i\tilde{b}})$, and the j-th column vector $(\tilde{m}_{1j}, \dots, \tilde{m}_{\tilde{a}j})$, respectively, $i = 1, \dots, \tilde{a}$, $j = 1, \dots, \tilde{b}$.

Example 6.1. Domain with no missing direct estimates.
If there are no missing direct estimates for a domain d, we have $\breve{m}_d = m$, $\breve{\boldsymbol{y}}_d = \acute{\boldsymbol{y}}_d = \boldsymbol{y}_d \in \mathbb{R}^m$ and $\acute{\boldsymbol{V}}_{ed}^{\text{inv}} = \boldsymbol{V}_{ed}^{-1} \in \mathbb{R}^{m \times m}$.

The BLUP of $\boldsymbol{u}_d$ under the MMFH model is equivalent to the BLUP under the MFH model, i.e.

$$\hat{\boldsymbol{u}}_d^{\text{BLUP}} = \text{E}[\boldsymbol{u}_d|\boldsymbol{y}_d] = \boldsymbol{\Phi}_d \boldsymbol{V}_{ed}^{-1}(\boldsymbol{y}_d - \boldsymbol{X}_d\hat{\boldsymbol{\beta}}^{\text{BLUE}}), \quad \boldsymbol{\Phi}_d = (\boldsymbol{V}_{ed}^{-1} + \boldsymbol{V}_{ud}^{-1})^{-1}. \tag{6.282}$$

Example 6.2. Domain with a direct estimate only for the first variable of interest.
If only for the first variable of interest a direct estimate is observed for a domain d, we have $\breve{m}_d = 1$, $\breve{\boldsymbol{y}}_d = y_{d1} \in \mathbb{R}$, $\acute{\boldsymbol{y}}_d = (y_{d1}, 0, \ldots, 0)^\top \in \mathbb{R}^m$, $\breve{\boldsymbol{V}}_{ed} = (\boldsymbol{V}_{ed})_{1,1} = \sigma_{ed1}^2 \in \mathbb{R}$.

The BLUP of $\boldsymbol{u}_d$ under the MMFH model is then given by

$$\hat{\boldsymbol{u}}_d^{\text{BLUP}} = \text{E}[\boldsymbol{u}_d|y_{d1}] = \boldsymbol{\Phi}_d \acute{\boldsymbol{V}}_{ed}^{\text{inv}}\Big((y_{d1}, 0, \ldots, 0)^\top - \boldsymbol{X}_d\hat{\boldsymbol{\beta}}^{\text{BLUE}}\Big), \tag{6.283}$$

where

$$\acute{\boldsymbol{V}}_{ed}^{\text{inv}} = \begin{pmatrix} \sigma_{ed1}^{-2} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \in \mathbb{R}^{m \times m}, \quad \boldsymbol{\Phi}_d = \left(\begin{pmatrix} \sigma_{ed1}^{-2} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} + \boldsymbol{V}_{ud}^{-1} \right)^{-1}. \tag{6.284}$$

Example 6.3. Domain with a direct estimate only for the first and third variable of interest.
If only for the first and third variable of interest a direct estimate is observed for a domain d, we have $\breve{m}_d = 2$, $\breve{\boldsymbol{y}}_d = (y_{d1}, y_{d3})^\top \in \mathbb{R}^2$, $\acute{\boldsymbol{y}}_d = (y_{d1}, 0, y_{d3}, \ldots, 0)^\top \in \mathbb{R}^m$,

$$\breve{\boldsymbol{V}}_{ed} = \begin{pmatrix} \sigma_{e_{d_1}}^2 & \sigma_{e_{d_1}}\sigma_{e_{d_3}} \\ \sigma_{e_{d_1}}\sigma_{e_{d_3}} & \sigma_{e_{d_3}}^2 \end{pmatrix} \in \mathbb{R}^{2\times 2}, \quad \acute{\boldsymbol{V}}_{ed}^{\text{inv}} = \begin{pmatrix} (\breve{\boldsymbol{V}}_{ed}^{-1})_{1,1} & 0 & (\breve{\boldsymbol{V}}_{ed}^{-1})_{1,3} & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ (\breve{\boldsymbol{V}}_{ed}^{-1})_{1,3} & 0 & (\breve{\boldsymbol{V}}_{ed}^{-1})_{3,3} & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix} \in \mathbb{R}^{m\times m}. \tag{6.285}$$

The BLUP of $\boldsymbol{u}_d$ under the MMFH model is then given by

$$\hat{\boldsymbol{u}}_d^{\text{BLUP}} = \text{E}[\boldsymbol{u}_d|(y_{d1}, y_{d3})^\top] = \boldsymbol{\Phi}_d \acute{\boldsymbol{V}}_{ed}^{\text{inv}}\Big((y_{d1}, 0, y_{d3}, \ldots, 0)^\top - \boldsymbol{X}_d\hat{\boldsymbol{\beta}}^{\text{BLUE}}\Big), \tag{6.286}$$

where

$$\boldsymbol{\Phi}_d = \left(\begin{pmatrix} (\breve{\boldsymbol{V}}_{ed}^{-1})_{1,1} & 0 & (\breve{\boldsymbol{V}}_{ed}^{-1})_{1,3} & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ (\breve{\boldsymbol{V}}_{ed}^{-1})_{1,3} & 0 & (\breve{\boldsymbol{V}}_{ed}^{-1})_{3,3} & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix} + \boldsymbol{V}_{ud}^{-1} \right)^{-1}. \tag{6.287}$$

6.5 Parameter estimation

This section presents the maximum likelihood (ML) and residual maximum likelihood (REML) method for estimating the model parameters under the introduced MMFH model. As the MFH model is a special case of the MMFH model, the algorithms can also be applied to get ML and REML parameter estimates under the MFH model. For the theory of ML and REML parameter estimation, we refer to Section 2.4.2. We proceed with the derivations in analogy to Morales et al. (2021, Section 19.4, 19.5), who give detailed derivations of ML and REML procedures for BFH models.

6.5.1 Maximum likelihood method

First and second partial derivatives of $\boldsymbol{V}_{ud}$
Before we get to the likelihood itself, we take a look at the derivatives of the random effect covariance matrix $\boldsymbol{V}_{ud}$. These are later used in the derivatives of the likelihood.

Recall that vector $\boldsymbol{\theta}$ of length $q = m\,(m+1)\,/2$ is given by (6.258), the first m elements of $\boldsymbol{\theta}$ refer to the random effect variances, and the $q-m$ last elements refer to the random effect correlations.

Let us define the subset of natural numbers $\mathbb{N}_m = \{1, 2, \dots, 0.5m(m-1)\}$ and the function $g_m : \mathbb{N}_m \mapsto \mathbb{N}^2$ such that

$$\begin{aligned}
g_m(1) &= (1,2), & g_m(2m-3) &= (2,m),\\
g_m(2) &= (1,3), & g_m(2m-2) &= (3,4),\\
&\vdots, & &\vdots\\
g_m(m-1) &= (1,m), & g_m(3m-6) &= (3,m),\\
g_m(m) &= (2,3), & &\vdots\\
&\vdots & g_m(0.5m(m-1)/2) &= (m-1,m).
\end{aligned}$$

We use function g_m to get the coordinates of the random effect correlations in covariance matrix $\boldsymbol{V}_{ud}$. For an example, take $m=3$. Then $g(1) = (1,2)$ are the column and row coordinates of ρ_{12} in $\boldsymbol{V}_{ud}$.

If $1 \leq a \leq m$, the components of the first partial derivatives of $\boldsymbol{V}_{ud}$ are

$$\left(\frac{\partial \boldsymbol{V}_{ud}}{\partial \theta_a}\right)_{i,j} = \left(\frac{\partial \boldsymbol{V}_{ud}}{\partial \theta_a}\right)_{j,i} = \begin{cases} 1 & \text{if } i=j=a,\\ (\rho_{ij}\sigma_{uj})/(2\sigma_{ui}) & \text{if } i=a, j\neq a,\\ 0 & \text{otherwise.}\end{cases} \tag{6.288}$$

If $m+1 \leq a \leq q$, the components of the first partial derivatives of $\boldsymbol{V}_{ud}$ are

$$\left(\frac{\partial \boldsymbol{V}_{ud}}{\partial \theta_a}\right)_{i,j} = \left(\frac{\partial \boldsymbol{V}_{ud}}{\partial \theta_a}\right)_{j,i} = \begin{cases} \sigma_{ui}\sigma_{uj} & \text{if } g_m(a-m) = (i,j),\\ 0 & \text{otherwise.}\end{cases} \tag{6.289}$$

If $1 \leq a \leq m$, the components of the second partial derivatives of $\boldsymbol{V}_{ud}$ are

$$\left(\frac{\partial^2 \boldsymbol{V}_{ud}}{\partial \theta_a^2}\right)_{i,j} = \left(\frac{\partial^2 \boldsymbol{V}_{ud}}{\partial \theta_a^2}\right)_{j,i} = \begin{cases} -(\rho_{ij}\sigma_{uj})/(4\sigma_{ui}^3) & \text{if } i = a, j \neq a, \\ 0 & \text{otherwise.} \end{cases} \tag{6.290}$$

If $1 \leq a \leq m$, $1 \leq b \leq m$, $a \neq b$, the components of the second partial derivatives of $\boldsymbol{V}_{ud}$ are

$$\left(\frac{\partial^2 \boldsymbol{V}_{ud}}{\partial \theta_a \partial \theta_b}\right)_{i,j} = \left(\frac{\partial^2 \boldsymbol{V}_{ud}}{\partial \theta_a \partial \theta_b}\right)_{j,i} = \begin{cases} (\rho_{ij})/(4\sigma_{ui}\sigma_{uj}) & \text{if } i = a, j = b, \\ 0 & \text{otherwise.} \end{cases} \tag{6.291}$$

If $1 \leq a \leq m$, $m+1 \leq b \leq q$, the components of the second partial derivatives of $\boldsymbol{V}_{ud}$ are

$$\left(\frac{\partial^2 \boldsymbol{V}_{ud}}{\partial \theta_a \partial \theta_b}\right)_{i,j} = \left(\frac{\partial^2 \boldsymbol{V}_{ud}}{\partial \theta_a \partial \theta_b}\right)_{j,i} = \begin{cases} \sigma_{uj}/(2\sigma_{ui}) & \text{if } g_m(b-m) = (i,j),\ i = a, \\ 0 & \text{otherwise.} \end{cases} \tag{6.292}$$

If $m+1 \leq a \leq q$, $m+1 \leq b \leq q$, the components of the second partial derivatives of $\boldsymbol{V}_{ud}$ are

$$\left(\frac{\partial^2 \boldsymbol{V}_{ud}}{\partial \theta_a \partial \theta_b}\right)_{i,j} = \left(\frac{\partial^2 \boldsymbol{V}_{ud}}{\partial \theta_a \partial \theta_b}\right)_{j,i} = 0. \tag{6.293}$$

For ease of exposition, we use the reduced matrix notation

$$\boldsymbol{V}_{uda} = \left(\frac{\partial \boldsymbol{V}_{ud}}{\partial \theta_a}\right), \quad \boldsymbol{V}_{udab} = \left(\frac{\partial^2 \boldsymbol{V}_{ud}}{\partial \theta_a \partial \theta_b}\right), \quad a, b = 1, \ldots, q. \tag{6.294}$$

Let the domain-specific derivatives of $\breve{\boldsymbol{V}}_{ud}$ be defined by $\breve{\boldsymbol{V}}_{uda} = \boldsymbol{\Lambda}_d \boldsymbol{V}_{uda} \boldsymbol{\Lambda}_d^\top \in \mathbb{R}^{\breve{m}_d \times \breve{m}_d}$ and $\breve{\boldsymbol{V}}_{udab} = \boldsymbol{\Lambda}_d \boldsymbol{V}_{udab} \boldsymbol{\Lambda}_d^\top \in \mathbb{R}^{\breve{m}_d \times \breve{m}_d}$, $d = 1, \ldots, D$, $a, b = 1, \ldots, q$. The vector of domain-specific model parameters for the observed variables of interest is $\breve{\boldsymbol{\psi}}_d = (\breve{\boldsymbol{\beta}}_d^\top, \breve{\boldsymbol{\theta}}_d^\top)^\top \in \mathbb{R}^{(\breve{p}_d + \breve{q}_d)}$, where $\breve{q}_d$ is the length of $\breve{\boldsymbol{\theta}}_d$, $d = 1, \ldots, D$.

Log-likelihood

The log-likelihood of observations $\breve{\boldsymbol{y}} \in \mathbb{R}^{\breve{M}}$ is $\ell = \sum_{d=1}^{D} \ell_d$, where

$$\ell_d = -\frac{\breve{m}_d}{2}\log(2\pi) - \frac{1}{2}\log\Big(\det\Big(\breve{\boldsymbol{V}}_d\Big)\Big) - \frac{1}{2}(\breve{\boldsymbol{y}}_d - \breve{\boldsymbol{X}}_d \breve{\boldsymbol{\beta}}_d)^\top \breve{\boldsymbol{V}}_d^{-1} (\breve{\boldsymbol{y}}_d - \breve{\boldsymbol{X}}_d \breve{\boldsymbol{\beta}}_d), \quad d = 1, \ldots, D. \tag{6.295}$$

First partial derivatives

The first partial derivatives of ℓ_d with respect to the components of $\boldsymbol{\psi} = (\boldsymbol{\beta}^\top, \boldsymbol{\theta}^\top)^\top \in \mathbb{R}^{(p+q)}$ are

$$\frac{\partial \ell_d}{\partial \breve{\boldsymbol{\beta}}_d} = \breve{\boldsymbol{X}}_d^\top \breve{\boldsymbol{V}}_d^{-1} (\breve{\boldsymbol{y}}_d - \breve{\boldsymbol{X}}_d \breve{\boldsymbol{\beta}}_d), \tag{6.296}$$

$$\frac{\partial \ell_d}{\partial \theta_a} = -\frac{1}{2}\operatorname{tr}\left(\breve{\boldsymbol{V}}_d^{-1} \breve{\boldsymbol{V}}_{uda}\right) + \frac{1}{2}(\breve{\boldsymbol{y}}_d - \breve{\boldsymbol{X}}_d \breve{\boldsymbol{\beta}}_d)^\top \breve{\boldsymbol{V}}_d^{-1} \breve{\boldsymbol{V}}_{uda} \breve{\boldsymbol{V}}_d^{-1} (\breve{\boldsymbol{y}}_d - \breve{\boldsymbol{X}}_d \breve{\boldsymbol{\beta}}_d), \quad a \in \mathcal{Q}_d. \tag{6.297}$$

The remaining first derivatives with respect to the parameters of $\boldsymbol{\psi}$ that are not in $\breve{\boldsymbol{\psi}}_d$ are equal to zero. Recall that $\mathcal{Q}_d$ is defined in (6.262).

Score vector
The components of the score vector are

$$s_t = \sum_{d=1}^{D} \frac{\partial l_d}{\partial \psi_t}, \quad t = 1, \dots, p+q. \tag{6.298}$$

The score vector is $\boldsymbol{s}(\boldsymbol{\psi}) = (\boldsymbol{s}_{\beta}^{\top}(\boldsymbol{\psi}), \boldsymbol{s}_{\boldsymbol{\theta}}^{\top}(\boldsymbol{\psi}))^{\top} \in \mathbb{R}^{(p+q)}$, where

$$\boldsymbol{s}_{\beta}(\boldsymbol{\psi}) = (s_1, \dots, s_p)^{\top}, \quad \boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{\psi})^{\top} = (s_{p+1}, \dots, s_{p+q})^{\top}. \tag{6.299}$$

Second partial derivatives
The second partial derivatives of ℓ_d with respect to the components of $\boldsymbol{\psi}$ are

$$\frac{\partial \ell_d^2}{\partial \breve{\boldsymbol{\beta}}_d \partial \breve{\boldsymbol{\beta}}_d^{\top}} = -\breve{\boldsymbol{X}}_d^{\top} \breve{\boldsymbol{V}}_d^{-1} \breve{\boldsymbol{X}}_d, \tag{6.300}$$

$$\frac{\partial \ell_d^2}{\partial \breve{\boldsymbol{\beta}}_d \partial \theta_a} = -\breve{\boldsymbol{X}}_d^{\top} \breve{\boldsymbol{V}}_d^{-1} \breve{\boldsymbol{V}}_{uda} \breve{\boldsymbol{V}}_d^{-1} (\breve{\boldsymbol{y}}_d - \breve{\boldsymbol{X}}_d \breve{\boldsymbol{\beta}}_d), \quad a \in \mathcal{Q}_d, \tag{6.301}$$

$$\begin{aligned} \frac{\partial \ell_d^2}{\partial \theta_a \partial \theta_b} = {} & \frac{1}{2} \operatorname{tr}\left(\breve{\boldsymbol{V}}_d^{-1} \breve{\boldsymbol{V}}_{uda} \breve{\boldsymbol{V}}_d^{-1} \breve{\boldsymbol{V}}_{udb} \right) - \frac{1}{2} \operatorname{tr}\left(\breve{\boldsymbol{V}}_d^{-1} \breve{\boldsymbol{V}}_{udab} \right) \\ & - (\breve{\boldsymbol{y}}_d - \breve{\boldsymbol{X}}_d \breve{\boldsymbol{\beta}}_d)^{\top} \breve{\boldsymbol{V}}_d^{-1} \breve{\boldsymbol{V}}_{uda} \breve{\boldsymbol{V}}_d^{-1} \breve{\boldsymbol{V}}_{udb} \breve{\boldsymbol{V}}_d^{-1} (\breve{\boldsymbol{y}}_d - \breve{\boldsymbol{X}}_d \breve{\boldsymbol{\beta}}_d) \\ & + \frac{1}{2} (\breve{\boldsymbol{y}}_d - \breve{\boldsymbol{X}}_d \breve{\boldsymbol{\beta}}_d)^{\top} \breve{\boldsymbol{V}}_d^{-1} \breve{\boldsymbol{V}}_{udab} \breve{\boldsymbol{V}}_d^{-1} (\breve{\boldsymbol{y}}_d - \breve{\boldsymbol{X}}_d \breve{\boldsymbol{\beta}}_d), \quad a, b \in \mathcal{Q}_d. \end{aligned} \tag{6.302}$$

The remaining second derivatives with respect to parameters of $\boldsymbol{\psi}$ that are not in $\breve{\boldsymbol{\psi}}_d$ are equal to zero.

Fisher information matrix
By changing the sign and taking the expectation of the second partial derivatives, we have the Fisher information matrix $\boldsymbol{F}_d = -\operatorname{E}\left[\frac{\partial^2 \ell_d}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}^{\top}}\right]$, with block-matrix components

$$\boldsymbol{F}_{d\breve{\beta}_d\breve{\beta}_d} = \breve{\boldsymbol{X}}_d^{\top} \breve{\boldsymbol{V}}_d^{-1} \breve{\boldsymbol{X}}_d, \quad \boldsymbol{F}_{d\theta_a\theta_b} = \frac{1}{2} \operatorname{tr}\left(\breve{\boldsymbol{V}}_d^{-1} \breve{\boldsymbol{V}}_{uda} \breve{\boldsymbol{V}}_d^{-1} \breve{\boldsymbol{V}}_{udb} \right), \quad a, b \in \mathcal{Q}_d. \tag{6.303}$$

The remaining block-matrix components of $\boldsymbol{F}_d$ with respect to the parameters of $\boldsymbol{\psi}$ that are not in $\breve{\boldsymbol{\psi}}_d$ are equal to zero.

The components of the Fisher information matrix are

$$\boldsymbol{F}_{st} = \sum_{d=1}^{D} \boldsymbol{F}_{dst}, \quad s, t = 1, \dots, p+q. \tag{6.304}$$

The Fisher information matrix is then given by

$$
\begin{aligned}
&\boldsymbol{F}(\boldsymbol{\psi}) \\
&= \begin{pmatrix}
\boldsymbol{F}_{\beta_{11}\beta_{11}} & \cdots & \boldsymbol{F}_{\beta_{11}\beta_{mp_m}} & \boldsymbol{F}_{\beta_{11}\sigma^2_{u1}} & \cdots & \boldsymbol{F}_{\beta_{11}\sigma^2_{um}} & \boldsymbol{F}_{\beta_{11}\rho_{12}} & \cdots & \boldsymbol{F}_{\beta_{11}\rho_{m-1,m}} \\
\vdots & \ddots & & & & & & & \\
\boldsymbol{F}_{\beta_{11}\beta_{mp_m}} & \cdots & \boldsymbol{F}_{\beta_{mp_m}\beta_{mp_m}} & & & \ddots & & & \\
\boldsymbol{F}_{\beta_{11}\sigma^2_{u1}} & & & \boldsymbol{F}_{\sigma^2_{u1}\sigma^2_{u1}} & & & & & \\
\vdots & & & & \ddots & & & & \vdots \\
\boldsymbol{F}_{\beta_{11}\sigma^2_{um}} & & \ddots & & & \boldsymbol{F}_{\sigma^2_{um}\sigma^2_{um}} & & & \\
\boldsymbol{F}_{\beta_{11}\rho_{12}} & & & & & & \boldsymbol{F}_{\rho_{12}\rho_{12}} & & \\
\vdots & & & & & & & \ddots & \\
\boldsymbol{F}_{\beta_{11}\rho_{m-1,m}} & & & & \cdots & & & & \boldsymbol{F}_{\rho_{m-1,m}\rho_{m-1,m}}
\end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{F}_{\boldsymbol{\beta}\boldsymbol{\beta}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{F}_{\boldsymbol{\theta}\boldsymbol{\theta}} \end{pmatrix} \in \mathbb{R}^{(p+q)\times(p+q)},
\end{aligned}
\tag{6.305}
$$

where $\boldsymbol{F}_{\boldsymbol{\beta}\boldsymbol{\beta}} \in \mathbb{R}^{p\times p}$ and $\boldsymbol{F}_{\boldsymbol{\theta}\boldsymbol{\theta}} \in \mathbb{R}^{q\times q}$.

ML Fisher-Scoring
The ML Fisher-Scoring procedure is given by Algorithm 6.1.

```
Algorithm 6.1 ML Fisher-Scoring
1. Set the initial values $\boldsymbol{\psi}^{(0)} = (\boldsymbol{\beta}^{(0)\top}, \boldsymbol{\theta}^{(0)\top})^\top \in \mathbb{R}^{(p+q)}$ and tolerance conditions $\varepsilon_s > 0$, $\forall s \in \{1, \dots, p+q\}$.
2. Repeat the following steps until the tolerance or the boundary conditions are fulfilled.
    a) Updating equations:
        $\boldsymbol{\beta}^{(r+1)} = \boldsymbol{\beta}^{(r)} + \boldsymbol{F}^{-1}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\theta}^{(r)}, \boldsymbol{\beta}^{(r)})\boldsymbol{s}_{\boldsymbol{\beta}}(\boldsymbol{\psi}^{(r)})$,
        $\boldsymbol{\theta}^{(r+1)} = \boldsymbol{\theta}^{(r)} + \boldsymbol{F}^{-1}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta}^{(r)}, \boldsymbol{\beta}^{(r+1)})\boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{\psi}^{(r)})$.
    b) Boundary conditions:
        If $\theta_a^{(r+1)} > 0$, $\forall a \in \{1, \dots, m\}$ and $|\theta_a^{(r+1)}| < 1$, $\forall a \in \{m+1, \dots, q\}$, continue.
        Otherwise, do $\hat{\boldsymbol{\psi}} = \boldsymbol{\psi}^{(r)}$ and stop.
    c) Tolerance conditions:
        If $|\psi_s^{(r+1)} - \psi_s^{(r)}| > \varepsilon_s$, $\forall s \in \{1, \dots, p+q\}$, continue. Otherwise, do $\hat{\boldsymbol{\psi}} = \boldsymbol{\psi}^{(r+1)}$
        and stop.
3. Output: $\hat{\boldsymbol{\psi}} = (\hat{\boldsymbol{\theta}}^\top, \hat{\boldsymbol{\beta}}^\top)^\top$, $\boldsymbol{F}^{-1}(\hat{\boldsymbol{\theta}})$.
```

As starting values for $\boldsymbol{\beta}$ and $(\theta_1, \dots, \theta_m)$, which correspond to $(\sigma^2_{u1}, \dots, \sigma^2_{um})$, we take the ML estimates of the corresponding univariate Fay-Herriot models for each target variable. As starting values for $(\theta_{m+1}, \dots, \theta_q)$, which correspond to the random effect correlations, we take the correlation of the two corresponding direct estimates in the data if possible and 0 else.

In Section 2.4.2, we saw that the ML estimators are consistent and follow an asymptotically normal distribution. The asymptotic distributions of the ML estimators $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\beta}}$,

$$\hat{\boldsymbol{\theta}} \sim N_q\Big(\boldsymbol{\theta}, \boldsymbol{F}^{-1}(\boldsymbol{\theta})\Big), \quad \hat{\boldsymbol{\beta}} \sim N_p\Big(\boldsymbol{\beta}, (\breve{\boldsymbol{X}}^\top \breve{\boldsymbol{V}}^{-1}(\boldsymbol{\theta})\breve{\boldsymbol{X}})^{-1}\Big), \tag{6.306}$$

can therefore be used to construct $(1-\alpha)$-level confidence intervals for the components θ_a of $\boldsymbol{\theta}$ and β_t of $\boldsymbol{\beta}$. The confidence intervals are given by

$$\hat{\theta}_a \pm z_{\alpha/2}\, \nu_{aa}^{1/2}, \quad \forall a \in \{1, \dots, q\}, \quad \hat{\beta}_t \pm z_{\alpha/2}\, q_{st}^{1/2}, \quad \forall t \in \{1, \dots, p\}, \tag{6.307}$$

where $\boldsymbol{F}^{-1}(\hat{\boldsymbol{\theta}}) = (\nu_{ab})_{a,b=1,\dots,q}$, $(\breve{\boldsymbol{X}}^\top \breve{\boldsymbol{V}}^{-1}(\hat{\boldsymbol{\theta}})\breve{\boldsymbol{X}})^{-1} = (q_{st})_{s,t=1,\dots,p}$ and z_α is the α-quantile of the $N(0,1)$ distribution. For $\hat{\boldsymbol{\beta}}_t = \boldsymbol{\beta}_0$, the p-value for testing the hypothesis $H_0 : \boldsymbol{\beta}_t = 0$ is

$$p\text{-value} = 2 \Pr_{H_0}(\hat{\boldsymbol{\beta}}_t > |\boldsymbol{\beta}_0|) = 2 \Pr(N(0,1) > |\boldsymbol{\beta}_0|/\sqrt{q_{tt}}\,). \tag{6.308}$$

6.5.2 Residual maximum likelihood method

Restricted log-likelihood in the MFH model
In the MFH model, where all direct estimates are available, the restricted log-likelihood is given by

$$\ell_{reml}(\boldsymbol{\theta}) = c - \frac{1}{2}\log(\det(\boldsymbol{V})) - \frac{1}{2}\log\Big(\det\Big(\boldsymbol{X}^\top \boldsymbol{V}^{-1}\boldsymbol{X}\Big)\Big) - \frac{1}{2}\boldsymbol{y}^\top \boldsymbol{P}\boldsymbol{y}, \tag{6.309}$$

where c is a constant term, $\boldsymbol{y} = (\boldsymbol{y}_1^\top, \dots, \boldsymbol{y}_D^\top)^\top \in \mathbb{R}^{mD}$, $\boldsymbol{V} \in \mathbb{R}^{mD \times mD}$ is the block-diagonal covariance matrix of $\boldsymbol{y}$ with blocks of size $m \times m$, $\boldsymbol{X} = (\boldsymbol{X}_1, \dots, \boldsymbol{X}_D)^\top \in \mathbb{R}^{mD \times p}$ and having full column rank, $\boldsymbol{P} = \boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{V}^{-1}$, and $\boldsymbol{PVP} = \boldsymbol{P}$ with $\boldsymbol{PX} = 0$.

Restricted log-likelihood in the MMFH model
Under the MMFH model, the REML log-likelihood takes the form (6.309), with vector $\boldsymbol{y}$ and matrices $\boldsymbol{X}$ and $\boldsymbol{V}$ reduced to those rows and columns that correspond to observed values $\breve{\boldsymbol{y}}$.

Recall that $\breve{M} = \sum_{d=1}^{D} \breve{m}_d$, $\breve{\boldsymbol{y}} = (\breve{\boldsymbol{y}}_1^\top, \dots, \breve{\boldsymbol{y}}_D^\top)^\top \in \mathbb{R}^{\breve{M}}$, $\breve{\boldsymbol{V}} \in \mathbb{R}^{\breve{M} \times \breve{M}}$ is the block-diagonal covariance matrix of $\breve{\boldsymbol{y}}$ with blocks of size $\breve{m}_d \times \breve{m}_d$, and $\breve{\boldsymbol{X}} = (\breve{\boldsymbol{X}}_1, \dots, \breve{\boldsymbol{X}}_D)^\top \in \mathbb{R}^{\breve{M} \times p}$. Let $\breve{\boldsymbol{P}} = \breve{\boldsymbol{V}}^{-1} - \breve{\boldsymbol{V}}^{-1}\breve{\boldsymbol{X}}(\breve{\boldsymbol{X}}^\top \breve{\boldsymbol{V}}^{-1}\breve{\boldsymbol{X}})^{-1}\breve{\boldsymbol{X}}^\top \breve{\boldsymbol{V}}^{-1} \in \mathbb{R}^{\breve{M} \times \breve{M}}$. For the MMFH model, we then have

$$\ell_{reml}(\boldsymbol{\theta}) = c - \frac{1}{2}\log\Big(\det\Big(\breve{\boldsymbol{V}}\Big)\Big) - \frac{1}{2}\log\Big(\det\Big(\breve{\boldsymbol{X}}^\top \breve{\boldsymbol{V}}^{-1}\breve{\boldsymbol{X}}\Big)\Big) - \frac{1}{2}\breve{\boldsymbol{y}}^\top \breve{\boldsymbol{P}}\breve{\boldsymbol{y}}, \tag{6.310}$$

where c is a constant term.

First partial derivatives
By applying the formulas

$$\frac{\partial \log\det\left(\breve{\boldsymbol{V}}\right)}{\partial\theta_a} = \operatorname{tr}\left(\breve{\boldsymbol{V}}^{-1}\frac{\partial\breve{\boldsymbol{V}}}{\partial\theta_a}\right), \quad \frac{\partial\breve{\boldsymbol{V}}^{-1}}{\partial\theta_a} = -\breve{\boldsymbol{V}}^{-1}\frac{\partial\breve{\boldsymbol{V}}}{\partial\theta_a}\breve{\boldsymbol{V}}^{-1}, \tag{6.311}$$

we calculate the first partial derivatives of ℓ_{reml} with respect to θ_a, i.e.

$$\frac{\partial\ell_{reml}(\boldsymbol{\theta})}{\partial\theta_a} = -\frac{1}{2}\operatorname{tr}\left(\breve{\boldsymbol{P}}\frac{\partial\breve{\boldsymbol{V}}}{\partial\theta_a}\right) - \frac{1}{2}\breve{\boldsymbol{y}}^\top\frac{\partial\breve{\boldsymbol{P}}}{\partial\theta_a}\breve{\boldsymbol{y}}, \quad a = 1,\dots,q. \tag{6.312}$$

Let us define $\breve{\boldsymbol{G}} = \breve{\boldsymbol{V}}^{-1}\breve{\boldsymbol{X}}(\breve{\boldsymbol{X}}^\top\breve{\boldsymbol{V}}^{-1}\breve{\boldsymbol{X}})^{-1}$, so that $\breve{\boldsymbol{P}} = (\boldsymbol{I}-\breve{\boldsymbol{G}}\breve{\boldsymbol{X}}^\top)\breve{\boldsymbol{V}}^{-1} = \breve{\boldsymbol{V}}^{-1}(\boldsymbol{I}-\breve{\boldsymbol{X}}\breve{\boldsymbol{G}}^\top)$. The first partial derivatives of $\breve{\boldsymbol{P}}$ with respect to θ_a are

$$\frac{\partial\breve{\boldsymbol{P}}}{\partial\theta_a} = -(\boldsymbol{I}-\breve{\boldsymbol{G}}\breve{\boldsymbol{X}}^\top)\breve{\boldsymbol{V}}^{-1}\frac{\partial\breve{\boldsymbol{V}}}{\partial\theta_a}\breve{\boldsymbol{V}}^{-1}(\boldsymbol{I}-\breve{\boldsymbol{G}}\breve{\boldsymbol{X}}^\top)^\top = -\breve{\boldsymbol{P}}\frac{\partial\breve{\boldsymbol{V}}}{\partial\theta_a}\breve{\boldsymbol{P}}, \quad a = 1,\dots,q. \tag{6.313}$$

Score vector
Therefore, the score vector is

$$\boldsymbol{s}(\boldsymbol{\theta}) = (s_1,\dots,s_q)^\top, \tag{6.314}$$

$$s_a = s_a(\boldsymbol{\theta}) = \frac{\partial\ell_{reml}}{\partial\theta_a} = -\frac{1}{2}\operatorname{tr}(\breve{\boldsymbol{P}}\breve{\boldsymbol{V}}_a) + \frac{1}{2}\breve{\boldsymbol{y}}^\top\breve{\boldsymbol{P}}\breve{\boldsymbol{V}}_a\breve{\boldsymbol{P}}\breve{\boldsymbol{y}}, \quad a = 1,\dots,q, \tag{6.315}$$

where $\breve{\boldsymbol{V}}_a = \partial\breve{\boldsymbol{V}}/\partial\theta_a = \underset{1\le d\le D}{\operatorname{diag}}(\breve{\boldsymbol{V}}_{uda})$ and the elements of $\breve{\boldsymbol{V}}_{uda}$ are given in Section 6.5.1.

Second partial derivatives
For $a, b \in \{1,\dots,q\}$, the second partial derivatives of the REML log-likelihood function are

$$\frac{\partial\ell^2_{reml}(\boldsymbol{\theta})}{\partial\theta_a\partial\theta_b} = \frac{1}{2}\operatorname{tr}\left(\breve{\boldsymbol{P}}\breve{\boldsymbol{V}}_a\breve{\boldsymbol{P}}\breve{\boldsymbol{V}}_b\right) - \operatorname{tr}\left(\breve{\boldsymbol{P}}\breve{\boldsymbol{V}}_{ab}\right) - \breve{\boldsymbol{y}}^\top\breve{\boldsymbol{P}}\breve{\boldsymbol{V}}_a\breve{\boldsymbol{P}}\breve{\boldsymbol{V}}_b\breve{\boldsymbol{P}}\breve{\boldsymbol{y}} + \frac{1}{2}\breve{\boldsymbol{y}}^\top\breve{\boldsymbol{P}}\breve{\boldsymbol{V}}_{ab}\breve{\boldsymbol{P}}\breve{\boldsymbol{y}}, \tag{6.316}$$

as $\breve{\boldsymbol{V}}_a$ is symmetric, $a = 1,\dots,q$.

Fisher information matrix
Note that $\breve{\boldsymbol{P}}\breve{\boldsymbol{X}} = 0$, $\breve{\boldsymbol{P}}\breve{\boldsymbol{V}} = \boldsymbol{I} - \breve{\boldsymbol{V}}^{-1}\breve{\boldsymbol{X}}\breve{\boldsymbol{Q}}\breve{\boldsymbol{X}}^\top$, where $\breve{\boldsymbol{Q}} = (\breve{\boldsymbol{X}}^\top\breve{\boldsymbol{V}}^{-1}\breve{\boldsymbol{X}})^{-1}$ and

$$\mathrm{E}[\breve{\boldsymbol{y}}^\top\boldsymbol{A}\ \breve{\boldsymbol{y}}] = \operatorname{tr}\left(\boldsymbol{A}\ \operatorname{Var}(\breve{\boldsymbol{y}})\right) + \mathrm{E}[\breve{\boldsymbol{y}}]^\top\boldsymbol{A}\ \mathrm{E}[\breve{\boldsymbol{y}}] \tag{6.317}$$

for an arbitrary $\breve{M}\times\breve{M}$-matrix $\boldsymbol{A}$. By changing the sign and taking the expectation of the second partial derivatives, we get the components of the Fisher information matrix, i.e.

$$\boldsymbol{F}_{ab} = \boldsymbol{F}_{ab}(\boldsymbol{\theta}) = \frac{1}{2}\operatorname{tr}\left(\breve{\boldsymbol{P}}\breve{\boldsymbol{V}}_a\breve{\boldsymbol{P}}\breve{\boldsymbol{V}}_b\right), \quad a, b = 1,\dots,q. \tag{6.318}$$

Therefore, the Fisher information matrix is

$$\boldsymbol{F}(\boldsymbol{\theta}) = (\boldsymbol{F}_{ab})_{a,b=1,\ldots,q}, \quad \boldsymbol{F}_{ab} = \boldsymbol{F}_{ab}(\boldsymbol{\theta}) = \frac{1}{2}\operatorname{tr}\Big(\breve{\boldsymbol{P}}\breve{\boldsymbol{V}}_a\breve{\boldsymbol{P}}\breve{\boldsymbol{V}}_b\Big), \quad a,b = 1,\ldots,q. \tag{6.319}$$

REML Fisher-Scoring

The REML Fisher-Scoring procedure is given by Algorithm 6.2. As starting values for $\boldsymbol{\beta}$ and $(\theta_1,\ldots,\theta_m)$, which correspond to $(\sigma^2_{u_1},\ldots,\sigma^2_{u_m})$, we take the REML estimates of the corresponding univariate Fay-Herriot models for each target variable. As starting values for $(\theta_{m+1},\ldots,\theta_q)$, which correspond to the random effect correlations, we take the correlation of the two corresponding direct estimates in the data if possible and 0 else.

Algorithm 6.2 REML Fisher-Scoring

```
1. Set the initial values $\boldsymbol{\psi}^{(0)} = (\boldsymbol{\beta}^{(0)\top}, \boldsymbol{\theta}^{(0)\top})^\top \in \mathbb{R}^{(p+q)}$ and tolerance conditions $\varepsilon_s > 0$,
   $\forall s \in \{1,\ldots,p+q\}$.
2. Repeat the following steps until the tolerance or the boundary conditions are fulfilled.
   a) Updating equation for $\boldsymbol{\theta}$:
          $\boldsymbol{\theta}^{(r+1)} = \boldsymbol{\theta}^{(r)} + \boldsymbol{F}^{-1}(\boldsymbol{\theta}^{(r)})\boldsymbol{s}(\boldsymbol{\theta}^{(r)}).$
   b) Boundary conditions:
      If $\theta_a^{(r+1)} > 0$, $\forall a \in \{1,\ldots,m\}$ and $|\theta_a^{(r+1)}| < 1$, $\forall a \in \{m+1,\ldots,q\}$, continue.
      Otherwise, do $\hat{\boldsymbol{\psi}} = \boldsymbol{\psi}^{(r)}$ and stop.
   c) Updating equation for $\boldsymbol{\beta}$:
          $\boldsymbol{\beta}^{(r+1)} = (\breve{\boldsymbol{X}}^\top\breve{\boldsymbol{V}}^{-1}(\boldsymbol{\theta}^{(r+1)})\breve{\boldsymbol{X}})^{-1}\breve{\boldsymbol{X}}^\top\breve{\boldsymbol{V}}^{-1}(\boldsymbol{\theta}^{(r+1)})\breve{\boldsymbol{y}}.$
   d) Tolerance conditions:
      If $|\psi_s^{(r+1)} - \psi_s^{(r)}| > \varepsilon_s$, $\forall s \in \{1,\ldots,p+q\}$, continue. Otherwise, do $\hat{\boldsymbol{\psi}} = \boldsymbol{\psi}^{(r+1)}$
      and stop.
3. Output: $\hat{\boldsymbol{\psi}} = (\hat{\boldsymbol{\theta}}^\top, \hat{\boldsymbol{\beta}}^\top)^\top$, $\boldsymbol{F}^{-1}(\hat{\boldsymbol{\theta}})$.
```

In Section 2.4.2, we saw that the REML estimators are consistent and follow an asymptotically normal distribution. The asymptotic distributions of the REML estimators $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\beta}}$,

$$\hat{\boldsymbol{\theta}} \sim N_q\Big(\boldsymbol{\theta}, \boldsymbol{F}^{-1}(\boldsymbol{\theta})\Big), \quad \hat{\boldsymbol{\beta}} \sim N_p\Big(\boldsymbol{\beta}, (\breve{\boldsymbol{X}}^\top\breve{\boldsymbol{V}}^{-1}(\boldsymbol{\theta})\breve{\boldsymbol{X}})^{-1}\Big), \tag{6.320}$$

can therefore be used to construct $(1-\alpha)$-level confidence intervals for the components θ_a of $\boldsymbol{\theta}$ and β_t of $\boldsymbol{\beta}$. The confidence intervals are given by

$$\hat{\theta}_a \pm z_{\alpha/2}\,\nu_{aa}^{1/2}, \quad \forall a \in \{1,\ldots,q\}, \quad \hat{\beta}_t \pm z_{\alpha/2}\,q_{st}^{1/2}, \quad \forall t \in \{1,\ldots,p\}, \tag{6.321}$$

where $\boldsymbol{F}^{-1}(\hat{\boldsymbol{\theta}}) = (\nu_{ab})_{a,b=1,\ldots,q}$, $(\breve{\boldsymbol{X}}^\top\breve{\boldsymbol{V}}^{-1}(\hat{\boldsymbol{\theta}})\breve{\boldsymbol{X}})^{-1} = (q_{st})_{s,t=1,\ldots,p}$ and z_α is the α-quantile of the $N(0,1)$ distribution. For $\hat{\boldsymbol{\beta}}_t = \boldsymbol{\beta}_0$, the p-value for testing the hypothesis $H_0 : \boldsymbol{\beta}_t = 0$

is

$$p\text{-value} = 2\Pr_{H_0}(\hat{\boldsymbol{\beta}}_t > |\boldsymbol{\beta}_0|) = 2\Pr(N(0,1) > |\boldsymbol{\beta}_0|/\sqrt{q_{tt}}\,). \tag{6.322}$$

6.6 Mean squared error

6.6.1 Best predictors

Recall that $\acute{\boldsymbol{V}}_{ed} = \boldsymbol{\Lambda}_d^\top \breve{\boldsymbol{V}}_{ed} \boldsymbol{\Lambda}_d \in \mathbb{R}^{m\times m}$ and $\acute{\boldsymbol{V}}_{ed}^{\text{inv}} = \boldsymbol{\Lambda}_d^\top \breve{\boldsymbol{V}}_{ed}^{-1} \boldsymbol{\Lambda}_d \in \mathbb{R}^{m\times m}$ contain the entries of $\breve{\boldsymbol{V}}_{ed}$ and $\breve{\boldsymbol{V}}_{ed}^{-1}$ for the variables observed in d and zeros otherwise, $d = 1,\dots,D$.

If $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ are known, the best predictor (BP) of $\boldsymbol{u}$ and $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$ are

$$\begin{aligned}
\hat{\boldsymbol{u}}_d^{\text{BP}} &= \boldsymbol{\Phi}_d \acute{\boldsymbol{V}}_{ed}^{\text{inv}} (\acute{\boldsymbol{y}}_d - \boldsymbol{X}_d\boldsymbol{\beta}) = \boldsymbol{\Phi}_d \acute{\boldsymbol{V}}_{ed}^{\text{inv}} (\boldsymbol{\Lambda}_d^\top \breve{\boldsymbol{y}}_d - \boldsymbol{X}_d\boldsymbol{\beta}), && (6.277)\\
\hat{\boldsymbol{\mu}}_d^{\text{BP}} &= \boldsymbol{X}_d\boldsymbol{\beta} + \hat{\boldsymbol{u}}_d^{\text{BP}}, \quad d = 1,\dots,D, && (6.278)
\end{aligned}$$

where

$$\boldsymbol{\Phi}_d = \left(\acute{\boldsymbol{V}}_{ed}^{\text{inv}} + \boldsymbol{V}_{ud}^{-1}\right)^{-1} = \begin{pmatrix} \phi_{d,11} & \cdots & \phi_{d,1m} \\ \vdots & \ddots & \vdots \\ \phi_{d,1m} & \cdots & \phi_{d,mm} \end{pmatrix}. \tag{6.323}$$

The variance matrix of $\hat{\boldsymbol{\mu}}_d^{\text{BP}}$, with $\acute{\boldsymbol{e}}_d = \boldsymbol{\Lambda}_d^\top \boldsymbol{\Lambda}_d \boldsymbol{e}_d \in \mathbb{R}^m$, is

$$\begin{aligned}
\operatorname{Var}(\hat{\boldsymbol{\mu}}_d^{\text{BP}}) = \operatorname{Var}(\hat{\boldsymbol{u}}_d^{\text{BP}}) = \operatorname{E}[\hat{\boldsymbol{u}}_d^{\text{BP}} (\hat{\boldsymbol{u}}_d^{\text{BP}})^\top] &= \boldsymbol{\Phi}_d \acute{\boldsymbol{V}}_{ed}^{\text{inv}} \operatorname{Var}(\boldsymbol{u}_d + \acute{\boldsymbol{e}}_d) \acute{\boldsymbol{V}}_{ed}^{\text{inv}} \boldsymbol{\Phi}_d \\
&= \boldsymbol{\Phi}_d \acute{\boldsymbol{V}}_{ed}^{\text{inv}} (\boldsymbol{V}_{ud} + \acute{\boldsymbol{V}}_{ed}) \acute{\boldsymbol{V}}_{ed}^{\text{inv}} \boldsymbol{\Phi}_d.
\end{aligned} \tag{6.324}$$

Further, the expectation matrix $\operatorname{E}[\hat{\boldsymbol{u}}_d^{\text{BP}} \boldsymbol{u}_d^\top]$ is

$$\operatorname{E}[\hat{\boldsymbol{u}}_d^{\text{BP}} \boldsymbol{u}_d^\top] = \boldsymbol{\Phi}_d \acute{\boldsymbol{V}}_{ed}^{\text{inv}} \boldsymbol{V}_{ud}. \tag{6.325}$$

As $\hat{\boldsymbol{\mu}}_d^{\text{BP}} - \boldsymbol{\mu}_d = \hat{\boldsymbol{u}}_d^{\text{BP}} - \boldsymbol{u}_d$, the MSE matrix of $\hat{\boldsymbol{\mu}}_d^{\text{BP}}$ is

$$\begin{aligned}
\operatorname{MSE}(\hat{\boldsymbol{\mu}}_d^{\text{BP}}) = \operatorname{MSE}(\hat{\boldsymbol{u}}_d^{\text{BP}}) &= \operatorname{E}\left[(\hat{\boldsymbol{u}}_d^{\text{BP}} - \boldsymbol{u}_d)(\hat{\boldsymbol{u}}_d^{\text{BP}} - \boldsymbol{u}_d)^\top\right] \\
&= \boldsymbol{\Phi}_d \acute{\boldsymbol{V}}_{ed}^{\text{inv}} (\boldsymbol{V}_{ud} + \acute{\boldsymbol{V}}_{ed}) \acute{\boldsymbol{V}}_{ed}^{\text{inv}} \boldsymbol{\Phi}_d + \boldsymbol{V}_{ud} - 2\boldsymbol{\Phi}_d \acute{\boldsymbol{V}}_{ed}^{\text{inv}} \boldsymbol{V}_{ud}.
\end{aligned} \tag{6.326}$$

6.6.2 Empirical best linear unbiased predictors

Recall that for ML/REML estimators $\hat{\boldsymbol{\theta}}$, the empirical BLUE (EBLUE) of $\boldsymbol{\beta}$ and the empirical BLUP (EBLUP) of $\boldsymbol{u}_d$ and $\boldsymbol{\mu}_d = \boldsymbol{X}_d\boldsymbol{\beta}_d + \boldsymbol{u}_d$ are given by

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= (\breve{\boldsymbol{X}}^\top \breve{\boldsymbol{V}}^{-1}(\hat{\boldsymbol{\theta}}) \breve{\boldsymbol{X}})^{-1} \breve{\boldsymbol{X}}^\top \breve{\boldsymbol{V}}^{-1}(\hat{\boldsymbol{\theta}}) \breve{\boldsymbol{y}}, && (6.279)\\
\hat{\boldsymbol{u}}_d^{\text{BLUP}} &= \boldsymbol{\Phi}_d \acute{\boldsymbol{V}}_{ed}^{\text{inv}} (\acute{\boldsymbol{y}}_d - \boldsymbol{X}_d\hat{\boldsymbol{\beta}}) = \boldsymbol{\Phi}_d \acute{\boldsymbol{V}}_{ed}^{\text{inv}} (\boldsymbol{\Lambda}_d^\top \breve{\boldsymbol{y}}_d - \boldsymbol{X}_d\hat{\boldsymbol{\beta}}), && (6.274)\\
\hat{\boldsymbol{\mu}}_d^{\text{EBLUP}} &= \boldsymbol{X}_d\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{u}}_d^{\text{EBLUP}}, \quad d = 1,\dots,D. && (6.275)
\end{aligned}$$

The MSE of $\hat{\boldsymbol{\mu}}_d^{\text{EBLUP}}$ is given by

$$
\begin{aligned}
&\text{MSE}(\hat{\boldsymbol{\mu}}_d^{\text{EBLUP}}) \\
&\quad = \text{E}[(\hat{\boldsymbol{\mu}}_d^{\text{EBLUP}} - \boldsymbol{\mu}_d)(\hat{\boldsymbol{\mu}}_d^{\text{EBLUP}} - \boldsymbol{\mu}_d)^\top] \\
&\quad = \text{E}\left[(\hat{\boldsymbol{\mu}}_d^{\text{BP}} - \boldsymbol{\mu}_d)(\hat{\boldsymbol{\mu}}_d^{\text{BP}} - \boldsymbol{\mu}_d)^\top\right] + \text{E}\left[(\hat{\boldsymbol{\mu}}_d^{\text{EBLUP}} - \hat{\boldsymbol{\mu}}_d^{\text{BP}})(\hat{\boldsymbol{\mu}}_d^{\text{EBLUP}} - \hat{\boldsymbol{\mu}}_d^{\text{BP}})^\top\right] \\
&\qquad + \text{E}\left[(\hat{\boldsymbol{\mu}}_d^{\text{EBLUP}} - \hat{\boldsymbol{\mu}}_d^{\text{BP}})(\hat{\boldsymbol{\mu}}_d^{\text{BP}} - \boldsymbol{\mu}_d)^\top\right] + \text{E}\left[(\hat{\boldsymbol{\mu}}_d^{\text{BP}} - \boldsymbol{\mu}_d)(\hat{\boldsymbol{\mu}}_d^{\text{EBLUP}} - \hat{\boldsymbol{\mu}}_d^{\text{BP}})^\top\right], \\
&\quad = \underbrace{\text{E}\left[(\hat{\boldsymbol{\mu}}_d^{\text{BP}} - \boldsymbol{\mu}_d)(\hat{\boldsymbol{\mu}}_d^{\text{BP}} - \boldsymbol{\mu}_d)^\top\right]}_{=\,\boldsymbol{M}_{1d}} + \underbrace{\text{E}\left[(\hat{\boldsymbol{\mu}}_d^{\text{EBLUP}} - \hat{\boldsymbol{\mu}}_d^{\text{BP}})(\hat{\boldsymbol{\mu}}_d^{\text{EBLUP}} - \hat{\boldsymbol{\mu}}_d^{\text{BP}})^\top\right]}_{=\,\boldsymbol{M}_{2d}}.
\end{aligned}
\tag{6.327}
$$

Kackar and Harville (1984) proved that for the ML/REML estimators of $\boldsymbol{\theta}$ and normally distributed sampling errors and random effects, terms $\text{E}\left[(\hat{\boldsymbol{\mu}}_d^{\text{EBLUP}} - \hat{\boldsymbol{\mu}}_d^{\text{BP}})(\hat{\boldsymbol{\mu}}_d^{\text{EBLUP}} - \hat{\boldsymbol{\mu}}_d^{\text{BP}})^\top\right]$ and $\text{E}\left[(\hat{\boldsymbol{\mu}}_d^{\text{BP}} - \boldsymbol{\mu}_d)(\hat{\boldsymbol{\mu}}_d^{\text{EBLUP}} - \hat{\boldsymbol{\mu}}_d^{\text{BP}})^\top\right]$ are zero.

Approximation of the first summand $\boldsymbol{M}_{1d}$
The first summand is $\boldsymbol{M}_{1d} = \text{E}\left[(\hat{\boldsymbol{\mu}}_d^{\text{BP}} - \boldsymbol{\mu}_d)(\hat{\boldsymbol{\mu}}_d^{\text{BP}} - \boldsymbol{\mu}_d)^\top\right] = \text{MSE}(\hat{\boldsymbol{\mu}}_d^{\text{BP}})$ and given by (6.326).

Approximation of the second summand $\boldsymbol{M}_{2d}$
The second summand is $\boldsymbol{M}_{2d} = \text{E}\left[(\hat{\boldsymbol{\mu}}_d^{\text{EBLUP}} - \hat{\boldsymbol{\mu}}_d^{\text{BP}})(\hat{\boldsymbol{\mu}}_d^{\text{EBLUP}} - \hat{\boldsymbol{\mu}}_d^{\text{BP}})^\top\right]$. The EBLUP is a function of the estimators $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})$ and $\acute{\boldsymbol{y}}_d$. For the sake of brevity, we write

$$
\boldsymbol{h}_d(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) = \hat{\boldsymbol{\mu}}_d^{\text{EBLUP}} = \boldsymbol{X}_d\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\Phi}}_d\acute{\boldsymbol{V}}_{ed}^{\text{inv}}(\acute{\boldsymbol{y}}_d - \boldsymbol{X}_d\hat{\boldsymbol{\beta}}),
\tag{6.328}
$$

where

$$
\hat{\boldsymbol{\Phi}}_d = \boldsymbol{\Phi}_d(\hat{\boldsymbol{\theta}}) = (\acute{\boldsymbol{V}}_{ed}^{\text{inv}} + \hat{\boldsymbol{V}}_{ud}^{-1})^{-1},
\tag{6.329}
$$

$$
\hat{\boldsymbol{V}}_{ud} = \boldsymbol{V}_{ud}(\hat{\boldsymbol{\theta}}) = \begin{pmatrix} \hat{\sigma}_{u_1}^2 & \hat{\rho}_{12}\hat{\sigma}_{u_1}\hat{\sigma}_{u_2} & \cdots & \hat{\rho}_{1m}\hat{\sigma}_{u_1}\hat{\sigma}_{u_m} \\ \hat{\rho}_{12}\hat{\sigma}_{u_1}\hat{\sigma}_{u_2} & \hat{\sigma}_{u_2}^2 & \cdots & \hat{\rho}_{2m}\hat{\sigma}_{u_2}\hat{\sigma}_{u_m} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\rho}_{1m}\hat{\sigma}_{u_1}\hat{\sigma}_{u_m} & \hat{\rho}_{2m}\hat{\sigma}_{u_2}\hat{\sigma}_{u_m} & \cdots & \hat{\sigma}_{u_m}^2 \end{pmatrix}.
\tag{6.330}
$$

For $a = 1, \ldots, q$, the derivatives of matrix $\boldsymbol{\Phi}_d(\boldsymbol{\theta})$, with respect to θ_a, are

$$
\begin{aligned}
\boldsymbol{\Phi}_{da} = \frac{\partial \boldsymbol{\Phi}_d}{\partial \theta_a} &= (\acute{\boldsymbol{V}}_{ed}^{\text{inv}} + \boldsymbol{V}_{ud}^{-1})^{-1}\boldsymbol{V}_{ud}^{-1}\boldsymbol{V}_{uda}\boldsymbol{V}_{ud}^{-1}(\acute{\boldsymbol{V}}_{ed}^{\text{inv}} + \boldsymbol{V}_{ud}^{-1})^{-1} \\
&= \begin{pmatrix} \phi_{da,11} & \cdots & \phi_{da,1m} \\ \vdots & \ddots & \vdots \\ \phi_{da,1m} & \cdots & \phi_{da,mm} \end{pmatrix}.
\end{aligned}
\tag{6.331}
$$

For $d = 1, \ldots, D$, $k = 1, \ldots, m$, we define $\acute{\boldsymbol{X}}_{dk} \in \mathbb{R}^{m \times p_k}$ with rows

$$
(\acute{\boldsymbol{X}}_{dk})_{l,*} = \begin{cases} \boldsymbol{x}_{dk} & \text{if } l = k, \\ 0 & \text{otherwise.} \end{cases}, \quad l = 1, \ldots, m.
\tag{6.332}
$$

For $k = 1, \dots, m$, $t = 1, \dots, p_k$, $a = 1, \dots, q$, the derivatives of $\boldsymbol{h}_d(\boldsymbol{\beta}, \boldsymbol{\theta})$ with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, are

$$\frac{\partial \boldsymbol{h}_d}{\partial \beta_{kt}} = \begin{pmatrix} h_{d\beta_{kt},1} \\ \vdots \\ h_{d\beta_{kt},m} \end{pmatrix} = \begin{cases} (\acute{\boldsymbol{X}}_{dk})_{*,t} - \boldsymbol{\Phi}_d \acute{\boldsymbol{V}}_{ed}^{\text{inv}} (\acute{\boldsymbol{X}}_{dk})_{*,t} = (\boldsymbol{I}_m - \boldsymbol{\Phi}_d \acute{\boldsymbol{V}}_{ed}^{\text{inv}})(\acute{\boldsymbol{X}}_{dk})_{*,t} & \text{if } k \in \mathcal{A}_d, \\ (\acute{\boldsymbol{X}}_{dk})_{*,t} & \text{otherwise,} \end{cases} \tag{6.333}$$

$$\frac{\partial \boldsymbol{h}_d}{\partial \theta_a} = \begin{pmatrix} h_{d\theta_a,1} \\ \vdots \\ h_{d\theta_a,m} \end{pmatrix} = \boldsymbol{\Phi}_{da} \acute{\boldsymbol{V}}_{ed}^{\text{inv}} (\acute{\boldsymbol{y}}_d - \boldsymbol{X}_d \boldsymbol{\beta}). \tag{6.334}$$

For $k, l = 1, \dots, m$, the vectors containing the derivatives of $\boldsymbol{h}_d(\boldsymbol{\beta}, \boldsymbol{\theta})$ and $\boldsymbol{\Phi}_d$, with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, are

$$\boldsymbol{h}_{d\boldsymbol{\beta}_k,l} = \underset{1 \le t \le p_k}{\text{col}} (h_{d\beta_{kt},l}) \in \mathbb{R}^{p_k}, \tag{6.335}$$

$$\boldsymbol{h}_{d\boldsymbol{\theta},l} = \underset{1 \le a \le q}{\text{col}} (h_{d\theta_a,l}) = \sum_{k=1}^{m} \frac{\acute{y}_{dk} - \boldsymbol{x}_{dk}^\top \boldsymbol{\beta}_k}{\sigma_{edk}^2} \boldsymbol{g}_{d\boldsymbol{\theta},lk} \lambda_{dk} \in \mathbb{R}^q, \tag{6.336}$$

$$\boldsymbol{g}_{d\boldsymbol{\theta},lk} = \underset{1 \le a \le q}{\text{col}} (\phi_{da,kl}) \in \mathbb{R}^q. \tag{6.337}$$

For the above formula, recall that λ_{dk} is defined in (6.260) and $\mathcal{A}_d$ is defined in (6.261). For $a, b, k, l = 1, \dots, m$, the corresponding matrices are defined as

$$\boldsymbol{H}_{d\boldsymbol{\beta}_k\boldsymbol{\beta}_l,ab} = \boldsymbol{h}_{d\boldsymbol{\beta}_k,a} \boldsymbol{h}_{d\boldsymbol{\beta}_l,b}^\top \in \mathbb{R}^{p_k \times p_l}, \ \boldsymbol{H}_{d\boldsymbol{\theta}\boldsymbol{\theta},ab} = \boldsymbol{h}_{d\boldsymbol{\theta},a} \boldsymbol{h}_{d\boldsymbol{\theta},b}^\top \in \mathbb{R}^{q \times q}, \quad \text{and} \tag{6.338}$$

$$\boldsymbol{H}_{d\boldsymbol{\beta}_k\boldsymbol{\theta},ab} = \boldsymbol{H}_{d\boldsymbol{\theta}\boldsymbol{\beta}_k,ab}^\top = \boldsymbol{h}_{d\boldsymbol{\beta}_k,a} \boldsymbol{h}_{d\boldsymbol{\theta},b}^\top \in \mathbb{R}^{p_k \times q}. \tag{6.339}$$

For $k, l = 1, \dots, m$, we furthermore define the matrices

$$\mathcal{H}_{d\boldsymbol{\beta}_k} = \underset{1 \le l \le m}{\text{row}} (\boldsymbol{h}_{d\boldsymbol{\beta}_k,l}) \in \mathbb{R}^{p_k \times m} \text{ and } \quad \mathcal{H}_{d\boldsymbol{\theta}} = \underset{1 \le l \le m}{\text{row}} (\boldsymbol{h}_{d\boldsymbol{\theta},l}) \in \mathbb{R}^{q \times m}. \tag{6.340}$$

A Taylor series expansion of $\boldsymbol{h}_d(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) \in \mathbb{R}^m$ around $(\boldsymbol{\beta}, \boldsymbol{\theta})$ yields

$$\begin{aligned} \boldsymbol{h}_d(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) = & \boldsymbol{h}_d(\boldsymbol{\beta}, \boldsymbol{\theta}) + \sum_{k=1}^{m} \sum_{t=1}^{p_k} \frac{\partial \boldsymbol{h}_d^\top(\boldsymbol{\beta}, \boldsymbol{\theta})}{\partial \beta_{kt}} (\hat{\beta}_{kt} - \beta_{kt}) + \sum_{a=1}^{q} \frac{\partial \boldsymbol{h}_d^\top(\boldsymbol{\beta}, \boldsymbol{\theta})}{\partial \theta_a} (\hat{\theta}_a - \theta_a) \\ & + o_P(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_2) + o_P(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_2). \end{aligned} \tag{6.341}$$

Therefore,

$$\boldsymbol{h}_d(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) = \boldsymbol{h}_d(\boldsymbol{\beta}, \boldsymbol{\theta}) + \sum_{k=1}^{m} \mathcal{H}_{d\boldsymbol{\beta}_k}^\top (\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k) + \mathcal{H}_{d\boldsymbol{\theta}}^\top (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + o_P(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_2) + o_P(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_2) \tag{6.342}$$

and

$$\begin{aligned}
\boldsymbol{M}_{2d} &= \mathrm{E}\left[(\hat{\boldsymbol{\mu}}_d^{\mathrm{EBLUP}} - \hat{\boldsymbol{\mu}}_d^{\mathrm{BLUP}})(\hat{\boldsymbol{\mu}}_d^{\mathrm{EBLUP}} - \hat{\boldsymbol{\mu}}_d^{\mathrm{BLUP}})^\top\right] \\
&= \mathrm{E}\left[\left(\boldsymbol{h}_d(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) - \boldsymbol{h}_d(\boldsymbol{\beta}, \boldsymbol{\theta})\right)\left(\boldsymbol{h}_d(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) - \boldsymbol{h}_d(\boldsymbol{\beta}, \boldsymbol{\theta})\right)^\top\right] \\
&= \underbrace{\sum_{k=1}^m \sum_{l=1}^m \mathrm{E}\left[\mathcal{H}_{d\beta_k}^\top(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k)(\hat{\boldsymbol{\beta}}_l - \boldsymbol{\beta}_l)^\top \mathcal{H}_{d\beta_l}\right]}_{=\ \boldsymbol{M}_{2.1d}} + \underbrace{\mathrm{E}\left[\mathcal{H}_{d\boldsymbol{\theta}}^\top(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top \mathcal{H}_{d\boldsymbol{\theta}}\right]}_{=\ \boldsymbol{M}_{2.2d}} \\
&\quad + \underbrace{\sum_{k=1}^m \mathrm{E}\left[\mathcal{H}_{d\beta_k}^\top(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top \mathcal{H}_{d\boldsymbol{\theta}}\right]}_{=\ \boldsymbol{M}_{2.3d}} + \underbrace{\sum_{k=1}^m \mathrm{E}\left[\mathcal{H}_{d\boldsymbol{\theta}}^\top(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k)^\top \mathcal{H}_{d\beta_k}\right]}_{=\ \boldsymbol{M}_{2.4d}} \\
&\quad + o_P(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_2^2) + o_P(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_2^2).
\end{aligned} \tag{6.343}$$

Approximation of parts of $\boldsymbol{M}_{2d}$: $\boldsymbol{M}_{2.1d}$
Concerning the first part of $\boldsymbol{M}_{2d}$,

$$\boldsymbol{M}_{2.1d} = \sum_{k=1}^m \sum_{l=1}^m \mathrm{E}\left[\mathcal{H}_{d\beta_k}^\top(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k)(\hat{\boldsymbol{\beta}}_l - \boldsymbol{\beta}_l)^\top \mathcal{H}_{d\beta_l}\right] = \sum_{k=1}^m \sum_{l=1}^m \boldsymbol{M}_{2.1dkl}, \tag{6.344}$$

we observe that

$$\begin{gathered}
\mathcal{H}_{d\beta_k}^\top(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k)(\hat{\boldsymbol{\beta}}_l - \boldsymbol{\beta}_l)^\top \mathcal{H}_{d\beta_l} = \\
\begin{pmatrix}
\boldsymbol{h}_{d\beta_k,1}^\top(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k)(\hat{\boldsymbol{\beta}}_l - \boldsymbol{\beta}_l)^\top \boldsymbol{h}_{d\beta_l,1} & \cdots & \boldsymbol{h}_{d\beta_k,1}^\top(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k)(\hat{\boldsymbol{\beta}}_l - \boldsymbol{\beta}_l)^\top \boldsymbol{h}_{d\beta_l,m} \\
\vdots & \ddots & \vdots \\
\boldsymbol{h}_{d\beta_k,m}^\top(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k)(\hat{\boldsymbol{\beta}}_l - \boldsymbol{\beta}_l)^\top \boldsymbol{h}_{d\beta_l,1} & \cdots & \boldsymbol{h}_{d\beta_k,m}^\top(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k)(\hat{\boldsymbol{\beta}}_l - \boldsymbol{\beta}_l)^\top \boldsymbol{h}_{d\beta_l,m}
\end{pmatrix}.
\end{gathered} \tag{6.345}$$

Note that $\boldsymbol{h}_{d\beta_k,a}^\top(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k)$ is a scalar. Then, we have

$$\begin{aligned}
\boldsymbol{h}_{d\beta_k,a}^\top(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k)(\hat{\boldsymbol{\beta}}_l - \boldsymbol{\beta}_l)^\top \boldsymbol{h}_{d\beta_l,b} &= (\hat{\boldsymbol{\beta}}_l - \boldsymbol{\beta}_l)^\top \boldsymbol{h}_{d\beta_l,b} \boldsymbol{h}_{d\beta_k,a}^\top (\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k) \\
&= (\hat{\boldsymbol{\beta}}_l - \boldsymbol{\beta}_l)^\top \boldsymbol{H}_{d\beta_l\beta_k,ba}(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k).
\end{aligned} \tag{6.346}$$

Therefore,

$$\boldsymbol{M}_{2.1d,kl} = \mathrm{E}\left[\begin{pmatrix} (\hat{\boldsymbol{\beta}}_l - \boldsymbol{\beta}_l)^\top \boldsymbol{H}_{d\beta_l\beta_k,11}(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k) & \cdots & (\hat{\boldsymbol{\beta}}_l - \boldsymbol{\beta}_l)^\top \boldsymbol{H}_{d\beta_l\beta_k,m1}(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k) \\ \vdots & \ddots & \vdots \\ (\hat{\boldsymbol{\beta}}_l - \boldsymbol{\beta}_l)^\top \boldsymbol{H}_{d\beta_l\beta_k,1m}(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k) & \cdots & (\hat{\boldsymbol{\beta}}_l - \boldsymbol{\beta}_l)^\top \boldsymbol{H}_{d\beta_l\beta_k,mm}(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k) \end{pmatrix}\right]. \tag{6.347}$$

We apply to $\boldsymbol{z} = \hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k$ the formula

$$\mathrm{E}[\boldsymbol{z}^\top \boldsymbol{A}\boldsymbol{z}] = \mathrm{tr}\left(\boldsymbol{A}\ \mathrm{Var}(\boldsymbol{z})\right) + \mathrm{E}[\boldsymbol{z}]^\top \boldsymbol{A}\, \mathrm{E}[\boldsymbol{z}] \tag{6.348}$$

for an arbitrary $p_k \times p_k$-matrix $\boldsymbol{A}$, $k = 1, \dots, m$. As the matrices $\boldsymbol{H}_{d\beta_l\beta_k,ab}$ are not random and $\mathrm{E}[\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k] = \mathcal{O}(\|\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k\|_2)$, we obtain

$$\begin{aligned} \boldsymbol{M}_{2.1d,kl} = &\begin{pmatrix} \mathrm{tr}\left(\boldsymbol{H}_{d\beta_l\beta_k,11}\ \mathrm{Cov}(\hat{\boldsymbol{\beta}}_k, \hat{\boldsymbol{\beta}}_l)\right) & \cdots & \mathrm{tr}\left(\boldsymbol{H}_{d\beta_l\beta_k,m1}\ \mathrm{Cov}(\hat{\boldsymbol{\beta}}_k, \hat{\boldsymbol{\beta}}_l)\right) \\ \vdots & \ddots & \vdots \\ \mathrm{tr}\left(\boldsymbol{H}_{d\beta_l\beta_k,1m}\ \mathrm{Cov}(\hat{\boldsymbol{\beta}}_k, \hat{\boldsymbol{\beta}}_l)\right) & \cdots & \mathrm{tr}\left(\boldsymbol{H}_{d\beta_l\beta_k,mm}\ \mathrm{Cov}(\hat{\boldsymbol{\beta}}_k, \hat{\boldsymbol{\beta}}_l)\right) \end{pmatrix} \\ &+ \mathcal{O}_{m\times m}(\|\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k\|_2 \|\hat{\boldsymbol{\beta}}_l - \boldsymbol{\beta}_l\|_2). \end{aligned} \tag{6.349}$$

Approximation of parts of $\boldsymbol{M}_{2d}$: $\boldsymbol{M}_{2.2d}$
The second part of $\boldsymbol{M}_{2d}$, $\boldsymbol{M}_{2.2d} = \mathrm{E}\left[\mathcal{H}_{d\boldsymbol{\theta}}^\top(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top \mathcal{H}_{d\boldsymbol{\theta}}\right]$, is

$$\boldsymbol{M}_{2.2d} = \mathrm{E}\left[\begin{pmatrix} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top \boldsymbol{H}_{d\boldsymbol{\theta}\boldsymbol{\theta},11}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) & \cdots & (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top \boldsymbol{H}_{d\boldsymbol{\theta}\boldsymbol{\theta},m1}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \\ \vdots & \ddots & \vdots \\ (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top \boldsymbol{H}_{d\boldsymbol{\theta}\boldsymbol{\theta},1m}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) & \cdots & (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top \boldsymbol{H}_{d\boldsymbol{\theta}\boldsymbol{\theta},mm}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \end{pmatrix}\right]. \tag{6.350}$$

We recall that $\boldsymbol{H}_{d\boldsymbol{\theta}\boldsymbol{\theta},ab}$ is random. This is why we calculate $\boldsymbol{M}_{2.2d}$ for the estimators $\hat{\boldsymbol{\theta}}^{-d}$ and $\hat{\boldsymbol{\beta}}_k^{-d}$, based on $\acute{\boldsymbol{y}}^{-d} = \underset{1\leq d'\leq D, d'\neq d}{\mathrm{col}}(\acute{\boldsymbol{y}}_1^\top, \dots, \acute{\boldsymbol{y}}_D^\top)^\top \in \mathbb{R}^{m(D-1)}$, $d = 1, \dots, D$, $k = 1, \dots, m$. As the target data $\boldsymbol{y}$ can be split into the independent parts $\acute{\boldsymbol{y}}^{-d}$ and $\acute{\boldsymbol{y}}_d$, for $a, b = 1, \dots, m$

we have that

$$
\begin{aligned}
&\boldsymbol{M}^{-d}_{2.2d,ab}\\
&= \mathrm{E}\left[(\hat{\boldsymbol{\theta}}^{-d}-\boldsymbol{\theta})^\top \boldsymbol{H}_{d\boldsymbol{\theta}\boldsymbol{\theta},ab}(\hat{\boldsymbol{\theta}}^{-d}-\boldsymbol{\theta})\right]\\
&= \mathrm{E}_{\acute{\boldsymbol{y}}^{-d}}\left[\mathrm{E}_{\acute{\boldsymbol{y}}_d}\left[(\hat{\boldsymbol{\theta}}^{-d}-\boldsymbol{\theta})^\top \boldsymbol{H}_{d\boldsymbol{\theta}\boldsymbol{\theta},ab}(\hat{\boldsymbol{\theta}}^{-d}-\boldsymbol{\theta})\right]\right]\\
&= \mathrm{E}_{\acute{\boldsymbol{y}}^{-d}}\left[\mathrm{E}_{\acute{\boldsymbol{y}}_d}\left[(\hat{\boldsymbol{\theta}}^{-d}-\boldsymbol{\theta})^\top \boldsymbol{h}_{d\boldsymbol{\theta},a}\boldsymbol{h}^\top_{d\boldsymbol{\theta},b}(\hat{\boldsymbol{\theta}}^{-d}-\boldsymbol{\theta})\right]\right]\\
&= \mathrm{E}_{\acute{\boldsymbol{y}}^{-d}}\Bigg[\mathrm{E}_{\acute{\boldsymbol{y}}_d}\Bigg[(\hat{\boldsymbol{\theta}}^{-d}-\boldsymbol{\theta})^\top\Big(\sum_{k=1}^{m}\frac{\acute{y}_{dk}-\boldsymbol{x}^\top_{dk}\boldsymbol{\beta}_k}{\sigma^2_{edk}}\boldsymbol{g}_{d\boldsymbol{\theta},ak}\lambda_{dk}\Big)\\
&\qquad\qquad\qquad\Big(\sum_{l=1}^{m}\frac{\acute{y}_{dl}-\boldsymbol{x}^\top_{dl}\boldsymbol{\beta}_l}{\sigma^2_{edl}}\boldsymbol{g}_{d\boldsymbol{\theta},bl}\lambda_{d_l}\Big)^\top(\hat{\boldsymbol{\theta}}^{-d}-\boldsymbol{\theta})\Bigg]\Bigg]\\
&= \sum_{k=1}^{m}\sum_{l=1}^{m}\mathrm{E}_{\acute{\boldsymbol{y}}^{-d}}\Big[(\hat{\boldsymbol{\theta}}^{-d}-\boldsymbol{\theta})^\top\boldsymbol{g}_{d\boldsymbol{\theta},ak}\lambda_{dk}\boldsymbol{g}^\top_{d\boldsymbol{\theta},bl}\lambda_{d_l}(\hat{\boldsymbol{\theta}}^{-d}-\boldsymbol{\theta})\\
&\qquad\qquad\qquad\sigma^{-2}_{edk}\sigma^{-2}_{edl}E_{\acute{\boldsymbol{y}}_d}\Big((\acute{y}_{dk}-\boldsymbol{x}^\top_{dk}\boldsymbol{\beta}_k)(\acute{y}_{dl}-\boldsymbol{x}^\top_{dl}\boldsymbol{\beta}_l)\Big)\Big]\\
&= \sum_{k=1}^{m}\sum_{l=1}^{m}\frac{(\boldsymbol{V}_d)_{k,l}}{\sigma^{-2}_{edk}\sigma^{-2}_{edl}}\operatorname{tr}\Big(\boldsymbol{g}_{d\boldsymbol{\theta},ak}\lambda_{dk}\boldsymbol{g}^\top_{d\boldsymbol{\theta},bl}\lambda_{d_l}\operatorname{Var}(\hat{\boldsymbol{\theta}}^{-d})\Big)+\mathcal{O}_{m\times m}(\|\hat{\boldsymbol{\theta}}^{-d}-\boldsymbol{\theta}\|^2_2).
\end{aligned}
\tag{6.351}
$$

Taking out $(\acute{\boldsymbol{y}}_d, \boldsymbol{X}_d)$ from the data file should not have great influence on $\operatorname{Var}(\hat{\boldsymbol{\theta}})$. Therefore, we approximate $\boldsymbol{M}_{2.2d,ab}$ by substituting $\hat{\boldsymbol{\theta}}^{-d}$ by $\hat{\boldsymbol{\theta}}$ in $\boldsymbol{M}^{-d}_{2.2d,ab}$, i.e.

$$
\boldsymbol{M}_{2.2d}=\begin{pmatrix}\boldsymbol{M}_{2.2d,11} & \cdots & \boldsymbol{M}_{2.2d,1m}\\ \vdots & \ddots & \vdots\\ \boldsymbol{M}_{2.2d,m1} & \cdots & \boldsymbol{M}_{2.2d,mm}\end{pmatrix}+\mathcal{O}_{m\times m}(\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\|^2_2). \tag{6.352}
$$

Approximation of parts of $\boldsymbol{M}_{2d}$: $\boldsymbol{M}_{2.3d}$, $\boldsymbol{M}_{2.4d}$

The third and fourth part of $\boldsymbol{M}_{2d}$ are $\boldsymbol{M}_{2.3d}$ and $\boldsymbol{M}_{2.4d}$, where

$$
\boldsymbol{M}_{2.3dk}=\boldsymbol{M}^\top_{2.4dk}=\mathrm{E}\left[\boldsymbol{h}^\top_{d\boldsymbol{\beta}_k}(\hat{\boldsymbol{\beta}}_k-\boldsymbol{\beta}_k)(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^\top\mathcal{H}_{d\boldsymbol{\theta}}\right],\quad k=1,\dots,m. \tag{6.353}
$$

We have

$$
\boldsymbol{M}_{2.3dk}=\mathrm{E}\left[\begin{pmatrix}(\hat{\boldsymbol{\beta}}_k-\boldsymbol{\beta}_k)^\top\boldsymbol{H}_{d\boldsymbol{\beta}_k\boldsymbol{\theta},11}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}) & \cdots & (\hat{\boldsymbol{\beta}}_k-\boldsymbol{\beta}_k)^\top\boldsymbol{H}_{d\boldsymbol{\beta}_k\boldsymbol{\theta},m1}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\\ \vdots & \ddots & \vdots\\ (\hat{\boldsymbol{\beta}}_k-\boldsymbol{\beta}_k)^\top\boldsymbol{H}_{d\boldsymbol{\beta}_k\boldsymbol{\theta},1m}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}) & \cdots & (\hat{\boldsymbol{\beta}}_k-\boldsymbol{\beta}_k)^\top\boldsymbol{H}_{d\boldsymbol{\beta}_k\boldsymbol{\theta},mm}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\end{pmatrix}\right]. \tag{6.354}
$$

We recall that $\boldsymbol{H}_{d\beta_k\boldsymbol{\theta},ab}$ is random. For $b=1$, it holds that

$$\begin{aligned}
\boldsymbol{M}_{2.3dk,a1}^{-d} &= \mathrm{E}\left[(\hat{\boldsymbol{\beta}}_k^{-d}-\boldsymbol{\beta}_k)^\top \boldsymbol{H}_{d\beta_k\boldsymbol{\theta},a1}(\hat{\boldsymbol{\theta}}^{-d}-\boldsymbol{\theta})\right] \\
&= \mathrm{E}_{\acute{\boldsymbol{y}}^{-d}}\left[\mathrm{E}_{\acute{\boldsymbol{y}}_d}\left[(\hat{\boldsymbol{\beta}}_k^{-d}-\boldsymbol{\beta}_k)^\top \boldsymbol{H}_{d\beta_k\boldsymbol{\theta},a1}(\hat{\boldsymbol{\theta}}^{-d}-\boldsymbol{\theta})\right]\right] \\
&= \mathrm{E}_{\acute{\boldsymbol{y}}^{-d}}\left[\mathrm{E}_{\acute{\boldsymbol{y}}_d}\left[(\hat{\boldsymbol{\beta}}_k^{-d}-\boldsymbol{\beta}_k)^\top \boldsymbol{h}_{d\beta_k,a}\boldsymbol{h}_{d\boldsymbol{\theta},1}^\top(\hat{\boldsymbol{\theta}}^{-d}-\boldsymbol{\theta})\right]\right] \\
&= \mathrm{E}_{\acute{\boldsymbol{y}}^{-d}}\left[\mathrm{E}_{\acute{\boldsymbol{y}}_d}\left[(\hat{\boldsymbol{\beta}}_k^{-d}-\boldsymbol{\beta}_k)^\top \boldsymbol{h}_{d\beta_k,a}\Big(\sum_{k=1}^m \frac{\acute{y}_{dk}-\boldsymbol{x}_{dk}^\top\boldsymbol{\beta}_k}{\sigma_{edk}^2}\boldsymbol{g}_{d\boldsymbol{\theta},1k}\lambda_{dk}\Big)^\top(\hat{\boldsymbol{\theta}}^{-d}-\boldsymbol{\theta})\right]\right] \\
&= \sum_{k=1}^m \mathrm{E}_{\acute{\boldsymbol{y}}^{-d}}\left[(\hat{\boldsymbol{\beta}}_k^{-d}-\boldsymbol{\beta}_k)^\top \boldsymbol{h}_{d\beta_k,a}\boldsymbol{g}_{d\boldsymbol{\theta},1k}^\top(\hat{\boldsymbol{\theta}}^{-d}-\boldsymbol{\theta})\sigma_{edk}^{-2}\,\mathrm{E}_{\acute{y}_{dk}}[\acute{y}_{dk}-\boldsymbol{x}_{dk}^\top\boldsymbol{\beta}_k]\right] \\
&= 0.
\end{aligned} \tag{6.355}$$

Similarly, for $b=2,\dots,m$ we get $\boldsymbol{M}_{2.3dk,ab}^{-d}=0$, $k=1,\dots,m$.

Approximation of $\mathrm{MSE}(\hat{\boldsymbol{\mu}}_d^{\mathrm{EBLUP}})$
We further assume that $\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|_2=\mathcal{O}(D^{-1/2})$ and $\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\|_2=\mathcal{O}(D^{-1/2})$ such that $\mathcal{O}_{m\times m}(\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\|_2)=\mathcal{O}_{m\times m}(D^{-1})$. For $d\in\mathcal{D}$, we have the following approximation to $\mathrm{MSE}(\hat{\boldsymbol{\mu}}_d^{\mathrm{EBLUP}})$.

$$\mathrm{MSE}(\hat{\boldsymbol{\mu}}_d^{\mathrm{EBLUP}})=\boldsymbol{G}_{d1}(\boldsymbol{\theta})+\boldsymbol{G}_{d2}(\boldsymbol{\theta})+\boldsymbol{G}_3(\boldsymbol{\theta})+\mathcal{O}_{m\times m}(D^{-1}), \tag{6.356}$$

where $\boldsymbol{G}_{d2}(\boldsymbol{\theta})=\boldsymbol{G}_{d2,11}(\boldsymbol{\theta})+\boldsymbol{G}_{d2,22}(\boldsymbol{\theta})+\boldsymbol{G}_{d2,12}(\boldsymbol{\theta})+\boldsymbol{G}_{d2,12}^\top(\boldsymbol{\theta})$ and, for $k,l=1,\dots,m$,

$$\begin{aligned}
\boldsymbol{G}_{d1}(\boldsymbol{\theta}) =&\boldsymbol{\Phi}_d(\boldsymbol{\theta})\acute{\boldsymbol{V}}_{ed}^{\mathrm{inv}}(\boldsymbol{\theta})\Big(\boldsymbol{V}_{ud}(\boldsymbol{\theta})+\acute{\boldsymbol{V}}_{ed}(\boldsymbol{\theta})\Big)\acute{\boldsymbol{V}}_{ed}^{\mathrm{inv}}(\boldsymbol{\theta})\boldsymbol{\Phi}_d(\boldsymbol{\theta}) \\
&+\boldsymbol{V}_{ud}(\boldsymbol{\theta})-2\boldsymbol{\Phi}_d(\boldsymbol{\theta})\acute{\boldsymbol{V}}_{ed}^{\mathrm{inv}}(\boldsymbol{\theta})\boldsymbol{V}_{ud}(\boldsymbol{\theta}),
\end{aligned} \tag{6.357}$$

$$\boldsymbol{G}_{d2,kl}(\boldsymbol{\theta})=\begin{pmatrix}
\mathrm{tr}\Big(\boldsymbol{H}_{d\beta_l\beta_k,11}(\boldsymbol{\theta})\ \mathrm{Cov}(\hat{\boldsymbol{\beta}}_k,\hat{\boldsymbol{\beta}}_l)\Big) & \cdots & \mathrm{tr}\Big(\boldsymbol{H}_{d\beta_l\beta_k,m1}(\boldsymbol{\theta})\ \mathrm{Cov}(\hat{\boldsymbol{\beta}}_k,\hat{\boldsymbol{\beta}}_l)\Big) \\
\vdots & \ddots & \vdots \\
\mathrm{tr}\Big(\boldsymbol{H}_{d\beta_l\beta_k,1m}(\boldsymbol{\theta})\ \mathrm{Cov}(\hat{\boldsymbol{\beta}}_k,\hat{\boldsymbol{\beta}}_l)\Big) & \cdots & \mathrm{tr}\Big(\boldsymbol{H}_{d\beta_l\beta_k,mm}(\boldsymbol{\theta})\ \mathrm{Cov}(\hat{\boldsymbol{\beta}}_k,\hat{\boldsymbol{\beta}}_l)\Big)
\end{pmatrix}, \tag{6.358}$$

$$\boldsymbol{G}_{d3}(\boldsymbol{\theta})=\begin{pmatrix}
\mathrm{tr}\Big(\boldsymbol{H}_{d\boldsymbol{\theta\theta},11}(\boldsymbol{\theta})\ \mathrm{Var}(\hat{\boldsymbol{\theta}})\Big) & \cdots & \mathrm{tr}\Big(\boldsymbol{H}_{d\boldsymbol{\theta\theta},m1}(\boldsymbol{\theta})\ \mathrm{Var}(\hat{\boldsymbol{\theta}})\Big) \\
\vdots & \ddots & \vdots \\
\mathrm{tr}\Big(\boldsymbol{H}_{d\boldsymbol{\theta\theta},1m}(\boldsymbol{\theta})\ \mathrm{Var}(\hat{\boldsymbol{\theta}})\Big) & \cdots & \mathrm{tr}\Big(\boldsymbol{H}_{d\boldsymbol{\theta\theta},mm}(\boldsymbol{\theta})\ \mathrm{Var}(\hat{\boldsymbol{\theta}})\Big)
\end{pmatrix}. \tag{6.359}$$

An estimator of $\mathrm{MSE}(\hat{\boldsymbol{\mu}}_d^{\mathrm{EBLUP}})$ is given by

$$\widehat{\mathrm{MSE}}(\hat{\boldsymbol{\mu}}_d^{\mathrm{EBLUP}})=\boldsymbol{G}_{d1}(\hat{\boldsymbol{\theta}})+\boldsymbol{G}_2(\hat{\boldsymbol{\theta}})+2\boldsymbol{G}_3(\hat{\boldsymbol{\theta}}), \tag{6.360}$$

similar to the MSE estimators proposed by Prasad and Rao (1990).

For ML parameter estimation, an additional bias term would have to be considered in the MSE formulas as shown for the FH model in (2.98). The derivation of the ML bias term under the MMFH model is not considered in this thesis and a potential subject of future research. For the domain prediction and MSE estimation, our focus is on REML parameter estimation.

As an alternative to the MSE estimation via (6.360), also a resampling estimator like the parametric bootstrap estimator presented in Algorithm 5.6 could be used, which, however, is computationally more demanding. In the resampling, the missing direct estimators would have to be explicitly accounted for.

6.7 Simulation

6.7.1 Motivation

We conduct model-based simulation studies to evaluate the performance of the parameter estimators, predictors, and the MSE estimator that we derived under the introduced MMFH model. By varying different parameters within the simulation, we aim to analyse the behaviour of the MMFH estimators and predictors under changing data scenarios.

In the simulation we consider three dependent variables. The correlation of sampling errors and random effects is crucial for multivariate FH models. To consider different correlation scenarios, we simulate the following: Dependent variables 1 and 3 represent the same variable based on two independent surveys, e.g. a survey in two different months without any sample overlaps. The sampling error correlation between the two is therefore zero and we expect the random effect correlation to be highly positive. Dependent variable 3 is estimated from the same survey as variable 1. Therefore, the sampling error correlation between the two is non-negative, and the sampling error correlation between variables 2 and 3 is zero.

We simulate a MMFH (6.263) model by including arbitrary missing data of interest in each domain. To keep the analysis simple, we consider missing values only for variable 1. For variable 2 and 3, we do not simulate any missing direct estimates.

To the simulated data with partially missing direct estimates we can apply different FH estimators. We apply the univariate FH estimators of the 3 variables, a trivariate FH estimator (MFH), and the MMFH estimator. The FH estimator for variable 1 and the MFH estimator can only consider the domains with no missing direct estimates for the parameter estimation and the calculation of EBLUPs. For the domains with missing direct estimates of variable 1, only synthetic FH (for variable 1) and MFH (for all 3 variables) estimates can be calculated. The simulation allows us not only to evaluate parameter estimation and EBLUPs under the proposed MMFH model, but also to contrast the performance of the different FH estimators and show the potential benefits of the proposed MMFH estimator.

6.7.2 Setup

We consider the particular multivariate FH model

$$\boldsymbol{y}_d = \boldsymbol{X}_d\boldsymbol{\beta} + \boldsymbol{u}_d + \boldsymbol{e}_d, \quad d = 1, \ldots, D. \tag{6.361}$$

with $m = 3$ dependent variables.

For the three variables, we consider one auxiliary variable plus intercept. For $k = 1, 2, 3$, set the number of auxiliary variables $p_k = 2$ and $\boldsymbol{\beta}_k = (\boldsymbol{\beta}_{k1}, \boldsymbol{\beta}_{k2})^\top = (2, 3)^\top$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top, \boldsymbol{\beta}_3^\top)^\top$. In total, there are $p = p_1 + p_2 + p_3 = 6$ fixed effects.

The auxiliary information is generated once and remains fixed throughout the simulation. For $d = 1, \ldots, D$, we take $\boldsymbol{X}_d = \text{diag}(\boldsymbol{x}_{d1}^\top, \boldsymbol{x}_{d2}^\top, \boldsymbol{x}_{d3}^\top)_{3\times 6}$, where, for $k = 1, 2, 3$, $\boldsymbol{x}_{dk} = (x_{dk1}, x_{dk2})^\top$, including intercept $x_{dk1} = 1$. Generate, for each $k = 1, 2, 3$, $x_{dk2} = U_{dk}$, $U_{dk} \overset{\text{ind}}{\sim} Unif(10, 100)$, where $Unif$ is the uniform distribution.

For $d = 1, \ldots, D$, the random effects $\boldsymbol{u}_d$ and sampling errors $\boldsymbol{e}_d$ of the model are generated in each iteration of the simulation. We take $\boldsymbol{u}_d \sim N_3(\boldsymbol{0}, \boldsymbol{V}_{ud})$ and $\boldsymbol{e}_d \sim N_3(\boldsymbol{0}, \boldsymbol{V}_{ed})$ with

$$\boldsymbol{V}_{ud} = \begin{pmatrix} \sigma_{u1}^2 & \rho_{12}\sigma_{u1}\sigma_{u2} & \rho_{13}\sigma_{u1}\sigma_{u3} \\ \rho_{12}\sigma_{u1}\sigma_{u2} & \sigma_{u2}^2 & \rho_{23}\sigma_{u2}\sigma_{u3} \\ \rho_{13}\sigma_{u1}\sigma_{u2} & \rho_{23}\sigma_{u2}\sigma_{u3} & \sigma_{u3}^2 \end{pmatrix}, \tag{6.362}$$

$$\boldsymbol{V}_{ed} = \begin{pmatrix} \sigma_{ed1}^2 & \rho_{ed12}\sigma_{ed1}\sigma_{ed2} & \rho_{ed13}\sigma_{ed1}\sigma_{ed3} \\ \rho_{ed12}\sigma_{ed1}\sigma_{ed2} & \sigma_{ed2}^2 & \rho_{ed23}\sigma_{ed2}\sigma_{ed3} \\ \rho_{ed13}\sigma_{ed1}\sigma_{ed2} & \rho_{ed13}\sigma_{ed2}\sigma_{ed3} & \sigma_{ed3}^2 \end{pmatrix}, \tag{6.363}$$

with $\sigma_{uk}^2 = 2$, $k = 1, 2, 3$, $\sigma_{edk}^2 = 3$, $d = 1, \ldots, D$. The variance components are given by vector $\boldsymbol{\theta}$ of length 6 with $\theta_k = \sigma_{uk}^2$, $k = 1, 2, 3$, $\theta_4 = \rho_{12}$, $\theta_5 = \rho_{13}$, and $\theta_6 = \rho_{23}$.

As we simulate dependent variables 1 and 2 to be from the same and variable 3 to be from an independent other survey, we set $\rho_{ed13} = \rho_{ed23} = 0$. For variable 1 and 2, we set sampling error correlation $\rho_{ed12} = -0.5$. For example, when we would consider the total of employed and unemployed in a domain as variables 1 and 2, we would expect their direct estimators to be highly negatively correlated. As we simulate dependent variables 1 and 3 to represent the same variable in two different time points, we set their random effect correlation relatively high with $\rho_{13} = 0.75$. We choose the random effects of variables 2 and 3 to be only moderately correlated with $\rho_{23} = 0.25$.

In the simulation, we vary the number of domains $D \in \{100, 200, 300\}$ and the correlation of the random effects $\rho_{12} \in \{-0.75, -0.5, -0.25, 0, 0.25, 0.5, 0.75\}$.

We simulate a MMFH model by assuming the direct estimates of variable 1 are missing for the first $D/2$ domains. The set of domains $\mathcal{D} = \{1, \ldots, D\}$ can then be partitioned into the two subsets:

$\mathcal{D}^{\text{V}_1.\text{mis}} = \{1, \ldots, D/2\}$ contains all $D/2$ domains for which variable 1 is missing.

$\mathcal{D}^{\mathrm{V}_1.\mathrm{obs}} = \{(D/2)+1,\dots,D\}$ contains all $D/2$ domains for which variable 1 is observed.

In a similar manner, we can define subsets $\mathcal{D}^{\mathrm{V}_2.\mathrm{obs}} = \mathcal{D}^{\mathrm{V}_3.\mathrm{obs}} = \mathcal{D}$ and $\mathcal{D}^{\mathrm{V}_2.\mathrm{mis}} = \mathcal{D}^{\mathrm{V}_3.\mathrm{mis}} = \emptyset$. Table 6.1 summarises which domain sets the different FH estimators use for parameter estimation and EBLUP calculation and for which domain sets only synthetic predictions can be calculated. FH_{V_k} is the univariate FH estimator for variable $k = 1,2,3$.

Table 6.1: Domain sets used by different FH estimators

FH estimator	Parameter estimation and EBLUPs	Synthetic predictions
FH_{V_1}	$\mathcal{D}^{\mathrm{V}_1.\mathrm{obs}}$	$\mathcal{D}^{\mathrm{V}_1.\mathrm{mis}}$
FH_{V_2}	$\mathcal{D}$	
FH_{V_3}	$\mathcal{D}$	
MFH	$\mathcal{D}^{\mathrm{V}_1.\mathrm{obs}}$	$\mathcal{D}^{\mathrm{V}_1.\mathrm{mis}}$
MMFH	$\mathcal{D}$	

For the univariate FH models, we use R (R Core Team, 2020) package sae (Molina & Marhuenda, 2015) with function mseFH for parameter estimation, the calculation of EBLUPs and MSE estimates. For the MFH and MMFH model, we estimate model parameters with the Fisher-Scoring Algorithms 6.1 and 6.2 and use the EBLUP and MSE formulas of the MMFH model presented in this chapter. Recall that the MFH model is a special case of the MMFH model with no missing direct estimates.

6.7.3 Simulation 1: Parameter estimation

Research question
Simulation 1 investigates the performance of the MMFH ML/REML parameter estimation presented in Section 6.5. In particular, we consider the relative Bias (RBias) and relative root MSE (RRMSE) of the estimators for increasing number of domains D. From theory, RBias and RRMSE of parameter estimators should decrease for increasing D as both ML and REML estimators are designed to be consistent.

The MMFH estimator is capable of using more domain information in the parameter estimation than the FH or MFH estimators. Another focus of Simulation 1 is therefore to compare the performance of the parameter estimation under the competing FH models for data with partially missing direct estimates.

Simulation settings
We set $\rho_{12} = 0.5$, $D \in \{100, 200, 300\}$, and conduct the simulation via Algorithm 6.3.

Algorithm 6.3 Steps of Simulation 1

```
The steps of Simulation 1 are
  1. Generate $\boldsymbol{x}_{dk}$, $d = 1, \dots, D$, $k = 1, 2, 3$.
  2. For $i = 1, \dots, I$, $I = 1,000$, do
       a) Generate $\boldsymbol{u}_d^{(i)} \sim N_3(\boldsymbol{0}, \boldsymbol{V}_{ud})$, $\boldsymbol{e}_d^{(i)} \sim N_3(\boldsymbol{0}, \boldsymbol{V}_{ed})$.
       b) For $d = 1, \dots, D$, calculate
              $\boldsymbol{\mu}_d^{(i)} = \boldsymbol{X}_d\boldsymbol{\beta} + \boldsymbol{u}_d^{(i)}, \quad \boldsymbol{y}_d^{(i)} = \boldsymbol{X}_d\boldsymbol{\beta} + \boldsymbol{u}_d^{(i)} + \boldsymbol{e}_d^{(i)}.$
       c) For $d \in \mathcal{D}^{\mathrm{V_1.mis}}$, set the direct estimates of variable 1, $y_{d1}^{(i)}$, missing.
       d) For every $\eta \in \{\boldsymbol{\beta}, \boldsymbol{\theta}\}$, calculate estimates $\hat{\eta}^{(i),a,b}$, $b \in \{\mathrm{ML}, \mathrm{REML}\}$:
            - FH models $\mathrm{FH}_{V_k}$, $k = 1, 2, 3$,   $a = \mathrm{FH}_{V_k}$,   based on $\mathcal{D}^{V.k.obs}$,
            - MFH model,                                      $a = \mathrm{MFH}$,       based on $\mathcal{D}^{\mathrm{V_1.obs}}$,
            - MMFH model,                                     $a = \mathrm{MMFH}$,      based on $\mathcal{D}$.
  3. For every $\eta \in \{\boldsymbol{\beta}, \boldsymbol{\theta}\}$, calculate the following performance measures for estimators
     $\hat{\eta}^{a,b}$, $a \in \{\mathrm{MMFH}, \mathrm{FH}_{V_k}, \mathrm{MFH}\}$, $b \in \{\mathrm{ML}, \mathrm{REML}\}$, $k = 1, 2, 3$,
  $\mathrm{RBias}(\hat{\eta}^{a,b}) = 100\frac{I^{-1}\sum_{i=1}^{I}(\hat{\eta}^{(i),a,b} - \eta)}{|\eta|}, \ \mathrm{RRMSE}(\hat{\eta}^{a,b}) = 100\frac{\left(I^{-1}\sum_{i=1}^{I}(\hat{\eta}^{(i),a,b} - \eta)^2\right)^{1/2}}{|\eta|}.$
```

Results

For the evaluation of the parameter estimation we focus on the parameters associated with variable 1, i.e. $\eta \in \{\beta_{11}, \beta_{12}, \sigma_{u1}^2, \rho_{12}, \rho_{13}\}$ since missing direct estimates are simulated for variable 1 only. Tables 6.2, 6.3, 6.4, and 6.5 present the RBias and RRMSE (in %) of the REML and ML parameter estimation under the different FH models for increasing number of domains D. Note that the (univariate) FH model does not give parameter estimates for ρ_{12} and ρ_{13}, wherefore the corresponding spots are left blank.

The MMFH ML and REML Fisher-Scoring algorithm work properly as the RRMSE of the parameter estimates decreases for increasing number of domains D. For ML, we see a bias in the variance component estimation which decreases as D increases, which is consistent with the ML theory. For REML, the bias is close to zero except for ρ_{12}, for which, however, the bias decreases as the number of domains increases. In terms of the RRMSE, the performance of the FH_{V_1}, MFH, and MMFH parameter estimation is quite similar for β_{11}, β_{12}, and σ_{u1}^2. For the estimation of the random effect correlations ρ_{12} and ρ_{13}, the MMFH model gives more efficient estimates (for ML and REML) than the MFH model as it takes into account all available information. Hence, for the parameter estimation the proposed MMFH model should be preferred over the FH and MFH model.

Table 6.2: RBias (in %) of REML parameter estimation

η	Value	Model	$D = 100$	$D = 200$	$D = 300$
β_{11}	2	FH_{V_1}	-2.265	-1.318	0.916
		MFH	-2.083	-0.982	0.938
		MMFH	-2.221	-1.067	0.886
β_{12}	3	FH_{V_1}	0.014	0.013	-0.008
		MFH	0.012	0.008	-0.008
		MMFH	0.013	0.009	-0.008
σ^2_{u1}	2	FH_{V_1}	0.164	-0.900	0.307
		MFH	0.297	-0.897	0.318
		MMFH	0.512	-0.833	0.274
ρ_{12}	0.5	FH_{V_1}			
		MFH	5.776	4.963	3.538
		MMFH	4.836	4.449	2.499
ρ_{13}	0.75	FH_{V_1}			
		MFH	-6.805	-0.323	0.222
		MMFH	-5.681	-0.394	-0.177

Table 6.3: RRMSE (in %) of REML parameter estimation

η	Value	Model	$D = 100$	$D = 200$	$D = 300$
β_{11}	2	FH_{V_1}	36.386	23.534	21.607
		MFH	36.174	24.047	21.203
		MMFH	36.261	24.263	21.229
β_{12}	3	FH_{V_1}	0.411	0.259	0.230
		MFH	0.408	0.264	0.226
		MMFH	0.408	0.264	0.226
σ^2_{u1}	2	FH_{V_1}	53.394	36.640	29.501
		MFH	53.224	36.620	29.500
		MMFH	53.647	36.659	29.435
ρ_{12}	0.5	FH_{V_1}			
		MFH	75.074	56.581	49.788
		MMFH	73.567	54.720	46.459
ρ_{13}	0.75	FH_{V_1}			
		MFH	38.718	28.318	23.029
		MMFH	37.788	27.586	23.094

Table 6.4: RBias (in %) of ML parameter estimation

η	Value	Model	$D = 100$	$D = 200$	$D = 300$
β_{11}	2	FH_{V_1}	-2.265	-1.318	0.916
		MFH	-1.815	-1.446	0.867
		MMFH	-1.801	-1.449	0.878
β_{12}	3	FH_{V_1}	0.014	0.013	-0.008
		MFH	0.009	0.014	-0.007
		MMFH	0.010	0.014	-0.007
σ^2_{u1}	2	FH_{V_1}	-9.722	-5.879	-3.030
		MFH	-9.017	-5.602	-2.833
		MMFH	-8.597	-5.409	-2.781
ρ_{12}	0.5	FH_{V_1}			
		MFH	11.710	9.028	6.665
		MMFH	9.500	7.423	4.671
ρ_{13}	0.75	FH_{V_1}			
		MFH	-4.066	2.065	2.245
		MMFH	-3.118	1.609	1.443

Table 6.5: RRMSE (in %) of ML parameter estimation

η	Value	Model	$D = 100$	$D = 200$	$D = 300$
β_{11}	2	FH_{V_1}	36.386	23.534	21.607
		MFH	35.524	22.801	20.291
		MMFH	35.536	22.597	20.199
β_{12}	3	FH_{V_1}	0.411	0.259	0.230
		MFH	0.398	0.246	0.212
		MMFH	0.397	0.246	0.212
σ^2_{u1}	2	FH_{V_1}	51.956	36.367	29.263
		MFH	51.722	36.375	29.284
		MMFH	52.140	36.406	29.242
ρ_{12}	0.5	FH_{V_1}			
		MFH	75.930	58.126	50.955
		MMFH	75.105	55.746	47.324
ρ_{13}	0.75	FH_{V_1}			
		MFH	39.067	28.589	23.135
		MMFH	37.832	27.625	23.120

6.7.4 Simulation 2: EBLUPs and MSE estimates

Research question
Simulation 2 investigates the performance of the EBLUPs under the introduced MMFH model and their MSE estimates under different correlation settings and for varying number of domains D. Furthermore, Simulation 2 investigates the performance of the proposed MSE estimators for MFH synthetic predictors (5.206). In addition, we compare the performance of the different FH estimators with the introduced MMFH estimator.

Simulation settings
We choose $\rho_{12} \in \{-0.75, -0.5, -0.25, 0, 0.25, 0.5, 0.75\}$, $D \in \{100, 200, 300\}$ and conduct the simulation via Algorithm 6.4. In this simulation, we only consider parameter estimation via REML Fisher-Scoring.

Results for the domains without missing values, $\mathcal{D}^{\mathrm{V}_1.\mathrm{obs}}$
We first focus on the domains in the set $\mathcal{D}^{\mathrm{V}_1.\mathrm{obs}}$, where all direct estimates are available. For this set, we can calculate EBLUPs for all three dependent variables with all three FH models (FH$_{V_k}$, $k = 1, 2, 3$, MFH, and MMFH). Note that the parameter estimations are based on different sets of domains due to the missing values of variable 1, Table 6.1 gives an overview.

The performance of the EBLUPs is evaluated for $\rho_{12} = 0.5$ with varying number of domains D. Tables 6.6 and 6.7 show the mean RBias and RRMSE (in %) of the EBLUPs calculated for the domains in $\mathcal{D}^{\mathrm{V}_1.\mathrm{obs}}$. Note that, unlike for the parameter estimation, we do not expect the RRMSE of the EBLUPs to decrease with increasing D as with increasing D also the number of predictions increases. For all three FH models, the RBias of the EBLUPs is close to zero. In terms of RRMSE, the MFH and MMFH models are more efficient than the univariate FH models. The performance of the MFH and MMFH model are close.

Tables 6.8 and 6.9 show the mean RBias and MSE of the MSE estimates. Both RBias and MSE tend towards zero for increasing D. However, we note that in terms of RBias, the MSE estimator of the univariate FH model shows a better behaviour than the MSE estimator of the MMFH model. Nevertheless, we consider the proposed MSE estimators for the MMFH model to be acceptable.

Results for variable 1 in domains with missing values, $\mathcal{D}^{\mathrm{V}_1.\mathrm{mis}}$
Next, we focus on the domains in the set $\mathcal{D}^{\mathrm{V}_1.\mathrm{mis}}$, i.e. those domains for which the direct estimates of variable 1 are considered missing. Again, the evaluation is done for $\rho_{12} = 0.5$ with varying number of domains D. The advantage of the proposed MMFH estimator is not only that it can incorporate the full information in the parameter estimation, but also that it allows to calculate EBLUPs for variable 1 in domains $\mathcal{D}^{\mathrm{V}_1.\mathrm{mis}}$, where the FH$_{V_1}$ and MFH model only allow to calculate synthetic predictions. Table 6.10 shows the performance of the different models for the point and MSE estimation of variable 1 only. The brackets behind the models indicate whether the point estimates are synthetic predictions or EBLUPs. For the point estimates, we see that the RBias of the estimators is close to zero. The RRMSE of the point estimates from the MMFH model (which are

Algorithm 6.4 Steps of Simulation 2

```
The steps of Simulation 2 are
  1. Generate $\boldsymbol{x}_{dk}$, $d = 1, \dots, D$, $k = 1, 2, 3$.
  2. For $i = 1, \dots, I$, $I = 1,000$, do
     a) For $d = 1, \dots, D$, generate $\boldsymbol{u}_d^{(i)} \sim N_3(\boldsymbol{0}, \boldsymbol{V}_{ud})$, $\boldsymbol{e}_d^{(i)} \sim N_3(\boldsymbol{0}, \boldsymbol{V}_{ed})$ and calculate
        $\boldsymbol{\mu}_d^{(i)} = \boldsymbol{X}_d\boldsymbol{\beta} + \boldsymbol{u}_d^{(i)}, \boldsymbol{y}_d^{(i)} = \boldsymbol{X}_d\boldsymbol{\beta} + \boldsymbol{u}_d^{(i)} + \boldsymbol{e}_d^{(i)}$.
     b) For $d \in \mathcal{D}^{\mathrm{V_1.mis}}$, set the direct estimates of variable 1, $y_{d1}^{(i)}$, missing.
     c) Marginal FH models $\mathrm{FH}_{V_k}$, $k = 1, 2, 3$,
        i. Calculate FH REML estimates of $\sigma_{uk}^2$ and $\boldsymbol{\beta}_k$, based on $\mathcal{D}^{V.k.obs}$.
        ii. For $d \in \mathcal{D}^{V.k.obs}$, calculate EBLUPs $\hat{\mu}_{dk}^{\mathrm{FH}_{V_k}(i)}$ (2.90) and MSE estimates
            $\widehat{\mathrm{MSE}}_{dk}^{\mathrm{FH}_{V_k}(i)} = \widehat{\mathrm{MSE}}(\hat{\mu}_{dk}^{\mathrm{FH}_{V_k}(i)})$ (2.96).
        iii. For $d \in \mathcal{D}^{V.k.mis}$, calculate synthetic predictors $\hat{\mu}_{dk}^{\mathrm{synFH}_{V_k}(i)}$ (2.90) and MSE
             estimates $\widehat{\mathrm{MSE}}_{dk}^{\mathrm{synFH}_{V_k}(i)} = \widehat{\mathrm{MSE}}(\hat{\mu}_{dk}^{\mathrm{synFH}_{V_k}(i)})$ using (6.255) with $m = 1$.
        iv. Combine all the estimates of the marginal FH models according to the
            scheme $\hat{\boldsymbol{\mu}}_d^{\mathrm{FH}(i)} = (\hat{\mu}_{d1}^{\mathrm{FH}_{V_1}(i)}, \hat{\mu}_{d2}^{\mathrm{FH}_{V_2}(i)}, \hat{\mu}_{d3}^{\mathrm{FH}_{V_3}(i)})^\top$.
     d) MFH model
        i. Calculate MFH REML estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ based on $\mathcal{D}^{\mathrm{V_1.obs}}$.
        ii. For $d \in \mathcal{D}^{\mathrm{V_1.obs}}$, calculate EBLUPs $\hat{\boldsymbol{\mu}}_d^{\mathrm{MFH}(i)}$ (5.199) and MSE estimates
            $\widehat{\mathrm{MSE}}_d^{\mathrm{MFH}(i)} = \widehat{\mathrm{MSE}}_d(\hat{\boldsymbol{\mu}}_d^{\mathrm{MFH}(i)})$ (5.206).
        iii. For $d \in \mathcal{D}^{\mathrm{V_1.mis}}$, calculate synthetic predictors in $\hat{\mu}_{dk}^{\mathrm{synMFH}(i)}$ (6.250) and
             MSE estimates $\widehat{\mathrm{MSE}}_d^{\mathrm{synMFH}(i)} = \widehat{\mathrm{MSE}}(\hat{\mu}_{dk}^{\mathrm{synMFH}(i)})$ (6.255).
     e) MMFH model
        i. Calculate MMFH REML estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ based on $\mathcal{D}$.
        ii. For $d \in \mathcal{D}$, calculate EBLUPs $\hat{\boldsymbol{\mu}}_d^{\mathrm{MMFH}(i)}$ (6.281) and MSE estimates
            $\widehat{\mathrm{MSE}}_d^{\mathrm{MMFH}(i)} = \widehat{\mathrm{MSE}}_d(\hat{\boldsymbol{\mu}}_d^{\mathrm{MMFH}(i)})$ (6.360).
  3. For $a \in \{\mathrm{FH}, \mathrm{synFH}, \mathrm{MFH}, \mathrm{synMFH}, \mathrm{MMFH}\}$, $d = 1, \dots, D$, calculate

     $\mathrm{MSE}(\hat{\boldsymbol{\mu}}_d^a) = I^{-1}\sum_{i=1}^{I}(\hat{\boldsymbol{\mu}}_d^{a(i)} - \boldsymbol{\mu}_d^{a(i)})^2, \ \mathrm{RBias}(\hat{\boldsymbol{\mu}}_d^a) = 100\frac{I^{-1}\sum_{i=1}^{I}(\hat{\boldsymbol{\mu}}_d^{a(i)} - \boldsymbol{\mu}_d^{a(i)})}{I^{-1}\sum_{i=1}^{I}\boldsymbol{\mu}_d^{a(i)}},$

     $\mathrm{RRMSE}(\hat{\boldsymbol{\mu}}_d^a) = 100\frac{\Big(\mathrm{MSE}(\hat{\boldsymbol{\mu}}_d)\Big)^{1/2}}{|I^{-1}\sum_{i=1}^{I}\boldsymbol{\mu}_d^{a(i)}|}, \ \widehat{\mathrm{MSE}}_d^{a*} = I^{-1}\sum_{i=1}^{I}\widehat{\mathrm{MSE}}_d^{a(i)},$

     $\mathrm{RBias}(\widehat{\mathrm{MSE}}_d^a) = 100\frac{\widehat{\mathrm{MSE}}_d^{a*} - \mathrm{MSE}(\hat{\boldsymbol{\mu}}_d^a)}{\mathrm{MSE}(\hat{\boldsymbol{\mu}}_d^a)}, \ \mathrm{MSE}(\widehat{\mathrm{MSE}}_d^a) = \widehat{\mathrm{MSE}}_d^{a*} - \mathrm{MSE}(\hat{\boldsymbol{\mu}}_d^a)^2.$
```

Table 6.6: Mean RBias (in %) of EBLUPs for $d \in \mathcal{D}^{\text{V}_1\text{.obs}}$

Variable	Model	$D = 100$	$D = 200$	$D = 300$
1	FH_{V_1}	0.014	0.021	0.013
	MFH	0.014	0.012	0.011
	MMFH	0.013	0.011	0.009
2	FH_{V_2}	0.007	0.021	0.016
	MFH	0.005	0.018	0.011
	MMFH	0.005	0.019	0.013
3	FH_{V_3}	0.016	0.018	0.013
	MFH	0.003	0.013	0.012
	MMFH	0.015	0.016	0.011

Table 6.7: Mean RRMSE (in %) of EBLUPs for $d \in \mathcal{D}^{\text{V}_1\text{.obs}}$

Variable	Model	$D = 100$	$D = 200$	$D = 300$
1	FH_{V_1}	1.009	1.075	0.969
	MFH	0.879	0.911	0.811
	MMFH	0.872	0.908	0.809
2	FH_{V_2}	0.873	0.945	0.967
	MFH	0.823	0.859	0.871
	MMFH	0.791	0.841	0.856
3	FH_{V_3}	0.967	1.043	0.998
	MFH	0.946	0.982	0.931
	MMFH	0.912	0.961	0.917

Table 6.8: Mean RBias (in %) of the MSE estimates for $d \in \mathcal{D}^{\text{V}_1\text{.obs}}$

Variable	Model	$D = 100$	$D = 200$	$D = 300$
1	FH_{V_1}	1.757	1.225	1.349
	MFH	18.535	7.650	3.879
	MMFH	16.108	6.314	3.231
2	FH_{V_2}	0.389	−0.068	1.076
	MFH	12.456	4.406	2.141
	MMFH	8.748	3.218	2.220
3	FH_{V_3}	0.745	0.794	−0.614
	MFH	10.163	4.971	1.748
	MMFH	7.114	4.227	0.688

Table 6.9: Mean MSE (in %) of the MSE estimates for $d \in \mathcal{D}^{\mathrm{V}_1\text{.obs}}$

Variable	Model	$D = 100$	$D = 200$	$D = 300$
1	FH_{V_1}	0.004	0.003	0.004
	MFH	0.037	0.006	0.003
	MMFH	0.027	0.005	0.002
2	FH_{V_2}	0.003	0.003	0.003
	MFH	0.021	0.004	0.002
	MMFH	0.010	0.003	0.002
3	FH_{V_3}	0.004	0.003	0.004
	MFH	0.018	0.005	0.003
	MMFH	0.009	0.004	0.002

EBLUPs) are always lower than the RRMSEs of the point estimates from the FH_{V_1} and MFH model (which are synthetic predictions). For partially missing direct estimates, applying the MMFH model to get EBLUPs for these values is therefore to be preferred over the synthetic predictions.

From Table 6.10, we also see that the MSE estimation of the MMFH model works properly for domain set $\mathcal{D}^{\mathrm{V}_1\text{.mis}}$. In Section 6.2.2, we also introduced an MSE estimator for synthetic predictions obtained from a MFH model. Table 6.10 shows that these MSE estimates work properly in terms of RBias and MSE.

Lastly, we look at the performance gains in terms of RRMSE from using the MMFH instead of the FH_{V_1} and MFH model in Table 6.11. The table displays the mean RRMSE of the models divided by the corresponding mean RRMSE of the MMFH model. The values are shown for $D = 200$ domains and varying random effect correlation ρ_{12}. The performance gain of using the MMFH model for predicting the values of variable 1, where direct estimates are missing, is high for all different scenarios of random effect correlation ρ_{12}. Even when the random effect correlation ρ_{12} is zero, the model still gains from using the non-zero random effect correlation of variables 1 and 3 and the sampling error correlation to variable 2 for the construction of the EBLUPs.

Table 6.10: Mean performance of the predictions and MSE estimates for variable 1 for $d \in \mathcal{D}^{\mathrm{V_1.mis}}$

Estimates	Measure	Model		$D = 100$	$D = 200$	$D = 300$
EBLUPS/ Synthetic predictions	RBias	FH$_{V_1}$	(Synthetic)	0.012	0.017	0.016
		MFH	(Synthetic)	0.014	0.017	0.017
		MMFH	(EBLUP)	0.019	0.009	0.011
	RRMSE	FH$_{V_1}$	(Synthetic)	1.167	1.250	1.207
		MFH	(Synthetic)	1.166	1.250	1.208
		MMFH	(EBLUP)	1.053	1.087	1.040
MSE	RBias	FH$_{V_1}$	(Synthetic)	1.122	1.873	0.712
		MFH	(Synthetic)	0.783	1.534	0.440
		MMFH	(EBLUP)	10.430	6.136	1.747
	MSE	FH$_{V_1}$	(Synthetic)	0.008	0.012	0.009
		MFH	(Synthetic)	0.008	0.011	0.009
		MMFH	(EBLUP)	0.037	0.014	0.005

Table 6.11: Mean RRMSE of FH synthetic predictions divided by mean RRMSE of MMFH EBLUPs for variable 1 for $d \in \mathcal{D}^{\mathrm{V_1.mis}}$

				ρ_{12}			
Model	-0.75	-0.5	-0.25	0	0.25	0.5	0.75
FH$_{V_1}$	1.223	1.198	1.135	1.106	1.108	1.149	1.242
MFH	1.219	1.195	1.133	1.105	1.107	1.150	1.244

6.8 Application

6.8.1 Data description

As an illustrative example, we apply the proposed MMFH model to publicly available county-level survey estimates of the *median annual income (dollars) Hispanic or Latino origin (of any race)* (`HC02_EST_VC12`) in 2010 and 2011 from the U.S. *American Community Survey* (ACS). The application is in part published in Burgard et al. (2021c).

The ACS is the largest official U.S. household survey. Every year, over 3.5 million randomly drawn households are contacted to participate in the survey. The focus of the ACS is on the estimation of social, housing, economic, and demographic characteristics. The U.S. Census Bureau provides ACS 1-year, 3-year, and 5-year estimates at different demographic and regional levels. A detailed description of the design and methodology of the survey can be found in U.S. Census Bureau (2014).

In the publications of U.S. official statistics, special emphasise is put on statistics for the population of Hispanic or Latino origin, e.g. Guzman (2019), Semega et al. (2019), and U.S. Census Bureau (2014). We therefore consider the *median annual income (dollars) Hispanic or Latino origin (of any race)* `HC02_EST_VC12` in 2010 and 2011 as the two dependent variables of a bivariate FH model. The domains of interest are the $D = 3,141$ U.S. counties. The ACS 1-year county-level estimates of `HC02_EST_VC12` in 2010 and 2011 can be downloaded from the U.S. Census Bureau website.[1] In addition to the survey estimates, their margins of error are also provided. The margins of error were calculated as $1.645 \times \sqrt{variance}$ (U.S. Census Bureau, 2014, Chapter 12.3).

For the 1-year county-level estimates of `HC02_EST_VC12`, it is noticeable that for many counties the survey estimates are missing, either for 2010 or 2011 or both. This is due to two reasons, for a detailed overview of the ACS data suppression we refer to U.S. Census Bureau (2016). First, the ACS 1-year estimates are only published for geographical entities with populations of minimum $65,000$. Second, the publications depend on the samples on which the estimates are based. For example, U.S. Census Bureau (2016) list minimum cell counts and thresholds for the margins of error as reasons for the suppression of 1-year estimates. As county-level sample sizes are random, it appears that survey estimates of `HC02_EST_VC12` in a county may be available in 2010, but not in 2011 and vice versa. In total, estimates of `HC02_EST_VC12` are available for 704 counties in 2010 and 684 counties in 2011. There are 58 counties for which the survey estimate of `HC02_EST_VC12` is missing in 2010, but available in 2011. The other way around, there are 78 counties for which the survey estimate of `HC02_EST_VC12` is missing in 2011, but available in 2010. For 762 counties at least one survey estimate of `HC02_EST_VC12` is available. There are only 626 counties for which `HC02_EST_VC12` is available in 2010 and 2011.

We see that the survey estimates of `HC02_EST_VC12` in 2010 and 2011 are partly missing or associated with large margins of error. This is exactly the situation for which we

[1] U.S. Census Bureau website https://data.census.gov/cedsci/, TableID: S1903.

introduced the MMFH model in this chapter. With the MMFH model, we can calculate model-based small area estimates of these county-level values.

6.8.2 Model choice

We calculate a MMFH model for the 762 counties for which at least one survey estimate of `HC02_EST_VC12` is available. We therefore have $D = 762$ domains of interest. For comparison, the general MFH model for the two variables could only be applied to the 626 counties for which `HC02_EST_VC12` is available in 2010 and 2011. Two univariate FH models could only take into account the 704 domains where `HC02_EST_VC12` 2010 or the 684 domains where `HC02_EST_VC12` 2011 is available respectively.

The MMFH model is calculated with $m = 2$ dependent variables, the ACS 1-year county-level estimates of variable `HC02_EST_VC12` in 2010 and 2011. The covariance matrices of the survey estimates are given by

$$\boldsymbol{V}_{ed} = \begin{pmatrix} \sigma_{ed1}^2 & 0 \\ 0 & \sigma_{ed2}^2 \end{pmatrix}, \; d = 1, \ldots, D, \tag{6.364}$$

where σ_{ed1}^2 and σ_{ed2}^2 are calculated as (margin of error/1.645)^2 of the corresponding estimates. The off-diagonal elements of $\boldsymbol{V}_{ed}$ are zero as the covariances of the sampling errors of the direct estimates between 2010 and 2011 are expected to be zero.

As auxiliary information for the MMFH model, we consider publicly available county-level data from the United States Census Bureau[2]. Note that the public availability of suitable county-level auxiliary information is limited, but should suffice for an illustration of the MMFH model. The model parameters are estimated via the REML Fisher-Scoring Algorithm 6.2. After considering different models for `HC02_EST_VC12` 2010 and 2011, we choose as auxiliary variables: the death rate in period 7/1/2010 to 6/30/2011 (`RDEATH2011`) and the civilian labour force unemployment rate 2010 RTE (`CLF040210D`) for both dependent variables.

6.8.3 Parameter estimates

We first discuss the estimated variances components and fixed effects of the fitted MMFH model. Table 6.12 shows the estimated variance components with additional 95% confidence intervals. None of the confidence intervals contains zero and the confidence intervals are relatively small, indicating that the random effects in the model contribute to variance identification of the dependent variables. As we model the same variable `HC02_EST_VC12` in two consecutive years as the dependent variables in the model, we would expect the

[2] The auxiliary data used are available at https://www.census.gov/. We consider: (1) U.S. county data files on https://www.census.gov/library/publications/2011/compendia/usa-counties-2011.html, and (2) county population totals and components of change on https://www.census.gov/data/tables/time-series/demo/popest/2010s-counties-total.html.

random effects correlation of HC02_EST_VC12 2010 and 2011 to be highly positive. This expectation is confirmed by model estimate $\hat{\rho}_{12} = 0.95$, shown in Table 6.12.

Table 6.13 displays the estimated fixed effects for HC02_EST_VC12 2010 and 2011. The p-values are all zero, rounded to the fifth digit. To put it differently, all estimated fixed effects are highly significant. The estimated fixed effects of HC02_EST_VC12 2010 and 2011 are very similar as would be expected as they represent the same variable in two consecutive years. From the estimated coefficients, we can see that counties with higher death rates and higher civilian labour force unemployment rate tend to have lower values of HC02_EST_VC12 which seems plausible. As this is an illustrative example and the choice of publicly available county-level data was limited, the coefficient should, however, be treated with caution.

Table 6.12: Variance component estimates and asymptotic 95% confidence intervals

	$\hat{\theta}$	Lower limit	Upper limit
$\hat{\sigma}_{u1}$	98,348,357	98,348,357	98,348,357
$\hat{\sigma}_{u2}$	92,409,379	92,409,379	92,409,379
$\hat{\rho}_{12}$	0.97	0.95	0.98

Table 6.13: Estimated fixed effects for HC02_EST_VC12 2010 and 2011

	HC02_EST_VC12 2010			
	$\hat{\beta}$	std.err.	t-value	p-value
(Intercept)	61,629.72	2,151.51	28.64	0
RDEATH2011	-1,757.83	214.64	-8.19	0
CLF040210D	-806.16	163.11	-4.94	0
	HC02_EST_VC12 2011			
	$\hat{\beta}$	std.err.	t-value	p-value
(Intercept)	61,122.45	2,076.02	29.44	0
RDEATH2011	-1,722.71	205.66	-8.38	0
CLF040210D	-824.08	158.03	-5.21	0

6.8.4 Model diagnostics

In Figure 6.1, different diagnostics are presented to check the validity of the calculated MMFH model. On the left and right hand side, the diagnostics are displayed for HC02_EST_VC12 2010 and 2011 respectively.

From theory, the design-based survey estimates may exhibit large variances, but are design-unbiased. Brown et al. (2001) therefore proposed to plot the direct estimates against the EBLUPs to see whether the model-based estimates systematically differ from the direct estimates. In row one, the direct estimates are plotted against the MMFH EBLUPs with an additional diagonal line $y = x$. For some very large direct estimates, the model-based EBLUPs are systematically lower. We note that these direct estimates are associated with high standard errors, wherefore the MMFH EBLUPs are smoothed more towards the model-based predictions. We therefore do not see indications of a model bias from the figures.

Next, we focus on the normality assumption of the model in row two of Figure 5.5. The MMFH EBLUPs are plotted against standardized residuals. From theory, we assume the residuals to be normally distributed with zero mean. Furthermore, the residuals should not indicate systematic differences for increasing EBLUP values. The residuals are calculated as $r_{dk}^{\text{EBLUP}} = \hat{\mu}_{dk}^{\text{Dir}} - \hat{\mu}_{dk}^{\text{EBLUP}}$, where $\hat{\mu}_{dk}^{\text{Dir}}$ are the ACS direct estimates and $\hat{\mu}_{dk}^{\text{EBLUP}}$ are the MMFH EBLUPs, $k = 1, 2$, $d = 1, \dots, D$. The standardized residuals are calculated as $(r_{dk}^{\text{EBLUP}} - D^{-1}\sum_{d=1}^{D} r_{dk}^{\text{EBLUP}})/std(r_{dk}^{\text{EBLUP}})$, where $std(r_{dk}^{\text{EBLUP}})$ is the standard deviation of the set of residuals, $k = 1, 2$, $d = 1, \dots, D$. For both variables `HC02_EST_VC12` 2010 and 2011 most residuals are within range. However, there are some standardized residuals with high absolute values. In total, there are 2.27% (16/704) values with absolute values larger than three for `HC02_EST_VC12` 2010. For `HC02_EST_VC12` 2011, there are 1.75% (12/684) values with absolute values larger than three. For the application, a further treatment of these outliers is necessary. This, however, is beyond the scope of this illustrative example of the proposed MMFH model and left as a potential future research topic. For robust SAE, we refer to Sinha and Rao (2009), Schmid and Münnich (2014) for robust SAE including spatial effects, and Baldermann et al. (2018) for the additional consideration of spatial non-stationarity. `R` (R Core Team, 2020) package `rsae` (Schoch, 2014) provides an implementation of robust FH models.

The plots in row three of Figure 5.5 show the standard error of the direct estimates versus the root MSEs (RMSEs) of the MMFH EBLUPs. Diagonal line $y = x$ is added to the plot. The MMFH EBPLUPs are always at least as efficient as the direct estimators as all values are along or below the diagonal line. For large standard errors of direct estimates, the EBLUPs give high efficiency gains over the direct estimators. For these cases, the MMFH model puts more weight on the estimated model than on the direct estimates.

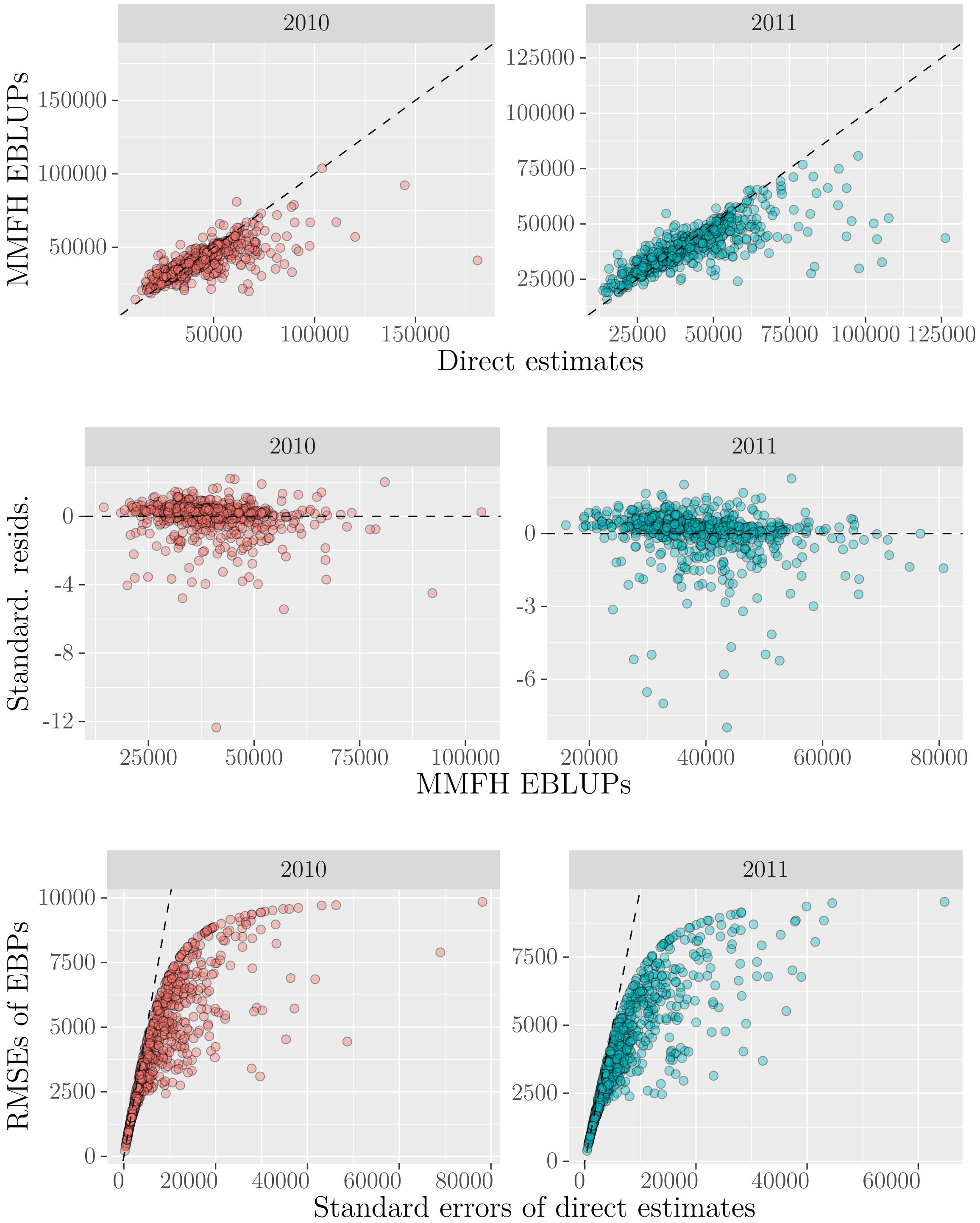

Figure 6.1: Model diagnostics for `HC02_EST_VC12` 2010 (left) and 2011 (right)

6.8.5 MMFH EBLUPs

Figure 6.2 displays the ACS 1-year estimates and the MMFH EBLUPs of HC02_EST_VC12 in 2010 and 2011. The values are shown only for the counties in the two neighbouring states of Indiana and Ohio instead of all U.S. states so that the values in the individual counties can be seen. The counties coloured white are those for which ACS 1-year estimates are not available in 2020 and 2011. For each year, we have outlined in red the counties for which an ACS 1-year estimate is missing for that year but is available for the other year. These are the counties for which the MMFH model presented in this chapter allows the calculation of EBLUPs as can be seen in the second row of the plot. We can see that the introduced MMFH estimator can provide a significant gain in information for concrete applications where survey estimates are often partially missing.

To evaluate whether the MMFH EBLUPs are plausible, especially for the counties with missing ACS 1-year estimates, we use additional information. Next to the 1-year estimates which were used in the MMFH model, also 3-year and 5-year ACS estimates of HC02_EST_VC12 are published by the U.S. Census Bureau. The 3- and 5-year estimates are based on sample information of three and five years and available for more counties than the 1-year estimates, compare the publication restrictions for them in U.S. Census Bureau (2016). We note that it is not recommended to directly compare ACS 1-, 3-, and 5-year estimates. Nevertheless, for evaluating whether the MMFH EBLUPs are realistic, especially the ones for missing ACS 1-year estimates, we consider the ACS 5-year estimates of HC02_EST_VC12 to be suitable as benchmarks.[3] The ACS 5-year estimates have lower sample variances and they are available for many counties where ACS 1-year estimates are missing but MMFH EBLUPs can be calculated.

ACS 5-year estimates of HC02_EST_VC12 are available for some counties for which ACS 1-year estimates are missing. However, also ACS 5-year estimates of HC02_EST_VC12 are not available for all U.S. counties. We therefore additionally use the Census ACS estimates for 2010 to validate the MMFH EBLUPs. Census estimates for variable HC02_EST_VC12 are not available. Instead, we use Census estimates of the *Median household income in the last 12 months (in 2009 inflation-adjusted dollars) in 2005-2009* (INC110209D). Variables HC02_EST_VC12 and INC110209D are close in definition and INC110209D estimates for 2005-2009 are available for all U.S. counties.[4]

We reorder the domains such that we can partition the set of domains $\mathcal{D} = \{1, \ldots, D\}$ into two subsets according to the domains with available and missing ACS 1-year estimates of HC02_EST_VC12 for each year:

2010: $\mathcal{D}^{\text{mis}\,2010} = \{1, \ldots, D^{\text{mis}}\}$ with $D^{\text{mis}\,2010} \leq D$,
$\mathcal{D}^{\text{obs}\,2010} = \{D^{\text{mis}\,2010} + 1, \ldots, D\}$.
2011: $\mathcal{D}^{\text{mis}\,2011} = \{1, \ldots, D^{\text{mis}}\}$ with $D^{\text{mis}\,2011} \leq D$,
$\mathcal{D}^{\text{obs}\,2011} = \{D^{\text{mis}\,2011} + 1, \ldots, D\}$.

[3] The ACS 5-year estimates are available at the U.S. Census Bureau website https://data.census.gov/cedsci/, TableID: S1903.

[4] The Census-ACS estimates are available at the U.S. Census Bureau website https://www.census.gov/library/publications/2011/compendia/usa-counties-2011.html.

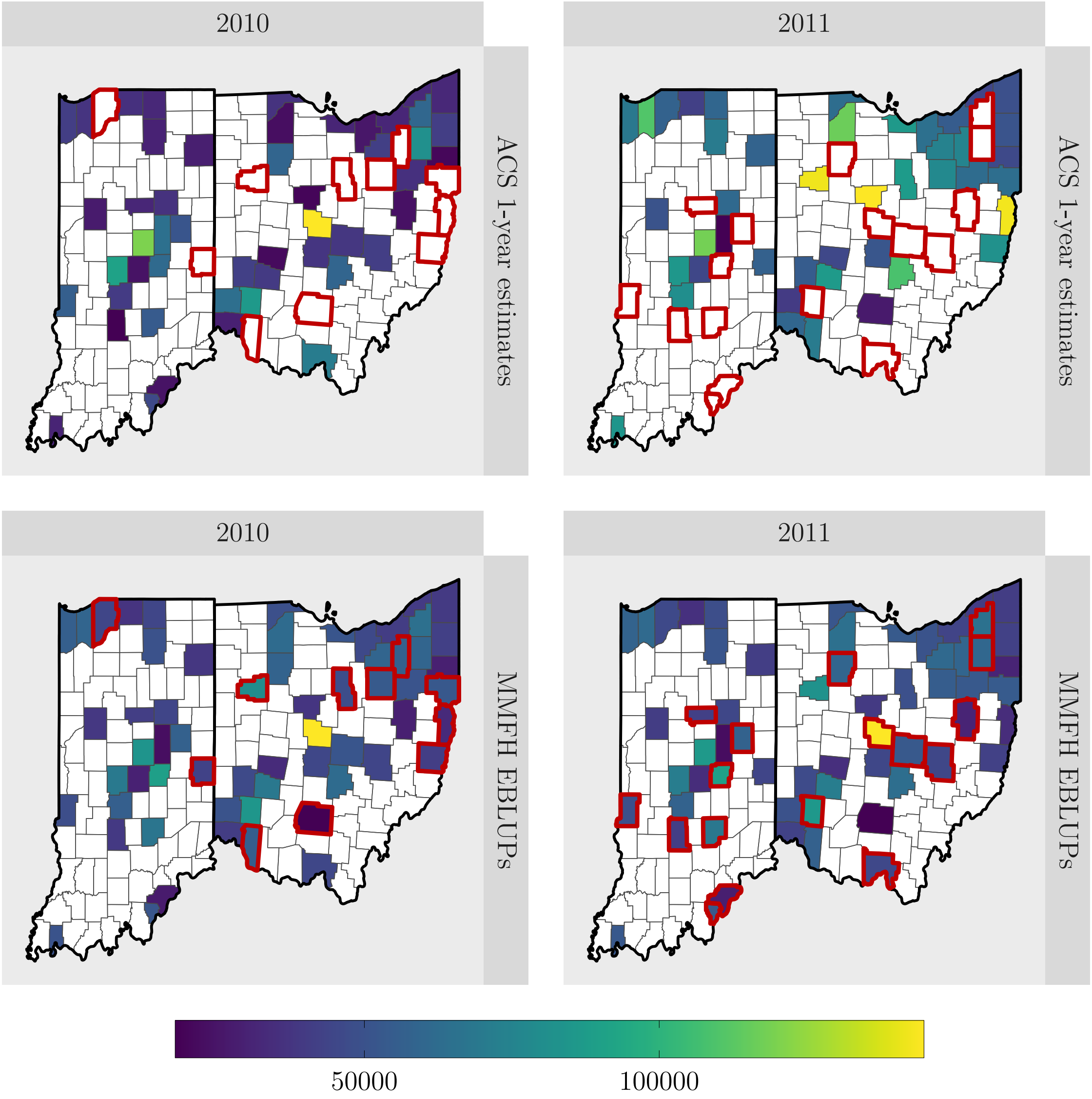

Figure 6.2: ACS 1-year estimates and MMFH EBLUPs of `HC02_EST_VC12` for the counties in Indiana and Ohio

Tables 6.14 and 6.15 show the quantiles of the relative difference of ACS 1-year estimates and MMFH EBLUPs to ACS 5-year estimates and Census estimates of `INC110209D`. For domains $\mathcal{D}^{\text{obs}\,2010}$ and $\mathcal{D}^{\text{obs}\,2011}$, the quantiles of the ACS 1-year estimates and MMFH EBLUPs are close. Taking into account the 95% quantile, the MMFH EBLUPs tend to give less extreme deviations from the 5-year and Census estimates compared to the ACS 1-year estimates, indicating that the model smoothes the predictions more to the model-based part for high sampling variances. For the domains for which no ACS 1-year estimate is available, the relative differences of the MMFH EBLUPs to the ACS 5-year estimates is higher. However, the relative differences of these estimates to the Census `INC110209D` estimates is similar for domains with fully and partially observed ACS 1-year estimates. We therefore see the results as indications that the MMFH EBLUPs of the partially missing values are plausible.

Table 6.14: Quantiles of the relative difference of estimates in 2010 and 2011 to ACS 5-year estimates in 2008-2012 and 2009-2013 of variable `HC02_EST_VC12` (in %)

			Quantiles				
Year	Estimates	Observations	5%	25%	50%	75%	95%
2010	ACS 1-year in $\mathcal{D}^{\text{obs}\,2010}$	704	-32	-12	-3	7	43
	MMFH EBLUPs in $\mathcal{D}^{\text{obs}\,2011}$	704	-32	-13	-5	2	23
	MMFH EBLUPs in $\mathcal{D}^{\text{mis}\,2010}$	58	-55	-26	-13	15	79
2011	ACS 1-year in $\mathcal{D}^{\text{obs}\,2011}$	684	-36	-15	-5	4	41
	MMFH EBLUPs in $\mathcal{D}^{\text{obs}\,2011}$	684	-28	-14	-7	-1	15
	MMFH EBLUPs in $\mathcal{D}^{\text{mis}\,2011}$	78	-45	-16	-6	5	34

Table 6.15: Quantiles of the relative difference of estimates in 2010 and 2011 to Census-ACS 2005-2009 estimates of variable `INC110209D` (in %)

			Quantiles				
Year	Estimates	Observations	5%	25%	50%	75%	95%
2010	ACS 1-year in $\mathcal{D}^{\text{obs}\,2010}$	704	-51	-35	-24	-10	33
	MMFH EBLUPs in $\mathcal{D}^{\text{obs}\,2010}$	704	-46	-33	-25	-16	2
	MMFH EBLUPs in $\mathcal{D}^{\text{mis}\,2010}$	58	-54	-38	-24	-13	8
2011	ACS 1-year in $\mathcal{D}^{\text{obs}\,2011}$	684	-52	-36	-25	-12	28
	MMFH EBLUPs in $\mathcal{D}^{\text{obs}\,2011}$	684	-45	-34	-26	-17	-0
	MMFH EBLUPs in $\mathcal{D}^{\text{mis}\,2011}$	78	-48	-33	-22	-9	6

6.9 Summary and outlook

In this chapter, we extended the multivariate Fay-Herriot model (MFH) to partially missing direct estimates, called MMFH model. MFH models take multi-variate domain survey estimates as input. The parameter estimation and calculation of EBLUPs under the MFH model can, however, only be based on those domains for which survey estimates are available for all dependent variables. In practical applications, such as the publicly available U.S. ACS data presented in Section 6.8, survey estimates are often at least partially missing for some domains of interest. The introduced MMFH model can take all domain information with at least one available survey estimate into account for the parameter estimation and production of EBLUPs. Under the MMFH model, we gave ML and REML fitting algorithms for the parameter estimation, derived EBLUPs, and presented an estimator for their MSEs. The MFH model, in contrast, can only give synthetic predictions whenever direct estimates are missing. For MFH synthetic predictions, we presented an MSE estimator.

In model-based Monte Carlo simulation studies, we showed the validity of the proposed parameter estimation algorithms, the EBLUP formulas and their MSE estimators under the MMFH model. Furthermore, we contrasted the performance and applicability of the MMFH model to the MFH model and the corresponding univariate FH models for each dependent variable. The studies showed that the MMFH EBLUPs for the missing direct estimates are more efficient than the MFH or FH synthetic predictions.

In an application to publicly available U.S. ACS data, we saw not only the practical necessity of the proposed MMFH model, but also evaluated the plausibility of the calculated EBLUPs for the missing direct estimates. Noticeable residuals in the application suggest that further research should investigate robust versions of the MMFH model.

The proposed MMFH model can be directly applied together with the theory in Chapter 5 for non-linear domain indicators under MFH models. Thereby, the proposed MMFH model and approximations of Chapter 5 allow to approximate best predictors of multi-variate domain indicators under partially missing direct estimates.

Chapter 7

Summary and Conclusions

We summarise the contributions of the thesis, which are presented in Chapters 3, 4, 5, and 6, and give an outlook on potential future research.

Chapter 3: Generation of a Longitudinal Employment Dataset for Simulations
In Chapter 3, we extended the cross-sectional RIFOSS dataset with monthly employment information. For that, we edited and aggregated the information in the SIAB dataset. Based on the transformed SIAB dataset, we calculated prediction models for monthly employment transitions. The RIFOSS dataset was then extended with monthly employment information using these prediction models. We validated the generated longitudinal employment information in the RIFOSS dataset and saw that they reflect the patterns in the SIAB dataset at the person-level and the aggregated level as well as aggregate patterns from official statistics. The generated longitudinal RIFOSS dataset served as the simulation population for the design-based studies of Chapter 4.

Although the data generation presented in this chapter was tailored to a specific application, some of the presented concepts provide insights for other application areas and data. This includes the presented discussion of the evaluation of probability predictions for imbalanced categorical data like the employment status. In the course of this discussion, we presented an extension of the Brier score, which we called weighted Brier score. To utilise as much information as possible from the SIAB dataset in the modelling process, we used a combination of the ensemble methods subagging and stacking to calculate the employment prediction models. Based on samples from the SIAB dataset, generalised additive models were calculated (subagging). The weights with which the predictions of these models were incorporated into the overall ensemble predictions were determined by optimised ensemble weights (stacking). The proposed weighted Brier score was used as a loss function in stacking, which constitutes quadratic programming.

In the evaluation of different prediction models, we saw that the proposed optimally weighted ensemble models showed small performance gains over the corresponding equally weighted ensembles and the ensemble input models. Comparing the optimally and equally weighted ensembles, the ensemble weight optimisation also had the advantage that some model weights were close to zero such that the associated individual models could be excluded from the ensemble. This is particularly useful when large amounts of data are generated frequently with these ensembles. For future research, we propose an investigation of sparse ensemble weight optimisation and the consideration of additional methods for handling imbalanced categorical data such as balancing techniques.

Chapter 4: Composite Estimation in the German Microcensus

In Chapter 4, we evaluated the use of composite estimators for the production of employment statistics in the German Microcensus, the design of which underwent major changes in 2020. Composite estimators incorporate additional information from previous samples available in rotating panel surveys (such as the Microcensus) to get more stable estimates. In the Chapter, we analysed the sample overlaps resulting from the sampling design of the Microcensus and how to use them in composite estimators. Furthermore, we presented adjustments to the formulas of the composite estimators to account for the regionally heterogeneous sample overlaps resulting from the Microcensus design.

In a design-based study, we evaluated the performance of the adjusted composite estimators for the production of employment statistics in the Microcensus. The study was conducted on the basis of the RIFOSS dataset, which was extended by longitudinal employment information in Chapter 3. The focus of the estimation was on monthly and quarterly statistics of levels and changes of employed and unemployed persons at the NUTS2-level. In the study, we evaluated the composite estimators with different sets of sample overlaps at different regional levels. The simulation results revealed that the presented adjustments of the composite estimators work properly. Based on the analysis, for the production of NUTS2-level employment estimates we recommend the use of the composite estimators MR2 and $\text{RC}_{\alpha=0.75}$ with sample overlap information at the NUTS2-level from the previous quarter. The simulation also provided insights into the influence of the sampling design and the magnitudes of the sample overlaps on the performance of composite estimators.

In the simulation, we neither considered non-response nor a rotation group bias as there was no historic information available on the two under the new German Microcensus design. In future research, it would be interesting to see how a potential rotation group bias influences the performance of the estimators. Potential future research also involves the investigation of variance estimation procedures for the composite estimators.

Chapter 5: Empirical Best Prediction in Multivariate Fay-Herriot Models

In Chapter 5, we turned to the approximation of multi-variable indicators in MFH models. As the best predictions (BPs) of these indicators in MFH models, in their general form, are given by multi-dimensional integrals, we proposed different approximations of them, including Gauss-Hermite quadrature, Monte Carlo integration with antithetic variates, and Quasi Monte Carlo integration with the Sobol and Halton sequence in combination with different integral forms. Furthermore, we proposed different parametric bootstrap procedures for the MSE estimation.

We conducted several model-based simulation studies tailored to data replicating survey estimates of the proportions of employed and unemployed as dependent variables and the unemployment rate as the indicator of interest. The simulation studies suggested that Gauss-Hermite quadrature is well suited to approximate the BPs under the MFH model for this indicator, already with only few function evaluations. In the simulation studies, we also contrasted the performance of the MFH approximations of the EBPs to the corresponding plug-in predictors. The analysis showed that, while the plug-in predictors have similar RRMSEs, they have larger biases than the approximations of the

EBPs, which is in line with the underlying theory. The studies further showed that the proposed MSE estimation of the EBP approximations works properly and that a separate estimation of components of the MSE gives the best results.

In an illustrative application, where we estimated unemployment rates in Spanish provinces crossed by age and sex classes based on publicly available Spanish labour force (SLFS) data, we showed the applicability of the proposed approach. The application also gave rise to potential future research. As there was no suitable auxiliary information publicly available, we estimated the auxiliary variables from a larger sample of micro-data from the SLFS. In the application, we treated the auxiliary information as if it was estimated without error. For future application, it would be interesting to combine the theory of multi-variable area-level domain indicators and measurement errors in the covariates.

Chapter 6: Multivariate Fay-Herriot Models under Missing Direct Estimates
In Chapter 6, we addressed another research gap in the practical application of MFH models: the consideration of partially missing survey estimates. For domains where even a single direct estimate of the dependent variables is missing, only synthetic predictions can be given by MFH models. We therefore introduced the MFH model with missing survey estimates, which we call MMFH model. For the model, we presented ML and REML Fisher scoring algorithms for parameter estimation, EBLUPs for those domains where at least one survey estimate is available, and formulas for the associated MSE estimation.

In model-based simulation studies, we validated the presented parameter estimation algorithms and formulas for EBLUPs and MSEs. We also contrasted the performance of the presented MMFH model with that of competing FH models. Especially for partially-missing survey estimates, the presented MMFH EBLUPs bring efficiency gains over the synthetic predictions of competing FH models. We illustrated the practical necessity and applicability of the MMFH model using publicly available data from the American Community Survey. In the application, we estimated the median annual income of the population with Hispanic or Latino origin for U.S. counties in 2010 and 2011. In the model validation, we noticed large residuals. Therefore, we consider a combination of the presented MMFH model with robust methods a possible future research topic.

Conclusion
This thesis contributes to the theory of estimation and prediction under incomplete survey data. It demonstrates how composite estimators can use the partially available information from sample overlaps in the German Microcensus to produce efficient labour force statistics (Chapter 4). In addition, the thesis shows how to adapt the formulas of the estimators to account for regionally heterogeneous sample overlaps. The methodological discussion and the procedures applied in the generation of the simulation data for this analysis (Chapter 3) contribute to our understanding of imbalanced data in the evaluation of probability predictions and the combination of ensemble methods with generalised additive models. Together, the developments in Chapters 5 and 6 allow for the approximation of BPs of multi-variable domain indicators under partially missing survey estimates in the proposed MMFH model, which contributes to the active field of research on FH models in small area estimation.

Bibliography

Afentakis, A., & Bihler, W. (2005). Das Hochrechnungsverfahren beim unterjährigen Mikrozensus ab 2005. *Wirtschaft und Statistik*, *10*, 1039–1049.

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). Wiley.

Agresti, A. (2019). *An introduction to categorical data analysis* (3rd ed.). Wiley.

Antoni, M., Ganzer, A., & Vom Berge, P. (2019). Sample of integrated labour market biographies regional file (SIAB-R) 1975–2017. FDZ-Datenreport, 04/2019 (en). The Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB). https://doi.org/10.5164/IAB.FDZD.1904.EN.V1.

Arima, S., Bell, W. R., Datta, G. S., Franco, C., & Liseo, B. (2017). Multivariate Fay-Herriot Bayesian estimation of small area means under functional measurement error. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *180*(4), 1191–1209.

Australian Bureau of Statistics. (2018). Labour statistics. Concepts, sources and methods. ABS Catalogue No. 6102.0.55.001. Australian Bureau of Statistics.

Bailar, B. A. (1975). The effects of rotation group bias on estimates from panel surveys. *Journal of the American Statistical Association*, *70*(349), 23–30.

Baldermann, C., Salvati, N., & Schmid, T. (2018). Robust small area estimation under spatial non-stationarity. *International Statistical Review*, *86*(1), 136–159.

Battese, G. E., Harter, R. M., & Fuller, W. A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, *83*(401), 28–36.

Beaumont, J.-F. (2005). Calibrated imputation in surveys under a quasi-model-assisted approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*(3), 445–458.

Beaumont, J.-F., & Bocci, C. (2005). A refinement of the regression composite estimator in the Labour Force Survey for change estimates. *SSC Annual Meeting. Proceedings of the Survey Methods Section*, 1–6.

Bell, P. A. (1999). The impact of sample rotation patterns and composite estimation on survey outcomes. Working Paper. ABS Catalogue number 1351.0, No. 99/1. Australian Bureau of Statistics.

Bell, P. A. (2001). Comparison of alternative Labour Force Survey estimators. *Survey Methodology*, *27*(1), 53–63.

Bell, S., & Robinson, S. (2020). Small area income and poverty estimates: 2019. U.S. Census Bureau. P30-08 https://www.census.gov/content/dam/Census/library/publications/2020/demo/p30-08.pdf.

Benavent, R., & Morales, D. (2016). Multivariate Fay-Herriot models for small area estimation. *Computational Statistics & Data Analysis*, *94*, 372–390.

Benavent, R., & Morales, D. (2021). Small area estimation under a temporal bivariate area-level linear mixed model with independent time effects. *Statistical Methods & Applications*, *30*(1), 195–222.

Berger, Y. G., & Priam, R. (2016). A simple variance estimator of change for rotating repeated surveys: An application to the European Union Statistics on Income and Living Conditions household surveys. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *179*(1), 251–272.

Berger, Y. G., Muñoz, J. F., & Rancourt, E. (2009). Variance estimation of survey estimates calibrated on estimated control totals. An application to the extended regression estimator and the regression composite estimator. *Computational Statistics & Data Analysis*, *53*(7), 2596–2604.

Bihler, W., & Zimmermann, D. (2016). Die neue Mikrozensusstichprobe ab 2016. *Wirtschaft und Statistik*, *6*, 20–30.

Bonnéry, D., Cheng, Y., & Lahiri, P. (2020). An evaluation of design-based properties of different composite estimators. *Statistics in Transition. New Series*, *21*(4), 166–190.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *26*(2), 211–252.

Bradley, A. A., Schwartz, S. S., & Hashino, T. (2008). Sampling uncertainty and confidence intervals for the Brier score and Brier skill score. *Weather and Forecasting*, *23*(5), 992–1006.

Breau, P., & Ernst, L. (1983). Alternative estimators to the current composite estimator. *Proceedings of the Survey Methods Section. American Statistical Association*, 397–402.

Breidt, F. J., & Opsomer, J. D. (2017). Model-assisted survey estimation with modern prediction techniques. *Statistical Science*, *32*(2), 190–205.

Breiman, L. (1996a). Bagging predictors. *Machine Learning*, *24*(2), 123–140.

Breiman, L. (1996b). Stacked regressions. *Machine Learning*, *24*(1), 49–64.

Breiman, L. (2001). Statistical modeling: The two cultures. *Statistical Science*, *16*(3), 199–215.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, *78*(1), 1–3.

Brown, G., Chambers, R., Heady, P., & Heasman, D. (2001). Evaluation of small area estimation methods. An application to unemployment estimates from the UK LFS. *Proceedings of Statistics Canada Symposium*.

Bühlmann, P. (2012). Bagging, boosting and ensemble methods. In J. Gentle, W. Härdle, & Y. Mori (Eds.), *Handbook of computational statistics* (2nd ed., pp. 985–1022). Springer.

Bühlmann, P., & Yu, B. (2002). Analyzing bagging. *The Annals of Statistics*, *30*(4), 927–961.

Burgard, J. P., Dörr, P., & Münnich, R. (2020a). Monte-Carlo simulation studies in survey statistics – an appraisal. Research Papers in Economics 04/20. Department of Economics. University of Trier.

Burgard, J. P., Esteban, M. D., Morales, D., & Pérez, A. (2020b). A Fay-Herriot model when auxiliary variables are measured with error. *TEST*, *29*(1), 166–195.

Burgard, J. P., Esteban, M. D., Morales, D., & Pérez, A. (2021a). Small area estimation under a measurement error bivariate Fay-Herriot model. *Statistical Methods & Applications*, *30*(1), 79–108.

Burgard, J. P., Krause, J., & Kreber, D. (2019a). Regularized area-level modelling for robust small area estimation in the presence of unknown covariate measurement errors. Research Papers in Economics 4/19. Department of Economics. University of Trier.

Burgard, J. P., Krause, J., Merkle, H., Münnich, R., & Schmaus, S. (2020c). Dynamische Mikrosimulationen zur Analyse und Planung regionaler Versorgungsstrukturen in der Pflege. In M. Hannapel & J. Kopp (Eds.), *Mikrosimulationen: Methodische Grundlagen und ausgewählte Anwendungsfelder* (pp. 283–313). Springer.

Burgard, J. P., Krause, J., & Morales, D. (2022). A measurement error Rao-Yu model for regional prevalence estimation over time using uncertain data obtained from dependent survey estimates. *TEST*, *31*, 204–234.

Burgard, J. P., Krause, J., & Schmaus, S. (2021b). Estimation of regional transition probabilities for spatial dynamic microsimulations from survey data lacking in regional detail. *Computational Statistics & Data Analysis*, *154*(107048).

Burgard, J. P., Morales, D., & Wölwer, A.-L. (2019b). Area-level small area estimation with missing values. Research Papers in Economics 14/19. Department of Economics. University of Trier.

Burgard, J. P., Morales, D., & Wölwer, A.-L. (2021c). Small area estimation of socioeconomic indicators for sampled and unsampled domains. *AStA Advances in Statistical Analysis*, *106*, 287–314.

Burgard, J. P., Münnich, R., & Zimmermann, T. (2014). The impact of sampling designs on small area estimates for business data. *Journal of Official Statistics*, *30*(4), 749–771.

Cantwell, P. J. (1990). Variance formulae for composite estimators in rotation designs. *Survey Methodology*, *16*(1), 153–163.

Cassel, C. M., Särndal, C.-E., & Wretman, J. H. (1976). Some results on generalized difference estimation and generalized regression estimation for finite populations. *Biometrika*, *63*(3), 615–620.

Chen, E. J., & Liu, T. P. (2002). Choices of alpha value in regression composite estimation for the Canadian Labour Force Survey: Impacts and evaluation. Methodology Branch Working Paper HSMD 2002-05E. Statistics Canada.

Cheng, Y., Huang, B., & Yu, Z. (2017). A note on iterative AK composite estimator for Current Population Survey. *Journal of Nonparametric Statistics*, *29*(2), 381–390.

Cheng, Y., Larsen, M. D., & Wakim, A. F. (2013). Comparisons of CPS unemployment estimates by rotation panel. *Proceedings of the Survey Research Methods Section. American Statistical Association*, 1825–1838.

Ciepiela, P., Gniado, M., Wesoowski, J., & Wojty, M. (2012). Dynamic K-composite estimator for an arbitrary rotation scheme. *Statistics in Transition. New Series*, *13*(1), 7–20.

Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). Wiley.

Collell, G., Prelec, D., & Patil, K. R. (2018). A simple plug-in bagging ensemble based on threshold-moving for classifying binary and multiclass imbalanced data. *Neurocomputing*, *275*, 330–340.

Cools, R. (1997). Constructing cubature formulae: The science behind the art. *Acta Numerica*, *6*, 1–54.

Cools, R. (2002). Advances in multidimensional integration. *Journal of Computational and Applied Mathematics*, *149*(1), 1–12.

Das, K., Jiang, J., & Rao, J. N. K. (2004). Mean squared error of empirical predictor. *The Annals of Statistics*, *32*(2), 818–840.

Datta, G. S., Ghosh, M., Nangia, N., & Natarajan, K. (1996). Estimation of median income of four-person families: A Bayesian approach. In D. A. Berry, K. M. Chaloner, & J. K. Geweke (Eds.), *Bayesian analysis in statistics and econometrics* (pp. 129–140). Wiley.

Datta, G. S., Fay, R. E., & Ghosh, M. (1991). Hierarchical and empirical multivariate Bayes analysis in small area estimation. *Proceedings of the Bureau of the Census Annual Research Conference*, 63–79.

Datta, G. S., & Lahiri, P. (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statistica Sinica*, *10*(2), 613–627.

Davis, P. J., & Polonsky, I. (1968). Numerical interpolation, differentiation and integration. In M. Abramowitz & I. A. Stegun (Eds.), *Handbook of mathematical functions with formulas, graphs, and mathematical tables* (pp. 875–924). Dover publications.

Davis, P. J., & Rabinowitz, P. (1984). *Methods of numerical integration* (2nd ed.). Academic Press.

De Bock, K. W., Coussement, K., & Cielen, D. (2019). An overview of multiple classifier systems based on generalized additive models. In E. Alfaro, M. Gámez, & N. García (Eds.), *Ensemble classification methods with applications in R* (pp. 175–186). Wiley.

De Bock, K. W., Coussement, K., & Van den Poel, D. (2010). Ensemble classification based on generalized additive models. *Computational Statistics & Data Analysis*, *54*(6), 1535–1546.

De Bock, K. W., & Van den Poel, D. (2012). Reconciling performance and interpretability in customer churn prediction using ensemble learning based on generalized additive models. *Expert Systems with Applications*, *39*(8), 6816–6826.

Dekking, F. M., Kraaikamp, C., Lopuhaä, H. P., & Meester, L. E. (2005). *A modern introduction to probability and statistics: Understanding why and how*. Springer.

Demidenko, E. (2013). *Mixed models: Theory and applications with R* (2nd ed.). Wiley.

Demidenko, E., & Spiegelman, D. (1997). A paradox: More measurement error can lead to more efficient estimates. *Communications in Statistics. Theory and Methods*, *26*(7), 1649–1675.

Destatis. (2018). Statistisches Jahrbuch. Deutschland und Internationales. Statistisches Bundesamt. Wiesbaden. https://www.destatis.de/DE/Themen/Querschnitt/Jahrbuch/statistisches-jahrbuch-2018-dl.pdf?___blob=publicationFile.

Destatis. (2020a). GV-ISys. Verzeichnis der Regional- und Gebietseinheiten. Definitionen und Beschreibungen. Statistisches Bundesamt. Wiesbaden. https://www.destatis.de/DE/Themen/Laender-Regionen/Regionales/Gemeindeverzeichnis/Administrativ/beschreibung-gebietseinheiten.pdf;jsessionid=474997AF38696EC2FE9344AB02AC168C.live712?___blob=publicationFile.

Destatis. (2020b). Mikrozensus 2019. Qualitätsbericht. Statistisches Bundesamt. Wiesbaden. https://www.destatis.de/DE/Methoden/Qualitaet/Qualitaetsberichte/Bevoelkerung/mikrozensus-2019.pdf?___blob=publicationFile.

Destatis. (2021). Mikrozensus 2020. Qualitätsbericht. Statistisches Bundesamt. Wiesbaden. https://www.destatis.de/DE/Methoden/Qualitaet/Qualitaetsberichte/Bevoelkerung/mikrozensus-2020.pdf?__blob=publicationFile.

Destatis & GESIS. (2012). Datenhandbuch zum Mikrozensus Scientific Use File 2008. https://www.forschungsdatenzentrum.de/sites/default/files/mz_2008_suf_svz.pdf.

Dever, J. A., & Valliant, R. (2010). A comparison of variance estimators for poststratification to estimated control totals. *Survey Methodology*, *36*(1), 45–56.

Deville, J.-C., & Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, *87*(418), 376–382.

Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. Chapman; Hall/CRC.

Erkens, G. (2012). Changes in panel bias in the U.S. Current Population Survey and its effects on labor force estimates. *Proceedings of the Survey Research Methods Section. American Statistical Association*, 4220–4232.

Esteban, M. D., Morales, D., Pérez, A., & Santamara, L. (2012). Two area-level time models for estimating small area poverty indicators. *Journal of the Indian Society of Agricultural Statistics*, *66*, 75–89.

Esteban, M. D., Lombardía, M. J., López-Vizcaíno, E., Morales, D., & Pérez, A. (2020). Small area estimation of proportions under area-level compositional mixed models. *TEST*, *29*(3), 793–818.

European Commission. (2016). EU labor force survey. Explanatory notes (to be applied from 2016Q1 onwards). Eurostat, Luxembourg. https://ec.europa.eu/eurostat/documents/1978984/6037342/EU-LFS-explanatory-notes-from-2016-onwards.pdf/0fd0fa60-b533-4a94-8766-fe3d78bcccad.

Fay, R. E. (1987). Application of multivariate regression of small domain estimation. In R. Platek, J. N. K. Rao, C. E. Särndal, & M. P. Singh (Eds.), *Small area statistics* (pp. 91–102). Wiley.

Fay, R. E., & Herriot, R. A. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association*, *74*(366), 269–277.

Fernández, A., García, S., Galar, M., Prati, R. C., Krawczyk, B., & Herrera, F. (2018). *Learning from imbalanced data sets*. Springer.

Freund, Y. (1995). Boosting a weak learning algorithm by majority. *Information and Computation*, *121*(2), 256–285.

Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, *55*, 119–139.

Fritsch, S., & Lüken, S. (2004). Erwerbstätigkeit in Deutschland – Methodische Grundlagen und Ergebnisse der Erwerbstätigenrechnung in den Volkswirtschaftlichen Gesamtrechnungen. *Wirtschaft und Statistik*, *2*, 139–147.

Fuller, W. A. (1990). Analysis of repeated surveys. *Survey Methodology*, *16*(2), 167–180.

Fuller, W. A., & Rao, J. N. K. (2001). A regression composite estimator with application to the Canadian Labour Force Survey. *Survey Methodology*, *27*(1), 45–51.

Galar, M., Fernández, A., Barrenechea, E., Bustince, H., & Herrera, F. (2011). A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybridbased approaches. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *42*(4), 463–484.

Gambino, J., Kennedy, B., & Singh, M. P. (2001). Regression composite estimation for the Canadian Labour Force Survey: Evaluation and implementation. *Survey Methodology*, *27*(1), 65–74.

Gatto, R., Loriga, S., & Spizzichino, A. (2009). Producing monthly estimates of labour market indicators exploiting the longitudinal dimension of the LFS microdata. *XXIV Convegno Nazionale di Economia del Lavoro – AIEL. Sassari 24–25 Settembre 2009*.

Gentle, J. E. (2003). *Random number generation and Monte Carlo methods* (2nd ed.). Springer.

Gentle, J. E. (2007). *Matrix algebra: Theory, computations, and applications in statistics*. Springer.

Ghosh, M. (2020). Small area estimation: Its evolution in five decades. *Statistics in Transition. New Series*, *21*(4), 1–22.

Ghosh, M., Nangia, N., & Kim, D. H. (1996). Estimation of median income of four-person families: A Bayesian time series approach. *Journal of the American Statistical Association*, *91*(436), 1423–1431.

Goldfarb, D., & Idnani, A. (1982). Dual and primal-dual methods for solving strictly convex quadratic programs. In J. P. Hennart (Ed.), *Numerical analysis. Lecture notes in mathematics* (pp. 226–239). Springer.

Goldfarb, D., & Idnani, A. (1983). A numerically stable dual method for solving strictly convex quadratic programs. *Mathematical Programming*, *27*(1), 1–33.

González, J., Tuerlinckx, F., De Boeck, P., & Cools, R. (2006). Numerical integration in logistic-normal models. *Computational Statistics & Data Analysis*, *51*(3), 1535–1548.

González-Manteiga, W., Lombardía, M. J., Molina, I., Morales, D., & Santamaría, L. (2008). Analytic and bootstrap approximations of prediction errors under a multivariate Fay-Herriot model. *Computational Statistics & Data Analysis*, *52*(12), 5242–5252.

Gurney, M., & Daly, J. F. (1965). A multivariate approach to estimation in periodic sample surveys. *Proceedings of the Social Statistics Section. American Statistical Association*, 242–257.

Guzman, G. G. (2019). Household income: 2018. Report Number ACSBR/18-01. U.S. Census Bureau. https://www.census.gov/content/dam/Census/library/publications/2019/acs/acsbr18-01.pdf.

Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *The Annals of Mathematical Statistics*, *35*(4), 1491–1523.

Halpern-Manners, A., & Warren, J. R. (2012). Panel conditioning in longitudinal studies: Evidence from labor force items in the Current Population Survey. *Demography*, *49*(4), 1499–1519.

Halton, J. H. (1960). On the efficiency of certain quasi-random sequences of points in evaluating multi-domensional integrals. *Numerische Mathematik*, *2*, 84–90.

Hansen, M. H., Hurwitz, W. N., Nisselson, H., & Steinberg, J. (1955). The redesign of the Census Current Population Survey. *Journal of the American Statistical Association*, *50*(271), 701–719.

Hartmann, M., & Riede, T. (2005). Erwerbslosigkeit nach dem Labour-Force-Konzept – Arbeitslosigkeit nach dem Sozialgesetzbuch: Gemeinsamkeiten und Unterschiede. *Wirtschaft und Statistik*, *4*, 303–310.

Hastie, T., & Tibshirani, R. (1986). Generalized additive models. *Statistical Science*, *1*(3), 297–318.

Hastie, T., & Tibshirani, R. (1990). *Generalized additive models*. Chapman; Hall.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(9), 1263–1284.

Henderson, C. R. (1963). Selection index and expected genetic advances. In W. D. Hanson & H. F. Robinson (Eds.), *Statistical genetics and plant breeding* (pp. 141–163). National Academy of Sciences – National Research Council.

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, *31*(2), 423–447.

Hochgürtel, T. (2013). Das künftige System der amtlichen Haushaltsstatistiken. *Wirtschaft und Statistik*, *7*, 457–466.

Hochstrasser, U. W. (1968). Orthogonal polynomials. In M. Abramowitz & I. A. Stegun (Eds.), *Handbook of mathematical functions with formulas, graphs, and mathematical tables* (pp. 771–802). Dover publications.

Honoré, B. E. (2002). Non-linear models with panel data. Working Paper CWP 13/02. Centre for Microdata Methods and Practice. Institute for Fiscal Studies, London.

Horvitz, D. G., & Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, *47*(260), 663–685.

Huang, E. T., & Bell, W. R. (2004). An empirical study on using ACS supplementary survey data in SAIPE state poverty models. *Proceedings of the Survey Research Methods Section. American Statistical Association*, 3677–3684.

Huang, E. T., & Bell, W. R. (2012). An empirical study on using previous American Community Survey data versus Census 2000 data in SAIPE models for poverty estimates. Research Report Series (Statistics 2012-04). U.S. Census Bureau. https://www.census.gov/content/dam/Census/library/working-papers/2012/adrm/rrs2012-04.pdf.

Hundenborn, J., & Enderer, J. (2019). Die Neuregelung des Mikrozensus ab 2020. *Wirtschaft und Statistik*, *6*, 9–17.

Isaki, C. T., & Fuller, W. A. (1982). Survey design under the regression superpopulation model. *Journal of the American Statistical Association*, *77*(377), 89–96.

Jäckel, P. (2002). *Monte Carlo methods in finance*. Wiley.

Jiang, J. (2007). *Linear and generalized linear mixed models and their applications*. Springer.

Jiang, J., & Lahiri, P. (2006). Mixed model prediction and small area estimation. *TEST*, *15*(1), 1–96.

Judd, K. L., Maliar, L., & Maliar, S. (2011). Supplement to "Numerically stable and accurate stochastic simulation approaches for solving dynamic economic models": Appendices. *Quantitative Economics*, *2*(2), 173–210.

Kackar, R. N., & Harville, D. A. (1981). Unbiasedness of two-stage estimation and prediction procedures for mixed linear models. *Communications in Statistics. Theory and Methods*, *10*(13), 1249–1261.

Kackar, R. N., & Harville, D. A. (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *Journal of the American Statistical Association*, *79*(388), 853–862.

Kalton, G. (2009). Designs for surveys over time. In D. Pfeffermann & C. R. Rao (Eds.), *Handbook of statistics* (pp. 89–108). Elsevier.

Kolb, J.-P. (2013). Methoden zur Erzeugung synthetischer Simulationsgesamtheiten. Dissertation. University of Trier. https://doi.org/10.25353/ubtr-xxxx-6d4e-ea71.

Kordos, J. (2014). Development of small area estimation in official statistics. *Statistics in Transition. New Series*, *17*(1), 105–132.

Körner, T., & Marder-Puch, K. (2015). Der Mikrozensus im Vergleich mit anderen Arbeitsmarktstatistiken – Ergebnisunterschiede und Hintergründe seit 2011. *Wirtschaft und Statistik*, *4*, 39–53.

Kott, P. S. (2006). Using calibration weighting to adjust for nonresponse and coverage errors. *Survey Methodology*, *32*(2), 133–142.

Kott, P. S. (2009). Calibration weighting: Combining probability samples and linear prediction models. In C. R. Rao (Ed.), *Handbook of statistics* (pp. 55–82). Elsevier.

Krause, J., Burgard, J. P., & Morales, D. (2022). Robust prediction of domain compositions from uncertain data using isometric logratio transformations in a penalized multivariate Fay-Herriot model. *Statistica Neerlandica*, *76*(1), 65–96.

Krueger, A. B., Mas, A., & Niu, X. (2017). The evolution of rotation group bias: Will the real unemployment rate please stand up? *Review of Economics and Statistics*, *99*(2), 258–264.

Kruppa, J., Liu, Y., Biau, G., Kohler, M., König, I. R., Malley, J. D., & Ziegler, A. (2014a). Probability estimation with machine learning methods for dichotomous and multicategory outcome: Theory. *Biometrical Journal*, *56*(4), 534–563.

Kruppa, J., Liu, Y., Diener, H.-C., Holste, T., Weimar, C., König, I. R., & Ziegler, A. (2014b). Probability estimation with machine learning methods for dichotomous and multicategory outcome: Applications. *Biometrical Journal*, *56*(4), 564–583.

Lahiri, P., & Pramanik, S. (2019). Evaluation of synthetic small-area estimators using design-based methods. *Austrian Journal of Statistics*, *48*(4), 43–57.

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, *38*(4), 963–974.

Lehtonen, R., & Veijanen, A. (2009). Design-based methods of estimation for domains and small areas. In C. R. Rao (Ed.), *Handbook of statistics* (pp. 219–249). Elsevier.

Lent, J., Miller, S. M., & Cantwell, P. (1994). Composite weights for the Current Population Surveys. *Proceedings of the Survey Research Methods Section. American Statistical Association*, 867–872.

Lent, J., Miller, S. M., & Duff, M. (1999). Effects of composite weights on some estimates from the Current Population Survey. *Journal of Official Statistics*, *15*(3), 431–448.

Leombruni, R., & Richiardi, M. (2006). LABORsim: An agent-based microsimulation of labour supply – An application to Italy. *Computational Economics*, *27*(1), 63–88.

Levell, P., & Shaw, J. (2016). Constructing full adult life-cycles from short panels. *International Journal of Microsimulation*, *9*(2), 5–40.

Li, J., & O'Donoghue, C. (2012). Simulating histories within dynamic microsimulation models. *International Journal of Microsimulation*, *5*(1), 52–76.

Li, J., & O'Donoghue, C. (2014). Evaluating binary alignment methods in microsimulation models. *Journal of Artificial Societies and Social Simulation*, *17*(1), 1–15.

Littell, R. C. (2002). Analysis of unbalanced mixed model data: A case study comparison of ANOVA versus REML/GLS. *Journal of Agricultural, Biological, and Environmental Statistics*, *7*(4), 472–490.

Littell, R. C., Pendergast, J., & Natarajan, R. (2004). Mixed models: Modelling covariance structure in the analysis of repeated measures data. In R. B. D'Agostino (Ed.), *Tutorials in biostatistics* (pp. 159–185). Wiley.

Liu, Q., & Pierce, D. A. (1994). A note on Gauss-Hermite quadrature. *Biometrika*, *81*(3), 624–629.

Lohr, S. L. (2010). *Sampling: Design and analysis* (2nd ed.). Brooks/Cole.

Lohr, S. L., & Ybarra, L. M. R. (2002). Area-level models using data from multiple surveys. *Proceedings: Modelling Survey Data for Social and Economic Research. Statistics Canada.*

Loriga, S. (2014). Calibration and regression composite estimation. *Training Regression Composite Estimation. 4–5 December. Eurostat. Luxemburg.*

Marchetti, S., & Secondi, L. (2017). Estimates of household consumption expenditure at provincial level in Italy by using small area estimation methods: real comparisons using purchasing power parities. *Social Indicators Research*, *131*(1), 215–234.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). Chapman; Hall.

McLay, J. M., Lay-Yee, R., Milne, B. J., & Davis, P. (2015). Regression-style models for parameter estimation in dynamic microsimulation: An empirical performance assessment. *International Journal of Microsimulation*, *8*(2), 83–127.

Molina, I., & Marhuenda, Y. (2015). SAE: An R package for small area estimation. *The R Journal*, *7*(1), 81–98.

Morales, D., Esteban, M. D., Pérez, A., & Hobza, T. (2021). *A course on small area estimation and mixed models*. Springer.

Münnich, R. (2008). Varianzschätzung in komplexen Erhebungen. *Austrian Journal of Statistics*, *37*(3&4), 319–334.

Münnich, R., & Burgard, J. P. (2012). On the influence of sampling design on small area estimates. *Journal of the Indian Society of Agricultural Statistics*, *66*(1), 145–156.

Münnich, R., Burgard, J. P., & Vogt, M. (2013). Small Area-Statistik: Methoden und Anwendungen. *AStA Wirtschafts- und Sozialstatistisches Archiv*, *6*, 149–191.

Münnich, R., Gabler, S., Ganninger, M., Burgard, J. P., & Kolb, J.-P. (2012a). Stichprobenoptimierung und Schätzung im Zensus 2011. *Statistik und Wissenschaft*, *21*.

Münnich, R., Sachs, E., & Wagner, M. (2012b). Calibration of estimator-weights via semismooth Newton method. *Journal of Global Optimization*, *52*(3), 471–485.

National Statistics Bureau of Bhutan and the World Bank. (2010). Small area estimation of poverty in rural Bhutan. Technical Report jointly prepared by National Statistics

Bureau of Bhutan and the World Bank https://documents1.worldbank.org/curated/en/221871468200950359/pdf/681790ESW0WHIT0erty0in0Rural0Bhutan.pdf.

Nelder, J. A. (1968). The combination of information in generally balanced designs. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *30*(2), 303–311.

Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *135*(3), 370–384.

Neyman, J., & Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica: Journal of the Econometric Society*, *16*(1), 1–32.

Nieuwenbroek, N., & Boonstra, H. J. (2002). Bascula 4.0 reference manual. BPA nr: 279-02-TMO. Statistics Netherlands.

Nocedal, J., & Wright, S. (2006). *Numerical optimization* (2nd ed.). Springer.

Park, M., & Fuller, W. A. (2005). Towards nonnegative regression weights for survey samples. *Survey Methodology*, *31*(1), 85–93.

Patterson, H. D., & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, *58*(3), 545–554.

Pfeffermann, D. (2002). Small area estimation – New developments and directions. *International Statistical Review*, *70*(1), 125–143.

Pfeffermann, D. (2013). New important developments in small area estimation. *Statistical Science*, *28*(1), 40–68.

Pfeffermann, D., & Ben-Hur, D. (2019). Estimation of randomisation mean square error in small area estimation. *International Statistical Review*, *87*(1), S31–S49.

Pink, B. (2007). Forthcoming changes to labour force statistics. ABS Catalogue No. 6292.0. Australian Bureau of Statistics.

Polikar, R. (2006). Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine*, *6*(3), 21–45.

Porter, A. T., Wikle, C. K., & Holan, S. H. (2015). Small area estimation via multivariate Fay-Herriot models with latent spatial dependence. *Australian & New Zealand Journal of Statistics*, *57*(1), 15–29.

Prasad, N. G. N., & Rao, J. N. K. (1990). The estimation of the mean squared error of small-area estimators. *Journal of the American Statistical Association*, *85*(409), 163–171.

Press, W. H., William, H., Teukolsky, S. A., Saul, A., Vetterling, W. T., & Flannery, B. P. (2007). *Numerical recipes: The art of scientific computing* (3rd ed.). Cambridge University Press.

Preston, J. (2015). Modified regression estimator for repeated business surveys with changing survey frames. *Survey Methodology*, *41*(1), 79–97.

R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. www.R-project.org.

Rao, J. N. K., & Molina, I. (2015). *Small area estimation* (2nd ed.). Wiley.

Rengers, M. (2004). Das international vereinbarte Labour-Force-Konzept. *Wirtschaft und Statistik*, *12*, 1369–1383.

Research Data Centres of the Statistical Offices of the Federation and the Federal States. (2018). Microcensus 2010. Campus file. https://doi.org/10.21242/12211.2010.00.00.5.2.0.

Richardson, R., Pacelli, L., Poggi, A., & Richiardi, M. (2018). Female labour force projections using microsimulation for six EU countries. *International Journal of Microsimulation*, *11*(2), 5–51.

Richiardi, M., & Poggi, A. (2014). Imputing individual effects in dynamic microsimulation models. an application to household formation and labour market participation in Italy. *International Journal of Microsimulation*, *7*(2), 3–39.

Riede, T. (2013). Weiterentwicklung des Systems der amtlichen Haushaltsstatistiken. In T. Riede, S. Bechthold, & N. Ott (Eds.), *Weiterentwicklung der amtlichen Haushaltsstatistiken* (pp. 13–30). SCIVERO.

Robinson, P. M., & Särndal, C.-E. (1983). Asymptotic properties of the generalized regression estimator in probability sampling. *Sankhy: The Indian Journal of Statistics, Series B*, *45*(2), 240–248.

Rupp, M. (2018). Optimization for multivariate and multi-domain methods in survey statistics. Dissertation. University of Trier. https://doi.org/10.25353/UBTR-8351-5432-14XX.

Salonen, R. (2014). Regression composite estimation for the Finnish LFS from a practical perspective. *9th Workshop on Labour Force Survey Methodology. 15th–16th of May. Rome.*

Salonen, R. (2016). Comparison of the single month estimates and the preliminary results of the Italian way of regression composite estimation method for the FI-LFS. *11th Workshop on Labour Force Survey Methodology. 27–28 April. Cardiff.*

Särndal, C.-E. (1980). On pi-inverse weighting versus best linear unbiased weighting in probability sampling. *Biometrika*, *67*(3), 639–650.

Särndal, C.-E. (2007). The calibration approach in survey theory and practice. *Survey Methodology*, *33*(2), 99–119.

Särndal, C.-E., Swensson, B., & Wretman, J. (1992). *Model-assisted survey sampling.* Springer.

Schapire, R. E. (1990). The strength of weak learnability. *Machine Learning*, *5*(2), 197–227.

Schimpl-Neimanns, B. (2011). Schätzung des Stichprobenfehlers in Mikrozensus Scientific Use Files ab 2005. *AStA Wirtschafts- und Sozialstatistisches Archiv*, *5*(1), 19–38.

Schmid, T., & Münnich, R. T. (2014). Spatial robust small area estimation. *Statistical Papers*, *55*(3), 653–670.

Schoch, T. (2014). rsae: Robust small area estimation. R package version 0.1-5. https://cran.r-project.org/web/packages/rsae/index.html.

Searle, S. R., Casella, G., & McCulloch, C. E. (2006). *Variance components.* Wiley.

Semega, J., Kollar, M., Creamer, J., & Mohanty, A. (2019). Income and poverty in the United States: 2018. Report Number P60-266. U.S. Census Bureau. https://www.census.gov/content/dam/Census/library/publications/2019/demo/p60-266.pdf.

Singh, A. C. (1996). Combining information in survey sampling by modified regression. *Proceedings of the Survey Research Methods Section. American Statistical Association*, 120–129.

Singh, A. C., Kenedy, B., & Wu, S. (2001a). Regression composite estimation for the Canadian Labour Force Survey with a rotating panel design. *Survey Methodology*, *27*(1), 33–44.

Singh, A. C., Kennedy, B., Wu, S., & Brisebois, F. (1997). Composite estimation for the Canadian Labour Force Survey. *Proceedings of the Survey Research Methods Section. American Statistical Association*, 300–305.

Singh, A. C., & Merkouris, P. (1995). Composite estimation by modified regression for repeated surveys. *Proceedings of the Survey Research Methods Section. American Statistical Association*, 420–425.

Singh, M. P., Hidiroglou, M. A., Gambino, J. G., & Kovaçevi, M. S. (2001b). Estimation methods and related systems at Statistics Canada. *International Statistical Review*, *69*(3), 461–485.

Sinha, S. K., & Rao, J. N. K. (2009). Robust small area estimation. *Canadian Journal of Statistics*, *37*(3), 381–399.

Skinner, C., & Wakefield, J. (2017). Introduction to the design and analysis of complex survey data. *Statistical Science*, *32*(2), 165–175.

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Sage.

Sobol', L. M. (1967). On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics*, *7*, 86–112.

Sobol', L. M. (1976). Uniformly distributed sequences with an additional uniform property. *USSR Computational Mathematics and Mathematical Physics*, *16*, 236–242.

Statistics Canada. (2017). Methodology of the Canadian Labour Force Survey. Catalogue no. 71-526-X.

Statistische Ämter des Bundes und der Länder. (2015). Zensus 2011. Methoden und Verfahren. Statistisches Bundesamt. Wiesbaden. https://www.zensus2011.de.

Steel, D., & McLaren, C. (2009). Design and analysis of surveys repeated over time. In C. R. Rao (Ed.), *Handbook of statistics* (pp. 289–313). Elsevier.

Stephensen, P. (2016). Logit scaling: A general method for alignment in microsimulation models. *International Journal of Microsimulation*, *9*(3), 89–102.

Stroud, A. H. (1971). *Approximate calculation of multiple integrals.* Prentice-Hall.

Templ, M. (2017). *Statistical disclosure control for microdata: Methods and applications in R.* Springer.

Turlach, B. A., & Weingessel, A. (2019). quadprog: Functions to solve quadratic programming problems. R package version 1.5-8. https://cran.r-project.org/web/packages/quadprog/index.html.

Ubaidillah, A., Notodiputro, K. A., Kurnia, A., & Mangku, I. W. (2019). Multivariate Fay-Herriot models for small area estimation with application to household consumption per capita expenditure in Indonesia. *Journal of Applied Statistics*, *46*(15), 2845–2861.

U.S. Bureau of Labor Statistics. (2006). Design and methodology. Current Population Survey. Technical Paper 66. https://cps.ipums.org/cps/resources/cpr/tp66.pdf.

U.S. Bureau of the Census. (2020). Statistical expertise & general research topics. U.S. Bureau of the Census. (FY 2021–FY 2025). Revised in January 2022 https://www.census.gov/content/dam/Census/topics/research/statistical-research/csrmfyp.pdf.

U.S. Census Bureau. (2014). American Community Survey. Design and methodology. Report Version 2.0. Washington, DC.

U.S. Census Bureau. (2016). American Community Survey. Data suppression. Washington, DC.

van Laarhoven, P. J. M., & Aarts, E. H. L. (1987). *Simulated annealing: Theory and applications*. Springer.

Wagner, J., Münnich, R., Hill, J., Stoffels, J., & Udelhoven, T. (2017). Non-parametric small area models using shape-constrained penalized B-splines. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *180*(4), 1089–1109.

Wallace, B. C., & Dahabreh, I. J. (2012). Class probability estimates are unreliable for imbalanced data (and how to fix them). *2012 IEEE 12th International Conference on Data Mining*, 695–704.

Wilks, D. S. (2010). Sampling distributions of the Brier score and Brier skill score under serial dependence. *Quarterly Journal of the Royal Meteorological Society*, *136*(653), 2109–2118.

Willenborg, L., & De Waal, T. (2012). *Elements of statistical disclosure control*. Springer.

Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, *5*(2), 241–259.

Wolter, K. (1985). *Introduction to variance estimation*. Springer.

Wood, J. (2008). On the covariance between related Horvitz-Thompson estimators. *Journal of Official Statistics*, *24*(1), 53–78.

Wood, S. N. (2003). Thin-plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *65*(1), 95–114.

Wood, S. N. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association*, *99*(467), 673–686.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*(1), 3–36.

Wood, S. N. (2017). *Generalized additive models: An introduction with R* (2nd ed.). Chapman; Hall/CRC.

Wood, S. N., Pya, N., & Säfken, B. (2016). Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association*, *111*(516), 1548–1563.

Wright, R. L. (1983). Finite population sampling with multivariate auxiliary information. *Journal of the American Statistical Association*, *78*(384), 879–884.

Wu, Y., Hu, D., Wu, M., & Hu, X. (2006). A numerical-integration perspective on Gaussian filters. *IEEE Transactions on Signal Processing*, *54*(8), 2910–2921.

Wuertz, D., Setz, T., & Chalabi, Y. (2017). fOptions: Rmetrics - pricing and evaluating basic options. R package version 3042.86. https://cran.r-project.org/web/packages/fOptions/index.html.

Xu, Z., Chang, X., Xu, F., & Zhang, H. (2012). $L_{1/2}$ Regularization: A thresholding representation theory and a fast solver. *IEEE Transactions on Neural Networks and Learning Systems*, *23*(7), 1013–1027.

Yang, Q., & Wu, X. (2006). 10 challenging problems in data mining research. *International Journal of Information Technology & Decision Making*, *5*(04), 597–604.

Yansaneh, I. S., & Fuller, W. A. (1998). Optimal recursive estimation for repeated surveys. *Survey Methodology*, *24*, 31–40.

Ybarra, L. M. R., & Lohr, S. L. (2008). Small area estimation when auxiliary information is measured with error. *Biometrika*, *95*(4), 919–931.

Zeng, J., Lin, S., Wang, Y., & Xu, Z. (2014). $L_{1/2}$ Regularization: Convergence of iterative half thresholding algorithm. *IEEE Transactions on Signal Processing*, *62*(9), 2317–2329.

Zimmermann, T. (2018). The interplay between sampling design and statistical modelling in small area estimation. Dissertation. University of Trier. https://doi.org/10.25353/UBTR-1671-5300-84XX.

Zimmermann, T. (2019). Einsatzmöglichkeiten von Small Area-Verfahren bei Kohortenschätzungen im Zensus 2021. *AStA Wirtschafts- und Sozialstatistisches Archiv*, *13*(2), 157–177.

Appendix A

Additional Material to Chapter 5

A.1 Simulation 1

With $\rho_e = 0.25$, for the plug-in predictors $\hat{R}_d^{\text{FH,BI}}$ and $\hat{R}_d^{\text{EBI}}$ the ARBias (in %) is 0.11 and 0.13 and the RRMSE (in %) is 8.33 and 7.82 respectively.

Table A.1: ARBias (in %), $\rho_e = 0.25$

		Number of function evaluations e						
Int.	Approx.	16	24/25	100	400	2,500	10,000	40,000
$\boldsymbol{I_1}$	MC	0.08	0.06	0.07	0.07	0.08	0.08	0.08
	MCA	0.08	0.08	0.08	0.08	0.08	0.08	0.08
	GH	0.08	0.08	0.08	0.08	0.08	0.08	0.08
	QMCH	0.53	0.44	0.14	0.01	0.06	0.07	0.08
	QMCS	0.51	0.21	0.05	0.04	0.07	0.08	0.08
$\boldsymbol{I_2}$	MC	0.30	0.29	0.12	0.10	0.08	0.08	0.08
	MCA	0.02	0.03	0.13	0.10	0.08	0.07	0.08
	GH	0.08	0.08	0.08	0.08	0.08	0.08	0.08
	QMCH	0.54	0.45	0.12	0.03	0.06	0.07	0.08
	QMCS	0.41	0.21	0.07	0.06	0.08	0.08	0.08
$\boldsymbol{I_3}$	MC	2.07	0.47	0.09	0.30	0.07	0.13	0.12
	MCA	1.59	0.70	0.48	0.17	0.02	0.02	0.11
	GH	0.06	0.08	0.08	0.08	0.08	0.08	0.08
	QMCH	1.43	1.13	0.32	0.09	0.06	0.07	0.08
	QMCS	1.65	0.19	0.27	0.14	0.05	0.07	0.08

With $\rho_e = 0$, for the plug-in predictors $\hat{R}_d^{\text{FH,BI}}$ and $\hat{R}_d^{\text{EBI}}$ the ARBias (in %) is 0.06 and 0.08 and the RRMSE (in %) is 8.35 and 8.07 respectively.

Table A.2: RRMSE (in %), $\rho_e = 0.25$

		Number of function evaluations e						
Int.	Approx.	16	24/25	100	400	2,500	10,000	40,000
	MC	7.97	7.96	7.89	7.85	7.82	7.82	7.82
	MCA	7.82	7.82	7.82	7.82	7.82	7.82	7.82
$\boldsymbol{I_1}$	GH	7.82	7.82	7.82	7.82	7.82	7.82	7.82
	QMCH	7.84	7.83	7.82	7.82	7.82	7.82	7.82
	QMCS	7.84	7.82	7.82	7.82	7.82	7.82	7.82
	MC	8.50	8.28	8.01	7.86	7.83	7.83	7.82
	MCA	8.31	8.09	7.93	7.83	7.82	7.82	7.82
$\boldsymbol{I_2}$	GH	7.84	7.82	7.82	7.82	7.82	7.82	7.82
	QMCH	8	7.97	7.85	7.83	7.82	7.82	7.82
	QMCS	7.98	7.88	7.85	7.83	7.82	7.82	7.82
	MC	43.39	32.15	15.97	11.68	8.75	7.98	7.87
	MCA	41.50	30.20	17.14	10.21	8.49	7.94	7.85
$\boldsymbol{I_3}$	GH	7.97	7.85	7.82	7.82	7.82	7.82	7.82
	QMCH	17.07	16.91	9.19	8.16	7.87	7.83	7.82
	QMCS	16.90	11.90	10.08	9.74	7.84	7.83	7.82

Table A.3: ARBias (in %), $\rho_e = 0$

		Number of function evaluations e						
Int.	Approx.	16	24/25	100	400	2,500	10,000	40,000
	MC	0.05	0.03	0.03	0.02	0.03	0.03	0.04
	MCA	0.03	0.03	0.03	0.03	0.03	0.03	0.03
$\boldsymbol{I_1}$	GH	0.03	0.03	0.03	0.03	0.03	0.03	0.03
	QMCH	0.55	0.48	0.18	0.04	0.02	0.03	0.03
	QMCS	0.58	0.16	0	0	0.03	0.03	0.03
	MC	0.06	0.02	0.01	0.02	0.03	0.03	0.04
	MCA	0.09	0.07	0.03	0	0.02	0.03	0.03
$\boldsymbol{I_2}$	GH	0.04	0.03	0.03	0.03	0.03	0.03	0.03
	QMCH	0.59	0.51	0.16	0.03	0.02	0.03	0.03
	QMCS	0.46	0.17	0	0	0.03	0.03	0.03
	MC	1.01	0.40	0.01	0.08	0.06	0.01	0.06
	MCA	0.67	1.11	0.20	0.07	0.03	0.04	0.03
$\boldsymbol{I_3}$	GH	0.02	0.04	0.03	0.03	0.03	0.03	0.03
	QMCH	0.86	0.84	0.26	0.06	0.01	0.03	0.03
	QMCS	1.30	0.03	0.13	0.04	0.04	0.03	0.04

Table A.4: RRMSE (in %), $\rho_e = 0$

		Number of function evaluations e						
Int.	Approx.	16	24/25	100	400	2,500	10,000	40,000
$\boldsymbol{I_1}$	MC	8.09	8.08	8.09	8.08	8.07	8.07	8.07
	MCA	8.07	8.07	8.07	8.07	8.07	8.07	8.07
	GH	8.07	8.07	8.07	8.07	8.07	8.07	8.07
	QMCH	8.09	8.08	8.07	8.07	8.07	8.07	8.07
	QMCS	8.09	8.07	8.07	8.07	8.07	8.07	8.07
$\boldsymbol{I_2}$	MC	8.62	8.44	8.19	8.10	8.09	8.07	8.07
	MCA	8.59	8.46	8.18	8.10	8.07	8.07	8.07
	GH	8.08	8.07	8.07	8.07	8.07	8.07	8.07
	QMCH	8.14	8.18	8.08	8.07	8.07	8.07	8.07
	QMCS	8.18	8.11	8.07	8.09	8.07	8.07	8.07
$\boldsymbol{I_3}$	MC	33.49	29.05	15.79	10.27	8.53	8.23	8.09
	MCA	26.23	21.18	13.51	9.63	8.29	8.12	8.10
	GH	8.16	8.07	8.07	8.07	8.07	8.07	8.07
	QMCH	16.50	15.54	9.49	8.31	8.09	8.07	8.07
	QMCS	15.93	11.14	9.74	8.25	8.07	8.07	8.07

A.2 Application

This section presents the maps of unemployment rate EBPs for the ages groups `AGE2` (25-54) and `AGE3` (55-64). We observe that unemployment rates are larger in the south of Spain for both sexes and that they are larger for females. The maps of estimated root-MSEs are also given.

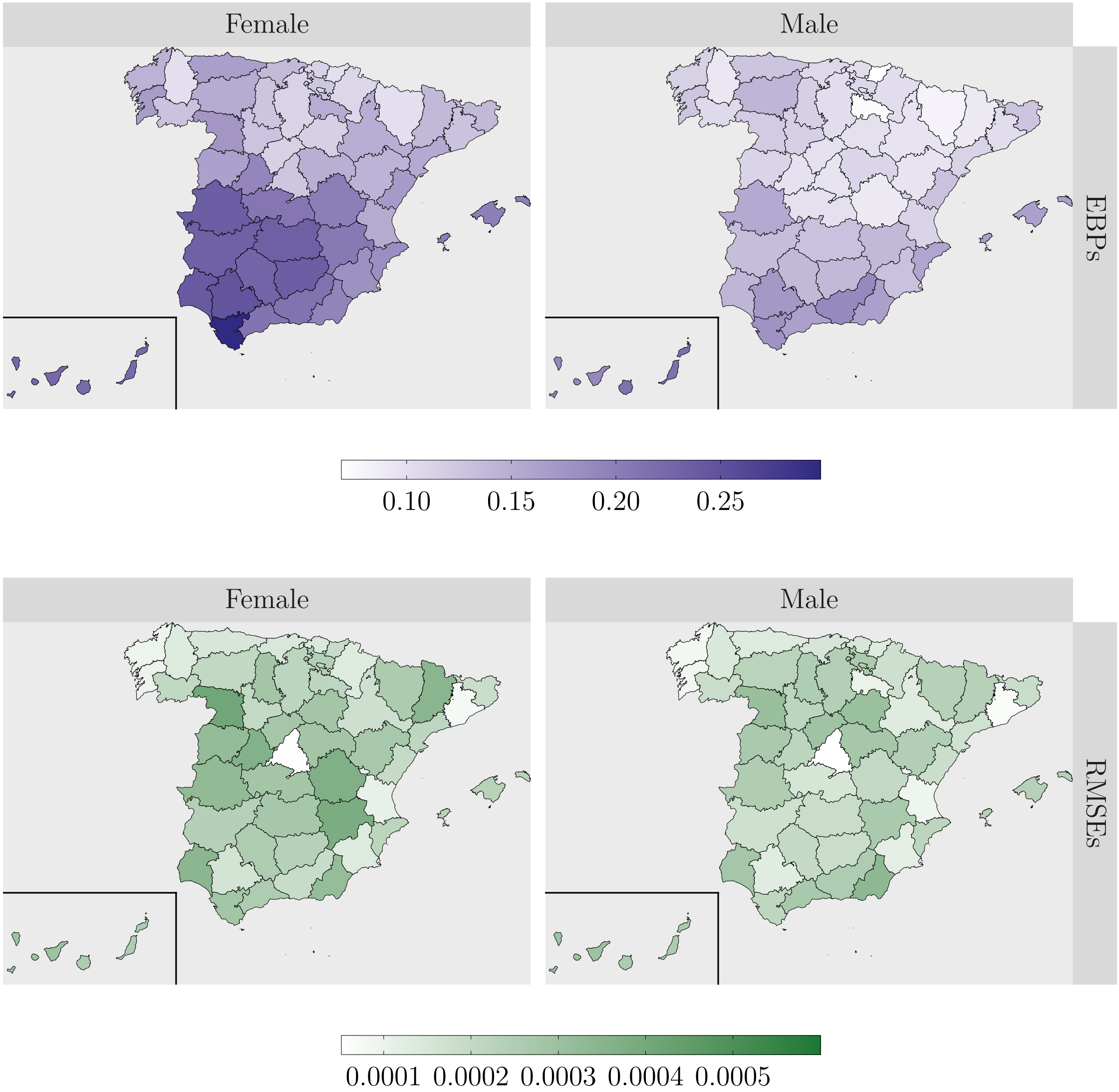

Figure A.1: BFH EBPs and their RMSEs for `AGE2` (25-54 years)

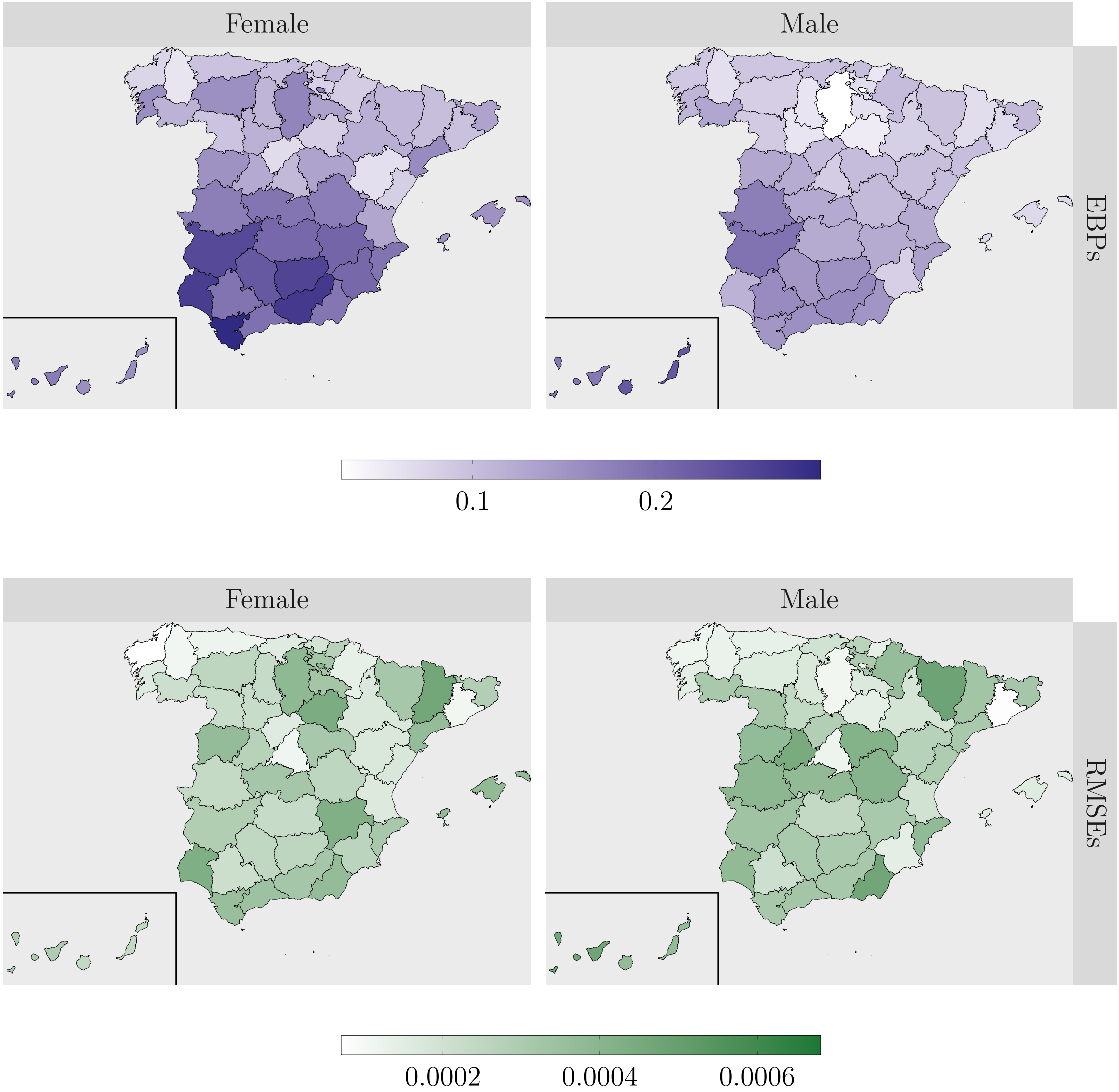

Figure A.2: BFH EBPs and their RMSEs for `AGE3` (55-64 years)